

ISSN 2713-3192
DOI 10.15622/ia.2026.25.5
<http://ia.spcras.ru>

TOM 25 № 5

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

INFORMATICS AND AUTOMATION



СПб ФИЦ РАН

Санкт-Петербург
2026



INFORMATICS AND AUTOMATION

Volume 25 № 5, 2026

A scientific and educational journal primarily specializing in computer science, automation, robotics, applied mathematics and interdisciplinary research

Founded in 2002

Frequency

6 times a year

Founder and Publisher:

St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS)

Editorial Office and Publisher Address:

39, 14th Line V.O., St. Petersburg, 199178, Russia

e-mail: ia@spcras.ru, website: <http://ia.spcras.ru>

Editor-in-Chief:

A. L. Ronzhin, Dr. Sci., Prof., St. Petersburg, Russia

Managing Editor: V. A. Ganicheva

Translator: Ya. N. Berezina

Art Editor: Yu. S. Cherepanova

The journal is registered with the Federal Service for Supervision of Communications, Information Technology, and Mass Media,
Registration Certificate (registration number) ПИ № ФС77-79228 dated September 25, 2020

The journal is indexed in the Scopus international database

The journal is included in the «List of Leading Peer-Reviewed Scientific Journals and Publications, in which the main scientific results of a dissertation for the degree of Doctor and Candidate of Sciences must be published»

The journal is published under the scientific and methodological guidance of the Department of Nanotechnology and Information Technologies of the Russian Academy of Sciences

© St. Petersburg Federal Research Center of the Russian Academy of Sciences, 2026

Reproduction in the press, as well as broadcasting or transmission by cable, of articles published in the print periodical «Informatics and Automation» on current economic, political, social, and religious issues is permitted, provided that the author's name and the print periodical «Informatics and Automation» are clearly indicated

Editorial Council

A. A. Ashimov	Academician of the National Academy of Sciences of the Republic of Kazakhstan, Dr. Sci., Prof., Almaty, Kazakhstan
I. A. Kalyaev	Academician of RAS, Dr. Sci., Prof., Taganrog, Russia
A. I. Rudskoi	Academician of RAS, Dr. Sci., Prof., St. Petersburg, Russia
V. Sgurev	Academician of the Bulgarian Academy of Sciences, Dr. Sci., Prof., Sofia, Bulgaria
B. Ya. Sovetov	Academician of RAE, Dr. Sci., Prof., St. Petersburg, Russia
V. A. Soyfer	Academician of RAS, Dr. Sci., Prof., Samara, Russia

Editorial Board

E. Azarov	Dr. Sci., Prof., Minsk, Belarus
M. V. Bobyr	Dr. Sci., Prof., Kursk, Russia
O. Yu. Gusikhin	Ph. D., Dearborn, USA
V. Delic	Dr. Sci., Prof., Novi Sad, Serbia
A. Dolgui	Dr. Sci., Prof., St. Etienne, France
M. Zelezný	Ph.D., Assoc. Prof., Plzen, Czech Republic
H. Kaya	Ph.D., Assoc. Prof., Utrecht, Netherlands
A. A. Karpov	Dr. Sci., Assoc. Prof., St. Petersburg, Russia
S. V. Kuleshov	Dr. Sci., St. Petersburg, Russia
E. A. Magid	Ph. D., Kazan, Russia
K. P. Markov	Ph.D., Assoc. Prof., Aizu, Japan
E. A. Martinez-Garcia	Ph.D., Professor, Chihuahua, Mexico
R. V. Meshcheryakov	Dr. Sci., Prof., Moscow, Russia
N. A. Moldovian	Dr. Sci., Prof., St. Petersburg, Russia
N. V. Hung	Ph.D., Ha Tinh, Viet Nam
V. V. Nikulin	Ph.D., Prof., New York, United States
V. Yu. Osipov	Dr. Sci., Prof., Deputy Editor-in-Chief, St. Petersburg, Russia
V. K. Pshikhov	Dr. Sci., Prof., Taganrog, Russia
J. Savage	Ph.D., Assoc. Prof., Mexico City, Mexico
H. Samani	Ph.D., Assoc. Prof., Plymouth, UK
M. Secujski	Ph.D., Assoc. Prof., Novi Sad, Serbia
A. V. Smirnov	Dr. Sci., Prof., St. Petersburg, Russia
B. V. Sokolov	Dr. Sci., Prof., St. Petersburg, Russia
K. H. Hsia	Ph.D., Professor, Douliou, Taiwan
M. M. Tatur	Dr. Sci., Prof., Minsk, Belarus
L. V. Utkin	Dr. Sci., Prof., St. Petersburg, Russia
M. N. Favorskaya	Dr. Sci., Prof., Krasnoyarsk, Russia
M. Fu	Ph.D., Prof., Hangzhou, China
A. D. Khomonenkov	Dr. Sci., Prof., St. Petersburg, Russia
L. B. Sheremetov	Dr. Sci., Mexico, Mexico

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

Том 25 № 5, 2026

Научный, научно-образовательный журнал с базовой специализацией в области информатики, автоматизации, робототехники, прикладной математики и междисциплинарных исследований.

Журнал основан в 2002 году

Периодичность

6 раз в год

Учредитель и издатель:

Федеральное государственное бюджетное учреждение науки
«Санкт-Петербургский Федеральный исследовательский центр Российской академии наук»
(СПб ФИЦ РАН)

Адрес редакции и издателя:

Россия, 199178, г. Санкт-Петербург, 14-я линия В.О., д. 39, лит. А
e-mail: ia@spcras.ru, сайт: <http://ia.spcras.ru>

Главный редактор:

А. Л. Ронжин, д-р техн. наук, проф., Санкт-Петербург, РФ

Выпускающий редактор: В. А. Ганичева

Переводчик: Я. Н. Березина

Художественный редактор: Ю. С. Черепанова

Журнал зарегистрирован Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций, свидетельство о регистрации (регистрационный номер) ПИ № ФС77-79228 от 25 сентября 2020 г.

Журнал индексируется в международной базе данных Scopus

Журнал входит в «Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертации на соискание ученой степени доктора и кандидата наук»

Журнал выпускается при научно-методическом руководстве Отделения нанотехнологий и информационных технологий Российской академии наук

© Федеральное государственное бюджетное учреждение науки
«Санкт-Петербургский Федеральный исследовательский центр
Российской академии наук», 2026

Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных в составе печатного периодического издания - журнала «ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ» статей по текущим экономическим, политическим, социальным и религиозным вопросам с обязательным указанием имени автора статьи и печатного периодического издания журнала «ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ»

Редакционный совет

А. А. Ашимов

академик Национальной академии наук Республики Казахстан,
д-р техн. наук, проф., Алматы, Казахстан

И. А. Каляев

академик РАН, д-р техн. наук, проф., Таганрог, РФ

А. И. Рудской

академик РАН, д-р техн. наук, проф., Санкт-Петербург, РФ

В. Сгурев

академик Болгарской академии наук, д-р техн. наук, проф., София,
Болгария

Б. Я. Советов

академик РАО, д-р техн. наук, проф., Санкт-Петербург, РФ

В. А. Сойфер

академик РАН, д-р техн. наук, проф., Самара, РФ

Редакционная коллегия

И. С. Азаров

д-р техн. наук, проф., Минск, Беларусь

М. В. Бобырь

д-р техн. наук, проф., Курск, РФ

О. Ю. Гусихин

д-р наук, Диаборн, США

В. Делич

д-р техн. наук, проф., Нови-Сад, Сербия

А. Б. Долгий

д-р наук, проф. Сент-Этьен, Франция

М. Железны

д-р наук, доцент, Пльзень, Чешская республика

Х. Кайя

д-р наук, доцент, Утрехт, Нидерланды

А. А. Карпов

д-р техн. наук, доцент, Санкт-Петербург, РФ

С. В. Кулешов

д-р техн. наук, Санкт-Петербург, РФ

Е. А. Магид

д-р наук, Казань, РФ

К. П. Марков

д-р наук, доцент, Аизу, Япония

Э. А. Мартинес-Гарсия

д-р наук, проф., Чиуауа, Мексика

Р. В. Мещеряков

д-р техн. наук, проф., Москва, РФ

Н. А. Молдован

д-р техн. наук, проф., Санкт-Петербург, РФ

В. Х. Нгуен

д-р наук, Хатинь, Вьетнам

В.В. Никулин

д-р наук, проф., Нью-Йорк, США

В.Ю. Осипов

д-р техн. наук, проф., зам. главного редактора, Санкт-Петербург, РФ

В. Х. Пшихопов

д-р техн. наук, проф., Таганрог, РФ

Х. К. Саваж

д-р техн. наук, доцент, Мехико, Мексика

Х. Самани

д-р наук, доцент, Плимут, Соединённое Королевство

М. Сечуйски

д-р техн. наук, доцент, Нови-Сад, Сербия

А. В. Смирнов

д-р техн. наук, проф., Санкт-Петербург, РФ

Б. В. Соколов

д-р техн. наук, проф., Санкт-Петербург, РФ

К. С. Ся

д-р наук, проф., Доулиу, Тайвань

М. М. Татур

д-р техн. наук, проф., Минск, Беларусь

Л. В. Уткин

д-р техн. наук, проф., Санкт-Петербург, РФ

М. Н. Фаворская

д-р техн. наук, проф., Красноярск, РФ

М. Фу

д-р наук, проф., Ханчжоу, Китай

А. Д. Хомоненко

д-р техн. наук, проф., Санкт-Петербург, РФ

Л. Б. Шереметов

д-р техн. наук, Мехико, Мексика

CONTENTS

Towards 70th Anniversary of Vasily Y. Osipov	1315
Towards 70th Anniversary of Alexander V. Smirnov	1318
Artificial Intelligence, Knowledge and Data Engineering	
A. Konstantinov, L. Utkin, S. Khomets A GUMBEL-SOFTMAX-BASED APPROACH FOR SELECTING IMPORTANT TOKENS IN SELF-ATTENTION MECHANISM	1322
A. Girin, N.y Teslya RECORD LINKAGE METHOD FOR ROAD TRAFFIC ACCIDENTS USING OFFICIAL DATA SOURCES AND NEWS PUBLICATIONS BASED ON SPATIO-TEMPORAL CLUSTERING	1356
S. Zateev, E. Kostyuchenko, A. Shelupanov COMPARATIVE ANALYSIS OF APPROACHES TO CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE DESIGN FOR KEYSTROKE DYNAMICS ANALYSIS	1385
A. Timofeeva, T. Avdeenko FEATURE SELECTION UNDER CLASS IMBALANCE: AN EMPIRICAL COMPARISON OF CORRELATION AND MUTUAL INFORMATION	1412
Y. Senkevich, M. Mishchenko SEARCH FOR SPECIFIC CHANGES IN GEOACOUSTIC EMISSION CHARACTERISTICS ASSOCIATED WITH STRONG EARTHQUAKES IN KAMCHATKA	1444
Information Security	
N. Miroshnikov, U. Tokareva, A. Shulzhenko A METHOD FOR DETECTING AI-TRANSFORMED SPEECH AUDIO RECORDINGS BASED ON ACOUSTIC AND STATISTICAL FEATURES	1476
E. Kozlova, D. Klimenka DETECTING ADVERSARIAL ATTACKS THROUGH ANOMALY IDENTIFICATION IN HIDDEN LAYERS OF A NEURAL NETWORK	1501
M. Firsova ONTOLOGY MODEL FOR SECURITY ANALYSIS OF INTERNET OF THINGS SYSTEMS	1520
D. Efanov, Y. Yelina DESIGN OF SELF-CHECKING DIGITAL DEVICES WITH IMPROVED TESTABILITY METRICS	1569
Mathematical Modeling and Applied Mathematics	
M.K. Do, V. Bogatyrev SEMI-MARKOV RELIABILITY MODEL OF A COMPUTER SYSTEM WITH CONTAINER VIRTUALIZATION	1604
S. Boev, A. Zakharov, A. Perlov, A. Tymoshenko A CLASSIFICATION MODEL FOR MOVING OBJECTS BASED ON RADAR PROFILE DATA FROM LARGE-APERTURE L-BAND RADARS, ACCOUNTING FOR STOCHASTIC NEGATIVE FACTORS OF VARIOUS NATURES	1640
S. Dvornikov, D. Vasilieva OIL SPILL IMAGE RECOGNITION BASED ON MULTIPLE-SCALE PROCESSING OF BRIGHTNESS DISTRIBUTION HISTOGRAMS	1670

СОДЕРЖАНИЕ

К 70-летию Юбилею Осипова Василия Юрьевича	1315
К 70-летию Юбилею Смирнова Александра Викторовича	1318
Искусственный интеллект, инженерия данных и знаний А.В. Константинов, Л.В. Уткин, С.П. Хоменц ПОДХОД НА ОСНОВЕ ОПЕРАЦИИ GUMBEL-SOFTMAX ДЛЯ ОТБОРА ЗНАЧИМЫХ ТОКЕНОВ В МЕХАНИЗМЕ САМОВНИМАНИЯ	1322
А.Р. Гирин, Н.Н. Тесля МЕТОД СОПОСТАВЛЕНИЯ ЗАПИСЕЙ О ДТП ИЗ ОФИЦИАЛЬНЫХ ИСТОЧНИКОВ И НОВОСТНЫХ ПУБЛИКАЦИЙ НА ОСНОВЕ ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ КЛАСТЕРИЗАЦИИ	1356
С.В. Затеев, Е.Ю. Косточенко, А.А. Шелупанов СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПОДХОДОВ К ПРОЕКТИРОВАНИЮ АРХИТЕКТУРЫ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА КЛАВИАТУРНОГО ПОЧЕРКА	1385
А.Ю. Тимофеева, Т.В. Авдеев ОТБОР ПРИЗНАКОВ В УСЛОВИЯХ НЕСБАЛАНСИРОВАННОСТИ КЛАССОВ: ЭМПИРИЧЕСКОЕ СРАВНЕНИЕ КОРРЕЛЯЦИИ И ВЗАИМНОЙ ИНФОРМАЦИИ	1412
Ю.И. Сенкевич, М.А. Мищенко МЕТОД ВЫДЕЛЕНИЯ ПРИЗНАКОВ СВЯЗИ СИГНАЛА ГЕОАКУСТИЧЕСКОЙ ЭМИССИИ С СИЛЬНЫМИ ЗЕМЛЕТРЯСЕНИЯМИ НА КАМЧАТКЕ	1444
Информационная безопасность Н.Ю. Мирошников, У.В. Токарева, А.Д. Шульженко СПОСОБ ОПРЕДЕЛЕНИЯ AI-ПРЕОБРАЗОВАННЫХ РЕЧЕВЫХ АУДИОЗАПИСЕЙ НА ОСНОВЕ АКУСТИЧЕСКИХ И СТАТИСТИЧЕСКИХ ПРИЗНАКОВ	1476
Е.И. Козлова, Д.М. Клименко ОБНАРУЖЕНИЕ СОСТЯЗАТЕЛЬНЫХ АТАК ЧЕРЕЗ ВЫЯВЛЕНИЕ АНОМАЛИЙ В СКРЫТЫХ СЛОЯХ НЕЙРОСЕТИ	1501
М.Г. Фирсова ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ ДЛЯ АНАЛИЗА ЗАЩИЩЕННОСТИ СИСТЕМ ИНТЕРНЕТА ВЕЩЕЙ	1520
Д.В. Ефанов, Е.И. Елина СИНТЕЗ САМОПРОВЕРЯЕМЫХ ЦИФРОВЫХ УСТРОЙСТВ С УЛУЧШЕННЫМИ ПОКАЗАТЕЛЯМИ КОНТРОЛЕПРИГОДНОСТИ	1569
Математическое моделирование и прикладная математика М.К. До, В.А. Богатырев ПОЛУМАРКОВСКАЯ МОДЕЛЬ НАДЕЖНОСТИ КОМПЬЮТЕРНОЙ СИСТЕМЫ С КОНТЕЙНЕРНОЙ ВИРТУАЛИЗАЦИЕЙ	1604
С.Ф. Боев, А.С. Захаров, А.Ю. Перлов, А.В. Тимошенко МОДЕЛЬ КЛАССИФИКАЦИИ ПОДВИЖНЫХ ОБЪЕКТОВ ПО ДАННЫМ РАДИОЛОКАЦИОННЫХ ПОРТРЕТОВ КРУПНОАПЕРТУРНЫХ РЛС L-ДИАПАЗОНА С УЧЕТОМ СТОХАСТИЧЕСКИХ НЕГАТИВНЫХ ФАКТОРОВ РАЗЛИЧНОЙ ПРИРОДЫ	1640
С.В. Дворников, Д.В. Васильева РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ РАЗЛИВОВ НЕФТЕПРОДУКТОВ НА ОСНОВЕ КРАТНОМАСШТАБНОЙ ОБРАБОТКИ ГИСТОГРАММ РАСПРЕДЕЛЕНИЯ ИХ ЯРКОСТИ	1670

К 70-ЛЕТНЕМУ ЮБИЛЕЮ ОСИПОВА ВАСИЛИЯ ЮРЬЕВИЧА



Доктор технических наук, профессор, директор СПИИРАН Осипов Василий Юрьевич, родился 25 октября 1956 г. в Новгородской области. В 1981 г. окончил Высшее военно-морское училище радиоэлектроники (ВВМУРЭ) им. А.С. Попова (н.в. ВМИРЭ им. А.С. Попова) по специальности радиотехнические средства с присвоением квалификации офицера с высшим военно-специальным образованием – военного радиоинженера. С 1981 по 1987 гг. проходил военную службу на Балтийском флоте. В 1987 г. поступил в адъюнктуру при ВВМУРЭ им. А.С. Попова (н.в. ВМИРЭ им. А.С. Попова), которую окончил, досрочно защитив кандидатскую диссертацию в 1990 г. Затем назначен на должность старшего преподавателя Калининградского высшего военно-морского училища (н.в. БВВМУ им. адмирала Ф.Ф. Ушакова), где проходил службу до конца 1991 г. В 1991 г. переведен в ВВМУРЭ им. А.С. Попова (н.в. ВМИРЭ им. А.С. Попова), где до 1997 г. был старшим преподавателем, доцентом и докторантом, а затем в 1997 г. назначен на должность начальника специальной кафедры, которой руководил до 2007 г. Докторскую диссертацию по техническим наукам на тему «Автоматизированный синтез программ специального назначения» защитил в 2000 г. Ученое звание профессора Осипову В.Ю. присвоено в 2001 г. Воинскую службу завершил в 2007 г. в воинском звании капитана 1 ранга. С 2007 по 2008 гг. работал профессором в ВМИРЭ

им. А.С. Попова. В 2008 г. принят на должность ведущего научного сотрудника в Санкт-Петербургском институте информатики и автоматизации Российской академии наук (СПИИРАН). В 2015 г. переведен на должность заведующего лабораторией информационно-вычислительных систем и технологий программирования СПИИРАН. В 2018 г. назначен на должность главного научного сотрудника, а в 2020 г. – на должность директора СПИИРАН.

Василий Юрьевич награжден медалями «300 лет Российскому флоту», «В память 300-летия Санкт-Петербурга», а также рядом ведомственных медалей. Является лауреат премии Правительства Санкт-Петербурга им. Г.А. Леонова за выдающиеся достижения в области кибернетики и искусственного интеллекта.

Осипов В.Ю. является автором более 170 научных публикаций, 40 из которых личные, включая статьи в отечественных и международных высокорейтинговых журналах, 5 монографий, 3 учебника; получил 8 патентов РФ на изобретения. Под руководством Василия Юрьевича подготовлены сотни офицеров с высшим военно-специальным образованием, 2 доктора наук и 8 кандидатов наук.

Ключевые научные результаты профессора Осипова В.Ю. относятся к следующим положениям: теория программного подавления электронно-вычислительных систем; оптимизация многошагового управления распределенными объектами; относительно конечные перестраиваемые операционные автоматы; автоматический дедуктивный синтез программ с циклами (включая самовоспроизводящиеся программы); методы многоуровневой интеллектуальной обработки информации в рекуррентных нейронных сетях [1 – 3]; модели рекуррентных нейронных сетей с управляемой ассоциативной обработкой сигналов, наделяемые прозрачными логическими одно и многоуровневыми линейными, спиральными, петлевыми и комбинированными структурами [4 – 6]; нейросетевые системы, ориентированные на решение задач распознавания, восстановления и прогнозирования разнородных событий с непрерывным обучением [7 – 15], и др.

Сотрудники СПб ФИЦ РАН, коллеги из многих организаций, его ученики и последователи, а также редакционная коллегия журнала «Информатика и автоматизация» (Труды СПИИРАН) поздравляют Осипова Василия Юрьевича с юбилеем и желают ему крепкого здоровья и дальнейших творческих успехов!

Список избранных публикаций

1. Осипов В.Ю. Устойчивость интеллектуальной обработки сигналов потоковыми рекуррентными нейронными сетями с непрерывным обучением. Информатика и автоматизация. 2025. Т. 24. №6. С. 1649–1682. DOI: 10.15622/ia.24.6.5.
2. Осипов В.Ю., Кулешов С.В., Милосердов Д.К., Зайцева А.А., Аксенов А.Ю. Рекуррентные нейронные сети с непрерывным обучением в задачах многофункциональной обработки новостных потоков. Информатика и автоматизация. 2022. Т. 21. №6. С. 1145–1168. DOI: 10.15622/ia.21.6.3.
3. Osipov V.Y., Osipova M. Space-time signal binding in recurrent neural networks with controlled elements. Neurocomputing. 2018. vol. 308. pp. 194–204. DOI: 10.1016/j.neucom.2018.05.009.
4. Осипов В.Ю. Рекуррентная нейронная сеть со структурой слоев в виде двойной спирали. Информационные технологии. 2014. №7. 56–60.
5. Осипов В.Ю. Метод управления синапсами в рекуррентной нейронной сети. Информационные технологии. 2013. №7. С. 61–66.
6. Осипов В.Ю. Устойчивость рекуррентных нейронных сетей с управляемыми синапсами. Информационные технологии. 2011. №9. С. 69–73.
7. Ogiy O.G., Osipov V.Y. Informative and Comprehensive Model for Managing the Fishery Complex Enterprises Labor Potential. Smart Innovation, Systems and Technologies. Springer. 2025. vol. 430. pp. 259–271. DOI: 10.1007/978-981-96-1206-2_20.
8. Осипов В.Ю. Интеллектуальная нейросетевая машина с функциями мышления. Информатика и автоматизация. 2024. Т. 23. №4. С. 1077–1109. DOI: 10.15622/ia.23.4.6.
9. Осипов В.Ю. Ассоциативно-пространственная адресация к памяти рекуррентных нейронных сетей. Информационные технологии. 2015. Т. 21. №8. С. 631–637.
10. Осипов В.Ю. Пределы памяти рекуррентных нейронных сетей со стиранием устаревшей информации. Научный вестник НГТУ. 2014. Т. 56. №3. С. 115–122.
11. Осипов В.Ю. Развитие процессов обработки информации в ассоциативных интеллектуальных машинах. Мехатроника, автоматизация, управление. 2014. №6. С. 8–13.
12. Осипов В.Ю. Аналоговые ассоциативные интеллектуальные системы. Труды СПИИРАН. 2013. №7(30). С. 141–155. DOI: 10.15622/sp.30.9.
13. Осипов В.Ю. Метод настройки ассоциативной интеллектуальной системы на входные сигналы. Информационные технологии. 2012. №9. С. 54–59.
14. Осипов В.Ю. Нейронная сеть с прошедшим, настоящим и будущим временем. Информационно-управляющие системы. 2011. №4. С. 30–33.
15. Осипов В.Ю. Ассоциативная интеллектуальная машина. Информационные технологии и вычислительные системы. 2010. №2. С. 59–67.

К 70-ЛЕТНЕМУ ЮБИЛЕЮ СМИРНОВА АЛЕКСАНДРА ВИКТОРОВИЧА



Доктор технических наук, профессор, заслуженный деятель науки Российской Федерации Смирнов Александр Викторович является выдающимся ученым, основателем нового направления «логистика знаний», включенных в список научных школ СПИИРАН (СПб ФИЦ РАН).

В 1979 г. Александр Викторович окончил Ленинградский политехнический институт имени М.И. Калинина (н.в. СПбПУ Петра Великого). В 1984 г. получил ученую степень кандидата технических наук по специальности «Техническая кибернетика и теория информации» в Санкт-Петербургском государственном электротехническом университете (СПбГЭТУ ЛЭТИ), а в 1994 г. получил ученую степень доктора технических наук по специальности «Применение вычислительной техники, математического моделирования и математических методов в научных исследованиях» в Санкт-Петербургском институте информатики и автоматизации Российской академии наук (СПИИРАН). Звание профессора Смирнову А.В. было присвоено в 1998 г. В настоящее время Александр Викторович является главным научным сотрудником Санкт-Петербургского Федерального исследовательского центра Российской академии наук (СПб ФИЦ РАН) и руководит лабораторией интегрированных систем автоматизации.

Профессор Смирнов А.В. удостоен государственных и ведомственных наград – медаль ордена «За заслуги перед Отечеством» II степени (1999 г.), почетное звание «Заслуженный деятель науки

Российской Федерации» (2007 г.), медаль «300 лет Российской Академии наук» (2024 г.).

Смирнов А.В. обладает обширным опытом в области исследования, разработки и проектирования комплексных компьютерных систем, математического моделирования, управления знаниями, поддержки принятия решений и управления гибкими цепочками поставок [1]. Александр Викторович работал над широким кругом проблем, таких как корпоративное управление знаниями, онтологически-ориентированная интеграция информации, управление контекстом, оперативная поддержка принятия решений, конфигурирование виртуальных предприятий и динамическая логистика [2, 3]. Полученные под его руководством результаты были неоднократно включены в список важнейших результатов РАН и позволили приступить к созданию персонализированных систем интеллектуальной поддержки принятия решений [4 – 6], которые имеют большое практическое значение для решения прикладных проблем как в области научных исследований и образования, так и в бизнесе и государственном управлении [7, 8]. Научные исследования Смирнова А.В. поддержаны многочисленными грантами ОНИТ РАН, РФФИ, РНФ и др. Основные результаты, полученные в ходе выполнения исследований, являются значимыми в контексте приоритетного направления «Н1. Переход к передовым технологиям проектирования и создания высокотехнологичной продукции, основанном на применении интеллектуальных производственных решений, роботизированных и высокопроизводительных вычислительных систем, новых материалов и химических соединений, результатов обработки больших объемов данных, технологий машинного обучения и искусственного интеллекта» Стратегии научно-технологического развития Российской Федерации (утвержденной Указом Президента РФ от 28 февраля 2024 г. № 145) [9, 10].

Профессор Смирнов А.В. признан международным научным сообществом в качестве эксперта в указанных областях исследований. В настоящее время является членом технического комитета IFAC по управлению производством (IFAC TC 5.1 on Manufacturing Plant Control), рабочей группы IFIP по управлению жизненным циклом изделий (IFIP TC WG5.1 on Global Product Development for the Whole Life-Cycle), технических комитетов IEEE по киберфизическим облачным системам (IEEE SMC TC on Cyber-Physical Cloud Systems) и по когнитивному ситуационному управлению (IEEE SMC TC on Cognitive Situation Management). Кроме того, Александр Викторович является почетным

членом Международной ассоциации «Институт систем и технологий информации, управления и коммуникаций» (Institute for Systems and Technologies of Information, Control and Communication) и членом Европейской академии по управлению производством (European Academy of Industrial Management).

Смирнов А.В. ведет активную педагогическую деятельность. В период руководства международной лабораторией «Интеллектуальные технологии для социо-киберфизических систем» (Университет ИТМО) профессор был одним из организаторов совместной магистерской программы по направлению «Бизнес информатика» Университета ИТМО и Университета г. Росток (Германия). Александр Викторович читал ряд учебных курсов в СПбПУ Петра Великого, СПбГЭТУ ЛЭТИ, а также в Университете г. Йенчепинг (Швеция). Под его руководством успешно защищено более 10 диссертаций, в том числе 3 на соискание ученой степени доктора технических наук. Является членом диссертационных советов СПб ФИЦ РАН и Университета ИТМО.

Смирнов А.В. является автором более 450 научных публикаций, среди которых более 10 монографий и учебных пособий. В настоящее время является членом редколлегии журналов: «Информационные технологии и вычислительные системы», «Труды ИСА РАН», «Искусственный интеллект и принятие решений» [4, 7, 10], «Информационно-управляющие системы» [6], «Информатика и автоматизация» [5, 11], «Journal of Intelligent Manufacturing», «International Journal of Multiagent and Grid Systems», «International Journal of Data Analysis Techniques and Strategies», «Management and Production Engineering Review», «International Journal of Product Lifecycle Management».

Сотрудники СПб ФИЦ РАН, коллеги из многих организаций, его ученики и последователи, а также редакционная коллегия журнала «Информатика и автоматизация» (Труды СПИИРАН) поздравляют Смирнова Александра Викторовича с юбилеем и желают ему крепкого здоровья и дальнейших творческих успехов!

Список избранных публикаций

1. Sandkuhl K., Shilov N., Smirnov A. Facilitating Digital Transformation: Success Factors and Multi-Aspect Ontologies. *International Journal of Integrated Supply Management*. 2020. vol. 13. no. 4. pp. 376–393. DOI: 10.1504/IJISM.2020.110739.
2. Sandkuhl K., Shilov N., Smirnov A., Seigerroth U. Multi-aspect Ontologies as Support for Quantified Products Lifecycle Management. *IFIP Advances in Information and Communication Technology*. Springer. 2025. vol. 740. pp. 15–29. DOI: 10.1007/978-3-031-93319-6_2.
3. Smirnov A., Levashova T., Ponomarev A., Shilov N. Methodology for Multi-Aspect Ontology Development: Ontology for Decision Support Based on Human-Machine Collective Intelligence. *IEEE Access*. 2021. vol. 9. pp. 135167–135185. DOI: 10.1109/ACCESS.2021.3116870.
4. Смирнов А.В., Пономарев А.В., Шилов Н.Г., Левашова Т.В., Агафонов А.А. Контекстно-зависимая поддержка принятия решений на основе дополненного интеллекта с использованием больших языковых моделей. Искусственный интеллект и принятие решений. 2025. №4. С. 3–17. DOI: 10.14357/20718594250401.
5. Смирнов А.В., Пономарев А.В., Шилов Н.Г., Левашова Т.В., Тесля Н.Н. Концепция построения коллаборативных систем поддержки принятия решений: подход и архитектура платформы. *Информатика и автоматизация*. 2024. Т. 23. №4. С. 1139–1172. DOI: 10.15622/ia.23.4.8.
6. Смирнов А.В., Левашова Т.В., Петров М.В. Базовый сценарий интеллектуальной поддержки принятия решений на основе моделей жизни пользователей в цифровой среде. Информационно-управляющие системы. 2021. №4(113). С. 47–60. DOI: 10.31799/1684-8853-2021-4-47-60.
7. Смирнов А.В., Левашова Т.В. Паттерны человеко-машинного сотрудничества в системах поддержки принятия решений. Искусственный интеллект и принятие решений. 2024. №2. С. 3–17. DOI: 10.14357/20718594240201.
8. Смирнов А.В., Левашова Т.В. Онтология паттернов человеко-машинного сотрудничества для поддержки принятия решений. Онтология проектирования. 2024. Т. 14. №3(53). С. 421–439. DOI: 10.18287/2223-9537-2024-14-3-421-439.
9. Smirnov A., Ponomarev A., Agafonov A. Ontology-Based Neuro-Symbolic AI: Effects on Prediction Quality and Explainability. *IEEE Access*. 2024. vol. 12. pp. 156609–156626. DOI: 10.1109/ACCESS.2024.3485185.
10. Смирнов А.В., Пономарев А.В., Шилов Н.Г., Левашова Т.В. Нейро-символический искусственный интеллект в коллаборативных системах поддержки принятия решений. Искусственный интеллект и принятие решений. 2022. №3. С. 36–50. DOI: 10.14357/20718594220303.
11. Шилов Н.Г., Пономарев А.В., Смирнов А.В. Анализ методов онтологически ориентированного нейро-символического интеллекта при коллаборативной поддержке принятия решений. *Информатика и автоматизация*. 2023. Т. 22. №3. С. 576–615. DOI: 10.15622/ia.22.3.4.

A.V. KONSTANTINOV, L.V. UTKIN, S.P. KHOMETZ
**A GUMBEL-SOFTMAX-BASED APPROACH FOR SELECTING
IMPORTANT TOKENS IN SELF-ATTENTION MECHANISM**

Konstantinov A.V., Utkin L.V., Khometz S.P. A Gumbel-Softmax-Based Approach for Selecting Important Tokens in Self-Attention Mechanism.

Abstract. The quadratic computational complexity of self-attention in transformers remains a significant bottleneck for efficiently processing long sequences. This paper introduces a plug-in method for reducing the computational cost of attention in an existing transformer backbone via differentiable mask learning, enabling dynamic, input-dependent token selection while preserving the model's ability to capture global contextual information. The proposed approach is designed for integration into an already trained transformer model with limited modification of the attention block. Rather than introducing another standalone efficient-attention architecture, the proposed approach is formulated as a plug-in token-selection module for an existing transformer backbone. It learns task-specific hard token-selection masks in a differentiable manner and is intended to reduce the effective attention cost of an already trained model at inference time. Our approach employs a Gumbel-Softmax-based mechanism to learn hard token-selection masks in a differentiable manner, offering a practical way to reduce inference time while preserving most of the baseline model quality in the setting considered. Experiments on four sentiment-analysis benchmarks with a DistilBERT backbone show that the proposed method achieves an 8–12% reduction in inference time while preserving most of the baseline quality in the setting considered. The learned masks show qualitatively interpretable selection patterns in the presented examples, often prioritizing sentiment-bearing tokens and supporting more efficient inference. The code implementing the model for experiments is publicly available at <https://github.com/NTAILab/GSBA-SIT-SAM>.

Keywords: transformer, self-attention, Gumbel-Softmax, token selection, differentiable mask.

1. Introduction. The transformer architecture [1, 2] has revolutionized natural language processing through its powerful self-attention mechanism, enabling models to capture long-range dependencies and contextual relationships in text. However, the quadratic computational complexity of the self-attention with respect to sequence length remains a fundamental limitation, particularly when processing long documents or deploying models in resource-constrained environments. While several approaches have attempted to address this challenge through sparse attention patterns or approximation methods, these solutions often compromise the model's ability to focus dynamically on the most relevant tokens for a given task. Existing approaches for reducing the quadratic complexity of self-attention can be broadly categorized into sparse attention methods, approximation-based approaches, and token pruning techniques. Comprehensive surveys of these directions can be found in [3–9]. These reviews present comprehensive studies of various algorithms, recent techniques and methods devised to extend the sequence length in transformers and LLMs.

In this work, we focus on reducing the effective computational cost of self-attention by dynamically selecting a subset of informative tokens on a per-input basis, while preserving the original pretrained backbone. The goal is not to propose another standalone efficient-transformer architecture, but to study a plug-in token-selection module that can be attached to an already trained model. Our method introduces selective token inclusion in the attention mechanism through differentiable mask learning, enabling the model to dynamically identify and prioritize the most informative tokens during both training and inference. Unlike static sparsity patterns or fixed window approaches, our technique preserves the transformer's ability to capture global dependencies while reducing the effective sequence length processed by each attention head.

It is important to distinguish the present setting from standalone efficient-transformer architectures such as Longformer, Linformer, Performer, or BigBird, which redesign the attention mechanism itself. In contrast, the present work studies a post-hoc token-selection module for an already trained backbone. For this reason, the main experimental comparison in this paper is performed against the same pretrained backbone without the proposed module, which isolates the effect of the token-selection mechanism itself.

The main contributions of our work are threefold:

1. **Differentiable Mask Learning:** we propose a Gumbel-Softmax-based mechanism for learning token-selection masks that enables hard selection of the most relevant tokens during forward passes while retaining differentiable training of the scoring module.

2. **Plug-in Token Selection for Existing Backbones:** our method is designed as a lightweight token-selection module that can be integrated into an already trained transformer backbone with limited modification of the attention block, offering a tunable trade-off between computational savings and predictive performance.

3. **Empirical Validation:** through experiments on several benchmark sentiment-analysis datasets (IMDb, SST-2, Rotten Tomatoes, and Yelp Polarity), we demonstrate that the proposed approach achieves an 8–12% reduction in inference time while preserving most of the baseline quality in the considered setting, outperforms random masking approaches, and exhibits stable training dynamics across different mask sizes and layer configurations.

Our results show particular advantages in processing long documents, where the method demonstrates strong performance on the Yelp Polarity dataset (average length 382 tokens) with only 50% of the tokens participating in attention computations. The learned masks show qualitatively interpretable

selection patterns in the presented examples, often prioritizing sentiment-bearing tokens while preserving important structural elements of the input.

The code implementing the model for experiments is publicly available at <https://github.com/NTAILab/GSBA-SIT-SAM>.

2. Related Work. Numerous approaches and methods have been developed to reduce the computational complexity of the attention mechanisms in transformers. In particular, Longformer [10] scales linearly with sequence length, making it easy to process documents of thousands of tokens or longer. Another model [11] uses sparse factorization of the attention matrix to simplify the attention. Linformer [12] applies the self-attention approximated by a low-rank matrix. Performer [13] is based on approximation of the softmax attention-kernels by using the Fast Attention Via positive Orthogonal Random features approach. Transformer-XL [14] consists of a segment-level recurrence mechanism. The Compressive Transformer [15] is an attentive sequence model which compresses past memories for long-range sequence learning. BigBird [16] applies a sparse attention mechanism that reduces this quadratic dependency to linear. Reformer [17] is based on using two techniques to improve the efficiency of transformers. The first replaces the dot-product attention with the locality-sensitive hashing. The second uses reversible residual layers instead of the standard residuals. The Token Pruning model is presented in [18], which introduces a novel token reduction method dubbed Learned Token Pruning which adaptively removes unimportant tokens. Two key challenges of standard transformer architectures, including the scaling input length and encoding structured inputs are considered in the ETC model [19]. Interesting solutions for reducing the transformer complexity are provided in [20–23]. In [24], a transformer architecture is proposed that combines the efficient parallelizable training of transformers with the efficient inference of RNNs. Popular approaches to make transformers faster and lighter are reviewed in [3]. Macformer is proposed in [25]. It employs random Maclaurin features to approximate various dot-product kernels. Linear transformers under conditions of corrupted training data and noise are studied in [26]. The self-attention expressed as a linear dot-product of kernel feature maps by using the associativity property of matrix products is proposed in [27]. The linear transformer called cosFormer is presented in [28]. A binarization paradigm customized to high-dimensional softmax attention via kernelized hashing and called EcoFormer is introduced in [29].

The above represents only a very small fraction of all currently proposed or existing approaches, algorithms, and models for reducing the computational complexity of transformers. Existing efficient-transformer methods reduce attention cost at different levels: some redesign the attention

pattern or kernel approximation, while others progressively prune tokens as representations pass through the layers. The present work is closer to learned token-selection methods, but differs in its focus on a fixed-budget key/value token-selection module that can be attached to an already trained backbone and then executed at inference time through actual gather-based pruning. In contrast to sparse and low-rank attention approximations, our method does not modify the attention formulation itself and can be applied as a post-hoc optimization to pretrained models.

3. Background: Randomized Baselines and Key/Value Selection.

Rather than replacing self-attention with an alternative mechanism of lower asymptotic complexity, we investigate complementary strategies that reduce the *effective* computational cost of attention by restricting the set of key/value tokens involved in the computation. This perspective offers two practical advantages. First, it enables converting existing pretrained transformer models into more efficient ones by replacing only the attention layers while reusing the original parameters; the resulting adaptation typically requires lightweight fine-tuning and is substantially cheaper than training a new architecture from scratch. Second, token-selection methods can be combined with orthogonal acceleration techniques (e.g., Flash Attention), yielding additive speedups.

Our analysis adopts the standard premise that, for many inputs, the contextual representation of a token can be recovered accurately by interacting with only a limited subset of informative “neighbors.”

3.1. Randomized baselines. We first consider a fully randomized approach that constructs subsets of keys (and the corresponding values) uniformly at random. This baseline is conceptually simple and exposes a fundamental limitation: under realistic compute budgets, random sampling is unlikely to include all task-critical tokens when their number is non-negligible.

3.2. From random sampling to learned key/value selection. The limitations of the randomized baseline motivate a second family of approaches, where the subset of key/value tokens is selected in a data-dependent manner based on the layer-wise token embeddings. The remainder of this section formalizes the randomized baseline and quantifies when it is expected to fail.

The first direction reduces complexity by approximating full attention using only a subset of keys and values and, if necessary, improving accuracy by repeating the procedure multiple times. Assume self-attention is applied to a sequence of length n . The main hyperparameter is the subset size k (the number of selected key/value tokens). We sample t subsets

of size k , compute attention for each subset, and then aggregate the results (e.g., by averaging). As t increases, the approximation quality improves but eventually saturates.

If the base attention has quadratic complexity $O(n^2)$, then computing attention with n queries and only k keys/values has complexity $O(nk)$ per repetition; after t repetitions, the total cost becomes $O(nkt)$. Thus, to obtain a computational advantage, one must satisfy:

$$kt < n. \quad (1)$$

If the sequence contains a absolutely relevant tokens that need to be considered collectively for each query token, then the probability of selecting all such tokens when taking a subset of size k is given by the hypergeometric distribution and equal to $\binom{n-a}{k-a} / \binom{n}{k}$. Then the probability of obtaining at least one subset where the result of applying the attention mechanism will be correct (using all a absolutely relevant tokens) becomes:

$$p = 1 - \left(1 - \frac{\binom{n-a}{k-a}}{\binom{n}{k}} \right)^t. \quad (2)$$

We consider a commonly used training sequence length of $n = 1024$, and compare probabilities of obtaining at least one correct subset (which will significantly affect the upper bound of the model's accuracy) for various parameter values: $k \in \{64, 128, 256\}$, $a \in \{1, 2\}$. These dependencies are shown in Fig. 1, where the red vertical line corresponds to the number of repetitions t beyond which computational efficiency superiority compared to the classical mechanism will be lost.

The results indicate that, under the condition of improved efficiency relative to the base attention mechanism, the probability of randomly selecting the necessary tokens (even for the minimal number of them) turns out to be substantially less than one. Fig. 2 shows the dependency of the probability on the number of absolutely relevant tokens a with the parameter $t = n/k$, i.e., in the case where parity in terms of efficiency arises between the considered model and the base attention mechanism.

The number of absolutely relevant tokens extremely significantly influences the probability of success, and for $a \geq 5$ the probability becomes extremely low for all considered subset sizes.

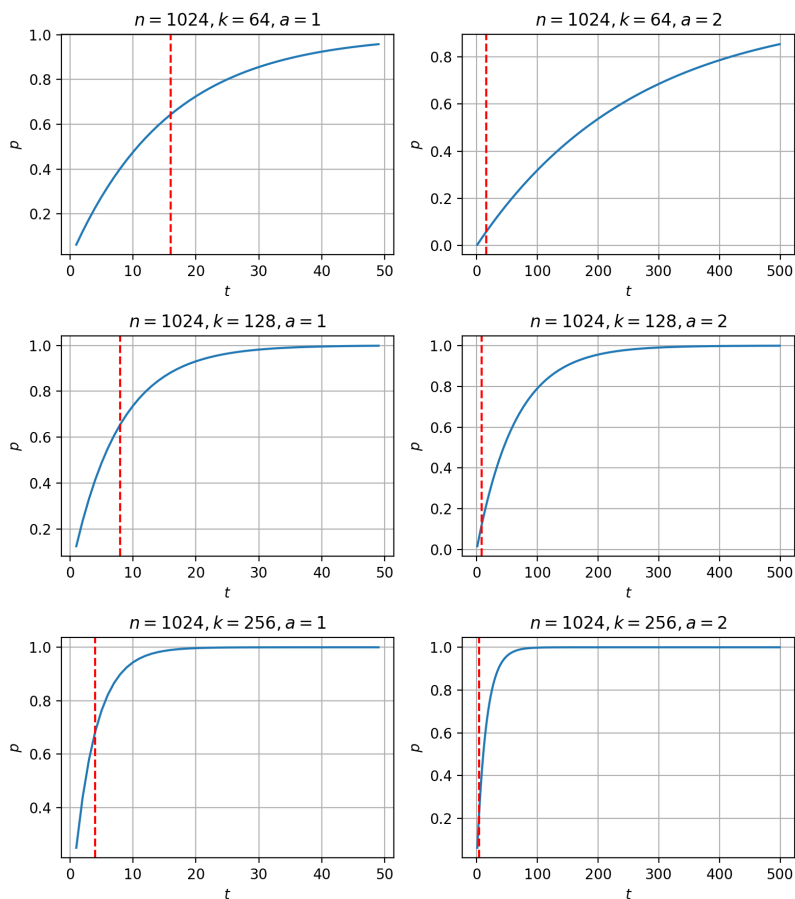


Fig. 1. Dependencies of the probability of a correct subset appearing

These results indicate that purely random subset selection becomes unreliable once the number of absolutely relevant tokens is moderate (e.g., $a \geq 5$): under realistic efficiency constraints, the probability of sampling a subset that contains all required tokens drops to a critically low level. Consequently, random sampling alone is unlikely to yield a method

that is simultaneously faster and accurate when correct attention behavior requires jointly considering many essential tokens.

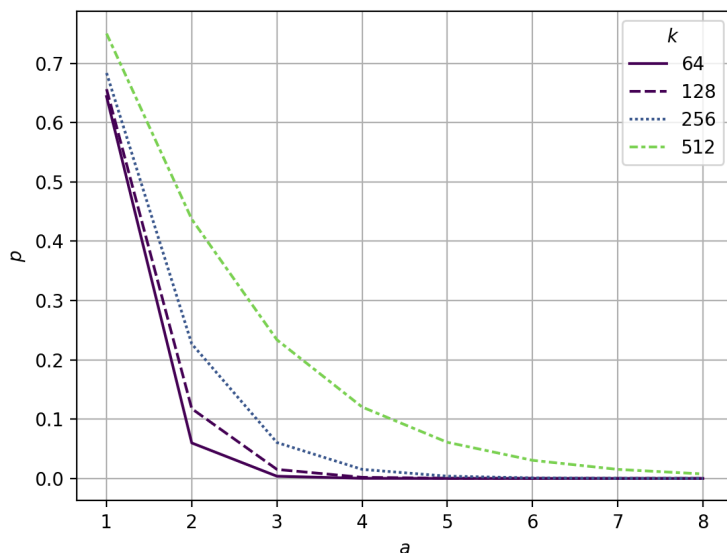


Fig. 2. Dependence of the probability of a correct subset at comparable efficiency

At the same time, the effective value of a is task- and layer-dependent. Therefore, we retain randomized sampling as a simple and informative baseline and use it for comparison throughout the experimental section.

Implication. The above analysis motivates replacing uniform random sampling with a *learned, data-dependent* key/value selection mechanism. In what follows, we introduce a differentiable token-selection method that constructs a single binary mask per layer (shared across queries), so that keys/values can be physically pruned at inference time while preserving end-to-end trainability.

4. Method: Differentiable Key/Value Token Selection.

4.1. Mask learning problem. To determine the most relevant tokens in a layer and subsequently build a model trainable end-to-end, various strategies were considered. A natural idea is to use the query vector itself to determine which keys should be considered for its processing. However, in this case, a separate mask is determined for each query, its values indicating whether a key vector is used for this query vector. This leads to the formation of a quadratic number of masks, one mask per query token, making significant

memory or computation savings impossible. Therefore, another approach was proposed.

An alternative is constructing a single mask for all query vectors. Such a mask does not depend on queries but relies on key vectors, and can be implemented, for example, as a random mask or a position-dependent mask (so-called padding mask). To avoid quadratic complexity in mask generation and simultaneously make the mask dependent on keys, an auxiliary model based on a small neural network is introduced. It works independently for each token embedding and produces a scalar importance score. Then these scores are used to construct a single binary mask, which is applied to all query vectors in the layer.

Let's consider the formulation of the subproblems of constructing and applying the mask based on key vectors. Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be the set of input embeddings of dimensionality d . The standard self-attention mechanism computes the attention matrix $A \in \mathbb{R}^{n \times n}$, which has a computational complexity $O(n^2)$. The goal of the developed method is to approximate full attention with a sparse analogue with a fixed computation budget $k \ll n$. The task of selecting an ordered set of indices $J \subset \{1, \dots, n\}$ of cardinality $|J| = k$ can be formulated as an optimization problem of finding a binary mask $M \in \{0, 1\}^n$, where $\sum_{i=1}^n M_i = k$. The following theoretically possible approaches to forming such a mask were analyzed:

1. *“Soft” selection of token embeddings (so-called “Soft Gating”).* The most obvious approach is to use a sigmoid activation to weight token importance: $\tilde{\mathbf{x}}_i = \sigma(f(\mathbf{x}_i)) \cdot \mathbf{x}_i$. The token embedding can be automatically reduced, which affects its subsequent participation in the attention mechanism (small attention weights, and consequently, small contribution to the result). However, this approach preserves the computational complexity $O(n^2)$, because these tokens are still included in the sequence and participate in the computation of the attention matrix A , despite making no actual contribution to the result. Moreover, the sigmoid function does not guarantee obtaining exactly k significant tokens, and the distribution of embeddings for the next layer may differ from the distribution during training, which negatively affects model accuracy.

2. *Stochastic Bernoulli sampling and reinforcement learning.* Building a mask is possible through independent sampling of each mask component from a Bernoulli distribution $m_i \sim \text{Bernoulli}(p_i)$, where the probability p_i is determined based on an auxiliary model. This approach has two drawbacks. First, it does not guarantee a fixed subset size k , which creates technical problems during batch processing since subsets of key matrices for different

training examples must have the same size (or an upper bound, in this case less than n). However, in this case, the dimensionality is dynamic, which not only complicates computations but requires introducing penalty components into the loss function; otherwise, the auxiliary model could be optimized such that all mask components have a probability of 1, meaning computational efficiency would not be improved. Second, the gradient estimate of the loss function with respect to parameters in this case is characterized by high variance, making training unstable (e.g., when using the REINFORCE algorithm).

3. *l_1 -regularization of attention weights (LASSO method).* Adding a penalty on the l_1 -norm of attention weights to encourage sparsity can lead to zero attention weights for many keys. However, from the perspective of neural network training, this constraint (on zero weights) is “soft”, meaning it is not guaranteed that sparsity will arise at the prediction stage. Furthermore, to determine which attention weights are close to zero, one first needs to fully compute the matrix A . Additionally, this approach does not guarantee obtaining exactly k non-zero elements and requires complex hyperparameter tuning of the penalty to balance sparsity and quality. Forming a mask based on near-zero elements requires applying a threshold function at the prediction stage, which significantly differs from computing the attention mechanism without a threshold function during training. Consequently, the distribution of embeddings for the next layer may differ from the distribution during training, which also negatively impacts model accuracy. Thus, using l_1 -regularization can be considered as an auxiliary penalty but is not a suitable solution for building token masks.

4. *Defining the derivative of the top- k token-selection operation through the extended “Gumbel-Softmax” mechanism.* To overcome the described limitations, a method was chosen that combines determinism of the output set size (selecting exactly k tokens with the highest scores) with the possibility of end-to-end differentiable training. Instead of independent evaluation of each token, a categorical distribution over all tokens in the sequence is introduced. This allows the selection task to be considered as sampling k elements without replacement, where the selection probability depends on the computed token significance score.

The fourth option was chosen as the main approach, namely defining the derivative of the top- k token-selection operation through the extended “Gumbel-Softmax” mechanism. It is important to note that during training and prediction, the result of this operation does not differ, and consequently does not reduce accuracy due to distribution shift for subsequent layers at the prediction stage.

4.2. Differentiable top- k selection via Gumbel-Softmax.

An auxiliary scoring network is introduced, which for each token $\mathbf{x}_i$ predicts an unnormalized logit μ_i . However, direct use of the $\text{argtop-}k$ operation is non-differentiable and does not allow training the scoring network via gradient descent. To estimate the gradient, a continuous relaxation via the Jacobian of the softmax operation is used.

During the forward pass at the *prediction (inference) stage*, a “hard” selection of indices is applied to form the ordered set J , which allows the physical discarding of unused tokens and reducing the complexity of attention computation to $O(n \cdot k)$. For the backward pass during training, a continuous approximation is used.

Let $\mathbf{m}$ be the mask of selected tokens obtained via the $\text{argtop-}k$ operation. This mask can be approximated using a probability vector π (corresponding to a discrete probability distribution over tokens), obtained via softmax with the temperature τ :

$$\pi_i = \text{GumbelSoftmax}_i(\mu) = \frac{\exp((\mu_i + g_i)/\tau)}{\sum_{j=1}^n \exp((\mu_j + g_j)/\tau)}, \quad (3)$$

where g_i is noise from a Gumbel distribution.

As $\tau \rightarrow 0$, the distribution π tends to a degenerate (discrete one-hot) distribution, and at high values of τ , to a uniform one. This property allows during training to smoothly transition from a near-uniform distribution at high temperature to a hard distribution at low temperature, while preserving the ability to compute the gradient of the loss function with respect to the scoring network parameters $\nabla_{\theta} L$.

Note that the classical Gumbel-Softmax approach allows randomly selecting an element in accordance with a given probability distribution by determining the index of the element with the maximum probability value: $\text{argmax } \pi_i$. To form the mask $\mathbf{m}$, it is necessary to select k elements without replacement, so instead of argmax , $\text{argtop-}k$ is used. The approximation of the mask $\mathbf{m}$ by the probability vector π can be done at best up to a multiplicative constant k , since $\sum_i \pi_i = 1$, $\sum_i m_i = k$. The temperature should not be too close to zero, because otherwise only one component of the vector π will be non-zero. With a correct choice of the τ parameter for estimating partial derivatives of the mask components m_i with respect to logits μ_j , it is valid to use the derivatives of the GumbelSoftmax function.

Thus, the proposed architecture solves three key problems:

1. Ensuring constant performance (due to fixed subset size). Exactly k tokens are always selected, ensuring predictable memory consumption and computation time.

2. Training the model with gradient descent. Using Gumbel-Softmax allows training the masking parameters jointly with the main model in an end-to-end manner.

3. Linear complexity of mask generation. The importance evaluation of key tokens occurs once for the entire attention layer in $O(n)$, reducing the overall computational complexity to $O(n(k+1))$.

4.3. Token-Importance Scoring Network. A key component of the architecture is the mechanism mapping each token embedding $\mathbf{x}_i \in \mathbb{R}^d$ to a scalar score of its significance, called the scoring network ϕ_θ . For parameterizing it, the following options were considered: self-attention mechanism, convolutional neural networks, fully connected neural networks with shared parameters.

The self-attention mechanism can be used as the scoring network ϕ_θ as follows. Each token $\mathbf{x}_i$ is assigned a number characterizing the relevance of this token for processing the entire sequence. Accordingly, vectors of dimension 1 are used as value vectors, which after linear combination with attention weights (in the ϕ_θ model) form logits of relevance scores. This option, besides the current token $\mathbf{x}_i$, considers the context of the entire sequence and is the most general option, but requires computing an attention matrix of size $n \times n$, leading back to the original quadratic complexity $O(n^2)$ and negating the performance gain. It is worth noting that for evaluating the absolute relevance of tokens, a simplified attention mechanism can be used: with a lower number of parameters or a linearized attention mechanism. However, the computation time of the scoring function ϕ_θ in that case would be comparable to computing the mechanism whose efficiency we aim to improve, negatively impacting performance.

The second option, convolutional neural networks, can allow considering local context (neighboring tokens) through the application of a one-dimensional convolution operation. This option has linear complexity $O(n \cdot w)$, where w is the window width, but has two significant drawbacks. First, applying convolutions, while allowing context consideration to some degree, has a fixed window width, which can lead to significant differences in such scores with minor permutations of sequence elements (where an element moves out of the window), complicating training and increasing the risk of overfitting. Second, applying convolution significantly limits the applicability to tasks with a constant sampling step. In other words, the convolution operation

is only applicable for analyzing sequences where the position of each element is its index, whereas the attention mechanism is also applied to more general tasks, such as analyzing time series with a variable sampling step. Positional encoding can define the spatial arrangement of objects, which is not directly related to token index, making the application of convolution impossible in this case.

The third option consists of using fully connected neural networks with shared parameters to determine the relevance score logit for each token individually. Namely, it was proposed to use a fully connected network that processes each token independently:

$$\mu_i = \phi_\theta(\mathbf{x}_i), \quad (4)$$

where ϕ_θ is a classical fully connected neural network with a scalar output whose parameters do not depend on the token index.

Note that since in a transformer containing multiple attention layers, the vector $\mathbf{x}_i$ at layer l is built considering context, i.e., aggregates information from other tokens at previous layers $1, \dots, l-1$, additional context consideration when applying the scoring network is not necessary. At the same time, an element-wise applied small fully connected neural network has minimal computational complexity, allowing mask generation with insignificant overhead.

Also, the neural network does not require learning an additive constant on the output, because the obtained logits $\mu = (\mu_1, \dots, \mu_n)$ are subsequently transformed into a discrete probability distribution π via Gumbel-Softmax, which is invariant to shift.

4.4. Token-Reduced Self-Attention. After forming the ordered set of indices J (where $|J| = k$), the attention operation must be performed. The standard formula for the attention mechanism is defined as [1, 2]:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V. \quad (5)$$

In a naive implementation, simply zeroing (masking) unselected elements by replacing corresponding components in the QK^T matrix with $-\infty$ does not lead to acceleration, because the matrix multiplication is still performed for matrices with dimensions $(n \times d), (d \times n)$. To achieve real acceleration, the proposed method applies an index gathering operation

(Gather). Let W_Q, W_K, W_V be the projection matrices of the attention layer (without loss of generality, single-head attention is considered). Full queries $Q = XW_Q$ are computed, but keys and values are computed only for the subset of indices J :

$$\hat{K} = \text{Gather}(X, J)W_K \in \mathbb{R}^{k \times d}, \quad \hat{V} = \text{Gather}(X, J)W_V \in \mathbb{R}^{k \times d}, \quad (6)$$

where the index gathering operation *Gather* is defined as concatenation of row vectors from the processed matrix with indices from the ordered set J :

$$\text{Gather}(X, J) = (\mathbf{x}_i)_{i \in J} = \begin{pmatrix} \mathbf{x}_{J_1} \\ \vdots \\ \mathbf{x}_{J_k} \end{pmatrix}. \quad (7)$$

The attention scores matrix is computed as:

$$\hat{S} = \frac{Q\hat{K}^T}{\sqrt{d}} \in \mathbb{R}^{n \times k}. \quad (8)$$

Note the dimensionality of the result: instead of a matrix of size $n \times n$, a matrix of size $n \times k$ is obtained. Under the condition $k \ll n$, the number of floating point operations (FLOPs) is reduced proportionally to the compression ratio $r = k/n$. Attention weights are calculated via softmax along the k dimension:

$$\hat{\alpha}_{ij} = \frac{\exp(S_{ij})}{\sum_{m=1}^k \exp(S_{im})}, \quad \text{where } j \in \{1, \dots, k\}. \quad (9)$$

The final result of the proposed attention mechanism is computed analogously to the classical mechanism:

$$\tilde{X} = \hat{A}\hat{V} \in \mathbb{R}^{n \times d}. \quad (10)$$

Thus, the computational complexity of the attention mechanism is reduced from quadratic $O(n^2d)$ to linear relative to the sequence length: $O(nkd)$.

4.5. Overall Model Architecture. The structure of the modified transformer layer *during training* is illustrated in Fig. 3. The left part of the scheme shows the modified attention block, where attention weights are sparsified by a hard mask and then renormalized; the right part shows the auxiliary block that generates this mask (scoring network followed by the Gumbel-Softmax/ $\text{argtop-}k$ procedure).

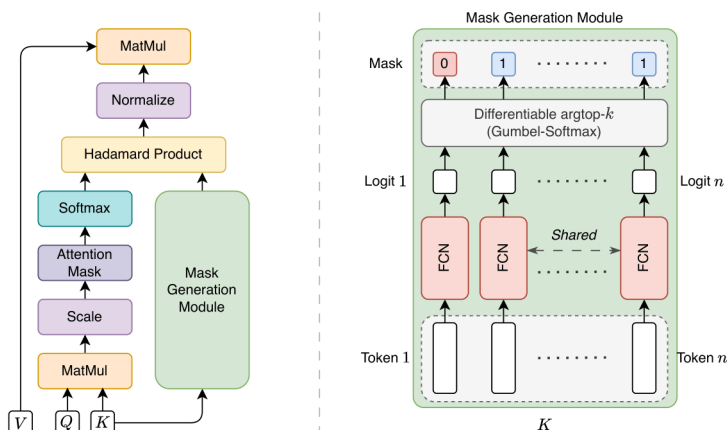


Fig. 3. Integration of the proposed model into the transformer architecture

In the mask-generation block (right part of Fig. 3), the input tokens are passed independently through a neural network, resulting in a logit for each token characterizing its significance. A softmax operation is applied to the logits to obtain a probability distribution, based on which a set of token indices (without replacement) of size k is generated.

At the *prediction stage*, the forward pass of each modified transformer layer is summarized in Algorithm 1.

Thus, the asymptotic computational complexity of the entire model at the prediction stage is reduced from $O(n^2)$ to $O(nk)$.

Although the proposed method reduces the attention computation from $O(n^2)$ to $O(nk)$ in the selected part of the computation, the observed end-to-end speedup is more moderate because attention is only one component of total inference time. The measured runtime also includes the overhead of mask generation, gather operations, and the remaining transformer computations outside the attention score matrix.

Algorithm 1 Prediction-stage forward pass of a modified transformer layer

Require: Token embeddings $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, subset size k , projections W_Q, W_K, W_V , scoring network ϕ_θ

Ensure: Output embeddings $\tilde{X} \in \mathbb{R}^{n \times d}$

- 1: $\mu \leftarrow (\phi_\theta(\mathbf{x}_i))_{i=1}^n$ ▷ Compute logits (token importance scores)
 - 2: $J \leftarrow \text{argtop}_k(\mu)$ ▷ Select keys/values and compute token-reduced attention
 - 3: $Q \leftarrow XW_Q$
 - 4: $\hat{K} \leftarrow \text{Gather}(X, J)W_K$
 - 5: $\hat{V} \leftarrow \text{Gather}(X, J)W_V$
 - 6: $\hat{S} \leftarrow \frac{Q\hat{K}^T}{\sqrt{d}}$
 - 7: $\hat{A} \leftarrow \text{softmax}(\hat{S})$ ▷ Row-wise over k keys
 - 8: **return** $\hat{A}\hat{V}$
-

To solve this problem, during training we compute the attention weights using all n key-value pairs (thus keeping the standard $O(n^2)$ computation for this step), and apply the hard mask $\mathbf{m}$ to the attention weights. Specifically, based on the discrete probability distribution π computed via Gumbel-Softmax, a “hard” mask $\mathbf{m} = (m_1, \dots, m_n)$ is constructed such that $m_i = 1 \iff i \in J$. Then the attention weights of the base mechanism are sparsified by element-wise multiplication with the mask followed by renormalization, so that non-relevant tokens receive strictly zero weights:

$$\tilde{\alpha}_{ij} = \frac{\alpha_{ij} \cdot m_j}{\sum_{g=1}^n \alpha_{ig} \cdot m_g}, \quad \tilde{A} = (\tilde{\alpha}_{ij}) \in \mathbb{R}^{n \times n}. \quad (11)$$

The result after multiplication by the value matrix will be completely equivalent to the result with the $\text{Gather}(X, J)$ operation:

$$\tilde{A}V = \hat{A}\hat{V} \in \mathbb{R}^{n \times d}, \quad (12)$$

since for each individual row it holds that:

$$\tilde{\alpha}_i V = \sum_{j=1}^n \frac{\alpha_{ij} \cdot m_j}{\sum_{g=1}^n \alpha_{ig} \cdot m_g} v_j = \sum_{j \in J} \hat{\alpha}_{ij} v_j = \hat{\alpha}_i \hat{V}. \quad (13)$$

Thus, this mechanism is an alternative implementation of the scheme used at the prediction stage, but the result functionally depends on the vector $\mathbf{m}$. Consequently, using automatic differentiation, derivatives with respect to the components of vector $\mathbf{m}$ can be calculated, which in turn depends on Gumbel-Softmax that takes the logit vector μ as input, depending on the parameters of the scoring network θ . Using the previously discussed method for defining the derivative of the operation generating the mask $\mathbf{m}$ and the automatic differentiation algorithm, partial derivatives of the loss function with respect to parameters θ are calculated.

At the *training stage*, the forward pass of each modified transformer layer is summarized in Algorithm 2.

During the backward pass at the training stage, all the same steps are used (in reverse order), except for step 2, where instead of the derivatives of the *argtop- k* operation, derivatives of the Gumbel-Softmax function with the same Gumbel noise as in the forward pass are used.

There is a complete correspondence between the presented training and prediction schemes from the perspective of computation results, due to the equivalence shown above of the attention mechanisms based on the non-differentiable operation $\text{Gather}(X, J)$ and the differentiable multiplication of attention weights by the mask (whose derivatives with respect to parameters are defined using the Gumbel-Softmax technique). Consequently, when applying the developed method to new data from the same distribution as the training data, the model's accuracy will at least not decrease due to a mismatch between training and prediction operations. At the same time, from a computational point of view, the proposed training scheme is less efficient. However, it is important to note the properties of this scheme that allow for more efficient training. First, for layers where the training of the auxiliary networks can be considered complete (no significant parameter updates occur), the forward pass procedure of the training stage can be replaced by the non-trainable but equivalent procedure for the prediction stage. Second, the proposed scheme with token relevance scoring does not explicitly take positional information into account, making it possible to train auxiliary networks only on short sequences where quadratic computational complexity is not critical, and to apply the efficient non-trainable procedure on long sequences.

Algorithm 2 Training-stage forward pass of a modified transformer layer

Require: Token embeddings $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, subset size k , temperature τ , projections W_Q, W_K, W_V , scoring network ϕ_θ

Ensure: Output embeddings $\tilde{X} \in \mathbb{R}^{n \times d}$

- 1: $\boldsymbol{\mu} \leftarrow (\phi_\theta(\mathbf{x}_i))_{i=1}^n$ ▷ Compute logits (token importance scores)
 - 2: $\mathbf{g} \sim \text{Gumbel}_n(0, 1)$
 - 3: $\boldsymbol{\pi} \leftarrow \text{softmax}((\boldsymbol{\mu} + \mathbf{g})/\tau)$
 - 4: $J \leftarrow \text{argtop}_k(\boldsymbol{\mu})$
 - 5: $\mathbf{m} \leftarrow (\mathbb{I}[j \in J])_{j=1}^n + \boldsymbol{\pi}.\text{detach}() - \boldsymbol{\pi}$ ▷ Backward pass uses the Jacobian of $\boldsymbol{\pi}$
 - 6: $Q \leftarrow XW_Q, K \leftarrow XW_K, V \leftarrow XW_V$ ▷ Any attention, e.g. SDPA
 - 7: $S \leftarrow \frac{QK^T}{\sqrt{d}}$
 - 8: $A \leftarrow \text{softmax}(S)$ ▷ Row-wise over n keys
 - 9: **for** $i \leftarrow 1$ **to** n **do**
 - 10: $\tilde{\boldsymbol{\alpha}}_i \leftarrow \frac{\boldsymbol{\alpha}_i \odot \mathbf{m}}{(\boldsymbol{\alpha}_i \odot \mathbf{m})^T \mathbf{1}}$ ▷ Apply mask and renormalize
 - 11: **end for**
 - 12: **return** $\tilde{A}V$
-

Thus, the developed method significantly reduces the asymptotic computational complexity of the algorithm from $O(n^2)$ to $O(nk)$, where k is a tunable hyperparameter, for an arbitrary base attention mechanism, by selecting the most relevant tokens. The selection is performed using a trainable small scoring neural network, the application of which has complexity $O(n)$ and does not require significant additional computation. This approach can be applied to improve efficiency both at the prediction stage and at the training stage, provided that suitable parameters are defined for the scoring neural network.

5. Experiments. In numerical experiments, our hypotheses are validated by considering the DistilBERT model [30] and its modified variants, including the addition of random and trainable masks to the attention mechanisms. In contrast to trainable masks derived in accordance with the proposed method, the random mask is based on the uniform random selection of k tokens from n ones, i.e., subset J is randomly formed. The random mask model does not require additional training time, since it does not compute the probability distribution $\boldsymbol{\pi}$ for selecting k tokens. It is used in experiments in order to illustrate a difference between models and outperformance of the proposed model.

The base model is a pre-trained DistilBERT¹, fine-tuned on a selected dataset for classification.

For the trainable mask, the overall training procedure consists of two stages. First, the base transformer is fine-tuned on the selected downstream dataset. Second, the parameters of the fine-tuned transformer are frozen, and further optimization is performed exclusively for the lightweight fully connected scoring network. For the random mask, no additional model training is conducted. Therefore, the additional training cost introduced by the proposed method is limited to the second stage rather than requiring retraining of the full backbone.

For reproducibility, we note that the experiments follow this two-stage training procedure throughout. The Adam optimizer with an initial learning rate of 10^{-5} and a linear scheduler is used for the second stage, and the Gumbel-Softmax temperature parameter is set to 10. The reported results correspond to the stated experimental setup without repeated-run variance estimates.

To compare the proposed models, four datasets from the sentiment analysis domain are selected. All datasets are obtained using the Hugging Face library.

The *IMDb* dataset [31] contains data for movie review classification (positive/negative). It includes 25,000 examples for training and 25,000 for testing. The average sentence length is 230 tokens. The dataset has an equal number of examples for each class.

The *SST-2* dataset [32] contains data for sentiment analysis of sentences (positive/negative sentiment) based on the dependency tree annotations. It includes both complete sentences and their parts. There are 67,349 examples for training and 1,821 for testing. The average sentence length is 10-20 tokens.

The *Rotten Tomatoes* dataset [33] contains data for classifying movie reviews (positive/negative). It differs from IMDb in its more “professional” evaluations, use of cinema-specific vocabulary, and complex mixed reviews. There are 8,530 examples for training and 1,066 for testing. The average sentence length is 20-30 tokens.

The *Yelp Polarity* dataset [34] includes data for classifying reviews on the Yelp platform. Reviews with 1-2 stars are considered negative, and those with 4-5 stars are positive. Examples in the dataset were filtered to increase the average sentence length. Only sentences with lengths between 300 and 512 words were retained. The filtered dataset contains 64,445 examples for training and 4,340 for testing. The average sentence length increased to 382 tokens.

¹From the Hugging Face Transformers library https://huggingface.co/docs/transformers/model_doc/distilbert

Results of numerical experiments with the above datasets are represented in Table 1, where three models are compared: the original DistilBERT, the DistilBERT with the random mask, and the DistilBERT with the trainable mask. For every model, we provide the mean time of processing the whole test set and the ROC AUC metric. The mask size is changed from 40% to 90% of the sentence length.

Table 1. Comparison results by using ROC AUC metric and runtime for models: DistilBERT and two proposed modifications with the trainable mask (TM) and the random mask (RM) of various sizes. All 6 layers are replaced

Dataset	Mask size	DistilBERT		DistilBERT+RM		DistilBERT+TM	
		Time	ROC AUC	Time	ROC AUC	Time	ROC AUC
IMDb	0.4	45.36	0.979	40.16	0.9414	40.58	0.9666
	0.5			41.24	0.9563	41.37	0.9734
	0.6			42.35	0.9658	42.63	0.9762
	0.7			43.79	0.9715	43.84	0.9772
	0.8			44.78	0.9754	45.11	0.9777
	0.9			46.05	0.9779	46.15	0.9779
SST-2	0.4	0.560	0.966	0.511	0.8454	0.533	0.8982
	0.5			0.537	0.8722	0.537	0.9139
	0.6			0.561	0.9082	0.551	0.9294
	0.7			0.566	0.9308	0.552	0.9550
	0.8			0.573	0.9493	0.556	0.9506
	0.9			0.577	0.9594	0.557	0.9638
Rotten Tomatoes	0.4	0.698	0.924	0.638	0.8011	0.622	0.8386
	0.5			0.647	0.8423	0.634	0.7927
	0.6			0.6567	0.8761	0.634	0.8814
	0.7			0.6654	0.8890	0.647	0.8961
	0.8			0.6814	0.9063	0.657	0.9239
	0.9			0.6846	0.9179	0.671	0.9241
Yelp Polarity	0.4	21.19	0.992	18.155	0.9722	18.491	0.9795
	0.5			18.692	0.9787	18.970	0.9882
	0.6			19.121	0.9839	19.449	0.9901
	0.7			19.649	0.9874	19.736	0.9912
	0.8			20.324	0.9894	20.100	0.9906
	0.9			21.012	0.9911	20.705	0.9915

Table 1 shows that as the mask size decreases, the performance gap between the trainable mask and random mask models widens significantly. For instance, on the SST-2 dataset with a mask size of 0.4, the trainable mask model achieves a ROC AUC of 0.8982, a substantial 5.3% absolute improvement over the random mask model (0.8454). This demonstrates that at high levels of sparsity, intelligent, learned selection becomes paramount. The trainable mask model effectively identifies and retains the most salient tokens for sentiment classification, while random removal severely degrades model reasoning. The fact that the trainable mask model nearly matches

the performance of the full DistilBERT model on the lengthy Yelp Polarity reviews with a mask size of just 0.5 (ROC AUC of 0.9882 vs. 0.9923) is particularly noteworthy, highlighting the method's strength in processing long-form content.

Fig. 4 illustrates the mask training process on the IMDB dataset. The graph illustrates the change in the loss function on the training and test sets, as well as the loss function value on the test set for the random mask. Hyperparameters: 6 layers are replaced, mask size is 60% of the maximum sentence length, a small neural network is used. The graph illustrates that training is performed correctly, and over time, the loss function value for the trainable mask becomes lower than the loss function for the random mask. The analysis of training dynamics also confirms the stability and effectiveness of the proposed learning procedure. The loss for the trainable mask model decreases steadily on both training and test sets, eventually falling below the constant loss value of the random mask.

Fig. 5 shows the dependence of the ROC AUC metric on the number of replaced layers for various mask sizes on the IMDB dataset for models with random and trainable masks.

Analysis of the graph illustrates that when using the trainable mask, the ROC AUC value remains almost unchanged regardless of the number of replaced layers, with fluctuations within 1%. In contrast, for the random mask, a significant decrease in ROC AUC is observed as the number of modified layers increases.

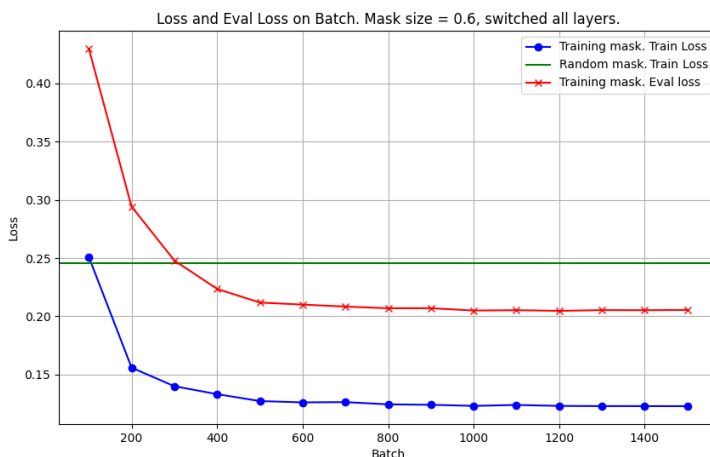


Fig. 4. Illustration of the mask training process using the IMDB dataset

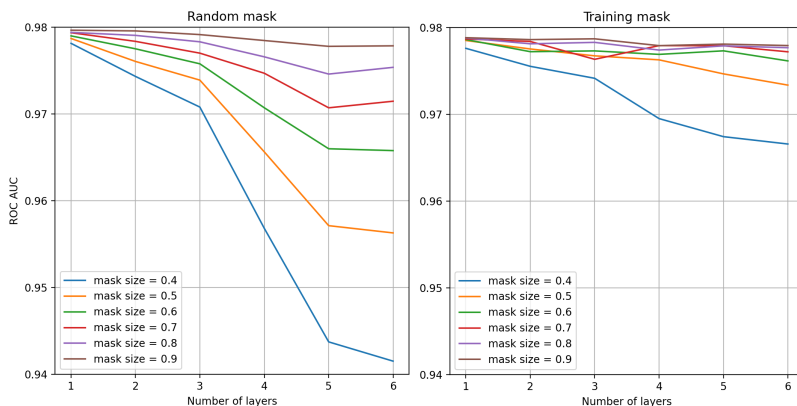


Fig. 5. Dependence of ROC-AUC metric on the number of replaced layers for random (the left graph) and trainable (the right graph) masks. The dataset IMDB is used

Figs. 6-11 display the dependence of the ROC AUC metric on the inference time compared to the unmodified original model for the IMDB and Yelp Polarity datasets. Fig. 6 shows results of experiments for the IMDB dataset conducted with a fixed mask size of 50% of the maximum sentence length. Attention layers are replaced starting from the first layer. For comparison purposes, Figs. 7-8 illustrates results of the same experiments when the fixed mask sizes are 40% (the left graph) and 60% (the right graph). The same experiments for the Yelp Polarity dataset are illustrated in Figs. 9-11.

The presented graphs demonstrate the balance between speed and prediction quality when using random and trainable masks. The results indicate that at comparable acceleration levels, the model with the trainable mask maintains significantly higher ROC AUC values compared to the random mask. Specifically, when all layers are replaced:

- The metric deviation does not exceed 1% from the original model for the trainable mask and 3% for the random mask;
- Time savings are achieved: 8% for the dataset IMDB and over 10% for the dataset Yelp Polarity.

The difference in acceleration is explained by the fact that the dataset Yelp Polarity contains longer sentences due to pre-filtering.

Additionally, the graphs illustrate increasing divergence in quality between the two mask types as the number of replaced layers grows: the ROC AUC gap progressively widens in favor of the trainable approach.

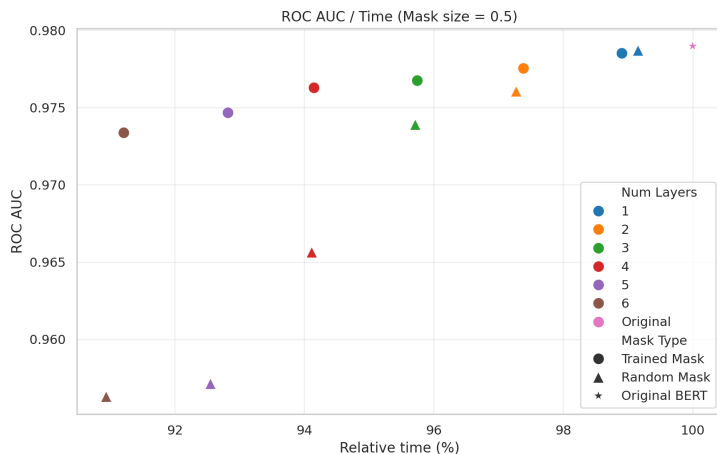


Fig. 6. Differences between the original model and its mask-modified variants with different numbers of replaced layers for the dataset IMDb by two parameters: reduction of ROC-AUC and achieved runtime reduction. The mask size is 0.5.

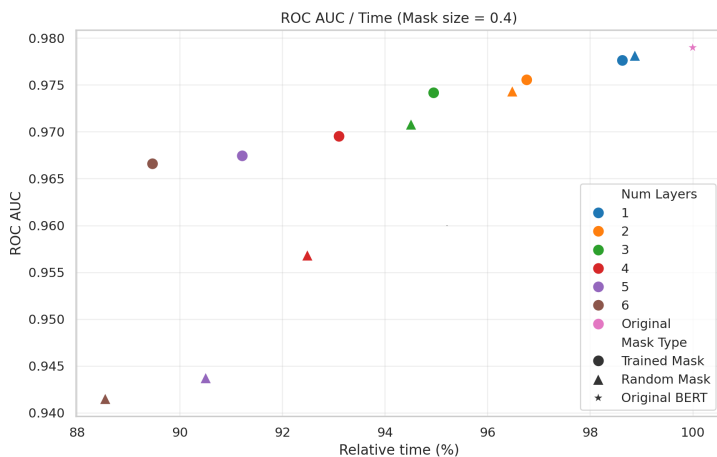


Fig. 7. Differences between the original model and its mask-modified variants with different numbers of replaced layers for the dataset IMDb by two parameters: reduction of ROC-AUC and achieved runtime reduction. The mask size is 0.4

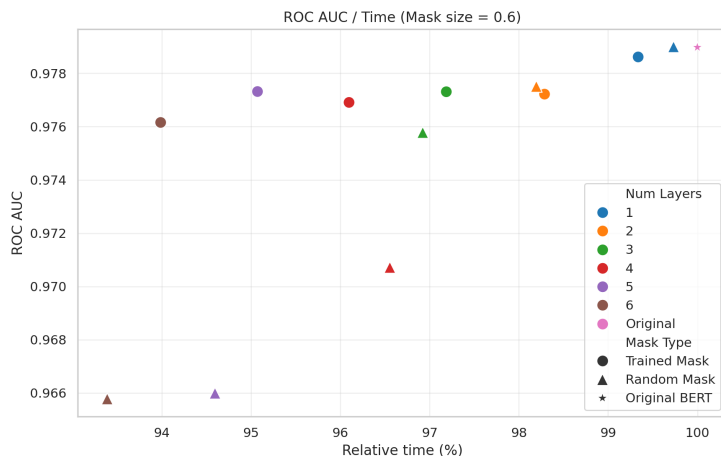


Fig. 8. Differences between the original model and its mask-modified variants with different numbers of replaced layers for the dataset IMDb by two parameters: reduction of ROC-AUC and achieved runtime reduction. The mask sizes is 0.6

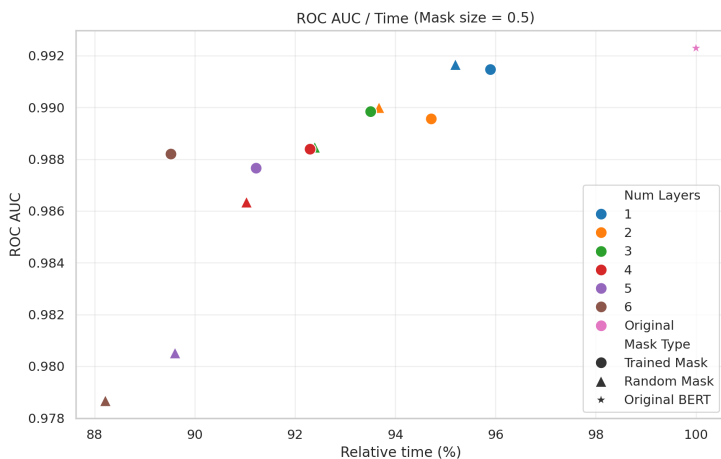


Fig. 9. Differences between the original model and its mask-modified variants with different numbers of replaced layers for the dataset Yelp Polarity by two parameters: reduction of ROC-AUC and achieved runtime reduction. The mask size is 0.5

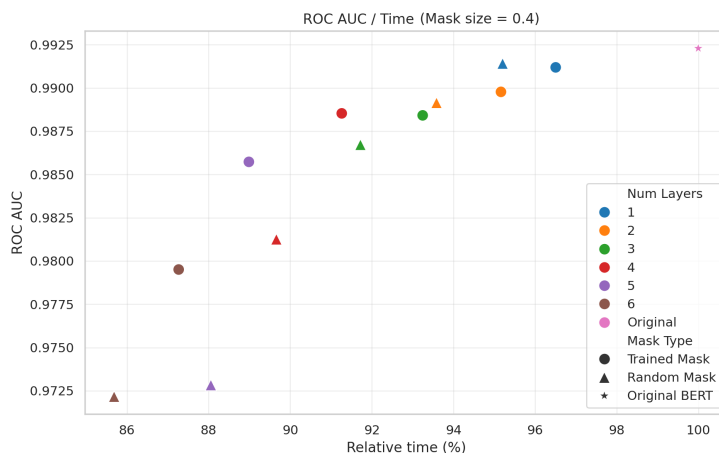


Fig. 10. Differences between the original model and its mask-modified variants with different numbers of replaced layers for the dataset Yelp Polarity by two parameters: reduction of ROC-AUC and achieved runtime reduction. The mask sizes is 0.4

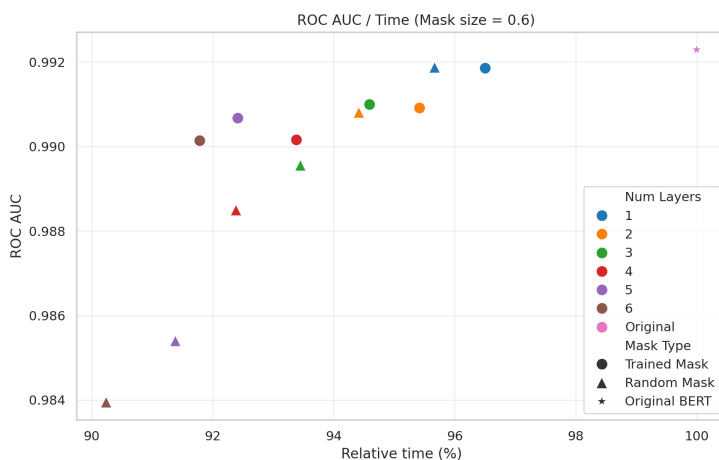


Fig. 11. Differences between the original model and its mask-modified variants with different numbers of replaced layers for the dataset Yelp Polarity by two parameters: reduction of ROC-AUC and achieved runtime reduction. The mask sizes are 0.6

Fig. 12 shows the softmax values calculated by the neural network for tokens from an example in the IMDb dataset. Only the first layer of the transformer architecture was modified in the experiment. As illustrated, the model correctly identifies significant tokens: adjectives expressing emotional tone receive the highest softmax values. With a mask size of 0.4, the selected key tokens are: “amazing”, “brilliant”, “capt”, “ivating”, “[SEP]”.

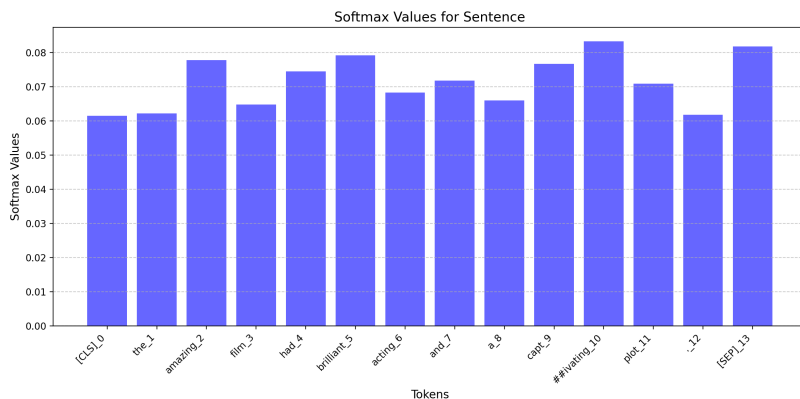


Fig. 12. Softmax values for tokens from an example in the IMDb dataset

Fig. 13 shows the dependence of classification accuracy on the number of tokens for the IMDb dataset. The mask size is 60%, corresponding to approximately 307 tokens. The graph illustrates that the number of tokens does not affect classification quality. This indicates that the model retains important information from sentences needed to determine review sentiment.

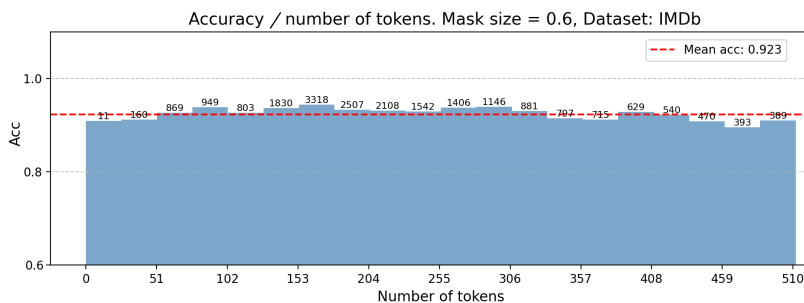


Fig. 13. Dependence of the classification accuracy on the number of tokens for dataset IMDb

In order to study different FCNs generating logits, we compare the following structures:

1. Linear (no hidden layers): consists of input and output layers with 768 and 1 parameters, respectively.
2. Medium: the number of hidden layers is 2; the number of parameters per hidden layer is 64.
3. Large: the number of hidden layers is 2; the number of parameters per hidden layer: 768.

The comparison shows that the medium version of the neural network is the optimal choice: it is less prone to overfitting than the Large version and outperforms the Linear model in accuracy. The comparison results are shown in Fig. 14, where the ROC AUC metric on the test set is measured for different values of the mask size. The analysis of training dynamics in Fig. 14 confirms the stability and effectiveness of the proposed learning procedure.

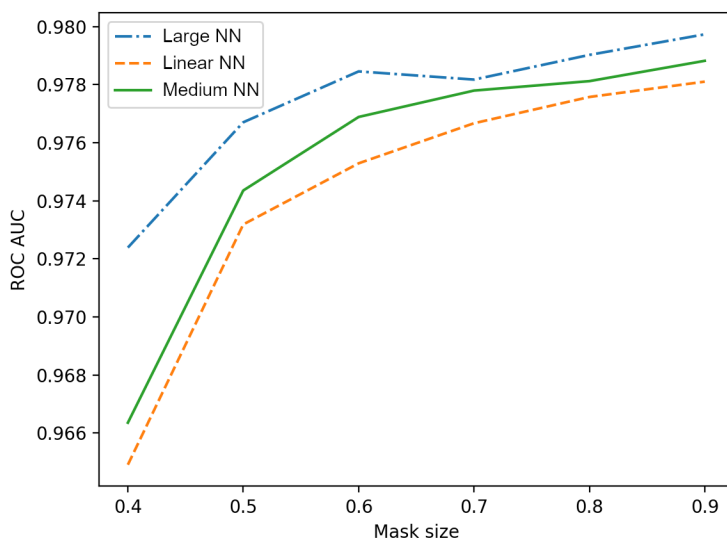


Fig. 14. Comparison of different FCNs using the dataset IMDB with the number of replaced layers equal to 6

The numerical experiments provides compelling evidence of the efficacy and advantages of our proposed differentiable mask learning approach. The consistent outperformance of the trainable mask over the random mask across all datasets and mask sizes underscores a critical finding: the reduction in computational cost achieved through token-selection need

not come at the expense of model accuracy, provided the selection is learned and optimized for the task.

6. Conclusion and Future Work. In this work, we proposed a differentiable token-selection mechanism for optimizing the self-attention computation in transformer models. The method introduces a lightweight auxiliary scoring network that learns to identify and select the most informative tokens via a Gumbel-Softmax-based relaxation, enabling hard token pruning while preserving differentiable training of the scoring module. By reducing the effective sequence length processed by attention layers, the proposed approach reduces the effective attention cost at inference time in the setting considered without modifying the underlying attention formulation.

A key advantage of the proposed method is that it is formulated as a plug-in token-selection module for an already trained transformer backbone. In the experimental setting considered here, the method is added to a pretrained model and optimized through an additional lightweight training stage for the scoring network, without redesigning the backbone architecture from scratch. The resulting attention mechanism achieves a controllable trade-off between computational efficiency and predictive performance through a single tunable parameter governing the number of selected tokens.

The experimental results indicate that the proposed method effectively reduces computational overhead while maintaining competitive performance in the considered setting. Across all datasets, the learned token-selection consistently outperformed random masking strategies, achieving an 8–12% reduction in inference time while preserving most of the baseline quality. The benefits were particularly pronounced for long sequences, where the method preserved high accuracy even when attending to only half of the input tokens. Moreover, the learned masks showed qualitatively interpretable behavior in the presented examples, often prioritizing semantically meaningful and sentiment-bearing tokens.

The present study has several limitations. First, the empirical evaluation is limited to DistilBERT and sentiment-classification benchmarks. Second, the paper focuses on comparison with the original backbone and with a random-mask baseline, because the main goal is to isolate the effect of the proposed plug-in module rather than to benchmark alternative transformer architectures. Third, repeated-run variance estimates are not included in the current experimental protocol. These limitations should be kept in mind when interpreting the scope of the reported results.

Overall, the results indicate that adaptive, learned token-selection provides a practical mechanism for reducing the effective attention cost at inference time in the considered experimental setting. Broader validation

on larger backbones and additional task families remains future work. Future work will also explore the interaction of the proposed method with alternative efficient-attention formulations, adaptive selection of the number of retained tokens, and richer context-aware scoring mechanisms.

References

1. Luong T., Pham H., Manning C. Effective approaches to attention-based neural machine translation. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 2015. pp. 1412–1421. DOI: 10.18653/v1/d15-1166.
2. Vaswani A., Shazeer N., Parmar N., et al. Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. 2017. pp. 6000–6010.
3. Fournier Q., Caron G.M., Aloise D. A Practical Survey on Faster and Lighter Transformers. *ACM Computing Surveys*. 2023. vol. 55. no. 14s. pp. 1–40. DOI: 10.1145/3586074.
4. Han Z., Gao C., Liu J., et al. Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey. *arXiv preprint arXiv:2403.14608*. 2024.
5. Huang H., Liang Z., Fang Z., et al. A Concise Review of Long Context in Large Language Models. *Proceedings of the International Conference on Algorithms, Software Engineering, and Network Security (ASENS'24)*. 2024. pp. 563–566. DOI: 10.1145/3677182.3677282.
6. Shao M., Basit A., Karri R., Shafique M. Survey of Different Large Language Model Architectures: Trends, Benchmarks, and Challenges. *IEEE Access*. 2024. vol. 12. pp. 188664–188706. DOI: 10.1109/ACCESS.2024.3482107.
7. Sun W., Hu J., Zhou Y., et al. Speed Always Wins: A Survey on Efficient Architectures for Large Language Models. *arXiv preprint arXiv:2508.09834*. 2025.
8. Wang X., Salmani M., Omid P., et al. Beyond the limits: a survey of techniques to extend the context length in large language models. *Proceedings of the Thirty Third International Joint Conference on Artificial Intelligence (IJCAI'24)*. 2024. pp. 8299–8307. DOI: 10.24963/ijcai.2024/917.
9. Zhou Z., Ning X., Hong K., et al. A Survey on Efficient Inference for Large Language Models. *arXiv preprint arXiv:2404.14294*. 2024.
10. Beltagy I., Peters M.E., Cohan A. Longformer: The Long-Document Transformer. *arXiv preprint arXiv:2004.05150*. 2020.
11. Child R., Gray S., Radford A., Sutskever I. Generating Long Sequences with Sparse Transformers. *arXiv preprint arXiv:1904.10509*. 2019.
12. Wang S., Li B.Z., Khabsa M., et al. Linformer: Self-attention with Linear Complexity. *arXiv preprint arXiv:2006.04768*. 2020.
13. Choromanski K.M., Likhoshesterov V.L., Dohan D., et al. Rethinking Attention with Performers. *Proceedings of the International Conference on Learning Representations (ICLR'21)*. 2021. pp. 1–38.
14. Dai Z., Yang Z., Yang Y., et al. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019. pp. 2978–2988. DOI: 10.18653/v1/P19-1285.
15. Rae J.W., Potapenko A., Jayakumar S.M., et al. Compressive Transformers for Long-Range Sequence Modelling. *Proceedings of the International Conference on Learning Representations (ICLR'20)*. 2020. pp. 1–19.
16. Zaheer M., Guruganesh G., Dubey A., et al. Big Bird: transformers for longer sequences. *Proceedings of the 34th International Conference on Neural Information Processing System (NIPS'20)*. 2020. vol. 21. pp. 17283–17297.

17. Kitaev N., Kaiser L., Levskaya A. Reformer: The Efficient Transformer. Proceedings of the International Conference on Learning Representations (ICLR'20). 2020. pp. 1–12.
18. Kim S., Shen S., Thorsley D., et al. Learned Token Pruning for Transformers. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'22). 2022. pp. 784–794. DOI: 10.1145/3534678.3539260.
19. Ravula A., Alberti C., Ainslie J., et al. ETC: Encoding Long and Structured Inputs in Transformers. Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020. 268–284.
20. Dao T., Fu D.Y., Ermon S., et al. FLASHATTENTION: fast and memory-efficient exact attention with IO-awareness. Proceedings of the 36th International Conference on Neural Information Processing System (NIPS'22). 2022. pp. 16344–16359.
21. Fu D.Y., Arora S., Grogan J., et al. Monarch MIXER: a simple sub-quadratic GEMM-based architecture. Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS'23). 2023. pp. 77546–77603.
22. Gu A., Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. Proceedings of the 1st Conference on Language Modeling. 2024. pp. 1–32.
23. Liu H., Abbeel P. Blockwise Parallel Transformer for Long Context Models. Proceedings of the 36th International Conference on Neural Large Information Processing Systems (NIPS'23). 2023. pp. 8828–8844. DOI: 10.52202/075280-0386.
24. Peng B., Alcaide E., Anthony Q., et al. RWKV: Reinventing RNNs for the Transformer Era. Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2023. pp. 1–30.
25. Guo Y., Ding L., Yuan Y., Wang G. Macformer: Transformer with Random Maclaurin Feature Attention. arXiv preprint arXiv:2408.11656. 2024.
26. Vladymyrov M., von Oswald J., Sandler M., Ge R. Linear Transformers are Versatile In-Context Learners. Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS'25). 2025. pp. 48784–48809. DOI: 10.52202/079017-1546.
27. Katharopoulos A., Vyas A., Pappas N., Fleuret F. Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention. Proceedings of the 37th International Conference on Machine Learning (PMLR). 2020. vol. 119. pp. 5156–5165.
28. Qin Z., Sun W., Deng H., et al. CosFormer: Rethinking Softmax In Attention. Proceedings of the International Conference on Learning Representations (ICLR'22). 2022. pp. 1–15.
29. Liu J., Pan Z., He H., et al. EcoFormer: Energy-Saving Attention with Linear Complexity. Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS'22). 2022. pp. 10295–10308.
30. Sanh V., Debut L., Chaumond J., Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and Lighter. arXiv preprint arXiv:1910.01108. 2019.
31. Maas A.L., Daly R.E., Pham P.T., et al. Learning word vectors for sentiment analysis. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011. vol. 1. pp. 142–150.
32. Socher R., Perelygin A., Wu J., et al. Recursive deep models for semantic compositionality over a sentiment treebank. Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2013. pp. 1631–1642.
33. Pang B., Lee L. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales. Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05). 2005. pp. 115–124. DOI: 10.3115/1219840.1219855.
34. Zhang X., Zhao J., Lecun Y. Character-level convolutional networks for text classification. Proceedings of the 29th International Conference on Neural Information Processing Systems (NIPS'15). 2015. pp. 649–657.

Konstantinov Andrey — Ph.D., Senior Lecturer, Higher School of Artificial Intelligence Technologies, Federal State Autonomous Educational Institution of Higher Education «Peter the Great St. Petersburg Polytechnic University» (SPbPU). Research interests: machine learning, survival analysis, explainable intelligence. The number of publications — 50. konstantinov_av@spbstu.ru; 29, Polytechnicheskaya st., 195251, St. Petersburg, Russia; office phone: +7(911)954-5565.

Utkin Lev — Ph.D., Dr.Sci., Professor, Higher School of Artificial Intelligence Technologies, Federal State Autonomous Educational Institution of Higher Education «Peter the Great St. Petersburg Polytechnic University» (SPbPU). Research interests: machine learning, survival analysis, theory of generalized probabilities, decision-making. The number of publications — 350. lev.utkin@gmail.com; 29, Polytechnicheskaya st., 195251, St. Petersburg, Russia; office phone: +7(921)344-6390.

Khomets Semen — Master's Student, Higher School of Artificial Intelligence Technologies, Federal State Autonomous Educational Institution of Higher Education «Peter the Great St. Petersburg Polytechnic University» (SPbPU). Research interests: machine learning. The number of publications — 2. homets_sp@spbstu.ru; 29, Polytechnicheskaya st., 195251, St. Petersburg, Russia; office phone: +7(928)135-4848.

Acknowledgements. The work was supported by the Ministry of Science and Higher Education of the Russian Federation within the framework of the state assignment FSEG-2024-0027.

А.В. КОНСТАНТИНОВ, Л.В. УТКИН, С.П. ХОМЕЦ
**ПОДХОД НА ОСНОВЕ ОПЕРАЦИИ GUMBEL-SOFTMAX ДЛЯ
ОТБОРА ЗНАЧИМЫХ ТОКЕНОВ В МЕХАНИЗМЕ
САМОВНИМАНИЯ**

Константинов А.В., Уткин Л. В., Хомец С.П. Подход на основе операции Gumbel-Softmax для отбора значимых токенов в механизме самовнимания.

Аннотация. Квадратичная вычислительная сложность механизма самовнимания в трансформерах остается существенным узким местом для эффективной обработки длинных последовательностей. В данной статье представлен метод снижения вычислительных затрат на использование механизма внимания с помощью дифференцируемого обучения масок, обеспечивающий динамический, зависящий от входных данных отбор токенов и при этом сохраняющий способность модели улавливать глобальную контекстную информацию. Предложенный подход не накладывает каких-либо конкретных архитектурных ограничений и может быть интегрирован в существующие модели трансформеров с минимальными изменениями. В отличие от существующих методов разреженного или приближенного внимания, предложенный подход обучается специфичным для задачи маскам отбора токенов полностью дифференцируемым образом, без наложения предопределенных шаблонов разреженности или архитектурных ограничений. Наш подход использует механизм на основе Gumbel-Softmax для обучения масок выбора токенов в сквозном режиме, предлагая решение, которое сокращает время вывода, сохраняя при этом точность модели. Эксперименты на четырех эталонных наборах данных, предназначенных для анализа настроения, показывают, что предложенный метод обеспечивает сокращение времени вывода на 8-12% при относительном снижении точности менее чем на 1% по сравнению с базовыми моделями. Обучаемые маски поддаются интерпретации, последовательно отдают приоритет токенам, содержащим эмоциональную окраску, и поддерживают эффективный, масштабируемый вывод с помощью трансформеров.

Ключевые слова: трансформер, самовнимание, Gumbel-Softmax, выбор токенов, дифференцируемая маска.

Литература

1. Luong T., Pham H., Manning C. Effective approaches to attention-based neural machine translation // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015. pp. 1412–1421. DOI: 10.18653/v1/d15-1166.
2. Vaswani A., Shazeer N., Parmar N., et al. Attention is all you need // Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). 2017. pp. 6000–6010.
3. Fournier Q., Caron G.M., Aloise D. A Practical Survey on Faster and Lighter Transformers // ACM Computing Surveys. 2023. vol. 55. no. 14s. pp. 1–40. DOI: 10.1145/3586074.
4. Han Z., Gao C., Liu J., et al. Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey // arXiv preprint arXiv:2403.14608. 2024.
5. Huang H., Liang Z., Fang Z., et al. A Concise Review of Long Context in Large Language Models // Proceedings of the International Conference on Algorithms, Software Engineering, and Network Security (ASENS'24). 2024. pp. 563–566. DOI: 10.1145/3677182.3677282.

6. Shao M., Basit A., Karri R., Shafique M. Survey of Different Large Language Model Architectures: Trends, Benchmarks, and Challenges // *IEEE Access*. 2024. vol. 12. pp. 188664–188706. DOI: 10.1109/ACCESS.2024.3482107.
7. Sun W., Hu J., Zhou Y., et al. Speed Always Wins: A Survey on Efficient Architectures for Large Language Models // *arXiv preprint arXiv:2508.09834*. 2025.
8. Wang X., Salmani M., Omid P., et al. Beyond the limits: a survey of techniques to extend the context length in large language models // *Proceedings of the Thirty Third International Joint Conference on Artificial Intelligence (IJCAI'24)*. 2024. pp. 8299–8307. DOI: 10.24963/ijcai.2024/917.
9. Zhou Z., Ning X., Hong K., et al. A Survey on Efficient Inference for Large Language Models // *arXiv preprint arXiv:2404.14294*. 2024.
10. Beltagy I., Peters M.E., Cohan A. Longformer: The Long-Document Transformer // *arXiv preprint arXiv:2004.05150*. 2020.
11. Child R., Gray S., Radford A., Sutskever I. Generating Long Sequences with Sparse Transformers // *arXiv preprint arXiv:1904.10509*. 2019.
12. Wang S., Li B.Z., Khabza M., et al. Linformer: Self-attention with Linear Complexity // *arXiv preprint arXiv:2006.04768*. 2020.
13. Choromanski K.M., Likhoshesterov V.L., Dohan D., et al. Rethinking Attention with Performers // *Proceedings of the International Conference on Learning Representations (ICLR'21)*. 2021. pp. 1–38.
14. Dai Z., Yang Z., Yang Y., et al. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context // *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019. pp. 2978–2988. DOI: 10.18653/v1/P19-1285.
15. Rae J.W., Potapenko A., Jayakumar S.M., et al. Compressive Transformers for Long-Range Sequence Modelling // *Proceedings of the International Conference on Learning Representations (ICLR'20)*. 2020. pp. 1–19.
16. Zaheer M., Guruganesh G., Dubey A., et al. Big Bird: transformers for longer sequences // *Proceedings of the 34th International Conference on Neural Information Processing System (NIPS'20)*. 2020. vol. 21. pp. 17283–17297.
17. Kitaev N., Kaiser L., Levskaya A. Reformer: The Efficient Transformer // *Proceedings of the International Conference on Learning Representations (ICLR'20)*. 2020. pp. 1–12.
18. Kim S., Shen S., Thorsley D., et al. Learned Token Pruning for Transformers // *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'22)*. 2022. pp. 784–794. DOI: 10.1145/3534678.3539260.
19. Ravula A., Alberti C., Ainslie J., et al. ETC: Encoding Long and Structured Inputs in Transformers // *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020. 268–284.
20. Dao T., Fu D.Y., Ermon S., et al. FLASHATTENTION: fast and memory-efficient exact attention with IO-awareness // *Proceedings of the 36th International Conference on Neural Information Processing System (NIPS'22)*. 2022. pp. 16344–16359.
21. Fu D.Y., Arora S., Grogan J., et al. Monarch MIXER: a simple sub-quadratic GEMM-based architecture // *Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS'23)*. 2023. pp. 77546–77603.
22. Gu A., Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces // *Proceedings of the 1st Conference on Language Modeling*. 2024. pp. 1–32.
23. Liu H., Abbeel P. Blockwise Parallel Transformer for Long Context Models // *Proceedings of the 36th International Conference on Neural Large Information Processing Systems (NIPS'23)*. 2023. pp. 8828–8844. DOI: 10.52202/075280-0386.

24. Peng B., Alcaide E., Anthony Q., et al. RWKV: Reinventing RNNs for the Transformer Era // Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2023. pp. 1–30.
25. Guo Y., Ding L., Yuan Y., Wang G. Macformer: Transformer with Random Maclaurin Feature Attention // arXiv preprint arXiv:2408.11656. 2024.
26. Vladymyrov M., von Oswald J., Sandler M., Ge R. Linear Transformers are Versatile In-Context Learners // Proceedings of the 37th International Conference on Neural Information Processing Systems (NIPS'25). 2025. pp. 48784–48809. DOI: 10.52202/079017-1546.
27. Katharopoulos A., Vyas A., Pappas N., Fleuret F. Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention // Proceedings of the 37th International Conference on Machine Learning (PMLR). 2020. vol. 119. pp. 5156–5165.
28. Qin Z., Sun W., Deng H., et al. CosFormer: Rethinking Softmax In Attention // Proceedings of the International Conference on Learning Representations (ICLR'22). 2022. pp. 1–15.
29. Liu J., Pan Z., He H., et al. EcoFormer: Energy-Saving Attention with Linear Complexity // Proceedings of the 36th International Conference on Neural Information Processing Systems (NIPS'22). 2022. pp. 10295–10308.
30. Sanh V., Debut L., Chaumond J., Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and Lighter // arXiv preprint arXiv:1910.01108. 2019.
31. Maas A.L., Daly R.E., Pham P.T., et al. Learning word vectors for sentiment analysis // Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011. vol. 1. pp. 142–150.
32. Socher R., Perelygin A., Wu J., et al. Recursive deep models for semantic compositionality over a sentiment treebank // Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2013. pp. 1631–1642.
33. Pang B., Lee L. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales // Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05). 2005. pp. 115–124. DOI: 10.3115/1219840.1219855.
34. Zhang X., Zhao J., Lecun Y. Character-level convolutional networks for text classification // Proceedings of the 29th International Conference on Neural Information Processing Systems (NIPS'15). 2015. pp. 649–657.

Константинов Андрей Владимирович — канд. физ.-мат. наук, старший преподаватель, высшая школа технологий искусственного интеллекта, Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого» (СПбПУ). Область научных интересов: машинное обучение, анализ выживаемости, объяснительный интеллект. Число научных публикаций — 50. konstantinov_av@spbstu.ru; ул. Политехническая, 29, 195251, Санкт-Петербург, Россия; р.т.: +7(911)954-5565.

Уткин Лев Владимирович — д-р техн. наук, профессор, высшая школа технологий искусственного интеллекта, Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого» (СПбПУ). Область научных интересов: машинное обучение, анализ выживаемости, теория обобщенных вероятностей, принятие решений. Число научных публикаций — 350. lev.utkin@gmail.com; ул. Политехническая, 29, 195251, Санкт-Петербург, Россия; р.т.: +7(921)344-6390.

Хомец Семен Павлович — магистрант, высшая школа технологий искусственного интеллекта, Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет

Петра Великого» (СПбПУ). Область научных интересов: машинное обучение. Число научных публикаций — 2. homets_sp@spbstu.ru; ул. Политехническая, 29, 195251, Санкт-Петербург, Россия; р.т.: +7(928)135-4848.

Поддержка исследований. Работа выполнена при поддержке Министерства науки и высшего образования РФ в рамках государственного задания FSEG-2024-0027.

А.Р. Гири́н, Н.Н. Тесля
**МЕТОД СОПОСТАВЛЕНИЯ ЗАПИСЕЙ О ДТП
ИЗ ОФИЦИАЛЬНЫХ ИСТОЧНИКОВ
И НОВОСТНЫХ ПУБЛИКАЦИЙ НА ОСНОВЕ
ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ КЛАСТЕРИЗАЦИИ**

Гирин А.Р., Тесля Н.Н. Метод сопоставления записей о ДТП из официальных источников и новостных публикаций на основе пространственно-временной кластеризации.

Аннотация. Принятие эффективных решений для снижения уровня аварийности на автомобильном транспорте основывается на результатах анализа мест концентрации ДТП. Полнота и точность данных о ДТП, используемых для анализа, могут быть повышены за счет интеграции нескольких источников. Важным этапом интеграции является сопоставление записей из различных источников, описывающих одни и те же события ДТП. Решение задачи сопоставления осложняется проблемой гетерогенности: записи о ДТП могут отличаться по структурированности и формату представления. В исследовании предложен метод решения задачи сопоставления записей о ДТП из двух гетерогенных источников: официальных ресурсов со структурированными карточками учета ДТП и новостных публикаций в виде неструктурированных текстов на естественном языке. Предложенный метод основан на гипотезе о пространственной и временной близости: записи из различных источников считаются соответствующими одному ДТП, если описанные в них события произошли на небольшом расстоянии и с незначительной временной разницей. Метод включает два последовательных этапа. На первом этапе из новостных публикаций извлекаются координаты и дата/время описываемого события с использованием экспертных правил, географического справочника и большой языковой модели LLaMA. На втором этапе выполняется сопоставление записей из официальных ресурсов и новостных публикаций на основе пространственно-временной кластеризации с применением алгоритма ST-DBSCAN. Метод позволяет дополнять описание контекста официально зарегистрированных ДТП и обогащать доступные для анализа данные за счет включения ДТП, сведения о которых отсутствуют в официальных источниках. Экспериментальная оценка метода проведена на реальных данных о ДТП в Санкт-Петербурге. Результаты показали, что комбинирование классических методов с большой языковой моделью повышает точность извлечения пространственно-временных атрибутов ДТП из новостных публикаций, а применение пространственно-временной кластеризации обеспечивает высокую точность сопоставления записей о ДТП (F1-мера = 0.84).

Ключевые слова: интеграция данных, сопоставление записей, дорожно-транспортные происшествия, газетир, экспертные правила, ST-DBSCAN, большие языковые модели, LLaMA.

1. Введение. Дорожно-транспортные происшествия (ДТП) представляют собой значительную социально-экономическую проблему, актуальную как на мировом уровне, так и в России. Согласно официальным данным, в 2024 году в стране произошло более 132 тыс. ДТП, в результате которых погибло свыше 14 тыс. человек и более 164 тыс. получили травмы [1]. Принятие эффективных управленческих решений, направленных на снижение уровня аварийности, основывается

на результатах анализа мест концентрации ДТП и выявлении причин их возникновения. Качество такого анализа и, как следствие, эффективность принимаемых решений напрямую зависят от полноты и точности доступных исследователям данных о ДТП [2].

В международной практике повышение полноты и точности данных о ДТП достигается за счет интеграции нескольких источников. Записи из различных источников, описывающие одни и те же события ДТП, сопоставляются между собой с последующим объединением. Это позволяет уточнять и обогащать описание контекста ДТП, а также включать в анализ данные о ранее не учтенных ДТП [3].

Как правило, в исследованиях основным источником данных о ДТП выступают записи из официальных ресурсов органов государственной власти. К числу дополнительных источников относятся данные бортовых устройств, записи устройств дорожной инфраструктуры, а также новостные публикации в СМИ и социальных сетях [4]. Новостные публикации являются особенно перспективным источником. Их распространенность и общедоступность обеспечивают широкий охват ДТП, в том числе событий, сведения о которых не публикуются на официальных ресурсах (например, ДТП без пострадавших). Кроме того, публикации содержат детализированное описание контекста ДТП, включая данные о местоположении, времени, участниках и возможных причинах ДТП, что представляет ценность для анализа [5, 6].

Интеграция записей из официальных ресурсов и новостных публикаций позволяет повысить качество данных о ДТП, используемых для анализа [7]. Ключевым этапом интеграции является решение задачи сопоставления записей из этих источников (*record linkage*), которая состоит в определении записей, описывающих одно и то же событие ДТП. Такое сопоставление осложняется проблемой гетерогенности источников. Если записи из официальных ресурсов имеют строгую структуру и регламентированный набор атрибутов, то новостные публикации представляют собой неструктурированный текст на естественном языке, что, наряду с проблемой отсутствия между источниками единого сквозного идентификатора описываемых событий ДТП, препятствует прямому сопоставлению записей [8].

Существующие методы решения задачи сопоставления записей могут быть разделены на несколько групп. К первой группе относятся классические детерминированные и вероятностные методы (*probabilistic record linkage*) [9, 10]. Детерминированные методы используют строгие правила или функции сходства строк (например, расстояние Левенштейна или Джаро-Винклера)

для сравнения атрибутов. Вероятностные методы, основанные на модели Фелледи-Сантера, оценивают вероятность совпадения записей на основе весов согласия и расхождения признаков. Основным преимуществом данных методов является их интерпретируемость и вычислительная эффективность, однако возможность их применения ограничена необходимостью наличия строго структурированных наборов атрибутов. Адаптация данных методов за счет предварительного извлечения полного набора атрибутов описываемого в тексте события ДТП сопряжена с риском накопления ошибок [11].

Отдельно может быть выделен метод сопоставления записей о ДТП на основе гипотезы о пространственной и временной близости: записи считаются соответствующими одному событию ДТП, если описанные в них события произошли на небольшом расстоянии друг от друга и с незначительной временной разницей. Гипотеза использована, например, при сопоставлении записей о ДТП из официального ресурса дорожной полиции и социального навигационного приложения Waze в городе Белу-Оризонти (Бразилия) [4]. Получение значений пространственно-временных атрибутов ДТП в данном исследовании осуществлялось на основе метаданных записей (геометок). Использование метаданных позволило исключить этап обработки текста публикации для извлечения атрибутов, однако применимость данного метода к решению поставленной задачи ограничена, так как метаданными сопровождается лишь малая часть новостных публикаций о ДТП в СМИ и социальных сетях.

Вторую группу составляют методы на основе классического машинного обучения, где задача сопоставления формулируется как задача бинарной классификации пар записей [12]. Несмотря на их способность выявлять более сложные зависимости между признаками по сравнению с детерминированными методами, их использование в предметной области исследования ограничено. Это обусловлено отсутствием обширных размеченных наборов данных с записями о ДТП, необходимых для обучения моделей, а также сложностью формирования репрезентативного вектора признаков из неструктурированного текста публикаций.

Третья группа включает методы глубокого обучения на основе предобученных языковых моделей, таких как BERT или RoBERTa, и специализированных систем, например, Ditto [13, 14]. Использование данных методов позволяет достигать высоких показателей F1-меры и обеспечивать устойчивость к пропускам или ошибкам в атрибутах. Однако высокая сложность внутренней архитектуры моделей затрудняет

интерпретацию логики принятия решений. Кроме того, данные методы требуют значительных объемов обучающих данных для тонкой настройки (fine-tuning) и часто показывают низкую точность на данных, не представленных в обучающей выборке [15].

Четвертую группу представляют методы, использующие генеративные большие языковые модели (LLMs), такие как GPT-4. Эти методы позволяют решать задачу сопоставления в режимах zero-shot (без примеров) или few-shot (с минимальным количеством примеров) за счет использования контекстного обучения и промпт-инжиниринга [16, 17]. Исследования показывают, что LLM способны генерировать структурированные объяснения принятых решений и обладают высокой способностью к обобщению на новых типах данных [18]. Тем не менее, методы на основе LLM характеризуются высокой чувствительностью к формулировкам запросов, а их применение может быть ограничено стоимостью API-запросов или высокими требованиями к вычислительным ресурсам при локальном развертывании моделей [19]. Кроме того, применение LLM может сопровождаться галлюцинациями модели.

Обзор существующих методов сопоставления записей позволяет сделать вывод о наличии значимых ограничений при их работе с неструктурированными текстами новостных публикаций. Детерминированные и вероятностные методы эффективны преимущественно при работе со структурированными данными. Использование классического машинного и глубокого обучения требует подготовки обширных размеченных наборов данных о ДТП, доступность которых в предметной области ограничена. Применение больших языковых моделей обеспечивает высокую точность сопоставления записей без предварительной настройки, однако такие модели чувствительны к формулировкам запросов и характеризуются высокими вычислительными затратами и возможными галлюцинациями модели.

Целью исследования является разработка метода сопоставления записей о ДТП из официальных ресурсов и новостных публикаций, позволяющего преодолеть ограничения существующих методов. В основу исследования положены следующие гипотезы:

1. Сопоставление записей о ДТП из разнородных источников может быть выполнено на основании сравнения значений их пространственно-временных атрибутов и проверки условия пространственной и временной близости описываемых в записях событий.

2. Пространственно-временные атрибуты могут быть извлечены непосредственно из текстов новостных публикаций, при этом точность

извлечения повышается за счет использования комбинированного метода: комбинации классических методов и большой языковой модели, способной интерпретировать сложные языковые конструкции.

На основе выдвинутых гипотез в исследовании предлагается двухэтапный метод сопоставления записей. На первом этапе из текстов новостных публикаций извлекаются координаты и дата/время описываемого события ДТП с использованием разработанной процедуры, объединяющей экспертные правила, географический справочник (газетир) и большую языковую модель LLaMA [20]. На втором этапе записи из официальных ресурсов и новостных публикаций сопоставляются на основе пространственно-временной кластеризации с использованием алгоритма ST-DBSCAN [21].

Описание и экспериментальная оценка метода выполнены на примере источников данных о ДТП в Санкт-Петербурге. В качестве официального ресурса использован сайт Госавтоинспекции [1], на котором размещаются карточки учета официально зарегистрированных ДТП. Источником новостных публикаций выступили региональные интернет-СМИ и тематические сообщества в социальной сети «ВКонтакте». На текущем этапе исследования рассмотрена исключительно текстовая часть новостных публикаций на русском языке, а в анализ включены только те записи, которые содержат описание единственного события ДТП. При этом предлагаемый метод не ограничен спецификой конкретных источников и может быть адаптирован для использования в других регионах.

Научная новизна предлагаемого подхода заключается в разработке гибридного метода сопоставления записей, основанного на концепции двухэтапной верификации пространственно-временных атрибутов. В отличие от существующих подходов, ориентированных на полное извлечение контекста ДТП, предложенный метод базируется на гипотезе о достаточности пространственно-временной близости записей для идентификации уникального события.

В отличие от современных (state-of-the-art) систем сопоставления (например, Ditto [13]), требующих наличия тысяч размеченных пар записей для обучения, предлагаемый метод ориентирован на работу в условиях отсутствия обучающих выборок. Прямое сравнение показывает, что использование LLM в режиме верификации в комбинации с классическими методами позволяет достичь точности сопоставления ($F1$ -мера = 0.84), сопоставимой с результатами специализированных трансформеров на эталонных наборах данных [18],

при существенном снижении трудозатрат на предварительную подготовку и разметку данных о ДТП.

2. Формализация предметной области исследования. В рамках исследования решается задача сопоставления записей о ДТП, полученных из официальных карточек учета ДТП и новостных публикаций в СМИ и социальных сетях. Задача сопоставления заключается в определении записей, описывающих одно и то же событие ДТП, и формально может быть поставлена в терминах классической задачи разрешения сущностей (Entity Resolution) [10].

Под сущностью понимается объект или событие реального мира, а разрешение сущностей заключается в сопоставлении записей из различных источников данных, описывающих одни и те же сущности. В предметной области исследования основной сущностью выступает ДТП. Источниками данных, содержащими описания ДТП, являются официальные карточки учета ДТП от Госавтоинспекции и новостные публикации в СМИ и социальных сетях. Для формальной постановки задачи сопоставления и описания предлагаемого метода ее решения вводится ряд обозначений.

2.1. Используемые обозначения. ДТП рассматривается как сложная сущность, состоящая из нескольких взаимосвязанных элементов: участников ДТП, транспортных средств и внешних условий (погодные, дорожные и другие условия). Состояние этих элементов в момент ДТП определяет контекст ДТП, который может быть представлен в виде множества пар «атрибут–значение»:

- Атрибуты участников ДТП (роль в ДТП, возраст, стаж вождения, состояние здоровья, наличие признаков опьянения, использование средств пассивной безопасности и другие);

- Атрибуты транспортных средств (категория, марка и модель, год выпуска, наличие технических неисправностей, характер и локализация повреждений и другие);

- Атрибуты внешней среды (погодные условия, тип и состояние дорожного покрытия, уровень естественной и искусственной освещенности, параметры транспортного потока, обустройство участка дороги и другие).

Пусть через E обозначается множество ДТП, произошедших в исследуемом регионе за фиксированный промежуток времени; через A множество атрибутов, используемых для описания контекста ДТП; через R множество доступных в исследовании записей, содержащих сведения о ДТП из множества E (примерами записи выступают официальные карточки учета ДТП и новостные публикации). В соответствии

с введенными ограничениями исследования, каждая запись содержит описание единственного события ДТП.

Поскольку формат записей может быть разным (от структурированных карточек учета ДТП до неструктурированного текста в новостных публикациях), значения атрибутов не всегда могут быть получены напрямую и требуют извлечения из записей. Пусть через $a(r)$ обозначается значение атрибута $a \in A$, извлеченное из записи $r \in R$. Для любой записи $r \in R$ через $e(r) \in E$ обозначается ДТП, сведения о котором содержатся в данной записи.

Записи $r_1, r_2 \in R$ считаются эквивалентными, если они описывают одно и то же событие ДТП, то есть для них выполняется условие $e(r_1) = e(r_2)$. Каждое реальное событие ДТП $e \in E$ характеризуется вектором параметров $\Theta_e = \langle L_e, T_e \rangle$, где L_e - точные географические координаты, а T_e - точное время события. Критерием истинного совпадения двух записей $r_1, r_2 \in R$ является равенство векторов их параметров:

$$e(r_1) = e(r_2) \iff \Theta_{e(r_1)} = \Theta_{e(r_2)}. \quad (1)$$

2.2. Описание источников данных о ДТП. Официальным источником данных о ДТП в России является сайт Госавтоинспекции, на котором публикуются карточки учета ДТП. Пусть через $R_{official} \subseteq R$ обозначается множество карточек учета ДТП для исследуемого региона за фиксированный промежуток времени. Структура карточек учета строго регламентирована и каждое официально зарегистрированное ДТП описывается только одной карточкой учета:

$$\forall r_1, r_2 \in R_{official}, e(r_1) = e(r_2) \implies r_1 = r_2. \quad (2)$$

В качестве альтернативного источника данных о ДТП рассматриваются новостные публикации в СМИ и социальных сетях. Пусть через $R_{news} \subseteq R$ обозначается множество записей о ДТП в исследуемом регионе за фиксированный промежуток времени, полученных из этих источников. В соответствии с введенными ограничениями исследования, каждая запись рассматривается исключительно как текстовое содержимое публикации. В отличие от карточек учета ДТП, одно и то же ДТП может описываться в нескольких публикациях. Для записей $r \in R_{news}$ определение

значений атрибутов $a(r)$ требует применения методов обработки естественного языка.

2.3. Формальная постановка задачи сопоставления записей.

С учетом введенных обозначений, задача сопоставления записей о ДТП, полученных из официальных карточек учета ДТП и новостных публикаций, формально ставится следующим образом.

Пусть $R = R_{official} \cup R_{news}$ множество всех доступных в исследовании записей о ДТП. Необходимо построить разбиение множества R на кластеры:

$$\mathcal{P} = \{R_1, \dots, R_n\}, \quad (3)$$

для которых выполняются следующие условия:

1. $\forall i \neq j \quad R_i \cap R_j = \emptyset$.
2. $\bigcup_{i=1}^n R_i = R$.
3. $\forall i \forall r_1, r_2 \in R_i \quad e(r_1) = e(r_2)$.
4. $\forall r_1, r_2 \in R \quad e(r_1) = e(r_2) \implies \exists k : r_1, r_2 \in R_k$.

$R = R_{official} \cup R_{news}$ – доступные описания событий; $e(r) \in E$ – событие, данные о котором описывает $r \in R$; $r_1, r_2 \in R$ считаются эквивалентными, если $e(r_1) = e(r_2)$.

Искомое разбиение $\mathcal{P}$ представляет собой фактор-множество множества R по отношению эквивалентности записей, где каждый кластер является классом эквивалентности, объединяющим все записи об одном уникальном событии ДТП.

В каждом кластере записей R_i может содержаться не более одной записи из $R_{official}$, поскольку в каждом кластере описывается единственное ДТП, а каждое ДТП может быть представлено только в одной карточке учета ДТП. Таким образом, искомое разбиение множества R будет состоять из кластеров двух типов:

1. Кластеры, содержащие одну запись из $R_{official}$ и ноль или более записей из R_{news} . Такие кластеры соответствуют ДТП, сведения о которых были официально опубликованы.

2. Кластеры, содержащие исключительно записи из R_{news} . Такие кластеры соответствуют ДТП, сведения о которых не были официально опубликованы, но были описаны в СМИ и социальных сетях.

3. Предлагаемый метод сопоставления записей. В рамках исследования решается задача сопоставления записей о ДТП. Задача формально поставлена как разбиение множества записей R на непересекающиеся кластеры, каждый из которых соответствует одному

уникальному ДТП. Построение такого разбиения требует возможности для любой пары записей проверить условие их эквивалентности (т.е. соответствия одному ДТП). На практике, однако, эта проверка не может быть выполнена напрямую в силу гетерогенности источников. Так как официальные карточки и новостные публикации формируются независимо, между ними отсутствует общий сквозной идентификатор, позволяющий однозначно установить эквивалентность записей.

Для преодоления этого ограничения предлагается метод сопоставления записей на основе гипотезы, согласно которой две записи считаются эквивалентными, если описанные в них события ДТП произошли на небольшом расстоянии друг от друга и с незначительной временной разницей (Рис. 1). В связи с этим из множества A выделяется ключевое подмножество атрибутов $a_{loc_coords}, a_{datetime} \in A$ (координаты и дата/время ДТП соответственно), совокупность значений которых является достаточной для условно уникальной идентификации ДТП.

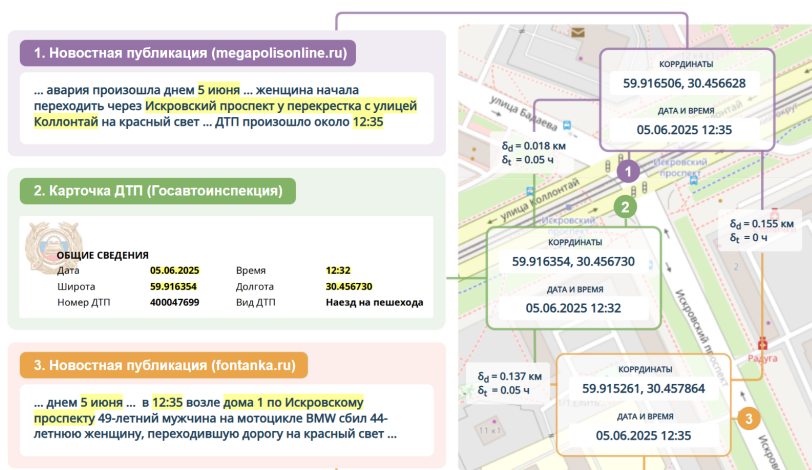


Рис. 1. Демонстрация близости значений пространственно-временных атрибутов для трех записей с описанием одного события ДТП

Практическое решение задачи сопоставления записей о ДТП сводится к последовательному решению следующих подзадач:

1. Разработка процедуры извлечения пространственно-временных атрибутов (a_{loc_coords} и $a_{datetime}$) из записей R .

2. Построение функции схожести записей, которая оценивает соответствие двух записей одному событию ДТП на основе сравнения значений их пространственно-временных атрибутов.

3. Разработка алгоритма сопоставления записей, который на основе построенной функции схожести формирует искомое разбиение множества записей R на кластеры.

Таким образом, предлагаемый метод сопоставления записей о ДТП заключается в последовательном выполнении двух этапов: извлечение пространственно-временных атрибутов из записей R с помощью разработанной процедуры и применение алгоритма сопоставления записей (Рис. 2).

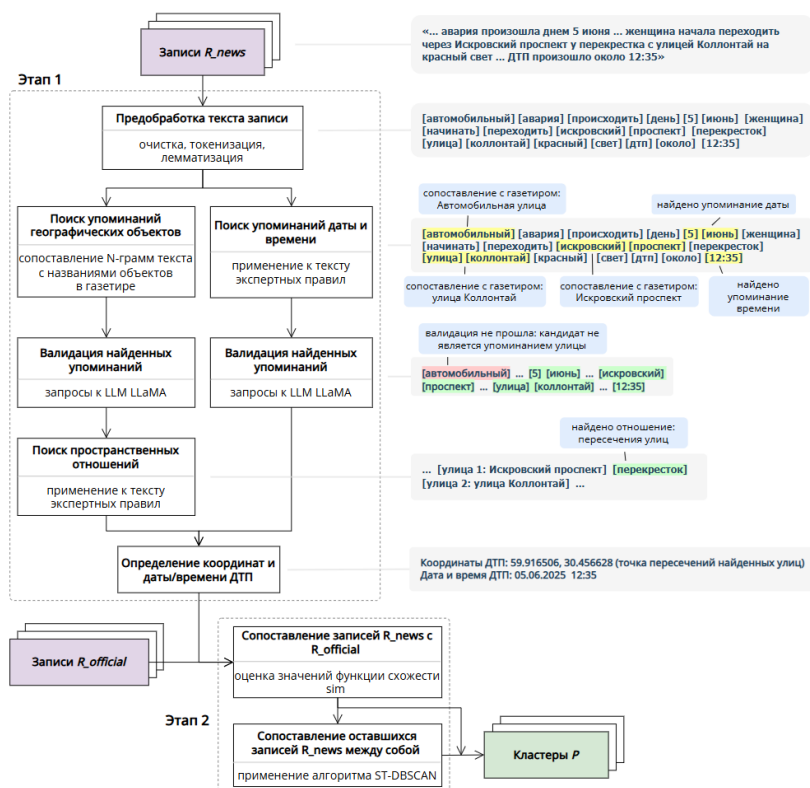


Рис. 2. Общая схема сопоставления записей о ДТП предлагаемым методом

3.1. Процедура извлечения пространственно-временных атрибутов из записей. Для записей из $R_{official}$ значения пространственно-временных атрибутов доступны напрямую из структурированных полей карточек учета ДТП. С другой стороны, определение значений атрибутов для записей R_{news} является нетривиальной задачей. Это обусловлено тем, что новостные публикации, как правило, не сопровождаются метаданными (например, геометками), поэтому значения атрибутов необходимо извлекать непосредственно из неструктурированного текста на русском языке.

Для извлечения пространственно-временных атрибутов из записей R_{news} предлагается процедура, основанная на совместном использовании классических методов (правил и географических справочников) для первичного поиска потенциальных упоминаний в тексте даты/времени и географических объектов, а также большой языковой модели LLM LLaMA для последующей валидации этих упоминаний. Валидация включает две проверки: исключение ложных срабатываний и определение релевантности упоминаний, то есть установление того, относится ли найденная дата/время или географический объект непосредственно к событию ДТП.

Применение методов на основе правил на этапе первичного поиска обеспечивает высокую прозрачность и интерпретируемость логики извлечения кандидатов (потенциальных упоминаний в тексте даты/времени и географических объектов) [22, 23]. Для локализации географических объектов в пределах ограниченной области, в которой названия объектов преимущественно не повторяются, эффективно использование географического справочника (газетира) [24, 25]. В то же время от использования методов классического машинного обучения [26, 27] и глубоких нейронных сетей [14] было решено отказаться ввиду отсутствия обширных размеченных корпусов данных и высоких трудозатрат на их аннотирование. Применение большой языковой модели LLM LLaMA мотивировано ее способностью учитывать семантический контекст всей публикации и интерпретировать сложные языковые конструкции. Это позволяет компенсировать ограничения классических методов [28].

Разработанная процедура включает два блока, отвечающих соответственно за извлечение даты/времени и местоположения события ДТП.

3.1.1. Извлечение даты и времени события ДТП. Извлечение даты и времени события ДТП из текста новостной публикации проводится в три последовательных этапа:

1. Предобработка текста. Входной текст приводится к стандартизованному формату, удобному для дальнейшего анализа. Этот этап включает очистку текста от нерелевантных символов, удаление лишних пробелов, приведение к нижнему регистру, разбиение на отдельные лексемы (токенизацию), приведение слов к их начальной форме (лемматизацию).

2. Применение правил. После предобработки к тексту применяется набор заранее составленных экспертных правил для поиска упоминаний даты и времени. Правила разработаны на основе анализа форматов представления даты и времени в русскоязычных текстах и охватывают как точные (например, «12:30»), так и приблизительные упоминания времени (например, «около часа дня»), а также абсолютные (например, «12 мая 2024 г.») и относительные (например, «вчера») форматы дат.

3. Валидация. Кандидаты (все упоминания даты и времени, найденные на предыдущем этапе) проходят валидацию. Для каждого кандидата последовательно проводятся две проверки с использованием специально разработанных запросов к LLM LLaMA:

- проверка корректности извлечения. Определяется, действительно ли кандидат (например, «16:40») соответствует какому-либо упоминанию даты и времени в исходном тексте, даже если оно представлено в другом формате (например, «без двадцати пять»);
- проверка релевантности событию ДТП. Определяется, относится ли кандидат непосредственно к событию ДТП, а не к другим сущностям, описанным в тексте.

Запросы к LLM LLaMA строятся с учетом принципов промпт-инжиниринга для обеспечения точности и воспроизводимости ответов [17]. Структура запроса включает входные данные (исходный текст, описание кандидата), постановку задачи и ограничения на формат ответа: строго бинарный ответ («ДА» или «НЕТ») для минимизации неоднозначности его интерпретации.

В исследовании использовалась модель LLaMA-3-8B-Instruct в квантованном формате (4 bit, метод GGUF), развернутая локально. Для обеспечения воспроизводимости результатов и минимизации вариативности ответов значение параметра температуры было установлено равным 0. Риск галлюцинаций минимизируется за счет архитектуры метода: LLM выполняет функцию бинарного классификатора для кандидатов, уже найденных экспертными правилами. Такой подход ограничивает модель контекстом исходной публикации и конкретными упоминаниями.

Для формирования итогового результата устраняются конфликты, возникающие при наличии нескольких подтвержденных упоминаний даты и времени. Например, точное время («07:30») конкретизирует более общее упоминание («утром»). Если неоднозначность устранить не удастся или если после валидации не остается ни одного подтвержденного кандидата, предполагается, что дата и время события ДТП не могут быть извлечены.

3.1.2. Извлечение местоположения события ДТП. Для поиска упоминаний в тексте географических объектов используется географический справочник (газетир), который формируется для исследуемого региона на основе открытых данных проекта OpenStreetMap. В него включаются объекты двух категорий: дорожная сеть (улицы, проспекты, шоссе и т.д.) и объекты инфраструктуры (станции метро, торговые центры и т.д.). Для каждого объекта в газетире хранятся: наименование, тип и координаты (точки для объектов инфраструктуры и полилинии для элементов дорожной сети).

Помимо идентификации отдельных географических объектов, процедура учитывает пространственные отношения между ними, что позволяет более точно определять местоположение ДТП. К основным типам таких отношений относятся: пересечение улиц (например, «на пересечении Невского проспекта и Садовой улицы»), участок дороги (например, «на Лиговском проспекте между Кузнечным переулком и Разъездной улицей»), привязка к объекту-ориентир (например, «на Московском проспекте у станции метро Электросила»). Примеры таких правил, формально описанные с использованием нотации на основе Бэкуса-Наура (EBNF), в упрощенном виде представлены в листинге 1.

Извлечение местоположения события ДТП включает следующие последовательные этапы:

1. Предобработка текста. Данный этап аналогичен предобработке, выполняемой при извлечении даты и времени.

2. Идентификация географических объектов. Упоминания географических объектов выявляются путем нечеткого сопоставления N-грамм текста с наименованиями объектов в газетире. В качестве метрики сходства используется расстояние Левенштейна с пороговым значением коэффициента подобия $s > 0.85$. Для повышения точности также учитываются типы объектов (например, слово «улица»), указанные в тексте рядом с упоминанием названия объекта.

3. Валидация. Все найденные кандидаты проходят двухэтапную валидацию для исключения ложных совпадений. Во-первых, LLM LLaMA

подтверждает, что найденное упоминание действительно является географическим объектом. Во-вторых, проверяется релевантность этого объекта непосредственно местоположению события ДТП. Подход к построению запросов к LLM LLaMA соответствует описанному в подразделе об извлечении даты и времени.

4. Идентификация пространственных отношений. Для успешно валидированных географических объектов выполняется поиск упоминаний пространственных отношений между ними с использованием заранее сформированных экспертных правил. Если отношение найдено, его координаты вычисляются с помощью геометрии объектов из газетера. Координаты перекрестка определяются как точка пересечения полилиний соответствующих дорог; координаты участка дороги, как соответствующий сегмент полилинии основной дороги; координаты объекта относительно ориентира, как точка (проекция ориентира на дорогу).

В результате выполнения процедуры формируется набор координат (точек и полилиний), соответствующих всем идентифицированным и валидированным упоминаниям. Для получения итоговой точки применяется иерархический подход. Приоритет отдается точечным координатам, полученным из упоминаний отдельных объектов (например, станций метро), перекрестков или привязок к ориентирам. При наличии нескольких таких точек вычисляется их геометрический центр.

```
RELATION ::= INTERSECTION | ROAD_SEGMENT | LANDMARK_REF;

INTERSECTION ::=
    ("на" | "у"), ("пересечении" | "перекрестке"),
    ROADWAY, "и", ROADWAY;

ROAD_SEGMENT ::=
    ROADWAY, (
        ("между", (ROADWAY | PLACE), "и", (ROADWAY | PLACE))
        |
        ("от", (ROADWAY | PLACE), "до", (ROADWAY | PLACE))
    );

LANDMARK_REF ::=
    ("на" | "по"), ROADWAY, ("напротив" | "рядом с"),
    PLACE;
```

Листинг 1. Примеры экспертных правил для идентификации пространственных отношений

3.2. Функция схожести записей. Для количественной оценки соответствия двух записей $r_1, r_2 \in R$ одному событию ДТП на основе сравнения извлеченных значений их пространственно-временных атрибутов определяется функция схожести $sim : R \times R \rightarrow [0, 1]$.

Для построения функции вводятся два пороговых значения: τ_t (максимальное допустимое расхождение по времени) и τ_d (максимальное допустимое расхождение по расстоянию), при превышении которых записи не могут считаться эквивалентными. На их основе определяются две вспомогательные функции, оценивающие близость записей по времени ($sim_{datetime}$) и по расстоянию (sim_{space}):

$$sim_{datetime}(r_1, r_2) = \begin{cases} 1 - \frac{\delta_t(r_1, r_2)}{\tau_t}, & \delta_t(r_1, r_2) \leq \tau_t, \\ 0 & \text{иначе} \end{cases} \quad (4)$$

где $\delta_t(r_1, r_2) = |a_{datetime}(r_1) - a_{datetime}(r_2)|$.

Функция схожести по расстоянию sim_{space} строится аналогично с использованием порогового значения τ_d и геодезического расстояния $\delta_d(r_1, r_2) = \|a_{loc_coords}(r_1) - a_{loc_coords}(r_2)\|$.

Итоговая функция схожести определяется как произведение $sim_{datetime}$ и sim_{space} . Построенная таким образом функция sim гарантирует, что схожесть записей будет высокой только в том случае, если они описывают события, которые произошли на небольшом расстоянии друг от друга и с незначительной временной разницей:

$$sim(r_1, r_2) = sim_{datetime}(r_1, r_2) \cdot sim_{space}(r_1, r_2). \quad (5)$$

Принимаются следующие формальные допущения об источниках шума в записях. Наблюдаемое значение атрибута $a(r)$ рассматривается как сумма истинного значения и ошибки (шума): $a(r) = \Theta_{e(r)} + \varepsilon(r)$. При этом возможны различные типы источников шума ε :

1. Для официальных записей ($R_{official}$) шум ε принимается пренебрежимо малым.

2. Для новостных публикаций (R_{news}) шум включает: неопределенность упоминаний даты и времени в тексте (использование неточных формулировок времени, например, «утром»),

погрешности методов извлечения атрибутов, а также возможные опечатки и фактические искажения в текстах публикаций.

Введение пороговых значений τ_t и τ_d для функции схожести позволяет компенсировать влияние данных источников шума при установлении эквивалентности записей.

Кусочно-линейный вид функций $sim_{datetime}$ и sim_{space} обусловлен необходимостью обеспечения устойчивости метода к зашумленным входным данным при извлечении атрибутов из неструктурированных текстов. В отличие от ступенчатых функций (жестких порогов), такой характер изменения схожести позволяет учитывать записи с умеренным уровнем $\varepsilon(r)$, не исключая их из процесса сопоставления, а пропорционально снижая значение схожести при увеличении расхождения. Это минимизирует риск ложного разделения кластеров при незначительных отклонениях в определении координат или времени события.

3.3. Алгоритм сопоставления записей. Для формирования искомого разбиения множества R на непересекающиеся кластеры, каждый из которых соответствует одному уникальному событию ДТП, предлагается алгоритм, осуществляющий распределение записей по кластерам на основе оценки близости значений их пространственно-временных атрибутов с использованием построенной функции схожести sim и введенных пороговых значений τ_t и τ_d .

Поскольку каждому официально зарегистрированному ДТП соответствует единственная запись в множестве $R_{official}$, а значения атрибутов $a_{datetime}$ и a_{loc_coords} для таких записей считаются точными, целесообразно рассматривать записи из $R_{official}$ как ядра формируемых кластеров. На первом этапе записи из R_{news} распределяются по данным кластерам на основе значений функции схожести sim с их ядрами, запись относится к тому кластеру, с ядром которого достигается максимальное положительное значение схожести. Если схожесть со всеми ядрами равна нулю, делается предположение, что запись описывает ДТП, сведения о котором не были официально опубликованы. На втором этапе такие нераспределенные записи подлежат кластеризации с использованием алгоритма ST-DBSCAN (модификации алгоритма DBSCAN, предназначенной для работы с пространственно-временными данными) [21]. Формальное описание алгоритма представлено в листинге 2.

Вход: $R = R_{\text{official}} \cup R_{\text{news}}$, пороговые значения τ_t, τ_d .
Выход: Разбиение $\mathcal{P} = \{R_1, \dots, R_l\}$ множества R .

Предварительный этап:

1. $C_{\text{init}} \leftarrow \{\{r\} \mid r \in R_{\text{official}}\}$ // Создание начальных кластеров
2. $R_{\text{unassigned}} \leftarrow \emptyset$

Этап 1: распределение записей из R_{news} по начальным кластерам

3. **for** each $r_{\text{news}} \in R_{\text{news}}$ **do**:
4. $r_{\text{official}}^* \leftarrow \arg \max_{r_{\text{official}} \in R_{\text{official}}} \text{sim}(r_{\text{news}}, r_{\text{official}})$
5. $\text{sim}_{\text{max}} \leftarrow \text{sim}(r_{\text{news}}, r_{\text{official}}^*)$
6. **if** $\text{sim}_{\text{max}} > 0$ **then**:
7. Найти кластер $R_i \in C_{\text{init}}$, такой что $r_{\text{official}}^* \in R_i$
8. $R_i \leftarrow R_i \cup \{r_{\text{news}}\}$
9. **else**:
10. $R_{\text{unassigned}} \leftarrow R_{\text{unassigned}} \cup \{r_{\text{news}}\}$
11. **end if**
12. **end for**

Этап 2: Кластеризация нераспределенных записей

13. // Параметры ST-DBSCAN:
14. $\text{eps}_{\text{space}} \leftarrow \tau_d$ // Максимальное расстояние между соседями
15. $\text{eps}_{\text{time}} \leftarrow \tau_t$ // Максимальная временная разница между соседями
16. $\text{min_samples} \leftarrow 2$ // Минимальное число соседей для формирования ядра кластера
17. $(C_{\text{clustered}}, R_{\text{noise}}) \leftarrow \text{ST-DBSCAN}(R_{\text{unassigned}}, \text{eps}_{\text{space}}, \text{eps}_{\text{time}}, \text{min_samples})$
18. $C_{\text{noise}} \leftarrow \{\{r\} \mid r \in R_{\text{noise}}\}$ // Каждая шумовая точка образует отдельный кластер
19. $\mathcal{P} \leftarrow C_{\text{init}} \cup C_{\text{clustered}} \cup C_{\text{noise}}$
20. **return** $\mathcal{P}$

Листинг 2. Алгоритм сопоставления записей

4. Экспериментальная оценка предлагаемого метода.

4.1. Методика экспериментальной оценки. Для экспериментальной оценки предлагаемого метода были собраны и размечены наборы данных о ДТП в Санкт-Петербурге за период с мая по июнь 2025 года. Эти наборы включают:

– Официальные карточки учета ДТП (R_{official}). Множество сформировано путем загрузки данных с официального сайта Госавтоинспекции по всем ДТП, зарегистрированным в исследуемом регионе за указанный период.

– Новостные публикации (R_{news}). Множество сформировано на основе публикаций из шести источников: трех региональных

интернет-СМИ (в частности, «Фонтанка.ру») и трех тематических сообществ в социальной сети «ВКонтакте» (в частности, «ДТП и ЧП | Санкт-Петербург»). Данные были получены в автоматическом режиме с использованием методов веб-скрейпинга.

После завершения сбора записей была выполнена их экспертная разметка, включающая два этапа:

1. Для каждой новостной публикации из R_{news} вручную определены эталонные значения пространственно-временных атрибутов (a_{loc_coords} и $a_{datetime}$) на основе содержащихся в тексте указаний на местоположение и время события ДТП. В итоговые наборы данных включались только те записи, которые описывают одно конкретное ДТП в исследуемом регионе за указанный период и для которых удалось однозначно установить значения обоих атрибутов.

2. Выполнено ручное разбиение всего множества записей $R = R_{official} \cup R_{news}$ на кластеры. Разметка проведена на основе комплексного анализа содержания записей, включающего сопоставление сопроводительных фото- и видеоматериалов, выявление уникальных деталей в описании контекста ДТП, позволяющих с высокой степенью уверенности установить соответствие нескольких записей одному и тому же событию ДТП.

Разметка выполнялась двумя экспертами независимо. Согласованность аннотаторов (оценка через коэффициент каппа Козна κ) составила 0,85 для этапа извлечения атрибутов и 0,83 для этапа сопоставления записей. Выявленные расхождения разрешались совместно.

После завершения разметки собранные данные были разделены на два непересекающихся набора: валидационный набор (записи за май 2025 г.), предназначенный для отладки экспертных правил, настройки запросов к большой языковой модели и подбора пороговых значений τ_t и τ_d ; тестовый набор (записи за июнь 2025 г.), предназначенный для оценки предлагаемого метода. Характеристики наборов данных представлены в таблице 1.

Таблица 1. Характеристики наборов данных для экспериментальной оценки предложенного метода

Набор	Период записей	$ R_{official} $	$ R_{news} $	Кластеры
Валидационный	Май 2025	271	364	339
Тестовый	Июнь 2025	300	389	351

Для экспериментальной оценки выполнена программная реализация предложенного метода. На предварительном этапе единократно формируется локальный географический справочник

для региона Санкт-Петербург на основе данных платформы OpenStreetMap посредством запросов к интерфейсу Overpass API. Полученный справочник включает данные об элементах дорожной сети (улицы, проспекты, шоссе и др.) и объектах инфраструктуры (станции метро, торговые центры и др.). Для каждого объекта сохраняются его уникальный идентификатор, тип, наименование и географические координаты.

На этапе обработки записей, согласно предлагаемому методу, осуществляется анализ множества R_{news} для извлечения значений атрибутов $a_{datetime}$ и a_{loc_coords} из новостных публикаций, после чего выполняется построение разбиения множества R на кластеры. Реализация метода выполнена на языке программирования Python. Хранение данных, включая географический справочник (газетир) и записи о ДТП, организовано в реляционной базе данных PostgreSQL с использованием расширения PostGIS. Инференс большой языковой модели выполнен с использованием библиотеки llama-cpp-python ($temperature = 0$, фиксированный $seed$).

4.1.1. Подбор пороговых значений τ_t и τ_d . Работа алгоритма сопоставления записей (листинг 2) определяется двумя пороговыми значениями τ_t и τ_d . Для подбора значений был использован размеченный валидационный набор данных. Методика подбора основана на анализе распределения пространственно-временных расхождений между записями, которые, согласно эталонной разметке, описывают одно и то же ДТП.

Был проведен анализ чувствительности алгоритма к гиперпараметрам τ_t и τ_d . Исследование выполнялось путем изменения порогов в диапазоне от 85-го до 99-го перцентиля эмпирических распределений внутрикластерных расхождений на валидационном наборе. Установлено, что при значениях ниже 90-го перцентиля наблюдается резкое падение полноты, а выше 97-го перцентиля снижение точности вследствие ложного объединения событий с близкими значениями пространственно-временных атрибутов. При выбранном 95-й перцентиле малые отклонения параметров приводят к незначительным изменениям F1-меры.

Таким образом, в качестве итоговых пороговых значений были выбраны 95-е перцентили соответствующих эмпирических распределений внутрикластерных расхождений по времени и по расстоянию.

4.1.2. Метрики для оценки процедуры извлечения пространственно-временных атрибутов. Работа процедуры извлечения атрибутов a_{loc_coords} и $a_{datetime}$ из записей оценивается путем

сравнения извлеченных значений с эталонными значениями из тестового набора. Используются следующие метрики: полнота извлечения (доля записей, для которых процедура определила значение атрибута), средняя абсолютная ошибка (MAE) для оценки погрешности извлечения атрибута $a_{datetime}$ и средняя ошибка в расстоянии (MDE) для оценки погрешности извлечения атрибута a_{loc_coords} .

Для проверки гипотезы о том, что точность извлечения пространственно-временных атрибутов повышается за счет комбинации классических методов и большой языковой модели, способной интерпретировать сложные языковые конструкции, оценка процедуры извлечения выполняется в двух сценариях:

1. Использование исключительно экспертных правил и географического справочника (газетира).
2. Использование разработанной процедуры, объединяющей экспертные правила, географический справочник (газетир) и большую языковую модель LLaMA.

4.1.3. Метрики для оценки алгоритма сопоставления записей.

Работа алгоритма сопоставления записей (листинг 2) оценивается путем сравнения сформированного им разбиения записей на кластеры с эталонным разбиением записей тестового набора. Для этого используются метрики Точности (Precision), Полноты (Recall) и F1-меры (F1-score), которые вычисляются на основе анализа всех возможных пар записей в наборе. Оценка алгоритма сопоставления записей проводится в двух сценариях:

1. Алгоритм выполняется на тестовом наборе с использованием эталонных (размеченных вручную) значений пространственно-временных атрибутов для записей из R_{news} .
2. Алгоритм выполняется на тестовом наборе с использованием значений пространственно-временных атрибутов, извлеченных разработанной процедурой в сценарии с использованием LLM LLaMA.

4.2. Результаты экспериментальной оценки. На основе анализа эмпирических распределений внутрикластерных расхождений, построенных на валидационном наборе данных, определены пороговые значения τ_t и τ_d для алгоритма сопоставления записей (листинг 2). В качестве значений выбраны 95-е процентиля соответствующих распределений:

1. Максимальное допустимое расхождение по времени: $\tau_t = 4.2$ ч.
2. Максимальное допустимое расхождение по расстоянию: $\tau_d = 0.87$ км.

Оценка работы процедуры извлечения атрибутов $a_{datetime}$ и a_{loc_coords} из записей множества R_{news} для двух сценариев представлена в таблице 2.

Таблица 2. Оценка процедуры извлечения атрибутов из новостных публикаций на тестовом наборе данных

Атрибут	Полнота извлечения	Средняя ошибка
Сценарий 1 (без LLM)		
$a_{datetime}$	0.84	$MAE_{time} = 1.7$ ч.
a_{loc_coords}	0.81	$MDE_{space} = 0.52$ км
Сценарий 2 (с LLM LLaMA)		
$a_{datetime}$	0.8	$MAE_{time} = 1.1$ ч.
a_{loc_coords}	0.76	$MDE_{space} = 0.35$ км

Результаты оценки разбиения записей на кластеры, сформированного разработанным алгоритмом (листинг 2), представлены в таблице 3. Достигнутое значение F1-меры (0.84) признается достаточным для решения задачи сопоставления записей, так как оно соответствует уровню точности современных систем разрешения сущностей, работающих с текстовыми данными из неструктурированных источников [18].

Таблица 3. Оценка работы алгоритма сопоставления записей на тестовом наборе данных

Сценарий эксперимента	Precision	Recall	F1-score
С использованием эталонных значений атрибутов $a_{datetime}$ и a_{loc_coords}	0.96	0.91	0.93
С использованием извлеченных значений атрибутов $a_{datetime}$ и a_{loc_coords}	0.87	0.82	0.84

Для оценки устойчивости предложенного метода к ошибкам входных данных была проведена следующая проверка на тестовом наборе. В значения координат и времени случайно выбранных записей (20% от выборки) был внесен искусственный шум, имитирующий ошибки извлечения: смещение до 200 метров и 30 минут соответственно. Результаты показали, что точность сопоставления снизилась незначительно (F1-мера уменьшилась с 0,84 до 0,82), что подтверждает слабую чувствительность метода к незначительным погрешностям работы LLM и экспертных правил.

5. Обсуждение и заключение. В рамках исследования предложен метод решения задачи сопоставления записей о ДТП из официальных ресурсов и новостных публикаций в СМИ и социальных сетях на основе

комбинированной процедуры извлечения пространственно-временных атрибутов и алгоритма пространственно-временной кластеризации.

В отличие от детерминированных и вероятностных методов, предложенный метод позволяет эффективно обрабатывать неструктурированные тексты новостных публикаций без необходимости предварительного формирования полного набора структурированных атрибутов. В сравнении с методами на основе классического машинного обучения и глубокого обучения, предложенный метод не требует подготовки обширных размеченных наборов данных. Это существенно снижает трудозатраты на предварительную разметку данных о ДТП и обеспечивает возможность применения метода в различных регионах и смежных предметных областях без необходимости дообучения или сложной настройки моделей под специфику новых данных. В отличие от прямого использования больших языковых моделей для решения задачи сопоставления записей, комбинированный метод обеспечивает более высокую стабильность и интерпретируемость результатов. Поскольку роль большой языковой модели ограничена верификацией конкретных значений извлеченных атрибутов, а итоговое решение об эквивалентности записей принимается детерминированным алгоритмом кластеризации на основе пространственно-временных пороговых значений, минимизируется риск влияния галлюцинаций модели на корректность сопоставления записей.

Результаты экспериментальной оценки подтверждают обоснованность выдвинутых в исследовании гипотез. Высокие значения метрик алгоритма сопоставления в сценарии с использованием эталонных данных ($F1\text{-score} = 0,93$) подтверждают первую гипотезу: пространственно-временная близость описываемых в записях событий ДТП является подходящим критерием для установления эквивалентности записей о ДТП. Это обосновывает возможность использования пространственно-временных атрибутов в качестве ключевых признаков для разрешения сущностей в условиях отсутствия общих идентификаторов между гетерогенными источниками данных. Сравнительный анализ сценариев извлечения подтверждает вторую гипотезу: комбинирование классических методов с большой языковой моделью LLaMA повышает точность определения пространственно-временных атрибутов событий ДТП из новостных публикаций за счет семантической валидации упоминаний. Это подтверждает эффективность использования языковых моделей в качестве инструментов верификации для повышения качества извлечения данных из неструктурированных текстов на естественном

языке классическими методами, что позволяет обогащать доступные для анализа данные без необходимости подготовки обширных размеченных наборов для обучения моделей.

Предложенный метод обладает следующими ограничениями:

- Метод не применим к новостным публикациям, содержащим описание нескольких событий ДТП, а также к публикациям, в которых отсутствуют достаточные данные для определения местоположения и времени ДТП;

- Эффективность сопоставления записей существенно зависит от точности работы процедуры извлечения пространственно-временных атрибутов из текстов;

- Эффективность сопоставления записей ограничена полнотой и актуальностью данных OpenStreetMap для исследуемого региона;

- В условиях аномально высокой плотности событий (например, несколько ДТП на одном перекрестке в пределах одного часа) использование исключительно пространственно-временных атрибутов может приводить к ошибочному объединению записей в один кластер;

- Использование большой языковой модели в процедуре извлечения данных накладывает ограничения на скорость обработки потока записей в режиме реального времени из-за значительных требований к вычислительным ресурсам;

- Возможно возникновение галлюцинаций в большой языковой модели. Однако использование локального географического справочника и экспертных правил на этапе формирования кандидатов позволяет минимизировать данный риск, ограничивая область работы LLM верификацией только реально существующих объектов исследуемого региона.

Дальнейшие исследования могут вестись в следующих направлениях:

- Разработка алгоритма декомпозиции текстов, описывающих несколько событий ДТП, для их последующего отдельного сопоставления с использованием предложенного метода;

- Автоматизация процесса обогащения официальных карточек учета ДТП дополнительными данными из кластеризованных новостных публикаций, включая фотоматериалы и описание причин ДТП, для формирования единого расширенного контекста событий;

- Проведение анализа воспроизводимости результатов и экспериментальная оценка адаптивности метода на данных из других регионов и смежных предметных областей. В рамках данного направления предполагается уточнение теоретических границ применимости метода,

в частности, определение критических значений плотности событий в пространстве и времени, при которых сохраняется высокая точность сопоставления записей;

– Создание на основе предложенного метода полнофункциональной аналитической системы, обеспечивающей автоматизированное извлечение записей из первичных источников, определение значений пространственно-временных атрибутов, построение разбиения множества записей на кластеры и визуализацию результатов интеграции данных.

Литература

1. Официальный сайт Госавтоинспекции МВД России. URL: <https://stat.gibdd.ru/> (дата обращения: 01.12.2025).
2. Rabbani M.B.A., Musarat M.A., Alaloul W., et al. Road accident data collection systems in developing and developed countries: a review // *International Journal of Integrated Engineering*. 2022. vol. 14. no. 1. pp. 336–352. DOI: 10.30880/ijie.2022.14.01.031.
3. Chand A., Jayesh S., Bhasi A.B. Road traffic accidents: An overview of data sources, analysis techniques and contributing factors // *Materials Today: Proceedings*. 2021. vol. 47. pp. 5135–5141. DOI: 10.1016/j.matpr.2021.05.415.
4. Dos Santos S.R., Jr Davis C.A., Smarzero R. Integration of data sources on traffic accidents // *Proceedings of the XVII GEOINFO*. 2016. pp. 93–104.
5. Ali F., Ali A., Imran M., et al. Traffic accident detection and condition analysis based on social networking data // *Accident Analysis Prevention*. 2021. vol. 151. 105973 p. DOI: 10.1016/j.aap.2021.105973.
6. Chang H., Li L., Huang J., et al. Tracking traffic congestion and accidents using social media data: A case study of Shanghai // *Accident Analysis Prevention*. 2022. vol. 169. 106618 p. DOI: 10.1016/j.aap.2022.106618.
7. Митряев И.С., Новиков А.Н., Кравченко А.А., Еремин С.В. Методы повышения точности извлечения информации из социальных сетей для интеллектуальных транспортных систем // *Мир транспорта и технологических машин*. 2024. №4-2(87). С. 114–121.
8. Girin A., Teslya N., Shilov N. Overview of Publicly-Available Data Sources on Road Traffic Accidents in Russia // *Proceedings of the 10th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. 2024. pp. 480–487. DOI: 10.5220/0012737100003702.
9. Fellegi I.P., Sunter A.B. A Theory for Record Linkage // *Journal of the American Statistical Association*. 1969. vol. 64. no. 328. pp. 1183–1210. DOI: 10.1080/01621459.1969.10501049.
10. Binette O., Steorts R.C. (Almost) all of entity resolution // *Science Advances*. 2022. vol. 8. no. 12. eabi8021 p. DOI: 10.1126/sciadv.abi8021.
11. Elmagarmid A.K., Ipeirotis P.G., Verykios V.S. Duplicate Record Detection: A Survey // *IEEE Transactions on Knowledge and Data Engineering*. 2007. vol. 19. no. 1. pp. 1–16. DOI: 10.1109/TKDE.2007.250581.
12. Mudgal S., Li H., Rekatsinaset T., et al. Deep Learning for Entity Matching: A Design Space Exploration // *Proceedings of the International Conference on Management of Data*. 2018. pp. 19–34.
13. Li Y., Li J., Suhara Y., et al. Effective entity matching with transformers // *The VLDB Journal*. 2023. vol. 32. pp. 1215–1235. DOI: 10.1007/s00778-023-00779-z.

14. Li Q., Li J., Sheng J., et al. A Survey on Deep Learning Event Extraction: Approaches and Applications // *IEEE Transactions on Neural Networks and Learning Systems*. 2024. vol. 35. no. 5. pp. 6301–6321. DOI: 10.1109/TNNLS.2022.3213168.
15. Paganelli M., Tiano D., Del Buono F., et al. How Transformers are Revolutionizing Entity Matching // *Proceedings of the 32nd Symposium on Advanced Database Systems*. 2024. vol. 3741. pp. 662–670.
16. Peeters R., Steiner A., Bizer C. Entity Matching using Large Language Models // *arXiv preprint arXiv:2310.11244*. 2023.
17. Marvin G., Hellen N., Jjing D., et al. Prompt Engineering in Large Language Models // *Algorithms for Intelligent Systems*. Springer. 2023. pp. 387–402. DOI: 10.1007/978-981-99-7962-2_30.
18. Peeters R., Bizer C. Using ChatGPT for Entity Matching // *Communications in Computer and Information Science*. Springer. 2023. vol. 1850. pp. 221–230. DOI: 10.1007/978-3-031-42941-5_20.
19. Li H., Li S., Hao F., et al. BoostER: Leveraging Large Language Models for Enhancing Entity Resolution // *Companion Proceedings of the ACM Web Conference*. 2024. pp. 1043–1046. DOI: 10.1145/3589335.3651245.
20. Touvron H., Lavril T., Izacard G., et al. LLaMA: Open and Efficient Foundation Language Models // *arXiv preprint arXiv:2302.13971*. 2023.
21. Birant D., Kut A. ST-DBSCAN: An algorithm for clustering spatial–temporal data // *Data Knowledge Engineering*. 2007. vol. 60. no. 1. pp. 208–221. DOI: 10.1016/j.datak.2006.01.013.
22. Strotgen J., Gertz M. Heildeltime: High quality rule-based extraction and normalization of temporal expressions // *Proceedings of the 5th International Workshop on Semantic Evaluation*. 2010. pp. 321–324.
23. Giridhar P., Abdelzaher T., George J., et al. On quality of event localization from social network feeds // *Proceedings of the IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. 2015. pp. 75–80. DOI: 10.1109/PERCOMW.2015.7133997.
24. Milusheva S., Marty R., Bedoya G., et al. Applying machine learning and geolocation techniques to social media data (Twitter) to develop a resource for urban planning // *PLOS One*. 2021. vol. 16. no. 2. e0244317 p. DOI: 10.1371/journal.pone.0244317.
25. Suat-Rojas N., Gutierrez-Osorio C., Pedraza C. Extraction and Analysis of Social Networks Data to Detect Traffic Accidents // *Information*. 2022. vol. 13. no. 1. 26 p. DOI: 10.3390/info13010026.
26. Hu X., Zhou Z., Li H., et al. Location Reference Recognition from Texts: A Survey and Comparison // *ACM Computing Surveys*. 2023. vol. 56. no. 5. pp. 1–37. DOI: 10.1145/3625819.
27. Xiang W., Wang B. A Survey of Event Extraction From Text // *IEEE Access*. 2019. vol. 7. pp. 173111–173137. DOI: 10.1109/ACCESS.2019.2956831.
28. Hu Y., Mai G., Cundy C., et al. Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages // *International Journal of Geographical Information Science*. 2023. vol. 37. no. 11. pp. 2289–2318. DOI: 10.1080/13658816.2023.2266495.

Гирин Алексей Романович — аспирант, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО» (ИТМО). Область научных интересов: интеграция данных, обработка текста на естественном языке, языковые модели. Число научных публикаций — 8.

agirin@itmo.ru; Кронверкский пр., 49А, 197101, Санкт-Петербург, Россия; р.т.: +7(812)480-0370.

Тесля Николай Николаевич — канд. техн. наук, старший научный сотрудник, лаборатория интегрированных систем автоматизации, Федеральное государственное бюджетное учреждение науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН). Область научных интересов: управление знаниями, онтологии, обработка текста на естественном языке, языковые модели. Число научных публикаций — 100. teslya@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)508-3311.

Поддержка исследований. Работа выполнена в рамках бюджетной темы FFZF-2025-0003.

A. GIRIN, N. TESLYA
**RECORD LINKAGE METHOD FOR ROAD TRAFFIC ACCIDENTS
USING OFFICIAL DATA SOURCES AND NEWS PUBLICATIONS
BASED ON SPATIO-TEMPORAL CLUSTERING**

Girin A., Teslya N. Record Linkage Method for Road Traffic Accidents Using Official Data Sources and News Publications Based on Spatio-Temporal Clustering.

Abstract. Effective decision-making for reducing the road accident rate in motor transport is based on the results of analyzing traffic accident hotspots. The completeness and accuracy of accident data used for analysis can be enhanced through the integration of multiple sources. A crucial step in integration is the matching of records from different sources that describe the same traffic accident events. Solving the matching problem is complicated by heterogeneity: accident records may differ in structure and format. This study proposes a method for matching accident records from two heterogeneous sources: official resources with structured traffic accident report cards and news publications in the form of unstructured natural language texts. The proposed method is based on the hypothesis of spatial and temporal proximity: records from different sources are considered to correspond to the same accident if the events described occurred within a short distance and with a negligible time difference. The method comprises two sequential stages. At the first stage, coordinates and date/time of the described event are extracted from news publications using expert rules, a geographic reference database, and the LLaMA large language model. At the second stage, records from official resources and news publications are matched using spatiotemporal clustering with the ST-DBSCAN algorithm. The method enables the enrichment of the contextual description of officially registered traffic accidents and enhances the data available for analysis by incorporating accidents not recorded in official sources. Experimental evaluation of the method was conducted on real-world traffic accident data from Saint Petersburg. The results demonstrated that combining classical methods with a large language model improves the accuracy of extracting spatiotemporal attributes of traffic accidents from news publications, while spatiotemporal clustering ensures high accuracy in matching accident records (F1-score = 0.84).

Keywords: data integration, record linkage, traffic accidents, gazetteer, expert rules, ST-DBSCAN, large language models, LLaMA.

References

1. Oficial'nyj sajt Gosavtoinspektzii MVD Rossii [Official web site of the State Traffic Inspectorate of the Ministry of Internal Affairs of Russia]. Available at: <https://stat.gibdd.ru/> (accessed 01.12.2025). (In Russ.).
2. Rabbani M.B.A., Musarat M.A., Alaloul W., et al. Road accident data collection systems in developing and developed countries: a review. *International Journal of Integrated Engineering*. 2022. vol. 14. no. 1. pp. 336–352. DOI: 10.30880/ijie.2022.14.01.031.
3. Chand A., Jayesh S., Bhasi A.B. Road traffic accidents: An overview of data sources, analysis techniques and contributing factors. *Materials Today: Proceedings*. 2021. vol. 47. pp. 5135–5141. DOI: 10.1016/j.matpr.2021.05.415.
4. Dos Santos S.R., Jr Davis C.A., Smarzaro R. Integration of data sources on traffic accidents. *Proceedings of the XVII GEOINFO*. 2016. pp. 93–104.

5. Ali F., Ali A., Imran M., et al. Traffic accident detection and condition analysis based on social networking data. *Accident Analysis Prevention*. 2021. vol. 151. 105973 p. DOI: 10.1016/j.aap.2021.105973.
6. Chang H., Li L., Huang J., et al. Tracking traffic congestion and accidents using social media data: A case study of Shanghai. *Accident Analysis Prevention*. 2022. vol. 169. 106618 p. DOI: 10.1016/j.aap.2022.106618.
7. Mitryaev I.S., Novikov A.N., Kravchenko A.A., Eremin S.V. [Methods for improving the accuracy of information extraction from social networks for intelligent transport systems]. *Mir transporta i tekhnologicheskikh mashin – The world of transport and technological machines*. 2024. no. 4-2(87). pp. 114–121. (In Russ.).
8. Girin A., Teslya N., Shilov N. Overview of Publicly-Available Data Sources on Road Traffic Accidents in Russia. *Proceedings of the 10th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*. 2024. pp. 480–487. DOI: 10.5220/0012737100003702.
9. Fellegi I.P., Sunter A.B. A Theory for Record Linkage. *Journal of the American Statistical Association*. 1969. vol. 64. no. 328. pp. 1183–1210. DOI: 10.1080/01621459.1969.10501049.
10. Binette O., Steorts R.C. (Almost) all of entity resolution. *Science Advances*. 2022. vol. 8. no. 12. eabi8021 p. DOI: 10.1126/sciadv.abi8021.
11. Elmagarmid A.K., Ipeirotis P.G., Verykios V.S. Duplicate Record Detection: A Survey. *IEEE Transactions on Knowledge and Data Engineering*. 2007. vol. 19. no. 1. pp. 1–16. DOI: 10.1109/TKDE.2007.250581.
12. Mudgal S., Li H., Rekatsinas T., et al. Deep Learning for Entity Matching: A Design Space Exploration. *Proceedings of the International Conference on Management of Data*. 2018. pp. 19–34.
13. Li Y., Li J., Suhara Y., et al. Effective entity matching with transformers. *The VLDB Journal*. 2023. vol. 32. pp. 1215–1235. DOI: 10.1007/s00778-023-00779-z.
14. Li Q., Li J., Sheng J., et al. A Survey on Deep Learning Event Extraction: Approaches and Applications. *IEEE Transactions on Neural Networks and Learning Systems*. 2024. vol. 35. no. 5. pp. 6301–6321. DOI: 10.1109/TNNLS.2022.3213168.
15. Paganelli M., Tiano D., Del Buono F., et al. How Transformers are Revolutionizing Entity Matching. *Proceedings of the 32nd Symposium on Advanced Database Systems*. 2024. vol. 3741. pp. 662–670.
16. Peeters R., Steiner A., Bizer C. Entity Matching using Large Language Models. *arXiv preprint arXiv:2310.11244*. 2023.
17. Marvin G., Hellen N., Jjing D., et al. Prompt Engineering in Large Language Models. *Algorithms for Intelligent Systems*. Springer. 2023. pp. 387–402. DOI: 10.1007/978-981-99-7962-2_30.
18. Peeters R., Bizer C. Using ChatGPT for Entity Matching. *Communications in Computer and Information Science*. Springer. 2023. vol. 1850. pp. 221–230. DOI: 10.1007/978-3-031-42941-5_20.
19. Li H., Li S., Hao F., et al. BoostER: Leveraging Large Language Models for Enhancing Entity Resolution. *Companion Proceedings of the ACM Web Conference*. 2024. pp. 1043–1046. DOI: 10.1145/3589335.3651245.
20. Touvron H., Lavril T., Izacard G., et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971*. 2023.
21. Birant D., Kut A. ST-DBSCAN: An algorithm for clustering spatial–temporal data. *Data Knowledge Engineering*. 2007. vol. 60. no. 1. pp. 208–221. DOI: 10.1016/j.datak.2006.01.013.

22. Strotgen J., Gertz M. Heildeltime: High quality rule-based extraction and normalization of temporal expressions. *Proceedings of the 5th International Workshop on Semantic Evaluation*. 2010. pp. 321–324.
23. Giridhar P., Abdelzaher T., George J., et al. On quality of event localization from social network feeds. *Proceedings of the IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. 2015. pp. 75–80. DOI: 10.1109/PERCOMW.2015.7133997.
24. Milusheva S., Marty R., Bedoya G., et al. Applying machine learning and geolocation techniques to social media data (Twitter) to develop a resource for urban planning. *PLOS One*. 2021. vol. 16. no. 2. e0244317 p. DOI: 10.1371/journal.pone.0244317.
25. Suat-Rojas N., Gutierrez-Osorio C., Pedraza C. Extraction and Analysis of Social Networks Data to Detect Traffic Accidents. *Information*. 2022. vol. 13. no. 1. 26 p. DOI: 10.3390/info13010026.
26. Hu X., Zhou Z., Li H., et al. Location Reference Recognition from Texts: A Survey and Comparison. *ACM Computing Surveys*. 2023. vol. 56. no. 5. pp. 1–37. DOI: 10.1145/3625819.
27. Xiang W., Wang B. A Survey of Event Extraction From Text. *IEEE Access*. 2019. vol. 7. pp. 173111–173137. DOI: 10.1109/ACCESS.2019.2956831.
28. Hu Y., Mai G., Cundy C., et al. Geo-knowledge-guided GPT models improve the extraction of location descriptions from disaster-related social media messages. *International Journal of Geographical Information Science*. 2023. vol. 37. no. 11. pp. 2289–2318. DOI: 10.1080/13658816.2023.2266495.

Girin Alexey — Graduate Student, ITMO University. Research interests: data integration, natural language text processing, language models. The number of publications — 8. agirin@itmo.ru; 49A, Kronverkskiy Ave., 197101, St. Petersburg, Russia; office phone: +7(812)480-0370.

Teslya Nikolay — Ph.D., Senior Researcher, Laboratory of Integrated Automation Systems, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: knowledge management, ontologies, natural language text processing, language models. The number of publications — 100. teslya@ias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)508-3311.

Acknowledgements. The work was supported out within the framework of the budget theme FFZF-2025-0003.

С.В. ЗАТЕЕВ, Е.Ю. КОСТЮЧЕНКО, А.А. ШЕЛУПАНОВ
**СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПОДХОДОВ
К ПРОЕКТИРОВАНИЮ АРХИТЕКТУРЫ СВЕРТОЧНЫХ
НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА КЛАВИАТУРНОГО
ПОЧЕРКА**

Затеев С.В., Костюченко Е.Ю., Шелупанов А.А. Сравнительный анализ подходов к проектированию архитектуры сверточных нейронных сетей для анализа клавиатурного почерка.

Аннотация. В условиях повсеместной цифровизации и учащения случаев компрометации конфиденциальной информации актуальность разработки надежных систем аутентификации, выступающих в качестве первоочередного защитного рубежа, становится критически важной. Перспективным решением проблем, связанных с утратой, кражей или несанкционированным использованием статических идентификаторов, представляются биометрические системы поведенческого типа. В частности, методы, основанные на анализе клавиатурного почерка (динамики работы с клавиатурой), позволяют осуществлять непрерывную и скрытую аутентификацию пользователя на протяжении всего сеанса работы, анализируя его уникальные и трудно подделываемые поведенческие паттерны. Данное направление характеризуется сочетанием высокого уровня безопасности с относительно умеренными затратами на внедрение и высоким удобством для конечного пользователя. Целью настоящего исследования является проведение сравнительного анализа различных подходов к проектированию и построению архитектур сверточных нейронных сетей, оптимизированных для обработки разреженных данных, описывающих динамику нажатий клавиш. В ходе экспериментальной работы было установлено, что сверточные нейронные сети демонстрируют исключительно высокую точность верификации, достигающую 99%, несмотря на то, что входные данные, сформированные в результате предобработки, обладают сильной разреженностью. Наилучшие результаты в рамках исследования были достигнуты при использовании архитектуры, основанной на механизме разделенной свертки, которая показала свою эффективность для работы с указанным специфическим типом данных.

Ключевые слова: аутентификация, поведенческая биометрия, клавиатурный почерк, разреженные данные, сверточные нейронные сети.

1. Введение. Современное общество характеризуется глубокой интеграцией информационных систем в различные сферы человеческой деятельности. В цифровой среде хранятся значительные массивы данных, включая персональную информацию и сведения, составляющие коммерческую тайну. Как следствие, задача обеспечения информационной безопасности, подразумевающая комплекс мер по защите данных от несанкционированного вмешательства, стала одним из ключевых направлений в области информационных технологий. Ее актуальность постоянно поддерживается прогрессом в методах кибератак, что требует адекватного и опережающего развития защитных механизмов.

В данном контексте системы аутентификации, выполняющие роль первоначального барьера на пути к информационным ресурсам, становятся критически важными. Их фундаментальная задача – верификация подлинности субъекта путем сопоставления предъявленных характеристик с эталонными образцами, зарегистрированными в системе.

Существующие методы аутентификации традиционно классифицируются по типу используемого идентификатора:

1. Знание: использование паролей или PIN-кодов. Несмотря на простоту реализации, этот подход уязвим к краже, фишингу и brute-force атакам, что ставит под сомнение его надежность.

2. Владение: применение физических носителей, таких как токены или карты доступа. Главные недостатки включают риск утери или кражи идентификатора, а также необходимость в дополнительной инфраструктуре, что повышает сложность и стоимость системы.

3. Биометрические данные: использование уникальных биологических или поведенческих характеристик. Биометрию принято разделять на физиологическую и поведенческую. К физиологической относят отпечатки пальцев, радужную оболочку глаза и другие неизменяемые признаки. Такие данные надежны, но статическая природа делает их уязвимыми для компрометации и воспроизведения. Поведенческая биометрия включает голос, походку и динамику работы с клавиатурой. Эти характеристики сложнее подделать благодаря отсутствию статичного шаблона. Однако их изменчивость под влиянием состояния пользователя может снижать стабильность и надежность.

Ключевой недостаток традиционных систем статической аутентификации – их неспособность обнаружить факт смены пользователя после первоначального входа. Биометрические системы, основанные на поведенческих характеристиках, предлагают решение этой проблемы, обеспечивая постоянный мониторинг.

Особый интерес в этом направлении представляют системы, анализирующие клавиатурный почерк – поведенческий биометрический идентификатор, основанный на уникальных паттернах взаимодействия пользователя с клавиатурой. Данный подход обладает рядом значительных преимуществ:

- ненавязчивость и прозрачность: не требуют дополнительного оборудования, используя стандартные устройства ввода;
- непрерывность: позволяет осуществлять верификацию на протяжении всего сеанса работы;

– многофакторность: анализирует комплекс параметров, таких как ритм печати, длительность нажатий и интервалы между ними.

Современные стандарты в области информационной безопасности признают клавиатурный почерк в качестве уникальной и перспективной характеристики для построения динамических систем защиты. В то время как аутентификация по статическим парольным фразам является хорошо изученной областью [1, 2, 3], исследования в сфере непрерывной поведенческой биометрии, в частности анализа клавиатурного почерка, находятся на стадии активного развития [3, 4]. Это направление представляется чрезвычайно перспективным, поскольку позволяет достичь высокого уровня безопасности, сочетая его с удобством для пользователя и экономической эффективностью внедрения.

Как показывает практика, в силу специфики формата данных, необходимого для обработки сверточными нейронными сетями, в итоговый набор данных попадает множество признаков, не обладающих должным уровнем информативности. Полученные матрицы характеризуются высокой степенью разреженности, что может приводить к снижению точности распознавания пользователей.

Таким образом, необходимо провести анализ эффективности использования сверточных нейронных сетей для решения задачи аутентификации по фактору клавиатурного почерка, построив архитектуру, которая бы грамотно обрабатывала такие разреженные данные. Кроме того, необходимо определиться с параметрами формирования многоканальных матриц, обеспечивающими лучшее качество работы модели сверточной нейронной сети.

2. Набор данных. Исходный набор данных, описывающий поток событий, фиксируемых при взаимодействии пользователя с клавиатурой, разбивается на интервалы, в соответствии с принципами, описанными в работах [5 – 9]. Интервалы формируются как пересекающиеся окна, содержащие фиксированное количество событий. В качестве параметров используются те, которые показали лучший результат в работах [5, 6]: количество событий в окне – 2000, пересечение – 50% от длины окна. На каждом интервале производится расчет признаков биграмм как в работе [10], однако в данном случае биграммы составляются из наиболее часто встречающихся клавиш [5]. Частоты употребления считаются по всей имеющейся выборке. Также в данной работе кроме среднего выборочного значения признака для каждого интервала используется расширенный набор статистических характеристик: среднее квадратичное отклонение, коэффициенты асимметрии и эксцесса. Кроме того, в данной работе

используются лишь три из пяти каналов, употребляемых в работе [10]. Так как продолжительность удержания отдельной клавиши и любой из 4-х признаков биграмм могут быть рассчитаны на основании оставшихся 3-х признаков набора биграмм. Рассчитанные признаки собираются в многоканальную матрицу по следующему принципу: измерения каждого канала – первая и вторая клавиши, составляющие биграмму, каналы – признаки биграмм. Признаки биграмм выделяются на основании принципов, изложенных в работах [5, 8, 11]. Размер матрицы определяется количеством используемых клавиш, чем больше размер, тем больше информации помещает в себя матрица, но увеличивается время ее обработки. Кроме того, чем реже используются клавиши, тем реже используются биграммы, содержащие их, а значит, появляется риск возникновения большого количества пустых значений в матрице, что может негативно сказаться на качестве работы модели. В данной работе осуществляется следующий принцип подбора значений: исследуются от 5 до 50 (с шагом 5) наиболее часто встречающихся клавиш.

Формирование многоканальной матрицы иллюстрирует рисунок 1. На основе каждого интервала выделяются 4 сегмента матрицы, каждый из которых заполняется значениями одной статистической характеристики (среднее выборочное в рамках интервала, среднее квадратичное отклонение, коэффициенты асимметрии и эксцесса). Каждый сегмент представляет собой матрицу размером $K \times K$ (K – количество наиболее часто встречающихся клавиш), содержащую 3 канала – по одному на каждый признак биграммы. Для набора данных, содержащих только средние значения, используется только один из сегментов (трехканальная матрица размера $K \times K$). Для набора с полным набором статистических характеристик проводится конкатенация четырех сегментов, которая укладывает их в блок 2×2 , как показано на рисунке 1, и увеличивает пространственный размер матрицы вдвое по каждой оси. При этом соответствующие каналы каждого сегмента конкатенируются между собой. В результате получается трехканальная матрица размером $2K \times 2K$.

Итоговый набор данных представляет собой совокупность интервалов из исходного набора данных, преобразованных к формату многоканальных матриц. На рисунке 2 представлено распределение интервалов по пользователям. Суммарно выделено 11 287 интервалов. Как можно заметить на рисунке, в данных нет очевидного дисбаланса классов. Один класс превосходит в 2 раза по количеству образцов средний уровень по набору данных, но это превосходство

нивелируется количеством классов. Превосходящий класс не занимает даже 10% от всего набора.

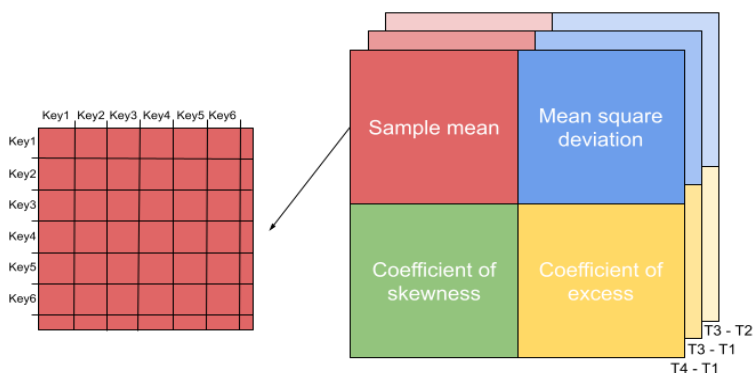


Рис. 1. Образование итоговой многоканальной матрицы

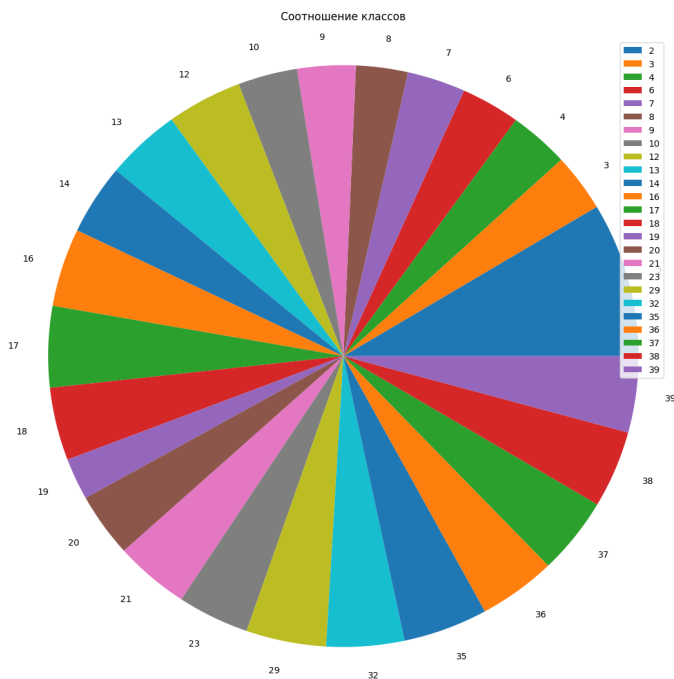


Рис. 2. Соотношение классов в итоговом наборе данных

3. Архитектуры модели

3.1. Классическая сверточная нейронная сеть. Сверточная нейронная сеть (CNN) – это специализированная архитектура глубокого обучения, разработанная для обработки данных с сеточной структурой, таких как изображения, временные ряды и пространственные данные. Основная идея CNN заключается в автоматическом извлечении иерархических признаков с помощью операций свертки и пулинга, что позволяет эффективно учитывать локальные зависимости в данных.

Архитектура используемой модели представлена на рисунке 3. Она состоит из N блоков свертки, каждый из которых состоит из двух сверток 3×3 и максимального пулинга 2×2 , количество каналов в каждом блоке прямо пропорционально порядковому номеру блока и кратно 32. Последний блок состоит из двух полносвязных слоев классификатора с 200 нейронами в первом полносвязном слое и 25 – во втором.

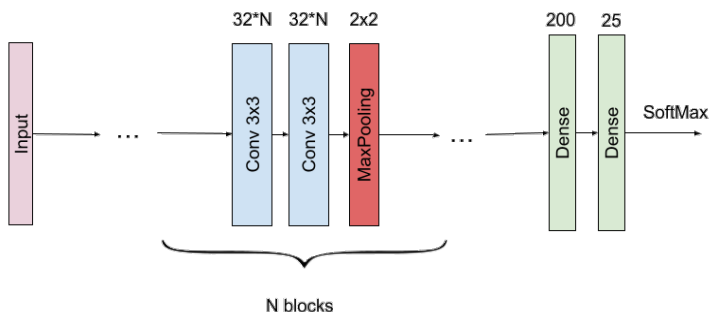


Рис. 3. Архитектура сверточной нейронной сети

Такая архитектура обусловлена желанием использовать максимум возможностей сверточных нейронных сетей для оценки информативности признаков, выделенных при помощи разных подходов. Размерность матриц в экспериментах варьируется от 5×5 (только средние значения и 5 наиболее часто употребляемых клавиш) до 100×100 (полный набор статистических характеристик, удваивающий размер матрицы, и 50 наиболее часто употребляемых клавиш). Архитектура, достаточно глубокая для обработки матрицы 100×100 , не сможет свернуть матрицу 5×5 . И наоборот: сеть, оптимальная для маленьких матриц, окажется недостаточно глубокой для анализа больших. Из этих соображений архитектура подбирает глубину сети в зависимости от размерности многоканальной матрицы.

Так как главное ограничение, вытекающее из маленького размера матрицы, является невозможность свернуть ее определенное количество раз при помощи пулинга, уменьшающего размерность в 2 раза, то и максимальная глубина сети рассчитывается как количество раз, которое размерность матрицы можно сократить вдвое, не превратив ее в единицу (1).

$$MaxDepth = \lfloor \log_2 MatrixDimension \rfloor - 1, \quad (1)$$

где $MatrixDimension$ – количество строк в квадратной матрице.

3.2. Дилатированная сверточная нейронная сеть.

Дилатированная сверточная нейронная сеть (DCNN) – это разновидность сверточной нейронной сети, использующая дилатированную свертку – операцию, применяемую к входным данным с пропуском элементов между ядром согласно параметру, называемому коэффициент дилатации. Коэффициент дилатации определяет, насколько растягивается ядро свертки. [12]

Архитектура используемой модели почти полностью идентична предыдущей и представлена на рисунке 4. Она состоит из N блоков свертки, каждый из которых состоит из двух сверток 3×3 и максимального пулинга 2×2 , количество каналов в каждом блоке прямо пропорционально порядковому номеру блока и кратно 32. Отличие от предыдущей архитектуры заключается в том, что первая из сверток в каждом блоке является дилатированной (разреженной). Последний блок состоит из двух полносвязных слоев классификатора с 200 нейронами в первом полносвязном слое и 25 – во втором. Расчет глубины осуществляется способом, обозначенным в (1).

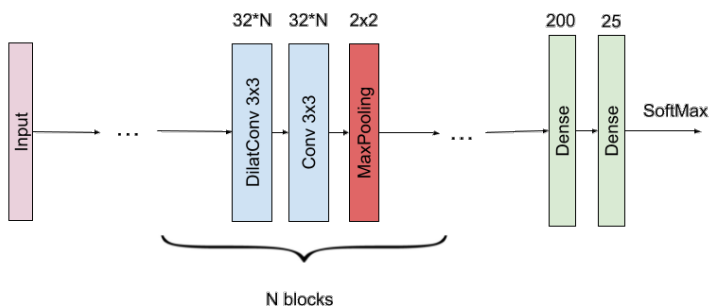


Рис. 4. Архитектура дилатированной сверточной сети

Логика использования такой архитектуры в рамках данной задачи предельно проста. В анализе изображений при помощи нейронных сетей свертка выполняет одну важную функцию – она вычисляет контекст пикселя изображения. Таким образом, при анализе учитываются не только значения того или иного пикселя, но и их взаимное расположение. Очевидно, что в задачах распознавания объектов, сегментации и т.д. наиважнейшие зависимости будут касаться ближайших пикселей (если текущий пиксель принадлежит искомому объекту или определенной закономерности, то и ближайшие ему пиксели с большой вероятностью будут принадлежать тому же объекту или той же закономерности). Однако, это не совсем характерно для текущей задачи. Строго говоря, собранные многоканальные матрицы кроме формы имеют мало общего с действительными изображениями, так как пространственные зависимости между пикселями в данном контексте носят абстрактный характер, в отличие от геометрических зависимостей в реальных изображениях. Размерности каналов матрицы соответствуют наиболее часто встречающимся клавишам. Но эти клавиши далеко не всегда входят в состав наиболее часто встречающихся биграмм. Как следствие, плотность последних в итоговых матрицах невелика, что и подтверждается высокой степенью разреженности данных. Поэтому важные зависимости далеко не всегда находятся между соседними пикселями и могут быть распределены по всей матрице. В связи с этим имеет смысл использовать разреженное ядро свертки для поиска более удаленных друг от друга зависимостей.

3.3. Сверточная нейронная сеть с комбинированной архитектурой. Предположительно, дилатированная свертка в чистом виде может быть неоптимальна для данной задачи, так как важные зависимости могут существовать как на локальном, так и на глобальном уровне. Поэтому была предложена комбинированная архитектура, объединяющая классическую и дилатированную свертки. Она представлена на рисунке 5.

Архитектура состоит из N блоков свертки, каждый из которых имеет следующий состав: распараллеленные слои классической свертки 3×3 и дилатированной свертки 3×3 с количеством каналов в каждом слое прямо пропорционально порядковому номеру блока и кратно 16, результаты обработки этими слоями конкатенируются и поступают на блок классической свертки 3×3 с количеством выходных каналов, прямо пропорциональным порядковому номеру блока и кратно 32. Последний блок состоит из двух полносвязных слоев классификатора с 200

нейронами в первом полносвязном слое и 25 – во втором. Расчет глубины осуществляется способом, обозначенным в (1).

Логика использования такой архитектуры следующий. Если важные зависимости могут существовать как на локальном, так и на глобальном уровне, значит, имеется больше логики в поиске зависимостей в признаках, расположенных в соседних ячейках матрицы. Однако, если в использовании дилатированной свертки есть смысл, то один из параллельных слоев будет анализировать соседние ячейки, а другой – расположенные на расстоянии одной ячейки друг от друга. Тем самым происходит одновременно анализ на большем пространстве, как при дилатированной свертке, но не теряются зависимости на меньшем пространстве.

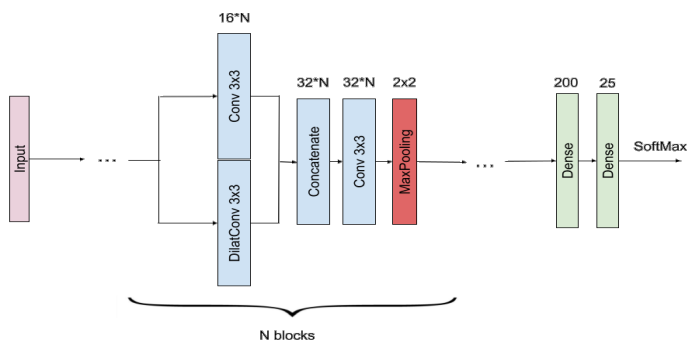


Рис. 5. Архитектура комбинированной сверточной сети

3.4. Сверточная нейронная сеть с увеличенным размером ядра.

Дилатированная свертка нацелена на поиск удаленных зависимостей, однако не исключено, что для данной задачи может потребоваться анализ более широкого локального контекста, чем обеспечивает ядро 3×3 . Альтернативным способом расширения анализируемого контекста является увеличение размера ядра свертки. Архитектура с ядром 5×5 представлена на рисунке 6. Она состоит из N блоков свертки, каждый из которых состоит из двух последовательных сверток 5×5 , 3×3 и максимального пулинга 2×2 , количество каналов в каждом блоке прямо пропорционально порядковому номеру блока и кратно 32. Последний блок состоит из двух полносвязных слоев классификатора с 200 нейронами в первом полносвязном слое и 25 – во втором. Расчет глубины осуществляется способом, обозначенным в (1).

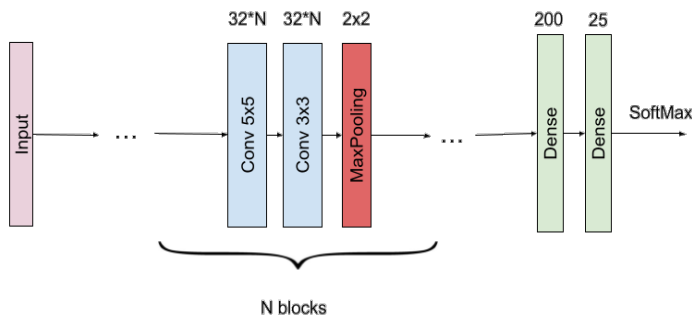


Рис. 6. Архитектура сверточной сети с увеличенным ядром

3.5. Поканальная разделенная сверточная нейронная сеть.

Поканальная разделенная сверточная нейронная сеть (Depthwise Separable Convolution) – это разновидность сверточной нейронной сети, использующая разделенную свертку – операцию, разделяющую стандартную свертку на два этапа: depthwise-свертку, где каждый входной канал обрабатывается отдельным фильтром, и pointwise-свертку (1x1), которая комбинирует результаты через линейную проекцию. Такой подход значительно сокращает количество параметров и вычислительную сложность при сохранении способности извлекать пространственные и межканальные зависимости [13].

Архитектура используемой модели почти полностью идентична классической и представлена на рисунке 7. Она состоит из N блоков свертки, каждый из которых состоит из двух сверток 3x3 и максимального пулинга 2x2, количество каналов в каждом блоке прямо пропорционально порядковому номеру блока и кратно 32. Отличие от предыдущей архитектуры заключается в том, что обе свертки в каждом блоке являются разделенными. Последний блок состоит из двух полносвязных слоев классификатора с 200 нейронами в первом полносвязном слое и 25 – во втором. Расчет глубины осуществляется способом, обозначенным в (1).

Логика использования такой архитектуры в рамках данной задачи заключается в следующем. Так как такие слои разделяют пространственную и межканальную свертки, предоставляется возможность обнаружить зависимости между признаками на различных каналах (т.е. различных признаков одних и тех же биграмм). Также pointwise-свертка позволяет сконцентрироваться не только на контексте признаков, но и на значениях самих признаков. Кроме того, такие архитектуры эффективны на малых данных, что соответствует текущей задаче (всего 11 тысяч образцов).

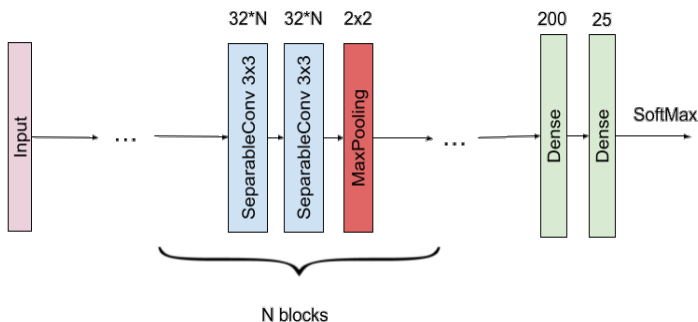


Рис. 7. Архитектура разделенной сверточной сети

3.6. Сверточная нейронная сеть с маскированием. В данной задаче маска – это матрица той же размерности, что и некоторый набор признаков внутри сверточной нейронной сети. Прежде чем этот набор признаков будет обработан очередной сверткой, рассчитывается маска, содержащая нули на тех же позициях, где содержат нули исходные данные, и единицы на остальных позициях. После обработки набора признаков сверткой каждый канал результата этой обработки поэлементно умножается на маску. Такой подход позволяет напрямую указать нейронной сети на каких позициях в наборе признаков находятся реальные данные, а на каких – пустые ячейки.

Архитектура используемой модели почти полностью идентична классической и представлена на рисунке 8. Она состоит из N блоков свертки, каждый из которых состоит из двух сверток 3×3 и максимального пулинга 2×2 , количество каналов в каждом блоке прямо пропорционально порядковому номеру блока и кратно 32. Отличие от предыдущей архитектуры заключается в том, что в начале каждого блока рассчитывается маска для набора признаков, а после каждой свертки осуществляется умножение результата обработки на эту маску. Последний блок состоит из двух полносвязных слоев классификатора с 200 нейронами в первом полносвязном слое и 25 – во втором. Расчет глубины осуществляется способом, обозначенным в (1).

Маскирование применяется для явного указания значимых областей в разреженных данных, предотвращая распространение нулевых значений через слои нейронной сети. В предлагаемом подходе бинарная маска инициализируется на входном слое каждого блока. На каждом сверточном слое маска поэлементно умножается на выходные карты признаков, эффективно обнуляя области без исходной информации. Это позволяет сети фокусироваться только

на релевантных паттернах и предотвращает обучение на артефактах, возникающих из-за разреженности данных. Реализованный механизм сквозного маскирования существенно улучшает устойчивость модели к зашумленным и неполным данным, что критически важно для задач биометрической идентификации по динамике нажатия клавиш.

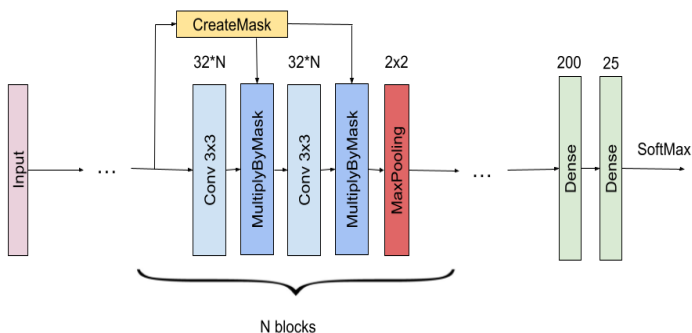


Рис. 8. Архитектура маскированной сверточной сети

4. Процедура разбиения данных и оценки устойчивости полученных результатов. Исходный набор данных изначально разбит на сессии для каждого пользователя. Ввиду ограниченного количества сессий и их существенно различной длины, разбиение на обучающую, валидационную и тестовую выборки выполнялось вручную с соблюдением следующих условий:

- для каждого пользователя выборки формировались из целых сессий, при этом все записи одной сессии относились строго к одной выборке (таким образом, исключалось попадание записей из одной сессии в разные выборки, загрязнение тестовой выборки не происходит);

- целевое соотношение объемов записей для каждого пользователя составляло примерно 80 % (обучающая), 10 % (валидационная) и 10 % (тестовая);

- на полученном фиксированном разбиении для каждой из рассматриваемых моделей проводилось 5 независимых запусков обучения, результатом модели считалось выборочное среднее по этим пяти запускам.

На основе указанной процедуры для каждой архитектуры выбиралась одна лучшая модель (по метрике на тестовой выборке).

Для оценки устойчивости полученных результатов и построения доверительных интервалов была применена процедура,

имитирующая множественное разбиение данных. Было сформировано 10 различных разбиений исходного набора данных на обучающую, валидационную и тестовую выборки, каждое из которых удовлетворяло тем же правилам. На каждом из 10 разбиений заново обучались все 6 моделей.

Итоговое сравнение лучших моделей разных архитектур выполнялось по выборке из 10 значений метрики с построением доверительных интервалов.

5. Результаты. Результаты оценки информативности наборов данных при помощи классической сверточной нейронной сети в виде тепловой карты представлены на рисунке 9. В таком виде можно нагляднее сравнить влияние использования статистических характеристик на качество распознавания, а также оценить целесообразность повышения глубины архитектуры для анализа более крупных матриц. На рисунке слева (Features: Mean) расположены результаты работы модели, обученной на наборе данных, использующем для выделения признаков в рамках интервала только средние значения, справа (Features: Complete) – результаты работы модели, обученной на наборе данных, использующем полный набор статистических характеристик. Тепловая карта имеет в качестве своих измерений количество используемых клавиш (Image Size) и количество блоков в нейронной сети (Network Depth).

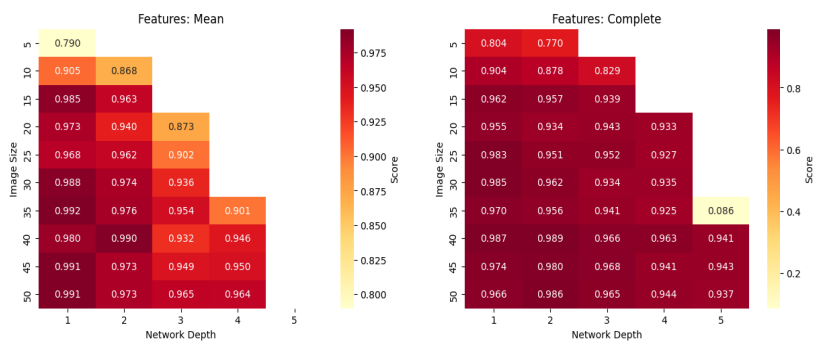


Рис. 9. Результаты оценки информативности наборов данных при помощи классической сверточной нейронной сети

Можно сделать следующие выводы:

- повышение размеров матрицы стабильно приводит к улучшению распознавания пользователей, однако увеличение размера свыше 25, это улучшение нельзя назвать качественным;

– наилучшие результаты распознавания принадлежат набору данных с использованием исключительно среднего выборочного значения, однако отличие от наилучшего результата у своего прямого аналога сложно назвать значительным (0.992 у набора со средним выборочным и 0.989 у набора с полным набором характеристик – отличие в третьем знаке), следовательно, для более точного сравнения стоит рассчитать статистическую значимость такого отличия;

– использование слишком глубоких сетей излишне: на рисунке 9 можно заметить, что лучшие результаты дают архитектуры с глубиной 1 и 2; особенно это заметно на данных, использующих исключительно среднее выборочное, даже архитектуры с глубиной 3 обрабатывают с ощутимо меньшим качеством, чем архитектуры с меньшей глубиной.

Результаты оценки информативности наборов данных при помощи дилатированной сверточной нейронной сети представлены на рисунке 10. Результаты подтверждают ранее выявленную закономерность: увеличение размера выше 25-30 не дает качественного прироста информативности.

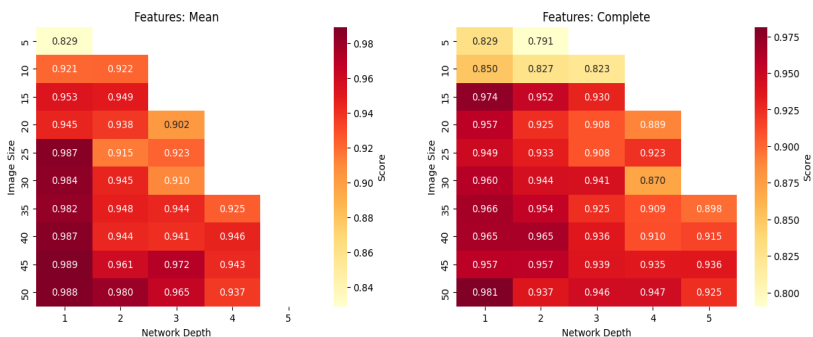


Рис. 10. Результаты оценки информативности наборов данных при помощи дилатированной сверточной нейронной сети

Выводы, следующие из результатов этого эксперимента:

– для дилатированной сверточной нейронной сети повышение размеров матрицы более 25-30, как и для классической, приводит к улучшению распознавания пользователей, однако это улучшение нельзя назвать качественным;

– наилучшие результаты распознавания принадлежат набору данных с использованием исключительно среднего выборочного значения (0.989 у набора со средним выборочным и 0.981 у набора

с полным набором характеристик – отличие в третьем знаке), однако различие сильнее, чем при использовании классической нейронной сети;

- использование слишком глубоких сетей и для дилатированных сверточных нейронных сетей излишне: модели с более глубокой архитектурой показывают результат хуже, причем разница между наилучшим и наихудшим результатами (в контексте сравнения только по глубине) выше, чем при использовании классических сверточных нейронных сетей;

- используемая архитектура дилатированных сверточных нейронных сетей уступает в качестве распознавания классическим нейронным сетям, что может говорить о том, что дилатированные свертки выгоднее было бы использовать не вместо классических, а вместе с ними.

Результаты оценки информативности наборов данных при помощи комбинированной сверточной нейронной сети представлены на рисунке 11. Как и в случае с классической CNN, увеличение размера выше 20 не дает качественного прироста информативности. Кроме того, с таким подходом архитектура с минимальной глубиной в большом количестве случаев перестает быть оптимальной.

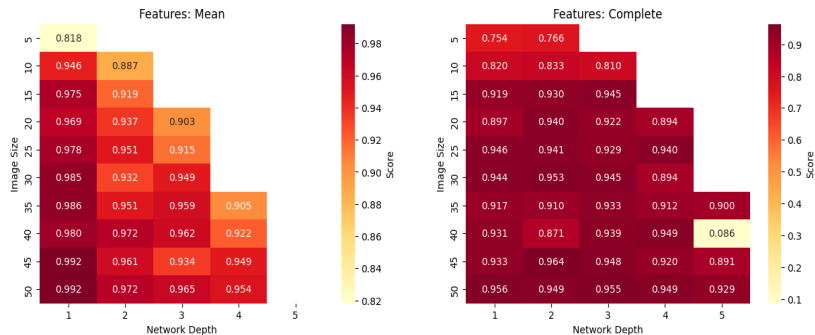


Рис. 11. Результаты оценки информативности наборов данных при помощи комбинированной сверточной нейронной сети

Результаты показывают, что:

- для комбинированной сверточной нейронной сети повышение размеров матрицы более 20 приводит к улучшению распознавания пользователей, однако это улучшение нельзя назвать качественным;
- наилучшие результаты распознавания принадлежат набору данных с использованием исключительно среднего выборочного значения (0.992 у набора со средним выборочным и 0.964 у набора

с полным набором характеристик – отличие в третьем знаке), однако различие сильнее, чем при использовании классической нейронной сети;

- использование слишком глубоких сетей и для комбинированных сверточных нейронных сетей излишне: модели с более глубокой архитектурой показывают результат хуже, причем разница между наилучшим и наихудшим результатами (в контексте сравнения только по глубине) выше, чем при использовании классических сверточных нейронных сетей;

- комбинированная архитектура сверточных нейронных сетей в среднем уступает в качестве распознавания классическим сверточным нейронным сетям, однако лучшие результаты соответствуют им.

Результаты оценки информативности наборов данных при помощи сверточной нейронной сети с увеличенной сверткой представлены на рисунке 12. Здесь также отчетливо прослеживается закономерность, отмеченная ранее.

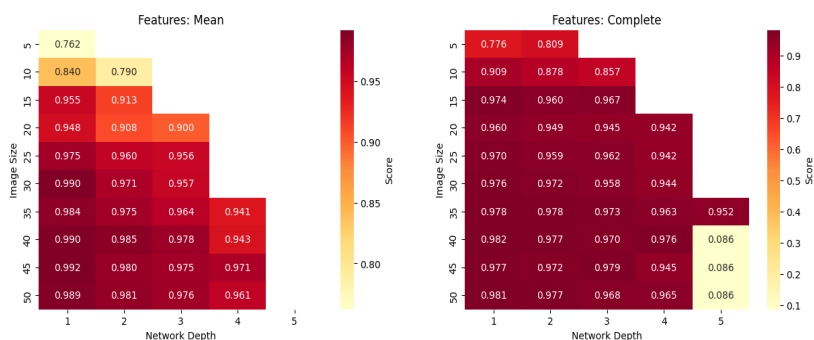


Рис. 12. Результаты оценки информативности наборов данных при помощи сверточной нейронной сети с увеличенным ядром

На рисунке 12 видно, что:

- для такой архитектуры сверточной нейронной сети повышение размеров матрицы более 25-30, как и для классической, приводит к улучшению распознавания пользователей, однако это улучшение нельзя назвать качественным;

- наилучшие результаты распознавания принадлежат набору данных с использованием исключительно среднего выборочного значения (0.992 у набора со средним выборочным и 0.981 у набора с полным набором характеристик – отличие в третьем знаке), однако различие сильнее, чем при использовании классической нейронной сети;

- использование слишком глубоких сетей излишне: модели с более глубокой архитектурой показывают результат хуже, причем

модели с глубиной в 5 блоков уже не могут адекватно обучиться, они определяют все образцы в один единственный класс;

- используемая архитектура сверточных нейронных сетей в общем не сильно уступает в качестве распознавания классическим нейронным сетям.

Результаты оценки информативности наборов данных при помощи сверточной нейронной сети представлены на рисунке 13. Качество предсказания показывает заметный рост при увеличении размеров матрицы на всем исследуемом промежутке.

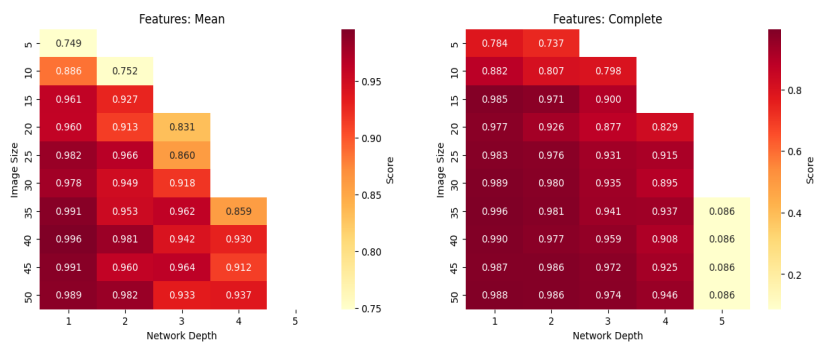


Рис. 13. Результаты оценки информативности наборов данных при помощи разделенной сверточной нейронной сети

Таким образом:

- для такой архитектуры сверточной нейронной сети повышение размеров матрицы приводит к заметному улучшению распознавания пользователей;

- наилучшие результаты распознавания принадлежат обоим наборам данных (0.996 у набора со средним выборочным и 0.996 у набора с полным набором характеристик), однако в данном случае набор данных с полным перечнем характеристик в среднем обрабатывает заметно лучше;

- использование слишком глубоких сетей излишне: модели с более глубокой архитектурой показывают результат хуже, причем все модели с глубиной в 5 блоков уже не могут адекватно обучиться, они определяют все образцы в один единственный класс;

- используемая архитектура сверточных нейронных сетей ощутимо превосходит классические нейронные сети в качестве распознавания.

Результаты оценки информативности наборов данных при помощи маскированной сверточной нейронной сети представлены

на рисунке 14. Качество предсказания показывает заметный рост на матрицах меньших размеров.

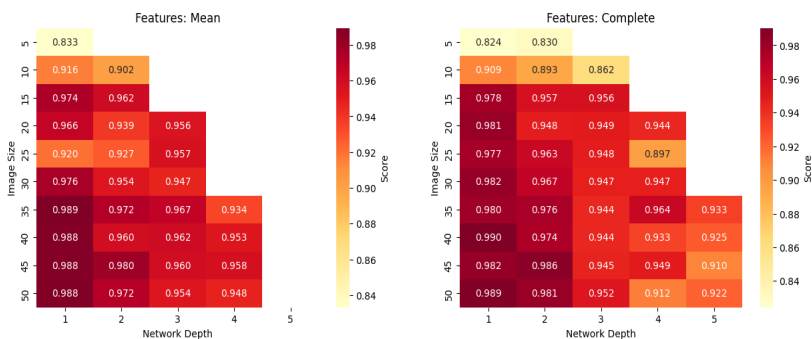


Рис. 14. Результаты оценки информативности наборов данных при помощи маскированной сверточной нейронной сети

На рисунке 14 видно, что:

- для такой архитектуры сверточной нейронной сети повышение размеров матрицы приводит к заметному улучшению распознавания пользователей;
- наилучшие результаты распознавания принадлежат набору данных с полным перечнем статистических характеристик (0.989 у набора со средним выборочным и 0.990 у набора с полным набором характеристик – отличие в третьем знаке), однако в данном случае набор данных с полным перечнем характеристик в среднем обрабатывает заметно лучше;
- использование слишком глубоких сетей излишне: модели с более глубокой архитектурой показывают результат хуже;
- используемая архитектура сверточных нейронных сетей не превосходит классические нейронные сети в качестве распознавания;
- использование масок увеличивает качество распознавания почти всех моделей, следовательно, как метод анализа разреженных данных более пригоден, чем методы с дилатированной сверткой, однако этот метод может требовать более тонкой настройки.

6. Обсуждение. Результаты оценки качества разных архитектур моделей сведены в единую таблицу, представленную на рисунке 15. В качестве значения в ячейке таблицы помещено максимальное значение среди моделей с разной глубиной, имеющих параметры обучающей выборки и архитектуру, соответствующих столбцам и строкам таблицы соответственно. В таблице отмечены лучшие результаты для каждой из архитектур.

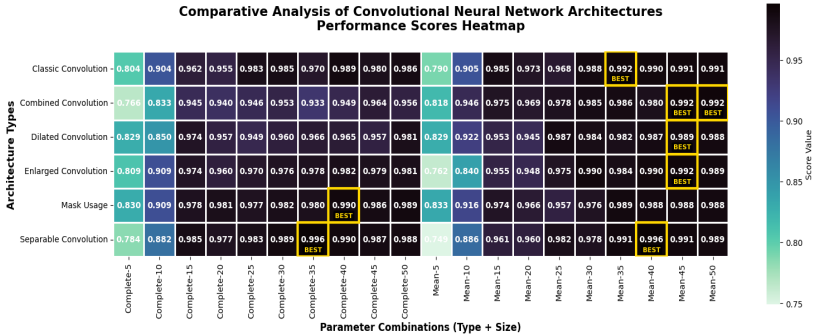


Рис. 15. Сравнение архитектур моделей

Стоит отметить несколько основных закономерностей. Увеличение размеров многоканальной матрицы путем увеличения количества анализируемых клавиш стабильно приводит к увеличению результатов вопреки выводам, сделанным при помощи оценки на основе энтропийных характеристик. Увеличение количества используемых статистических характеристик, наоборот, приводит к настолько сильной разреженности матриц, что точность предсказания на них ощутимо падает. Однако, использование определенных архитектур (архитектур с разделенной сверткой и с масками) эту проблему решает, и такие архитектуры справляются с полным набором характеристик лучше, чем с исключительно средним выборочным.

На рисунке 16 представлено сравнение наилучших представителей разных архитектур между собой с учетом доверительных интервалов, рассчитанных на основе 10 разных разбиений исходного набора данных методом, описанным в разделе 4. Для расчета доверительных интервалов была использована (2).

$$Score = MeanScore \pm t_{n-1, 0.95} * \frac{\sigma}{\sqrt{n}} \quad (2)$$

где σ – стандартное отклонение, n – количество значений в выборке ($=10$), t – коэффициент Стьюдента.

Как можно заметить, размеры доверительных интервалов слишком велики, чтобы можно было однозначно сравнить модели между собой. Однако, радиусы интервалов достаточно малы, чтобы сравнить эти модели с аналогами из других работ. Ключевые аналоги в области использования сверточных нейронных сетей, а также

лучшие модели, использующие иные методы машинного обучения, сведены в таблицу 1 для сравнения.

Из таблицы видно, что в среднем классическая, комбинированная и разделенная сверточная модель отрабатывают лучше, чем большинство аналогов, если верить метрике точности. Если ориентироваться по метрике EER, модели, собранные в данной работе также превосходят большинство аналогов.

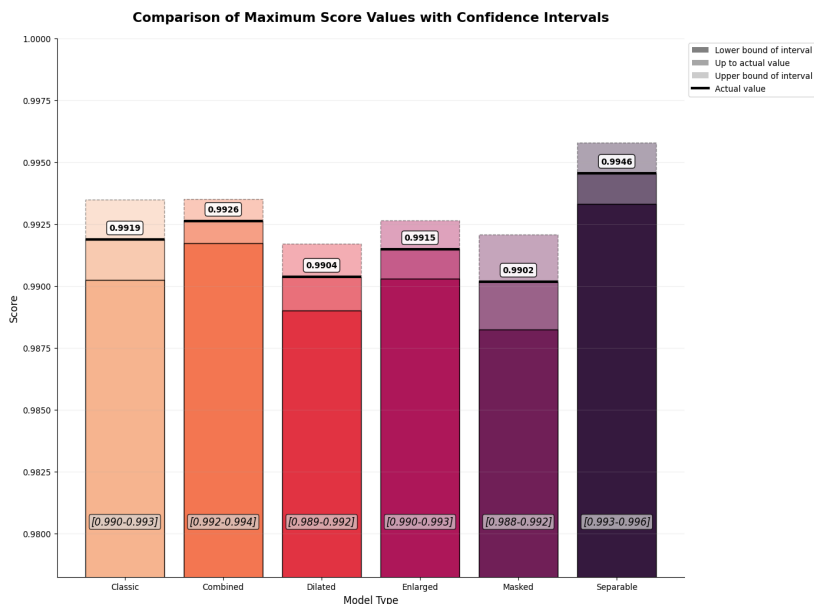


Рис. 16. Сравнение наилучших представителей разных архитектур сверточных нейронных сетей

7. Заключение. Таким образом, был проведен анализ эффективности использования сверточных нейронных сетей для решения задачи аутентификации по фактору клавиатурного почерка. По результатам анализа можно сделать следующие выводы. Классическая и комбинированная CNN показали стабильно высокие результаты, сопоставимые с лучшими аналогами. Дилатированная CNN и CNN с увеличенным ядром не превзошли классический подход, что указывает на большую важность локальных зависимостей в данных для данной задачи. Сеть с маскированием доказала свою пригодность для работы с разреженными данными, но не показала превосходства над классической CNN, требуя при этом более тонкой настройки.

Разделенная сверточная сеть продемонстрировала наилучший результат, достигнув точности 99,5% и наименьшего EER (0,2%). Ее эффективность объясняется способностью эффективно выделять зависимости между разными каналами (признаками биграмм) и высокой эффективностью на малых выборках данных.

Таблица 1. Сравнение моделей с аналогами

Источник	Метод	Точность	EER
-	Классическая сверточная нейронная сеть	99,2%	0,37%
-	Дилатированная сверточная нейронная сеть	99%	0,52%
-	Комбинированная сверточная нейронная сеть	99,3%	0,31%
-	Сверточная нейронная сеть с увеличенным ядром	99,1%	0,4%
-	Разделенная сверточная нейронная сеть	99,5%	0,2%
-	Сверточная нейронная сеть с использованием масок	99%	0,33%
[14]	XGBoost	96,39%	-
[15]	LightGBM	94,68%	-
[16]	Случайные леса	99%	-
[17]	Перцептрон	98%	-
[18]	Многослойный перцептрон	97,18%	0,02%
[19]	Статистический профиль на основе взвешенного Манхэттенского расстояния	-	5,47%
[20]	Сиамская нейронная сеть + LSTM	-	2,2%
[21]	CNN-RNN	97,77%	0,01%
[22]	Трансформер на основе CNN-GRU	96,68%	3,33%
[23]	Трансформер TypeFormer	-	3,25%
[24]	Трансформер DoubleStrokeNet	-	0,75%
[10]	Классическая сверточная нейронная сеть	92%	-
[25]	Сверточная нейронная сеть + Рекуррентная сеть	98,7%	-
[26]	Гибридная архитектура CNN-GRU	99,3%	4%
[27]	Оптимизированная CNN с предобработкой и каскадным классификатором	99,95%	0,65%

Также были определены параметры формирования многоканальных матриц, обеспечивающие лучшее качество работы модели сверточной нейронной сети. Установлено, что увеличение размера матрицы (количества анализируемых клавиш) сверх 25-30 не дает существенного прироста качества. Наилучшие результаты в большинстве архитектур были достигнуты при использовании только среднего выборочного значения, в то время как расширенный набор статистических характеристик (среднеквадратичное отклонение,

асимметрия, эксцесс) часто приводил к чрезмерному разреживанию данных и снижению точности.

Кроме того, в результате экспериментального исследования был сделан ряд других выводов.

- Эффективность классических CNN. Классическая сверточная нейронная сеть показала высокую эффективность, достигнув точности 99.2%. Это подтвердило способность CNN успешно обрабатывать разреженные данные, извлекать информативные признаки и игнорировать малоинформативные, несмотря на высокую разреженность матриц.

- Ограничение глубины сетей. Для всех протестированных архитектур было выявлено, что использование излишне глубоких сетей (глубиной более 2 блоков) является контрпродуктивным и приводит к значительному снижению качества распознавания.

- Сравнение с аналогами. Проведенное сравнение с существующими аналогами показало, что разработанные модели, в особенности классическая CNN и разделенная CNN, демонстрируют конкурентоспособные и превосходящие результаты как по метрике точности, так и по EER. Разделенная CNN уступает лишь одному известному аналогу по точности, но превосходит его по EER.

Перспективы дальнейших исследований видятся в более тонкой настройке архитектуры разделенной сверточной сети, исследовании методов компенсации разреженности при использовании полного набора статистических характеристик.

Литература

1. Risto H.N., Graven O.H. Fixed-text keystroke dynamics authentication data set – collection and analysis // *Annals of Telecommunications*. 2024. vol. 79. pp. 731–743. DOI: 10.1007/s12243-024-01039-z.
2. Shadman R., Wahab A.A., Manno M., Lukaszewski M., Hou D., Hussain F. Keystroke Dynamics: Concepts, Techniques, and Applications // *ACM Computing Surveys*. 2025. vol. 57. no. 11. pp. 1–35. DOI: 10.1145/3733103.
3. Shelupanov A., Evsyutin O., Konev A., Kostyuchenko E., Kruchinin D., Nikiforov D. Information Security Methods – Modern Research Directions // *Symmetry*. 2019. vol. 11. no. 2. 150 p. DOI: 10.3390/sym11020150.
4. Давыденко С.А., Костюченко Е.Ю., Новиков С.Н. Оценивание информативности признаков в наборах данных для проведения продлённой аутентификации // *Информатика и автоматизация*. 2024. Т. 23. №1. С. 65–100. DOI: 10.15622/ia.23.1.3.
5. Zateev S., Shelupanov A. Keyboard Handwriting as a Factor of User Authentication // *Lecture Notes in Networks and Systems*. Springer. 2024. vol. 942. pp. 155–167. DOI: 10.1007/978-3-031-55349-3_13.
6. Затеев С.В. Продленная аутентификация на основе анализа клавиатурного почерка // *Теория и практика обеспечения информационной безопасности: Сб. научн. тр. XX Международной конференции студентов, аспирантов и молодых ученых*. 2023. Т. 7. С. 116–123.

7. Monaco J.V., Bakelman N., Cha S.-H. Tappert C.C. Developing a keystroke biometric system for continual authentication of computer users // *Proceedings of European Intelligence and Security Informatics Conference (EISIC)*. 2012. pp. 210–216. DOI: 10.1109/EISIC.2012.58.
8. Altwajry N. Authentication by Keystroke Dynamics: The Influence of Typing Language // *Applied Sciences*. 2023. vol. 13. no. 20. 11478 p. DOI: 10.3390/app132011478.
9. Bailey K.O., Okolika J.S., Peterson G.L. User identification and authentication using multi-modal behavioral biometrics // *Computers & Security*. 2014. vol. 43. pp. 77–89. DOI: 10.1016/j.cose.2014.03.005.
10. Sharma A., Jurecek M., Stamp M. Keystroke Dynamics for User Identification // *Machine Learning, Deep Learning and AI for Cybersecurity*. Springer. 2025. pp. 601–622. DOI: 10.1007/978-3-031-83157-7_21.
11. Wycislik L., Wylezek P., Momot A. The Improved Biometric Identification of Keystroke Dynamics Based on Deep Learning Approaches // *Sensors*. 2024. vol. 24. no. 12. 3763 p. DOI: 10.3390/s24123763.
12. Spata M.O., Russo V.M., Ortis A., Battiato S. A New Pipeline for Snooping Keystroke Based on Deep Learning Algorithm // *IEEE Access*. 2025. vol. 13. pp. 24498–24514. DOI: 10.1109/ACCESS.2025.3536877.
13. Kaur D., De Carvalho G.H.S., Dividino R., Woungang I., Anpalagan A. Novel Explainable CNN-LightGBM Model for Smartphone Continuous Authentication // *IEEE Access*. 2026. vol. 14. pp. 45826–45843. DOI: 10.1109/ACCESS.2026.3672372.
14. Chang H.-C., Li J., Wu C.-S., Stamp M. Machine Learning and Deep Learning for Fixed-Text Keystroke Dynamics // *Advances in Information Security*. Springer. 2022. vol. 54. pp. 309–329. DOI: 10.1007/978-3-030-97087-1_13.
15. Muralidharan A., Eaman A., Hassan E. A Comparative Analysis of Machine Learning Models for Behavioral Biometric Authentication using Keystroke Dynamics // *Proceedings of the 20th International Conference on Future Networks and Communications*. Procedia Computer Science. 2025. vol. 265. pp. 116–123. DOI: 10.1016/j.procs.2025.07.163.
16. Krishnamoorthy S., Rueda L., Saad S., Elmiligi H. Identification of User Behavioral Biometrics for Authentication Using Keystroke Dynamics and Machine Learning // *Proceedings of 2nd International Conference on Biometric Engineering and Applications (ICBEA '18)*. 2018. pp. 50–57. DOI: 10.1145/3230820.3230829.
17. Bleha S.A., Knopp J., Obaidat M.S. Performance of the Perceptron algorithm for the classification of computer users // *Proceedings of ACM/SIGAPP Symposium on Applied Computing: Technological Challenges of the 1990s*. 1992. pp. 863–866.
18. Stylios I., Skalkos A., Kokolakis S., Karyda M. BioPrivacy: Development of a Keystroke Dynamics Continuous Authentication System // *Lecture Notes in Computer Science*. Springer. 2021. vol. 13106. pp. 158–170. DOI: 10.1007/978-3-030-95484-0_10.
19. Wahab A., Hou D., Schuckers S., Barbir, A. Utilizing Keystroke Dynamics as Additional Security Measure to Protect Account Recovery Mechanism // *Proceedings of the 7th International Conference on Information Systems Security and Privacy (ICISSP)*. 2021. vol. 1. pp. 33–42. DOI: 10.5220/0010191200330042.
20. Acien A., Morales A., Monaco J.V., Vera-Rodriguez R., Fierrez J. TypeNet: Deep learning keystroke biometrics // *IEEE Transactions on Biometrics, Behavior, and Identity Science*. 2021. vol. 4. no. 1. pp. 57–70. DOI: 10.1109/TBIOM.2021.3112540.
21. Li J., Chang H.-C., Stamp M. Free-Text Keystroke Dynamics for User Authentication // *Advances in Information Security*. Springer. 2022. vol. 54. pp. 357–380. DOI: 10.1007/978-3-030-97087-1_15.
22. Gonzalez N., Stragapede G., Vera-Rodriguez R., Tolosana R. Type2Branch: Keystroke Biometrics Based on a Dual-Branch Architecture With Attention

- Mechanisms and Set2set Loss // IEEE Transactions on Information Forensics and Security. 2025. vol. 20. pp. 5859–5871. DOI: 10.1109/TIFS.2025.3574992.
23. Stragapede G., Delgado-Santos P., Tolosana R., Vera-Rodriguez R., Guest R., Morales A. TypeFormer: Transformers for mobile keystroke biometrics // Neural Computing and Applications. 2024. vol. 36. pp. 18531–18545. DOI: 10.1007/s00521-024-10140-2.
24. Neacsu T., Poncu T., Ruseti S., Dascalu M. DoubleStrokeNet: Bigram-level keystroke authentication // Electronics. 2023. vol. 12. no. 20. 4309 p. DOI: 10.3390/electronics12204309.
25. Kasprowski P., Borowska Z., Harezlak K. Biometric Identification Based on Keystroke Dynamics // Sensors. 2022. vol. 22. no. 9. 3158 p. DOI: 10.3390/s22093158.
26. Chang H.-C., Li J., Stamp M. Machine Learning-Based Analysis of Free-Text Keystroke Dynamics // Advances in Information Security. Springer. 2022. vol. 54. pp. 331–356. DOI: 10.1007/978-3-030-97087-1_14.
27. AbdelRaouf H., Chelloug S.A., Muthanna A., Semaary N., Amin K., Ibrahim M. Efficient Convolutional Neural Network-Based Keystroke Dynamics for Boosting User Authentication // Sensors. 2023. vol. 23. no. 10. 4898 p. DOI: 10.3390/s23104898.

Затеев Станислав Вадимович — младший научный сотрудник, научно-инжиниринговый центр «интеллектуальные системы доверенного взаимодействия», Федеральное государственное автономное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ФГАОУ ВО «ТУСУР»). Область научных интересов: искусственный интеллект и машинное обучение, анализ данных. Число научных публикаций — 4. zateev.stanislav@gmail.com; пр. Ленина, 40, 634050, Томск, Россия; р.т.: +7(931)491-7404.

Костюченко Евгений Юрьевич — канд. техн. наук, заведующий лабораторией, лаборатория съёма, анализа и управления биологическими сигналами; доцент, кафедра комплексной информационной безопасности электронно-вычислительных систем, Федеральное государственное автономное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ФГАОУ ВО «ТУСУР»). Область научных интересов: искусственный интеллект и машинное обучение, обработка речи, биометрия, анализ данных. Число научных публикаций — 154. key@fb.tusur.ru; пр. Ленина, 40, 634050, Томск, Россия; р.т.: +7(382)270-1529.

Шелупанов Александр Александрович — д-р техн. наук, профессор, член-корреспондент РАН, директор, институт системной интеграции и безопасности; заведующий кафедрой, кафедра комплексной информационной безопасности электронно-вычислительных систем, Федеральное государственное автономное образовательное учреждение высшего образования «Томский государственный университет систем управления и радиоэлектроники» (ФГАОУ ВО «ТУСУР»). Область научных интересов: информационная безопасность, математическое моделирование. Число научных публикаций — 437. saa@tusur.ru; пр. Ленина, 40, 634050, Томск, Россия; р.т.: +7(382)290-7155.

Поддержка исследований. Работа выполнена при поддержке Министерства науки и высшего образования РФ: FEWM-2026-0009 (ТУСУР).

S. ZATEEV, E. KOSTYUCHENKO, A. SHELUPANOV
**COMPARATIVE ANALYSIS OF APPROACHES TO
CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE
DESIGN FOR KEYSTROKE DYNAMICS ANALYSIS**

Zateev S., Kostyuchenko E., Shelupanov A. Comparative Analysis of Approaches to Convolutional Neural Network Architecture Design for Keystroke Dynamics Analysis.

Abstract. In the context of widespread digitalization and the increasing frequency of confidential information compromise incidents, the relevance of developing reliable authentication systems serving as the primary line of defense becomes critically important. Behavioral biometric systems present a promising solution to the problems associated with loss, theft, or unauthorized use of static identifiers. In particular, methods based on keystroke dynamics analysis enable continuous and transparent user authentication throughout an entire work session by analyzing unique and difficult-to-forge behavioral patterns. This research direction is characterized by a combination of a high level of security with relatively moderate implementation costs and high user convenience. The aim of this study is to conduct a comparative analysis of various approaches to the design and construction of convolutional neural network architectures optimized for processing sparse data describing keystroke dynamics. The experimental work revealed that convolutional neural networks demonstrate exceptionally high verification accuracy, reaching 99%, despite the fact that the input data generated during preprocessing exhibit significant sparsity. The best results within the framework of this study were achieved using an architecture based on a depthwise separable convolution mechanism, which proved to be effective for handling this specific data type.

Keywords: authentication, behavioral biometrics, keystroke dynamics, sparse data, convolutional neural networks.

References

1. Risto H.N., Graven O.H. Fixed-text keystroke dynamics authentication data set – collection and analysis. *Annals of Telecommunications*. 2024. vol. 79. pp. 731–743. DOI: 10.1007/s12243-024-01039-z.
2. Shadman R., Wahab A.A., Manno M., Lukaszewski M., Hou D., Hussain F. Keystroke Dynamics: Concepts, Techniques, and Applications. *ACM Computing Surveys*. 2025. vol. 57. no. 11. pp. 1–35. DOI: 10.1145/3733103.
3. Shelupanov A., Evsyutin O., Konev A., Kostyuchenko E., Kruchinin D., Nikiforov D. Information Security Methods – Modern Research Directions. *Symmetry*. 2019. vol. 11. no. 2. 150 p. DOI: 10.3390/sym11020150.
4. Davydenko S., Kostyuchenko E., Novikov S. [Evaluation of the Informativeness of Features in Datasets for Continuous Verification]. *Informatika i avtomatizatsiya – Informatics and Automation*. 2024. vol. 23. no. 1. pp. 65–100. DOI: 10.15622/ia.23.1.3. (In Russ.).
5. Zateev S., Shelupanov A. Keyboard Handwriting as a Factor of User Authentication. *Lecture Notes in Networks and Systems*. Springer. 2024. vol. 942. pp. 155–167. DOI: 10.1007/978-3-031-55349-3_13.
6. Zateev S. [Extended User Authentication Based on Keyboard Handwriting Analysis]. *Teoriya i praktika obespecheniya informatsionnoy bezopasnosti: Sb. nauchn. tr. XX Mezhdunarodnoy konferentsii studentov, aspirantov i molodykh uchenykh [Theory and practice of information security : Collected papers]*. 2023. vol. 7. pp. 116–123. (In Russ.).

7. Monaco J.V., Bakelman N., Cha S.-H. Tappert C.C. Developing a keystroke biometric system for continual authentication of computer users. *Proceedings of European Intelligence and Security Informatics Conference (EISIC)*. 2012. pp. 210–216. DOI: 10.1109/EISIC.2012.58.
8. Altwaijry N. Authentication by Keystroke Dynamics: The Influence of Typing Language. *Applied Sciences*. 2023. vol. 13. no. 20. 11478 p. DOI: 10.3390/app132011478.
9. Bailey K.O., Okolica J.S., Peterson G.L. User identification and authentication using multi-modal behavioral biometrics. *Computers & Security*. 2014. vol. 43. pp. 77–89. DOI: 10.1016/j.cose.2014.03.005.
10. Sharma A., Jurecek M., Stamp M. Keystroke Dynamics for User Identification. *Machine Learning, Deep Learning and AI for Cybersecurity*. Springer. 2025. pp. 601–622. DOI: 10.1007/978-3-031-83157-7_21.
11. Wycislik L., Wylezek P., Momot A. The Improved Biometric Identification of Keystroke Dynamics Based on Deep Learning Approaches. *Sensors*. 2024. vol. 24. no. 12. 3763 p. DOI: 10.3390/s24123763.
12. Spata M.O., Russo V.M., Ortis A., Battiato S. A New Pipeline for Snooping Keystroke Based on Deep Learning Algorithm. *IEEE Access*. 2025. vol. 13. pp. 24498–24514. DOI: 10.1109/ACCESS.2025.3536877.
13. Kaur D., De Carvalho G.H.S., Dividino R., Woungang I., Anpalagan A. Novel Explainable CNN-LightGBM Model for Smartphone Continuous Authentication. *IEEE Access*. 2026. vol. 14. pp. 45826–45843. DOI: 10.1109/ACCESS.2026.3672372.
14. Chang H.-C., Li J., Wu C.-S., Stamp M. Machine Learning and Deep Learning for Fixed-Text Keystroke Dynamics. *Advances in Information Security*. Springer. 2022. vol. 54. pp. 309–329. DOI: 10.1007/978-3-030-97087-1_13.
15. Muralidharan A., Eaman A., Hassan E. A Comparative Analysis of Machine Learning Models for Behavioral Biometric Authentication using Keystroke Dynamics. *Proceedings of the 20th International Conference on Future Networks and Communications*. *Procedia Computer Science*. 2025. vol. 265. pp. 116–123. DOI: 10.1016/j.procs.2025.07.163.
16. Krishnamoorthy S., Rueda L., Saad S., Elmiligi H. Identification of User Behavioral Biometrics for Authentication Using Keystroke Dynamics and Machine Learning. *Proceedings of 2nd International Conference on Biometric Engineering and Applications (ICBEA '18)*. 2018. pp. 50–57. DOI: 10.1145/3230820.3230829.
17. Bleha S.A., Knopp J., Obaidat M.S. Performance of the Perceptron algorithm for the classification of computer users. *Proceedings of ACM/SIGAPP Symposium on Applied Computing: Technological Challenges of the 1990s*. 1992. pp. 863–866.
18. Stylios I., Skalkos A., Kokolakis S., Karyda M. BioPrivacy: Development of a Keystroke Dynamics Continuous Authentication System. *Lecture Notes in Computer Science*. Springer. 2021. vol. 13106. pp. 158–170. DOI: 10.1007/978-3-030-95484-0_10.
19. Wahab A., Hou D., Schuckers S., Barbir, A. Utilizing Keystroke Dynamics as Additional Security Measure to Protect Account Recovery Mechanism. *Proceedings of the 7th International Conference on Information Systems Security and Privacy (ICISSP)*. 2021. vol. 1. pp. 33–42. DOI: 10.5220/0010191200330042.
20. Acien A., Morales A., Monaco J.V., Vera-Rodriguez R., Fierrez J. TypeNet: Deep learning keystroke biometrics. *IEEE Transactions on Biometrics, Behavior, and Identity Science*. 2021. vol. 4. no. 1. pp. 57–70. DOI: 10.1109/TBIOM.2021.3112540.
21. Li J., Chang H.-C., Stamp M. Free-Text Keystroke Dynamics for User Authentication. *Advances in Information Security*. Springer. 2022. vol. 54. pp. 357–380. DOI: 10.1007/978-3-030-97087-1_15.
22. Gonzalez N., Stragapede G., Vera-Rodriguez R., Tolosana R. Type2Branch: Keystroke Biometrics Based on a Dual-Branch Architecture With Attention

- Mechanisms and Set2set Loss. IEEE Transactions on Information Forensics and Security. 2025. vol. 20. pp. 5859–5871. DOI: 10.1109/TIFS.2025.3574992.
23. Stragapede G., Delgado-Santos P., Tolosana R., Vera-Rodriguez R., Guest R., Morales A. TypeFormer: Transformers for mobile keystroke biometrics. Neural Computing and Applications. 2024. vol. 36. pp. 18531–18545. DOI: 10.1007/s00521-024-10140-2.
 24. Neacsu T., Poncu T., Ruseti S., Dascalu M. DoubleStrokeNet: Bigram-level keystroke authentication. Electronics. 2023. vol. 12. no. 20. 4309 p. DOI: 10.3390/electronics12204309.
 25. Kasprowski P., Borowska Z., Harezlak K. Biometric Identification Based on Keystroke Dynamics. Sensors. 2022. vol. 22. no. 9. 3158 p. DOI: 10.3390/s22093158.
 26. Chang H.-C., Li J., Stamp M. Machine Learning-Based Analysis of Free-Text Keystroke Dynamics. Advances in Information Security. Springer. 2022. vol. 54. pp. 331–356. DOI: 10.1007/978-3-030-97087-1_14.
 27. AbdelRaouf H., Chelloug S.A., Muthanna A., Semary N., Amin K., Ibrahim M. Efficient Convolutional Neural Network-Based Keystroke Dynamics for Boosting User Authentication. Sensors. 2023. vol. 23. no. 10. 4898 p. DOI: 10.3390/s23104898.

Zateev Stanislav — Junior Research Associate, Scientific and Engineering Center "Intelligent Trusted Interaction Systems", Federal State Autonomous Educational Institution of Higher Education "Tomsk State University of Control Systems and Radioelectronics" (TUSUR). Research interests: artificial intelligence and machine learning, data analysis. The number of publications — 4. zateev.stanislav@gmail.com; 40, Lenin Ave., 634050, Tomsk, Russia; office phone: +7(931)491-7404.

Kostyuchenko Evgeniy — Ph.D., Head of the Laboratory, Laboratory for the Removal, Analysis and Control of Biological Signals; Associate Professor, Department of Integrated Information Security of Electronic Computing Systems, Federal State Autonomous Educational Institution of Higher Education "Tomsk State University of Control Systems and Radioelectronics" (TUSUR). Research interests: artificial intelligence and machine learning, speech processing, biometrics, data analysis. The number of publications — 154. key@fb.tusur.ru; 40, Lenin Ave., 634050, Tomsk, Russia; office phone: +7(382)270-1529.

Shelupanov Alexander — Ph.D., Dr.Sci., Professor, Corresponding Member of RAS, Director, Institute of System Integration and Security; Head of the Department, Department of Integrated Information Security of Electronic Computing Systems, Federal State Autonomous Educational Institution of Higher Education "Tomsk State University of Control Systems and Radioelectronics" (TUSUR). Research interests: information security, mathematical modeling. The number of publications — 437. saa@tusur.ru; 40, Lenin Ave., 634050, Tomsk, Russia; office phone: +7(382)290-7155.

Acknowledgements. The work was supported by the Ministry of Science and Higher Education of the Russian Federation: FEWM-2026-0009 (TUSUR).

А.Ю. ТИМОФЕЕВА, Т.В. АВДЕЕНКО
**ОТБОР ПРИЗНАКОВ В УСЛОВИЯХ
НЕСБАЛАНСИРОВАННОСТИ КЛАССОВ: ЭМПИРИЧЕСКОЕ
СРАВНЕНИЕ КОРРЕЛЯЦИИ И ВЗАИМНОЙ ИНФОРМАЦИИ**

Тимофеева А.Ю., Авдеенко Т.В. Отбор признаков в условиях несбалансированности классов: эмпирическое сравнение корреляции и взаимной информации.

Аннотация. Рассматривается проблема несбалансированности данных в задачах предварительного отбора признаков с использованием скрининга с гарантированной независимостью. В рамках данного подхода анализируется выбор меры зависимости, применяемой для ранжирования признаков. В работе исследуются два наиболее распространенных показателя взаимосвязи – коэффициент корреляции и взаимная информация. Коэффициент корреляции характеризует линейную зависимость между переменными, его оценки просты в вычислении и обладают хорошими статистическими свойствами. Взаимная информация является более универсальной мерой, способной выявлять нелинейные зависимости, однако ее оценивание требует восстановления распределений по выборочным данным, что приводит к повышенной вычислительной сложности и чувствительности к выбору параметров. На основе серии вычислительных экспериментов показано, что методы оценивания взаимной информации на основе k ближайших соседей уступают ядерным методам с точки зрения устойчивости и качества скрининга. Установлено, что во всех рассмотренных сценариях, включая различные уровни несбалансированности данных, корреляционные меры обеспечивают более стабильные оценки и более надежное ранжирование признаков по сравнению с взаимной информацией. Практическая применимость результатов продемонстрирована на реальных данных электроэнцефалографии, связанных с анализом арифметических способностей. Показано, что двухшаговая процедура, включающая скрининг на основе коэффициента корреляции и последующее моделирование с использованием LASSO-регрессии, обеспечивает существенное снижение ошибки кросс-валидации по сравнению с подходами, основанными на взаимной информации и отсутствием предварительного отбора. Использование главной компоненты, построенной по отобранным признакам, позволяет повысить точность прогноза. Полученные результаты свидетельствуют о том, что в условиях малых выборок, мультиколлинеарности и несбалансированности данных корреляционные меры являются более предпочтительными для задач предварительного отбора признаков с точки зрения устойчивости и воспроизводимости.

Ключевые слова: несбалансированные данные, взаимная информация, коэффициент корреляции, скрининг с гарантированной независимостью, ядерные оценки, k ближайших соседей, LASSO-регрессия.

1. Введение. В задачах высокоразмерного анализа данных, когда число признаков существенно превышает объем выборки, предварительное сокращение размерности является необходимым этапом построения устойчивых статистических моделей. Одним из наиболее распространенных подходов к решению этой задачи является отбор, или скрининг, признаков (feature screening), при котором исходные переменные ранжируются по степени их связи

с откликом, после чего отбирается подмножество наиболее информативных признаков [1].

Применение процедур скрининга позволяет существенно сократить размерность пространства признаков и повысить устойчивость последующего моделирования [2]. Их значимость особенно возрастает в условиях малых выборок [3] и мультиколлинеарности [4], при которых использование полного набора признаков приводит к переобучению и ухудшению обобщающей способности моделей.

Среди методов предварительного отбора признаков широкое распространение получил подход скрининга с гарантированной независимостью (Sure Independence Screening, SIS) [5], в рамках которого каждый признак оценивается независимо от других. Несмотря на концептуальную простоту, данный метод демонстрирует высокую эффективность в задачах с ультравысокой размерностью, особенно в сочетании с процедурами регуляризации, такими как LASSO. Ключевым элементом таких методов является выбор меры зависимости, используемой для ранжирования признаков.

В практике анализа данных в качестве мер зависимости наиболее часто используются коэффициенты корреляции и взаимная информация [6]. Коэффициент корреляции Пирсона применяется для оценки линейной зависимости между количественными переменными. В случае бинарного отклика используется точно-бисериальный коэффициент корреляции [7], представляющий собой естественное обобщение корреляции на случай смешанных типов данных и интерпретируемый как корреляция между непрерывным признаком и индикаторной переменной, кодирующей принадлежность к классу. К достоинствам корреляционных показателей относятся простота вычисления, отсутствие настраиваемых параметров и устойчивость при малых объемах выборки, что делает их особенно удобными для задач скрининга.

Взаимная информация является более общей мерой зависимости, основанной на энтропийных характеристиках распределения. Она позволяет выявлять как линейные, так и нелинейные взаимосвязи без предположений о функциональной форме зависимости. Вместе с тем ее практическое применение связано с необходимостью оценивания по выборочным данным, что может приводить к высокой вариации оценок, особенно при ограниченном объеме выборки.

Несмотря на широкое использование указанных подходов, вопрос о выборе меры зависимости для процедур скрининга остается открытым. Дополнительную сложность представляет

несбалансированность данных, характерная для многих прикладных задач [8]. Она проявляется в существенной асимметрии распределения категориального отклика, при которой миноритарные классы представлены значительно меньшим числом наблюдений. Подобные ситуации возникают, в частности, в задачах обнаружения мошенничества [9], медицинской диагностики [10] и интеллектуального анализа образовательных данных [11]. При этом высокая общая точность классификации не гарантирует адекватного качества предсказания для класса меньшинства [12].

В контексте отбора признаков несбалансированность данных оказывает влияние на точность оценивания мер зависимости, поскольку асимметрия распределения отклика изменяет их статистические свойства. Это может приводить к увеличению дисперсии оценок и снижению надежности ранжирования признаков. Несмотря на значительное количество работ, посвященных обработке несбалансированных данных в задачах классификации [3, 13, 14], влияние данного фактора на выбор меры зависимости для процедур скрининга изучено недостаточно.

В частности, остается открытым вопрос о сравнительной вариативности различных показателей взаимосвязи в условиях асимметричного распределения отклика. Существующие исследования в основном сосредоточены либо на сравнении методов классификации [14], либо на разработке алгоритмов оценивания взаимной информации [15], тогда как систематический анализ пригодности различных мер зависимости именно для задач предварительного отбора признаков представлен ограниченно.

Целью настоящей работы является сравнительный анализ корреляционных и информационно-теоретических мер зависимости в задачах отбора признаков при наличии несбалансированных данных. В работе исследуются статистические свойства оценок, включая их вариацию и устойчивость, а также влияние выбора меры зависимости на качество процедур скрининга.

Для достижения поставленной цели проводится серия модельных экспериментов. В качестве базовой была выбрана линейная модель, позволяющая контролируемо варьировать уровень зависимости и степень несбалансированности данных. Такая постановка позволяет исследовать влияние вариации и смещения оценок независимо от сложности функциональной формы связи. Использование более сложных нелинейных моделей затрудняет интерпретацию результатов, поскольку в этом случае одновременно изменяются как свойства самой зависимости, так и свойства оцениваемых показателей.

Кроме того, выполняется анализ реальных данных электроэнцефалографии (ЭЭГ), характеризующихся высокой размерностью и сложной структурой зависимостей, что позволяет оценить практическую применимость рассматриваемых подходов.

Основной результат работы состоит в том, что в условиях слабых зависимостей, мультиколлинеарности, ограниченного объема выборки и несбалансированности данных корреляционные меры демонстрируют более стабильное и воспроизводимое поведение по сравнению с оценками взаимной информации, несмотря на их меньшую теоретическую универсальность.

Структура работы следующая. В разделе 2 описываются методологические детали рассматриваемых подходов. Раздел 3 посвящен вычислительным экспериментам. В разделе 4 приводятся результаты анализа реальных данных электроэнцефалографии. В заключительных разделах представлены обсуждение полученных результатов и основные выводы.

2. Методы исследования.

2.1. Рассматриваемые меры зависимости и методы их оценивания. В работе рассматриваются два класса показателей взаимосвязи между признаком и откликом: корреляционные и информационно-теоретические меры.

Для количественного отклика используется коэффициент корреляции Пирсона. В случае бинарного отклика применяется точно-бисериальный коэффициент корреляции, который может быть интерпретирован как коэффициент корреляции между непрерывной переменной и бинарным индикатором, кодирующим принадлежность наблюдения к одному из классов.

Следует отметить, что даже в рамках нормальной модели существует множество оценок коэффициента корреляции, среди которых только одна является несмещенной и обладает минимальной дисперсией [16]. Тем не менее на практике, как правило, используется классическая выборочная оценка коэффициента корреляции Пирсона r [16]. Результаты проведенных в работе экспериментов показывают, что ее смещение является незначительным, а дисперсия близка к нижней границе, задаваемой неравенством Рао-Крамера.

Оценка точно-бисериального коэффициента корреляции [2] эквивалентна коэффициенту корреляции Пирсона между непрерывным признаком и бинарной переменной, принимающей значения 0 и 1. Однако данный показатель обладает ограничением: его максимальное абсолютное значение может быть меньше единицы, что зависит

от степени несбалансированности классов [17]. В связи с этим для сопоставления результатов целесообразно использовать бисериальный коэффициент корреляции, характеризующий связь между латентной непрерывной переменной, лежащей в основе бинарного отклика, и наблюдаемым признаком.

Простая оценка r_{rb} бисериального коэффициента может быть получена путем корректировки точно-бисериального коэффициента на множитель, зависящий от степени несбалансированности классов [17]. Альтернативно данный коэффициент может быть оценен методом максимального правдоподобия (обозначим соответствующую оценку как r_b), как описано в [18]. В последнем случае процедура оценивания сводится к решению задачи оптимизации, что повышает вычислительные затраты.

В качестве альтернативы корреляционным мерам рассматривается взаимная информация, позволяющая выявлять как линейные, так и нелинейные зависимости между переменными. В отличие от коэффициентов корреляции, оценивание взаимной информации в непрерывном случае требует задания дополнительных параметров, в частности параметра сглаживания при построении оценок плотностей распределения.

Для оценивания взаимной информации используются два основных подхода [19]. Первый базируется на ядерном оценивании плотностей распределения, при котором взаимная информация вычисляется на основе сглаженных оценок совместного и маргинальных распределений. Данный подход является достаточно гибким и применим как к непрерывным, так и к смешанным данным. Для снижения систематической ошибки может применяться коррекция смещения. В настоящей работе используется реализация данного подхода в пакете *mpmi* среды R [20], где коррекция смещения осуществляется методом складного ножа. Ширина окна для ядерного сглаживания определяется с использованием функции *dpik()* из пакета *KernSmooth*. Обозначим через mi_{ks}^{cc} и mi_{ks}^{cd} соответствующие оценки взаимной информации для непрерывного случая и для случая дискретного отклика.

Второй подход, предложенный в [21], основан на оценивании энтропии с использованием расстояний до k ближайших соседей (k -nearest neighbors). В работе [22] данный метод адаптирован для анализа взаимосвязи между непрерывной и бинарной переменными. Оба варианта реализованы в пакете *knnmi* среды R [23]. Однако исследования свойств оценок для непрерывного отклика,

рассчитанных с помощью функции *mutual_inf_cc()* пакета *knnmi*, показали, что они сильно смещаются к нулю с ростом k , и это смещение не соответствует, описанному в статье [21]. Такие свойства делают их непригодными для скрининга с гарантированной независимостью, поэтому был выбран пакет *YEAB* [24], где функция *mut_info_knn()* дает приемлемое смещение и показывает гораздо лучшие результаты, чем *mutual_inf_cc()* в пакете *knnmi*. Для дискретного отклика использовалась функция *mutual_inf_cd()* пакета *knnmi*.

Существенной особенностью данного подхода является зависимость результатов от выбора параметра k : с увеличением k возрастает систематическая ошибка оценивания, тогда как вариация уменьшается. В [21] рекомендуется выбирать значения параметра k в диапазоне от 2 до 4. Обозначим соответствующие оценки взаимной информации через mi_{knn}^{cc} для непрерывного случая и mi_{knn}^{cd} для случая дискретного отклика.

Рассмотренные подходы широко применяются на практике, однако их чувствительность к выбору параметров и зависимость от структуры данных могут существенно влиять на качество оценивания. Ожидается, что данная чувствительность проявляется особенно существенно в условиях ограниченного объема выборки и несбалансированности классов.

2.2. Процедура скрининга и критерии оценки.

Для предварительного отбора признаков используется процедура скрининга с гарантированной независимостью (Sure Independence Screening, SIS). Данный подход реализует двухшаговую схему. На первом этапе производится фильтрация признаков на основе их индивидуальной взаимосвязи с откликом, что позволяет исключить заведомо нерелевантные переменные. На втором этапе отобранное подмножество анализируется с применением методов регуляризации, в частности LASSO, учитывающих совместное влияние признаков.

Таким образом, ключевую роль на этапе скрининга играют статистические свойства оценок показателей взаимосвязи, используемых для ранжирования признаков, прежде всего их дисперсия и смещение.

Формально [5], в рамках обобщенной линейной модели, связывающей отклик Y с признаками $X_1, \dots, X_p$, процедура SIS предполагает выделение подмножества признаков вида:

$$\hat{M}_{\gamma_n} = \{1 \leq j \leq p: |\hat{\beta}_j^M| \geq \gamma_n\},$$

где $\hat{\beta}_j^M$ – оценка параметра обобщенной линейной модели при переменной X_j , γ_n – заданное пороговое значение, зависящее от объема выборки n . Обычно для удобства на практике задается не γ_n , а число отбираемых переменных $d_n = \left\lfloor \frac{n}{\log(n)} \right\rfloor$. Кроме того, вместо оценок параметров $\hat{\beta}_j^M$ рассматриваются оценки различных показателей взаимосвязи $\hat{\theta}_j$, например, коэффициента корреляции Пирсона, точно-бисериального коэффициента корреляции [2], взаимной информации [6]. Тогда по результатам отбора получаем подмножество:

$$\hat{M}_{d_n} = \{1 \leq j \leq p: \text{первые } d_n \text{ наибольших } |\hat{\theta}_j|\}.$$

В [25] показано, что при надлежащих условиях регулярности эта процедура обладает свойством надежного скрининга:

$$P(M_* \subset M_{\gamma_n}) \rightarrow 1 \text{ при } n \rightarrow \infty,$$

где истинная разреженная модель описывается множеством $M_* = \{1 \leq j \leq p: \beta_j^* \neq 0\}$, β_j^* соответствуют истинным значениям параметров обобщенной линейной модели. Тем самым скрининг асимптотически гарантирует достоверность того, что ни один важный признак не будет потерян при отборе.

В настоящей работе исследуется, в какой степени данное свойство сохраняется в условиях несбалансированности данных и ограниченного объема выборки. Для этого используются два типа критериев. Во-первых, анализируются статистические характеристики оценок показателей взаимосвязи, включая их рассеяние и смещение, при различных уровнях зависимости и степени несбалансированности отклика. Во-вторых, непосредственно оценивается качество процедуры отбора – вероятность включения релевантных признаков в отобранное подмножество.

Дополнительно на реальных данных электроэнцефалографии проводится анализ эффективности двухшаговой процедуры (SIS + LASSO) в задаче прогнозирования арифметических способностей, что позволяет сопоставить теоретические выводы с практическими результатами.

3. Модельные эксперименты.

3.1. План эксперимента. Для анализа статистических свойств рассматриваемых мер зависимости проведена серия модельных экспериментов в контролируемых условиях. Основная цель –

исследовать поведение оценок при варьировании силы связи между переменными и степени несбалансированности классов. Используются упрощенные линейные модели, часто применимые в анализе данных при отсутствии предположений о наличии явной нелинейности и невозможности восстановления зависимостей более сложной формы из-за ограниченности объема выборки.

Рассматривается ситуация, в которой отклик определяется одним релевантным признаком, тогда как остальные признаки являются шумовыми. Генерация данных осуществляется в два этапа.

На первом этапе моделируется непрерывная латентная переменная η , линейно связанная с информативным признаком ξ :

$$\eta = \xi + \varepsilon, \quad (1)$$

где ξ моделировалась как стандартно нормально распределенная величина, ε имеет нормальное распределение с нулевым средним и дисперсией $\rho^{-2} - 1$.

На втором этапе формируется наблюдаемый отклик: либо в виде непрерывной переменной η , либо в виде бинарной переменной ζ , полученной бинаризацией η по порогу, соответствующему выборочному квантилю уровня $P(\zeta = 0)$.

Таким образом, степень несбалансированности классов регулируется порогом бинаризации $P(\zeta = 0)$, что позволяет варьировать распределение отклика от симметричного до существенно асимметричного.

В экспериментах варьируются следующие параметры:

- уровень зависимости $\rho \in [0.01, 0.6]$ с шагом 0.01;
- степень несбалансированности классов $P(\zeta = 0) \in [0.5, 0.95]$

с шагом 0.05.

Для каждого набора параметров проводится многократное моделирование: характеристики оценок усредняются по 10^6 повторениям для корреляционных показателей (в силу их вычислительной простоты) и по 10^4 повторениям для взаимной информации. Для повышения сопоставимости результатов используются одни и те же значения случайных ошибок при различных значениях ρ .

Для моделирования задачи отбора признаков дополнительно рассматривается простая модель:

$$Y = \sum_{j=1}^m f(X_j) + \sum_{k=1}^{p-m} \varepsilon_k, \quad (2)$$

где $X_1, \dots, X_m$ – релевантные признаки с единичной дисперсией, $\varepsilon_1, \dots, \varepsilon_{p-m}$ – нерелевантные признаки с дисперсией 0.001, все признаки распределены нормально и независимо с нулевыми математическими ожиданиями, p – общее число признаков.

Сила связи между откликом и переменными регулировалась путем варьирования значений m, p . Для модели (2) при $f(X_j) = X_j$ справедливо:

$$\rho_{YX_j} = (m + 0.001 \cdot (p - m))^{-1/2}, j = 1, \dots, m, \rho_{Y\varepsilon_k} = \sqrt{0.001} \rho_{YX_1}, \\ k = 1, \dots, p - m,$$

при $f(X_j) = e^{X_j}$ получаем:

$$DY = e(e - 1)m + 0.001 \cdot (p - m), \rho_{YX_j} = \sqrt{e / DY}, j = 1, \dots, m, \\ \rho_{Y\varepsilon_k} = \sqrt{0.001 / DY}, k = 1, \dots, p - m.$$

Рассматриваются три сценария:

1. Умеренная линейная связь: $f(X_j) = X_j$, $m = 5, p = 1000$, $\rho_{YX_j} \approx 0.408, \rho_{Y\varepsilon_k} \approx 0.013$.
2. Слабая линейная связь: $f(X_j) = X_j$, $m = 10, p = 3000$, $\rho_{YX_j} \approx 0.277, \rho_{Y\varepsilon_k} \approx 0.009$.
3. Умеренная нелинейная связь: $f(X_j) = e^{X_j}$, $m = 5, p = 1000$, $\rho_{YX_j} \approx 0.334, \rho_{Y\varepsilon_k} \approx 0.006$.

Объем выборки n фиксируется равным 300. Результаты усредняются по 1000 повторениям. Для моделирования дискретного отклика признак Y в модели (2) был бинаризован с порогом, равным выборочному квантилю уровня $P(\zeta = 0)$.

3.2. Сравнимые показатели взаимосвязи. В рамках экспериментов сравниваются следующие меры зависимости:

- коэффициент корреляции Пирсона r (для непрерывного отклика),
- бисериальный коэффициент корреляции r_b (для бинарного отклика),
- точечно-бисериальный коэффициент корреляции r_{pb} (для бинарного отклика),
- взаимная информация с ядерным оцениванием плотности mi_{ks}^{cc} (для непрерывного отклика) и mi_{ks}^{cd} (для бинарного отклика),

– взаимная информация, оцененная методом ближайших соседей, mi_{knn}^{cc} (для непрерывного отклика) и mi_{knn}^{cd} (для бинарного отклика).

Для методов оценивания взаимной информации параметры сглаживания (ширина окна, число соседей) выбираются из стандартных диапазонов, используемых в прикладных исследованиях, и фиксируются в рамках каждого сценария.

Истинные значения взаимной информации в модели с непрерывным откликом MI^{cont} задавались из известного соотношения [21] для двумерного нормального распределения:

$$MI^{cont}(\rho) = -\frac{1}{2} \log(1 - \rho^2).$$

В случае бинарного отклика значения взаимной информации вычисляются с помощью численного интегрирования по определению [22] с учетом вероятностной структуры модели [18].

Результаты моделирования для случая непрерывного отклика сопоставлялись с теоретической дисперсией эффективных несмещенных оценок коэффициента корреляции для двумерного нормального распределения с неизвестными параметрами сдвига и масштаба [16]:

$$D_{ef}^{cont}(\rho) = \frac{(1-\rho^2)^2}{n}. \quad (3)$$

Это отношение указывает нижнюю границу дисперсии для несмещенных оценок коэффициента корреляции при непрерывном отклике. Сопоставление с ней полученных в ходе экспериментов оценок условного рассеяния показывает, насколько близки оценки r к эффективным.

3.3. Характеристики оценки. Качество рассматриваемых методов оценивается с использованием следующих характеристик:

- оценка условного рассеяния $\widehat{D}_{\hat{\theta}}$ оценки $\hat{\theta}$, характеризующая средний квадрат отклонений оценок от истинного значения параметра θ ;
- оценка условного смещения $\widehat{b}_{\hat{\theta}}$ относительно истинных значений θ , характеризующая систематическую ошибку;
- оценка коэффициента вариации $V\widehat{K}_{\hat{\theta}}$ оценки $\hat{\theta}$ как отношение квадратного корня из рассеяния $\widehat{D}_{\hat{\theta}}$ к истинному значению параметра θ , показывает относительный уровень флуктуации оценок;

– вероятность включения релевантных признаков в отобранное подмножество, характеризующее качество ранжирования признаков.

Особое внимание уделяется области слабых зависимостей, поскольку именно в этом режиме процедуры скрининга наиболее чувствительны к вариативности оценок и риску потери релевантных признаков.

3.4. Результаты. На рисунке 1 представлены результаты моделирования для модели (1) с непрерывным откликом. Для обеспечения сопоставимости методов анализ проводится в терминах коэффициентов вариации, представленных в логарифмической шкале.

Коэффициент корреляции Пирсона демонстрирует практически полное совпадение с теоретической нижней границей вариации, рассчитанной на основе (3), что свидетельствует о его эффективности. Таким образом, выборочная оценка корреляции фактически достигает предельной точности, допустимой в рамках рассматриваемой модели.

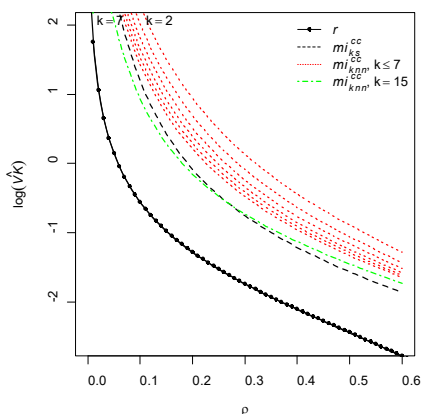


Рис. 1. Оценки коэффициента вариации оценок коэффициента корреляции и взаимной информации, в логарифмической шкале, непрерывный отклик

В отличие от этого, оценки взаимной информации демонстрируют существенно более сложное и менее стабильное поведение, зависящее от используемого метода оценивания. Ядерные оценки характеризуются монотонным снижением вариации с ростом коэффициента корреляции ρ . При этом для малых значений тесноты связи наблюдается значительное положительное смещение, которое уменьшается по мере усиления зависимости и становится отрицательным при больших значениях $\rho \geq 0.55$ (Рис. 2).

Оценки взаимной информации, основанные на расстояниях до k ближайших соседей, демонстрируют меньший коэффициент вариации по сравнению с ядерными оценками вблизи нулевой корреляции ($\rho < 0.25$) при достаточно больших значениях параметра k ($k=15$). Это достигается благодаря тому, что при малых ρ оценки k ближайших соседей имеют меньшую дисперсию и их смещение ближе к нулю (Рис. 2) по сравнению с ядерными оценками.

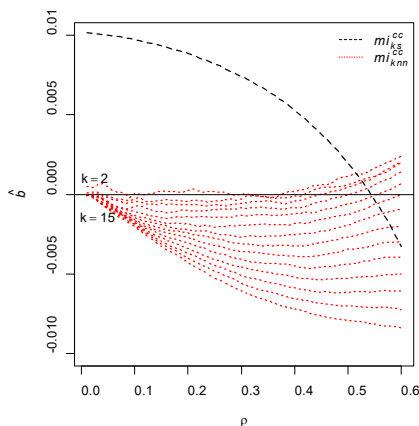


Рис. 2. Оценка смещения оценки взаимной информации, в зависимости от ρ , непрерывный отклик

Для случая бинарного отклика (Рис. 3) наблюдаются аналогичные зависимости. В условиях симметричного распределения отклика увеличение k приводит к снижению рассеяния оценок при слабой зависимости, однако сопровождается ее ростом при более высоких значениях ρ . При существенной несбалансированности данных увеличение k до 20 приводит к снижению вариации на всем диапазоне значений ρ . Тем не менее, во всех рассмотренных сценариях ядерные оценки взаимной информации демонстрируют более устойчивое поведение по сравнению с методами на основе ближайших соседей. Следовательно, при необходимости использования взаимной информации в процедурах скрининга предпочтение следует отдавать ядерным оценкам.

Сравнение оценок бисериального и точно-бисериального коэффициентов корреляции (Рис. 4) показывает, что их вариативность практически совпадает, особенно в области слабых зависимостей. При этом точно-бисериальный коэффициент обладает рядом практических преимуществ, включая вычислительную простоту

и более низкую вариативность при малых значениях ρ (вертикальными линиями отмечены значения ρ , ниже которых оценки точно-бисериальной корреляции обладают меньшим рассеянием). Это позволяет рекомендовать его в качестве основного инструмента для скрининга признаков.

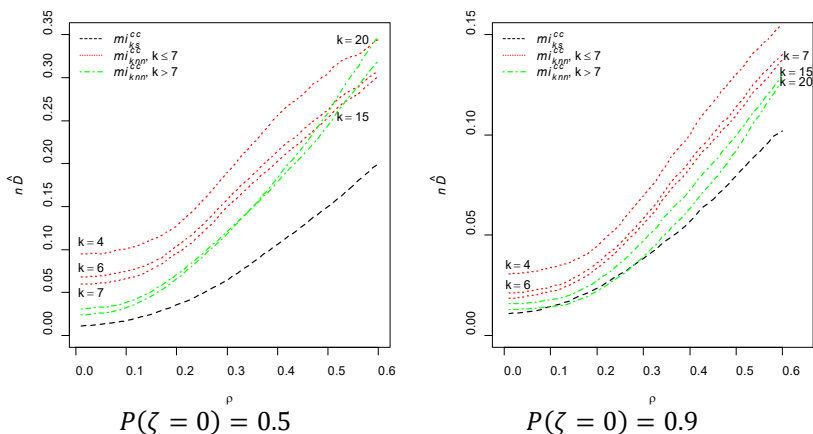


Рис. 3. Оценки рассеяния оценок взаимной информации, умноженные на объем выборки, бинарный отклик

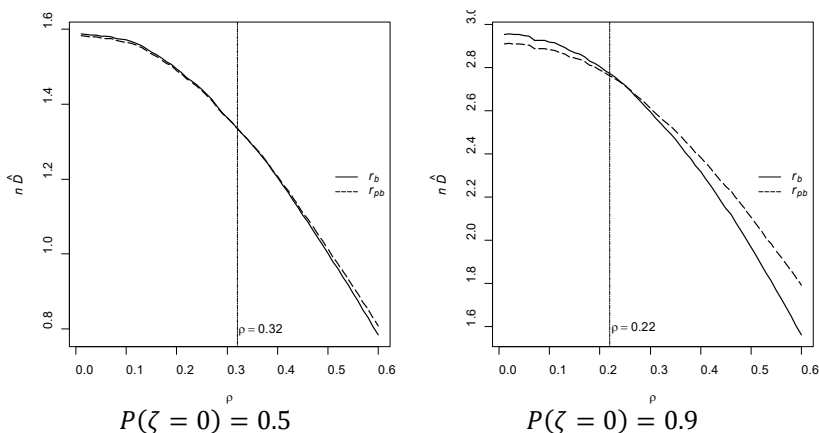


Рис. 4. Оценки рассеяния оценок коэффициента корреляции, умноженные на объем выборки, бинарный отклик

Таким образом, в ходе скрининга с гарантированной независимостью приемлемо использовать точно-бисериальный

коэффициент корреляции и ядерные оценки взаимной информации. Дальнейший анализ направлен на исследование влияния несбалансированности данных. На рисунке 5 представлено отношение коэффициентов вариации для бинарного и непрерывного отклика. Значения ρ взяты с шагом 0.05, стрелками указаны направления уменьшения значений ρ . Установлено, что с ростом асимметрии распределения отклика вариативность оценок возрастает для всех рассматриваемых мер зависимости. Это указывает на принципиальную значимость учета несбалансированности при построении процедур скрининга.

На рисунке 5 прослеживается устойчивая тенденция увеличения отношения коэффициентов вариации (для бинарного отклика по сравнению с непрерывным) по мере роста несбалансированности классов. Это указывает на существенное влияние асимметрии распределения отклика на качество процедур скрининга с гарантированной независимостью.

Для обеих рассматриваемых мер зависимости более слабые связи характеризуются меньшим относительным увеличением вариативности по сравнению с сильными зависимостями. Данный эффект является благоприятным для процедур скрининга, поскольку способствует более надежному исключению нерелевантных признаков.

В сравнении с бисериальной корреляцией ядерные оценки взаимной информации демонстрируют более низкое отношение вариации при переходе от непрерывного отклика к бинарному. Отрицательные значения на рисунке 5 свидетельствуют о том, что в условиях слабой корреляции и умеренной несбалансированности ядерные оценки взаимной информации при бинарном отклике могут обладать относительным преимуществом по сравнению с непрерывным случаем.

Прямое сравнение ядерных оценок взаимной информации mi_{ks}^{cd} и точно-бисериального коэффициента корреляции (Рис. 6) показывает, что взаимная информация систематически уступает корреляции по величине коэффициента вариации (при $\rho \rightarrow 0$ соотношение вариации ядерных оценок взаимной информации и точно-бисериального коэффициента корреляции стремительно растет). Более того, с ростом несбалансированности данных это различие существенно увеличивается. Таким образом, в условиях бинарного отклика использование взаимной информации приводит к менее стабильному ранжированию признаков.

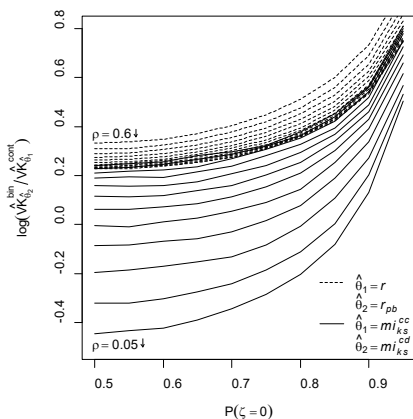


Рис. 5. Отношение оценок коэффициентов вариации ядерных оценок взаимной информации и коэффициентов корреляции при бинарном отклике к случаю непрерывного отклика в зависимости от вероятности нулевого класса, в логарифмической шкале

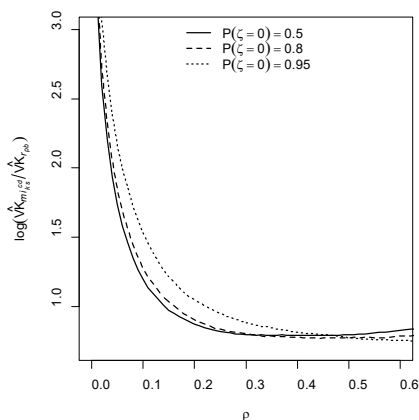


Рис. 6. Отношение оценок коэффициентов вариации оценок MI к ρ , в зависимости от ρ , в логарифмической шкале, бинарный отклик

В процедурах скрининга с гарантированной независимостью ключевым требованием является высокая вероятность включения релевантных признаков в отобранное подмножество (практическая достоверность). Увеличение дисперсии оценок показателей взаимосвязи приводит к снижению этой вероятности, ухудшая качество отбора. Существенную роль при этом играет также смещение оценок. Завышенные оценки увеличивают вероятность включения

нерелевантных признаков, что, в частности, наблюдается для ядерных оценок взаимной информации mi_{ks}^{cc} в области слабых зависимостей ($\rho < 0.55$). В свою очередь, занижение оценок mi_{knn}^{cc} , характерное для методов на основе k ближайших соседей при больших значениях k, приводит к росту вероятности пропуска значимых переменных.

Стоит отметить, что могут быть получены отрицательные оценки взаимной информации. Для целей скрининга они обнулялись.

Анализ качества скрининга подтверждает ранее полученные выводы. Для модели (1) с бинаризацией отклика в качестве порогового значения для коэффициента корреляции фиксируется уровень $\gamma_n = 0.15$, тогда как соответствующие пороги для взаимной информации зависят от доли отрицательного класса и определяются численно на основе ее определения. Предполагается, что истинная корреляция между переменными ξ и η равна 0.3. Вероятность включения релевантного признака $P(|\hat{\theta}| \geq \gamma_n)$ оценивается методом моделирования как доля реализаций, в которых значение выбранного показателя взаимосвязи превышает заданный порог.

Для модели (1) с непрерывным откликом вероятность включения релевантного признака при использовании коэффициента корреляции составляет 0.9969, тогда как для ядерных оценок взаимной информации – 0.992, что соответствует уровню практической достоверности. В то же время методы на основе k ближайших соседей демонстрируют худшие результаты (0.928 при k = 20).

Результаты для бинарного отклика (таблица 1) показывают, что с увеличением несбалансированности данных вероятность включения релевантного признака заметно снижается и перестает соответствовать уровню практической достоверности. Бисериальный и точечно-бисериальный коэффициенты корреляции дают практически совпадающие результаты, что позволяет отдать предпочтение точечно-бисериальному коэффициенту ввиду его вычислительной простоты.

Ядерные оценки взаимной информации существенно превосходят методы на основе расстояний до k ближайших соседей, однако уступают корреляционным мерам. В частности, при высокой степени несбалансированности (вероятность положительного класса 0.1) вероятность пропуска релевантного признака при использовании ядерных оценок взаимной информации составляет 0.122, тогда как для бисериальной корреляции она более чем в два раза ниже и равна 0.056.

На первом этапе скрининга с гарантированной независимостью для модели (2) пороговые значения показателей взаимосвязи γ_n задавались исходя из фиксированного размера

отбираемого подмножества признаков $d_n = \left\lfloor \frac{n}{\log(n)} \right\rfloor = 52$. При слабой линейной зависимости (сценарий В) и использовании коэффициента корреляции Пирсона все релевантные признаки включались в отобранное подмножество в 92.6% случаев, тогда как при умеренной линейной зависимости (сценарий А) – во всех реализациях, при нелинейной зависимости (сценарий С) – в 99.4% случаев. Это указывает на то, что при преобладании слабых связей в исходном наборе признаков даже скрининг на основе коэффициента корреляции не обеспечивает гарантированного включения всех значимых переменных.

Таблица 1. Оценка $P(|\hat{\theta}| \geq \gamma_n)$ при различных методах оценивания показателей взаимосвязи и разном уровне несбалансированности данных

Метод оценивания	$P(\zeta = 0)$				
	0.5	0.6	0.7	0.8	0.9
r_b	0.985	0.983	0.980	0.971	0.944
r_{pb}	0.985	0.983	0.980	0.971	0.943
mi_{ks}^{cd}	0.958	0.956	0.950	0.931	0.878
$mi_{knn}^{cd}, k=4$	0.795	0.798	0.793	0.786	0.756
$mi_{knn}^{cd}, k=6$	0.834	0.833	0.838	0.833	0.790
$mi_{knn}^{cd}, k=15$	0.913	0.911	0.914	0.890	0.798
$mi_{knn}^{cd}, k=20$	0.929	0.927	0.923	0.897	0.786

Для сценария В увеличение размера отбираемого подмножества приводит к заметному улучшению результатов. В частности, при увеличении размера до $d_n = 2 \left\lfloor \frac{n}{\log(n)} \right\rfloor = 104$ и до $d_n = 3 \left\lfloor \frac{n}{\log(n)} \right\rfloor = 156$ (в соответствии с подходом [2]) вероятность включения всех релевантных признаков возрастает до 0.971 и 0.985 соответственно, что может рассматриваться как практически приемлемый уровень.

Для сценария С при увеличении размера отбираемого подмножества вероятность включения всех релевантных признаков на основе коэффициента корреляции повышается: при $d_n = 104$ до 0.997, при $d_n = 156$ до 0.998. Для такого сценария с нелинейной зависимостью ядерные оценки взаимной информации проявляют себя еще лучше: при $d_n = 52$ вероятность включения всех релевантных признаков оценена как 0.999, а при $d_n = 104$ уже составляет 1.

В остальных сценариях использование ядерных оценок взаимной информации показывает существенно менее устойчивые результаты. В случае умеренной линейной зависимости доля реализаций, в которых все релевантные признаки включены в подмножество из 52 переменных,

составляет 0.991, тогда как при слабой зависимости она резко снижается до 0.013. Увеличение размера подмножества до 156 приводит лишь к незначительному росту этого показателя (до 0.107) в условиях слабой связи. Таким образом, в сценарии В скрининг с гарантированной независимостью на основе ядерных оценок взаимной информации оказывается практически непригодным.

Результаты для модели (2) при бинаризации отклика представлены в таблице 2, где $P\left(\bigcap_{j=1}^m \left|\hat{\theta}_{YX_j}\right| \geq \gamma_n\right)$ обозначает вероятность того, что оценки показателей взаимосвязи одновременно для всех релевантных признаков превышают заданный порог. Для моделей А и С в условиях умеренной зависимости получены приемлемые результаты. Точечно-бисериальный коэффициент корреляции обеспечивает практически достоверное включение всех релевантных признаков даже при очень высокой степени асимметрии распределения отклика. При низкой доле положительного класса (0.1) ядерные оценки взаимной информации демонстрируют неудовлетворительные результаты в модели А, тогда как в условиях нелинейных зависимостей (модель С) плохой результат дает и точечно-бисериальный коэффициент корреляции. Как и следовало ожидать, увеличение степени несбалансированности приводит к ухудшению качества скрининга.

Несмотря на умеренные нелинейные зависимости в модели С при достаточно сбалансированных данных (вероятность отрицательного класса от 0.5 до 0.6) точечно-бисериальный коэффициент корреляции обеспечивает такое же качество скрининга, как и взаимная информация. При увеличении вероятности отрицательного класса до 0.8 наблюдаются незначительные преимущества взаимной информации, тогда как при 0.9 – различия в пользу точечно-бисериального коэффициента корреляции, но они очень малы. Следовательно, при умеренной нелинейной связи скрининг на основе коэффициента корреляции предпочтителен, поскольку обеспечивает сопоставимое с взаимной информацией качество отбора, но вычислительно гораздо менее затратный.

В целом результаты моделирования выявляют устойчивые различия в поведении рассматриваемых мер зависимости. Корреляционные показатели обеспечивают низкую вариативность оценок и стабильное ранжирование признаков в условиях умеренной зависимости, включая сценарии с высокой несбалансированностью данных. В то же время при слабых зависимостях эффективность скрининга снижается даже для корреляционных мер.

Напротив, методы оценивания взаимной информации характеризуются существенно большей вариативностью, особенно в области слабых зависимостей, причем данный эффект усиливается с ростом несбалансированности данных.

Таблица 2. Оценка $P\left(\bigcap_{j=1}^m |\hat{\theta}_{YX_j}| \geq \gamma_n\right)$ при различных методах оценивания показателей взаимосвязи и разном уровне несбалансированности данных

Метод оценивания	Сценарий, d_n	$P(\zeta = 0)$				
		0.5	0.6	0.7	0.8	0.9
r_{pb}	A, 52	1	0.999	0.998	0.996	0.95
r_{pb}	A, 104	1	1	1	1	0.984
r_{pb}	A, 156	1	1	1	1	0.994
mi_{ks}^{cd}	A, 52	0.982	0.983	0.974	0.92	0.603
mi_{ks}^{cd}	A, 104	0.995	0.993	0.991	0.963	0.787
mi_{ks}^{cd}	A, 156	0.998	0.998	0.996	0.981	0.865
r_{pb}	B, 52	0.406	0.375	0.27	0.121	0.004
r_{pb}	B, 104	0.642	0.603	0.509	0.273	0.025
r_{pb}	B, 156	0.743	0.709	0.627	0.406	0.081
mi_{ks}^{cd}	B, 52	0.015	0.016	0.002	0	0
mi_{ks}^{cd}	B, 104	0.073	0.067	0.024	0.009	0
mi_{ks}^{cd}	B, 156	0.139	0.115	0.06	0.022	0
r_{pb}	C, 52	1	0.998	0.993	0.951	0.652
r_{pb}	C, 104	1	1	0.997	0.974	0.795
r_{pb}	C, 156	1	1	1	0.984	0.856
mi_{ks}^{cd}	C, 52	1	0.998	0.995	0.962	0.655
mi_{ks}^{cd}	C, 104	1	1	0.998	0.976	0.781
mi_{ks}^{cd}	C, 156	1	1	0.999	0.988	0.847

В целом проведенные эксперименты выявляют устойчивое преимущество корреляционных мер зависимости. Они обеспечивают низкую вариативность оценок и высокую вероятность корректного отбора релевантных признаков в широком диапазоне условий. Взаимная информация, несмотря на теоретическую универсальность, характеризуется повышенной вариативностью и чувствительностью к параметрам оценивания, что приводит к снижению качества скрининга, особенно в условиях слабых зависимостей и несбалансированных данных.

3.5. Выводы. Проведенные вычислительные эксперименты позволяют сделать следующие обобщающие выводы.

Во-первых, корреляционные меры зависимости демонстрируют наименьшую вариативность оценок и обеспечивают наиболее стабильное ранжирование признаков в широком диапазоне параметров, включая условия слабых зависимостей.

Во-вторых, оценки взаимной информации характеризуются существенно большей вариативностью, особенно в области слабых зависимостей, что приводит к снижению вероятности включения релевантных признаков в процедуры скрининга.

В-третьих, методы оценивания взаимной информации на основе k ближайших соседей проявляют высокую чувствительность к выбору параметра k , вследствие чего уступают ядерным методам и не обеспечивают надежного отбора признаков.

В-четвертых, несбалансированность данных приводит к росту вариативности оценок для всех рассматриваемых мер зависимости, однако данный эффект значительно сильнее выражен для взаимной информации, что дополнительно снижает эффективность соответствующих процедур скрининга.

4. Исследование на реальных данных: анализ ЭЭГ

4.1. Описание данных. Для демонстрации практической применимости рассмотренных подходов используется набор данных ЭЭГ, характеризующийся высокой размерностью признакового пространства и сложной структурой зависимостей между признаками.

Данные заимствованы из [26] и содержат записи ЭЭГ, а также оценки арифметических способностей 36 испытуемых, полученные в процессе выполнения когнитивных заданий. Таким образом, анализ проводится в условиях малого объема выборки.

Предварительная обработка данных выполнена в работе [27]. В качестве признаков использованы метрики связности графов, построенных на основе пяти методов оценки функциональной коннективности (Coh – когерентность, ImCoh – мнимая когерентность, PLI – индекс фазового лага, DPLI – направленный индекс фазового лага, WPLI – взвешенный индекс фазового лага) в пяти частотных диапазонах $(\delta, \theta, \alpha, \beta, \gamma)$. Для каждой комбинации метода и диапазона формировался граф коннективности, определяемый значимыми взаимодействиями между областями мозга, после чего для каждого графа вычислялись семь метрик связности: APL, CPL – средняя и характерная длина пути, D – диаметр, CC – коэффициент кластеризации, Mod – модулярность, EC – центральность по собственному вектору, Cl – центральность по близости. Таким образом, исходно формировалось $5 \times 5 \times 7 = 175$ признаков.

Дополнительно признаки рассчитывались для девяти уровней порогов значимости (квантилей от 0.1 до 0.9 абсолютных значений показателей коннективности для каждого индивида), что увеличило размерность признакового пространства до 1575 переменных. При этом для малых значений порога отдельные метрики связности (в частности, средняя и медианная длина пути, диаметр) принимали практически постоянные значения. Для предотвращения искажения оценок взаимосвязи из анализа были исключены признаки, в которых одно значение встречалось с частотой 60% и более. В результате итоговая размерность пространства признаков составила 1141.

Таким образом, число признаков существенно превышает объем выборки, что делает задачу особенно чувствительной к качеству предварительного отбора.

Преимуществом данного набора данных является возможность рассмотрения двух типов отклика. В качестве непрерывного отклика используется показатель арифметической продуктивности – число правильно решенных заданий за единицу времени. Кроме того, на основе этого показателя формируется бинарный отклик, отражающий уровень математических способностей: испытуемые классифицируются в зависимости от того, превышает ли число решенных задач пороговое значение 10. Доля испытуемых с результатом, не превышающим данный порог, составляет 27.7%, что соответствует выраженной несбалансированности классов.

Перед проведением анализа данные были стандартизованы.

Используемые признаки представляют собой агрегированные характеристики графов функциональной коннективности мозга. Для них линейные аппроксимации широко применяются в исследованиях индивидуальных различий и когнитивных способностей, что подтверждается работами, посвященными анализу связности ЭЭГ [28, 29]. Поэтому применение известной двухшаговой процедуры скрининга с гарантированной независимостью [2] на основе линейной модели с регуляризацией представляется оправданным.

4.2. План исследования. На первом этапе выполняется скрининг признаков с использованием различных мер зависимости: точно-бисериального коэффициента корреляции и взаимной информации, оцененной ядерным методом. Признаки ранжируются по величине соответствующего показателя, после чего отбирается подмножество наиболее значимых переменных. Пороговые значения параметров взаимосвязи γ_n задаются косвенно – через фиксированный размер отбираемого подмножества признаков $d_n = \left\lfloor \frac{n}{\log(n)} \right\rfloor = 10$.

На втором этапе на отобранных признаках строится предсказательная модель с использованием LASSO-регрессии, позволяющей учитывать совместное влияние переменных и выполнять регуляризацию. Поскольку исходное пространство признаков характеризуется выраженной мультиколлинеарностью [27], предварительно на основе отобранных признаков строятся главные компоненты. Это обеспечивает ортогональность входных факторов и повышает устойчивость оценивания, одновременно позволяя учитывать структуру взаимосвязей между признаками. Использование главных компонент также приводит к дополнительному снижению размерности, что важно для интерпретируемости результатов.

4.3. Метрики оценивания. Оценка качества моделей проводится с использованием процедуры перекрестной проверки по принципу «leave-one-out» (LOOCV), что позволяет минимизировать риск переобучения при малом объеме выборки.

Для непрерывного отклика в рамках LOOCV рассчитывается среднеквадратическая ошибка (MSE). Для бинарного отклика используется обобщенная линейная модель с LASSO-регуляризацией, и в качестве метрики качества рассматривается среднее отклонение, вычисляемое по результатам LOOCV:

$$Deviance = -\frac{2}{n} \sum_{i=1}^n y_i \log \hat{p}_i + (1 - y_i) \log(1 - \hat{p}_i),$$

где y_i – наблюдаемое значение бинарного отклика, а $\hat{p}_i$ – оценка вероятности принадлежности i -го объекта к положительному классу, полученная по модели, обученной без учета данного наблюдения.

Дополнительно анализируется устойчивость процедуры отбора признаков при варьировании размера отобранного подмножества, а также при использовании различных мер зависимости.

4.4. Результаты. На рисунке 7 представлены зависимости среднеквадратической ошибки от параметра регуляризации λ для моделей без предварительного отбора признаков (all), а также с отбором на основе коэффициента корреляции Пирсона и ядерных оценок взаимной информации. Штриховыми линиями соответствующего цвета отображены отклонения от среднего на одну стандартную ошибку.

Во всех случаях предварительный скрининг приводит к снижению ошибки по сравнению с использованием полного набора признаков. Однако значимое улучшение по сравнению с базовой моделью (без признаков) достигается только при использовании

корреляционного критерия: в этом случае ошибка снижается почти в два раза. При выборе параметра регуляризации по правилу одного стандартного отклонения итоговая модель включает лишь одну главную компоненту, которая демонстрирует сильную положительную связь с откликом (коэффициент корреляции Пирсона равен 0.781).

Стабильность результатов скрининга на основе коэффициента корреляции исследована при варьировании размера подмножества признаков d_n в диапазоне от 10 до 2 (Рис. 8). Для интерпретации вклада отдельных признаков ($\hat{\beta}_j^M$) оценки коэффициентов LASSO-регрессии при главной компоненте пересчитывались с учетом нагрузок исходных признаков. Полученные результаты свидетельствуют о том, что ряд характеристик структуры графа коннективности, в частности коэффициент кластеризации CC [27], положительно связан с арифметическими способностями.

Для бинарного отклика (Рис. 9) получены качественно сходные результаты при варьировании параметра регуляризации λ . Предварительный скрининг также снижает значение функции потерь *Deviance* по сравнению с использованием всех признаков (all). Однако для ядерных оценок взаимной информации улучшение достигается только при включении нескольких главных компонент (3-4), что существенно усложняет интерпретацию модели. В то же время при использовании точно-бисериального коэффициента корреляции минимум отклонения достигается уже при одной компоненте.

Анализ устойчивости оценок влияния признаков в бинарном случае (рисунок 10) показывает, что знаки эффектов распределяются примерно поровну между положительными и отрицательными. При этом наблюдается согласованность с непрерывным случаем для ряда характеристик: например, диаметр графа на основе мнимой когерентности в α -диапазоне при пороге значимости 0.5 ($D \propto ImCoh 0.5$) и модулярность ($Mod \propto DPLI 0.7$) оказывают отрицательное влияние, тогда как центральность по собственному вектору ($EC \propto DPLI$) на основе направленного индекса фазового лага в δ -диапазоне – положительное.

В целом результаты показывают, что выбор меры зависимости оказывает существенное влияние на итоговое качество модели. Скрининг на основе корреляции обеспечивает более низкую прогнозную ошибку, а также демонстрирует большую устойчивость к вариациям выборки. В то же время использование взаимной информации, хотя и улучшает результаты по сравнению с отсутствием отбора, не приводит к значимому снижению ошибки относительно базового уровня.

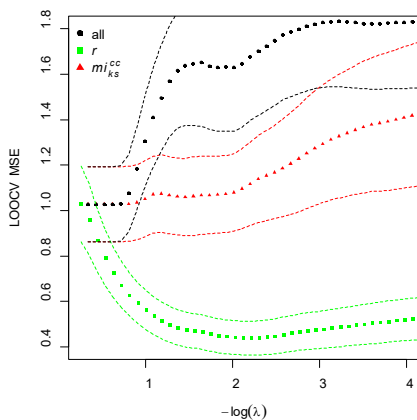


Рис. 7. Среднеквадратическая ошибка LASSO-регрессии при разных вариантах отбора признаков

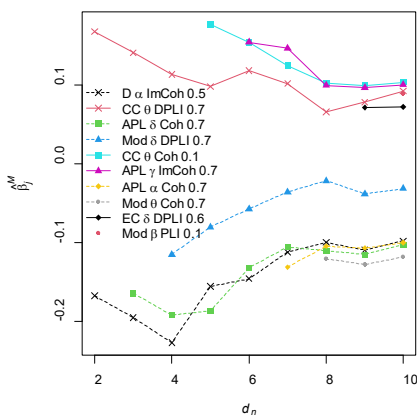


Рис. 8. Результаты скрининга с гарантированной независимостью на основе коэффициента корреляции Пирсона

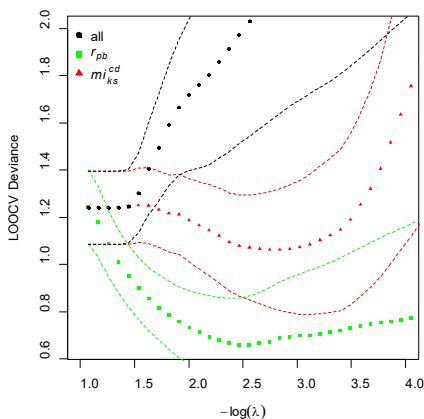


Рис. 9. Среднее отклонение обобщенной линейной LASSO-регрессии при разных вариантах отбора признаков

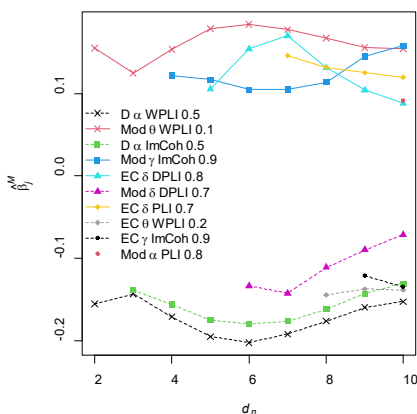


Рис. 10. Результаты скрининга с гарантированной независимостью на основе точно-бисериального коэффициента корреляции

4.5. Выводы. Полученные результаты подтверждают выводы модельных экспериментов и позволяют сформулировать следующие практические положения.

Во-первых, в условиях высокой размерности признакового пространства и несбалансированности данных корреляционные меры являются более надежным инструментом для предварительного отбора признаков, обеспечивая устойчивые и воспроизводимые результаты.

Во-вторых, применение взаимной информации требует осторожности, поскольку процедуры ее оценивания являются

вычислительно более сложными и чувствительными к выбору параметров, что снижает устойчивость результатов.

В-третьих, преимущества взаимной информации как универсальной меры зависимости не всегда реализуются на практике, особенно при малом объеме выборки, слабых зависимостях и выраженной несбалансированности данных.

5. Обсуждение. Полученные результаты демонстрируют согласованную картину как на модельных, так и на реальных данных: выбор меры зависимости оказывает существенное влияние на эффективность процедур скрининга признаков, особенно в условиях слабых зависимостей и несбалансированности данных.

Ключевым результатом является устойчиво более высокая вариативность оценок взаимной информации по сравнению с корреляционными показателями. Это приводит к снижению надежности ранжирования признаков и увеличивает вероятность пропуска релевантных переменных при использовании скрининга с гарантированной независимостью. Данный эффект особенно критичен в задачах, где решения принимаются на основе малых различий между слабыми сигналами.

Наблюдаемые различия объясняются особенностями процедур оценивания. Взаимная информация требует дополнительного этапа оценивания распределений, что связано с выбором настраиваемых параметров (например, числа ближайших соседей или ширины окна в ядерных методах). Эти параметры определяют компромисс между смещением и дисперсией и делают оценки чувствительными к объему выборки и структуре данных. В отличие от этого, корреляционные меры имеют аналитические оценки и не требуют настройки, что обеспечивает их более стабильное поведение.

Несбалансированность данных усиливает выявленные различия. Асимметрия распределения отклика приводит к увеличению неопределенности оценок взаимной информации, тогда как корреляционные показатели оказываются менее чувствительными к данному фактору. Это согласуется с результатами моделирования, где рост степени несбалансированности сопровождался увеличением вариативности оценок взаимной информации и снижением качества скрининга.

Полученные результаты показывают, что теоретическая универсальность взаимной информации не всегда трансформируется в практические преимущества в задачах предварительного отбора признаков. В условиях ограниченного объема выборки и слабых зависимостей более простые меры могут обеспечивать более надежные

и воспроизводимые результаты. Это подчеркивает необходимость учета статистических свойств оценок при выборе методов анализа.

В то же время следует отметить ограничения проведенного исследования. Рассмотренные модели охватывают широкий, но не исчерпывающий класс зависимостей: линейные либо адекватно аппроксимируемые линейной моделью. В частности, при наличии выраженных нелинейных эффектов взаимная информация потенциально может обладать преимуществами. Кроме того, результаты могут зависеть от конкретных методов оценивания и выбора параметров. Эти ограничения требуют проведения дополнительных исследований, выходящих за рамки настоящей работы.

6. Заключение. В работе проведен сравнительный анализ корреляционных и информационно-теоретических мер зависимости в задачах скрининга признаков при наличии несбалансированных данных.

На основе результатов модельных экспериментов и анализа реальных данных показано, что корреляционные меры обеспечивают более стабильные и воспроизводимые оценки в условиях слабых зависимостей, ограниченного объема выборки и асимметричного распределения отклика. Взаимная информация, несмотря на свою теоретическую универсальность, характеризуется повышенной вариативностью и чувствительностью к параметрам оценивания, что приводит к снижению эффективности процедур предварительного отбора признаков.

Полученные результаты указывают на то, что при выборе меры зависимости для задач скрининга необходимо учитывать не только способность выявлять сложные зависимости, но и статистическую устойчивость оценок. В рассматриваемых условиях простые корреляционные показатели оказываются предпочтительными с точки зрения надежности и воспроизводимости результатов.

Дальнейшие исследования могут быть направлены на анализ более сложных типов зависимостей и на разработку методов оценивания взаимной информации, обладающих повышенной устойчивостью в условиях малых выборок и несбалансированных данных.

Литература

1. Liu J.Y., Zhong W., Li R.Z. A selective overview of feature screening for ultrahigh-dimensional data // Science China Mathematics. 2015. vol. 58. pp. 1–22. DOI: 10.1007/s11425-015-5062-9.
2. Jiang J., Shang J. Feature Screening for High-Dimensional Variable Selection in Generalized Linear Models // Entropy. 2023. vol. 25. no. 6. 851 p. DOI: 10.3390/e25060851.

3. Wasikowski M., Chen X. Combating the Small Sample Class Imbalance Problem Using Feature Selection // IEEE Transactions on Knowledge and Data Engineering. 2010. vol. 22. no. 10. pp. 1388–1400. DOI: 10.1109/TKDE.2009.187.
4. Chan J.Y.-L., Leow S.M.H., Bea K.T., et al. Mitigating the Multicollinearity Problem and its Machine Learning Approach: A Review // Mathematics. 2022. vol. 10. no. 8. 1283 p. DOI: 10.3390/math10081283.
5. Fan J., Fan Y. High Dimensional Classification Using Features Annealed Independence Rules // Annals of Statistics. 2008. vol. 36. no. 6. pp. 2605–2637. DOI: 10.1214/07-AOS504.
6. Xu Y., Qian Q. i-SISSO: Mutual information-based improved sure independent screening and sparsifying operator algorithm // Engineering Applications of Artificial Intelligence. 2022. vol. 116. 105442 p. DOI: 10.1016/j.engappai.2022.105442.
7. Shukla G., Singh S., Dhule C., et al. Point biserial correlation symbiotic organism search nanoengineering based drug delivery for tumor diagnosis // Scientific Reports. 2024. vol. 14. 6530 p. DOI: 10.1038/s41598-024-55159-6.
8. Huyen C. Designing Machine Learning Systems // Sebastopol: O'ReillyMedia. 2022. 443 p.
9. Mqadi N., Naicker N., Adeliyi T. Solving Misclassification of the Credit Card Imbalance Problem Using Near Miss // Mathematical Problems in Engineering. 2021. no. 1. 1–16 pp. DOI: 10.1155/2021/7194728.
10. Tasci E., Zhuge Y., Camphausen K., Krauze, A.V. Bias and Class Imbalance in Oncologic Data – Towards Inclusive and Transferrable AI in Large Scale Oncology Data Sets // Cancers. 2022. vol. 14. no. 12. 2897 p. DOI: 10.3390/cancers14122897.
11. Wongvorachan T., He S., Bulut O. A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining // Information. 2023. vol. 14. no. 1. 54 p. DOI: 10.3390/info14010054.
12. Kumar A., Goel S., Sinha N., Bhardwaj A. A Review on Unbalanced Data Classification // Proceedings of International Joint Conference on Advances in Computational Intelligence (IJCACI 2021). 2022. pp. 197–208. DOI: 10.1007/978-981-19-0332-8_14.
13. Mladenic D., Grobelnik M. Feature Selection for Unbalanced Class Distribution and Naive Bayes // Proceedings of the Sixteenth International Conference on Machine Learning (ICML '99). 1999. pp. 258–267.
14. Wang L., Han M., Li X., et al. Review of Classification Methods on Unbalanced Data Sets // IEEE Access. 2021. vol. 9. pp. 64606–64628. DOI: 10.1109/ACCESS.2021.3074243.
15. Belghazi M.I., Baratin A., Rajeshwar S., et al. Mutual Information Neural Estimation // Proceedings of the 35th International Conference on Machine Learning (PMLR). 2018. vol. 80. pp. 531–540.
16. Джонсон Н.Л., Коц С., Балакришнан Н. Одномерные непрерывные распределения: ч. 2 / Пер. 2-го англ. изд. Волковой О.И., Стригуновой М.С., Шиховой Н.А. / Под ред. Чепурина Е.В. // М.: БИНОМ. Лаборатория знаний. 2012. 600 с.
17. Cheng Y., Liu H. A short note on the maximal point-biserial correlation under non-normality // British Journal of Mathematical and Statistical Psychology. 2016. vol. 69. no. 3. pp. 344–351. DOI: 10.1111/bmsp.12075.
18. Timofeeva A.Y. Data Type Detection for Choosing an Appropriate Correlation Coefficient in the Bivariate Case // Proceedings of the Second Young Scientist's International Workshop on Trends in Information Processing (YSIP2). 2017. vol. 1837. pp. 188–194.
19. Czyz P., Grabowski F., Vogt J., Beerenwinkel N., Marx A. Beyond Normal: On the Evaluation of Mutual Information Estimators // Advances in Neural Information Processing Systems. 2023. vol. 36. pp. 1–34. DOI: 10.52202/075280-0742.

20. Mpmi: Mixed-Pair Mutual Information Estimators. URL: <https://CRAN.R-project.org/package=mpmi> (дата обращения: 20.04.2026).
21. Kraskov A., Stogbauer H., Grassberger P. Estimating mutual information // *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics*. 2004. vol. 69. no. 6. 066138 p. DOI: 10.1103/PhysRevE.69.066138.
22. Ross B.C. Mutual Information between Discrete and Continuous Data Sets // *PloS ONE*. 2014. vol. 9. no. 2. e87357 p. DOI: 10.1371/journal.pone.0087357.
23. Knnmi: k-Nearest Neighbor Mutual Information Estimator. URL: <https://CRAN.R-project.org/package=knnmi> (дата обращения: 20.04.2026).
24. YEAB: Analyze Data from Analysis of Behavior Experiments. URL: <https://CRAN.R-project.org/package=YEAB> (дата обращения: 20.04.2026).
25. Fan J., Song R. Sure Independence screening in generalized linear models with NP-dimensionality // *Annals of Statistics*. 2010. vol. 38. no. 6. pp. 3567–3604. DOI: 10.1214/10-AOS798.
26. Zyma I., Tukaev S., Seleznev I., et al. Electroencephalograms during Mental Arithmetic Task Performance // *Data*. 2019. vol. 4. no. 1. pp. 14. DOI: 10.3390/data4010014.
27. Timofeeva A.Y., Avdeenko T.V., Alkov S.S. A Study of Robust Partial Correlation between Electroencephalogram Connectivity and Human Arithmetic Abilities // *Pattern Recognition and Image Analysis*. 2025. vol. 35. pp. 189–200. DOI: 10.1134/S1054661825700105.
28. Zakharov I., Tabueva A., Adamovich T., et al. Alpha Band Resting-State EEG Connectivity is Associated with Non-verbal Intelligence // *Frontiers in Human Neuroscience*. 2020. vol. 14. no. 10. DOI: 10.3389/fnhum.2020.00010.
29. Авдеевко Т.В., Тимофеева А.Ю., Муртазина М.Ш., Разумникова О.М. Выявление устойчивых связей между показателями коннективности ЭЭГ и компонентами интеллекта // *Вестник Южно-Уральского государственного университета*. 2022. Т. 11. №4. С. 19–36.

Тимофеева Анастасия Юрьевна — канд. экон. наук, доцент, кафедра теоретической и прикладной информатики, Новосибирский государственный технический университет (НГТУ). Область научных интересов: развитие методов проектирования признаков для построения интерпретируемых моделей машинного обучения. Число научных публикаций — 100. a.timofeeva@corp.nstu.ru; пр. Карла Маркса, 20, 630073, Новосибирск, Россия; р.т.: +7(383)346-0600.

Авдеевко Татьяна Владимировна — д-р техн. наук, профессор, кафедра теоретической и прикладной информатики, Новосибирский государственный технический университет (НГТУ). Область научных интересов: моделирование и идентификация, математические и информационные методы поддержки принятия решений, инженерия знаний. Число научных публикаций — 250. avdeenko2101@gmail.com; пр. Карла Маркса, 20, 630073, Новосибирск, Россия; р.т.: +7(383)346-0600.

A. TIMOFEEVA, T. AVDEENKO

FEATURE SELECTION UNDER CLASS IMBALANCE: AN EMPIRICAL COMPARISON OF CORRELATION AND MUTUAL INFORMATION

Timofeeva A., Avdeenko T. Feature Selection under Class Imbalance: An Empirical Comparison of Correlation and Mutual Information.

Abstract. The problem of data imbalance in feature prescreening tasks using sure independence screening is considered. Within this framework, the choice of the dependence measure used for feature ranking is analyzed. This study investigates two of the most common association measures – the correlation coefficient and mutual information. The correlation coefficient characterizes linear dependence between variables; its estimates are computationally simple and possess favorable statistical properties. Mutual information is a more universal measure capable of detecting nonlinear dependencies; however, its estimation requires the reconstruction of distributions from sample data, which leads to increased computational complexity and sensitivity to parameter selection. Based on a series of computational experiments, it is shown that mutual information estimation methods based on k-nearest neighbors are inferior to kernel-based methods in terms of stability and screening quality. It is established that across all considered scenarios, including various levels of data imbalance, correlation-based measures provide more stable estimates and more reliable feature ranking compared to mutual information. The practical applicability of the results is demonstrated on real electroencephalography (EEG) data related to the analysis of arithmetic abilities. It is shown that a two-step procedure comprising screening based on the correlation coefficient followed by modeling using LASSO regression provides a substantial reduction in cross-validation error compared to approaches based on mutual information and those without prescreening. The use of the principal component constructed from the selected features further improves prediction accuracy. The obtained results indicate that in the context of small sample sizes, multicollinearity, and data imbalance, correlation-based measures are preferable for feature prescreening tasks in terms of stability and reproducibility.

Keywords: imbalanced data, mutual information, correlation coefficient, sure independence screening, kernel estimates, k-nearest neighbors, LASSO regression.

References

1. Liu J.Y., Zhong W., Li R.Z. A selective overview of feature screening for ultrahigh-dimensional data. *Science China Mathematics*. 2015. vol. 58. pp. 1–22. DOI: 10.1007/s11425-015-5062-9.
2. Jiang J., Shang J. Feature Screening for High-Dimensional Variable Selection in Generalized Linear Models. *Entropy*. 2023. vol. 25. no. 6. 851 p. DOI: 10.3390/e25060851.
3. Wasikowski M., Chen X. Combating the Small Sample Class Imbalance Problem Using Feature Selection. *IEEE Transactions on Knowledge and Data Engineering*. 2010. vol. 22. no. 10. pp. 1388–1400. DOI: 10.1109/TKDE.2009.187.
4. Chan J.Y.-L., Leow S.M.H., Bea K.T., et al. Mitigating the Multicollinearity Problem and its Machine Learning Approach: A Review. *Mathematics*. 2022. vol. 10. no. 8. 1283 p. DOI: 10.3390/math10081283.
5. Fan J., Fan Y. High Dimensional Classification Using Features Annealed Independence Rules. *Annals of Statistics*. 2008. vol. 36. no. 6. pp. 2605–2637. DOI: 10.1214/07-AOS504.

6. Xu Y., Qian Q. i-SISSO: Mutual information-based improved sure independent screening and sparsifying operator algorithm. *Engineering Applications of Artificial Intelligence*. 2022. vol. 116. 105442 p. DOI: 10.1016/j.engappai.2022.105442.
7. Shukla G., Singh S., Dhule C., et al. Point biserial correlation symbiotic organism search nanoengineering based drug delivery for tumor diagnosis. *Scientific Reports*. 2024. vol. 14. 6530 p. DOI: 10.1038/s41598-024-55159-6.
8. Huyen C. *Designing Machine Learning Systems*. Sebastopol: O'ReillyMedia. 2022. 443 p.
9. Mqadi N., Naicker N., Adeliyi T. Solving Misclassification of the Credit Card Imbalance Problem Using Near Miss. *Mathematical Problems in Engineering*. 2021. no. 1. 1–16 pp. DOI: 10.1155/2021/7194728.
10. Tasci E., Zhuge Y., Camphausen K., Krauze, A.V. Bias and Class Imbalance in Oncologic Data – Towards Inclusive and Transferrable AI in Large Scale Oncology Data Sets. *Cancers*. 2022. vol. 14. no. 12. 2897 p. DOI: 10.3390/cancers14122897.
11. Wongvorachan T., He S., Bulut O. A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. *Information*. 2023. vol. 14. no. 1. 54 p. DOI: 10.3390/info14010054.
12. Kumar A., Goel S., Sinha N., Bhardwaj A. A Review on Unbalanced Data Classification. *Proceedings of International Joint Conference on Advances in Computational Intelligence (IJCACI 2021)*. 2022. pp. 197–208. DOI: 10.1007/978-981-19-0332-8_14.
13. Mladenic D., Grobelnik M. Feature Selection for Unbalanced Class Distribution and Naive Bayes. *Proceedings of the Sixteenth International Conference on Machine Learning (ICML '99)*. 1999. pp. 258–267.
14. Wang L., Han M., Li X., et al. Review of Classification Methods on Unbalanced Data Sets. *IEEE Access*. 2021. vol. 9. pp. 64606–64628. DOI: 10.1109/ACCESS.2021.3074243.
15. Belghazi M.I., Baratin A., Rajeshwar S., et al. Mutual Information Neural Estimation. *Proceedings of the 35th International Conference on Machine Learning (PMLR)*. 2018. vol. 80. pp. 531–540.
16. Johnson N.L., Kots S., Balakrishnan N. *Odnomernyy'e nepreryvny'e raspredeleniya: ch. 2 [One-dimensional continuous distributions: part 2] (Per. 2-go angl. izd. Volkovoj O.I., Strigunovoj M.S., Shixovoj N.A. / Russ. ed.: Chepurina E.V). M.: BINOM. Laboratoriya znaniy. 2012. 600 p. (In Russ.)*.
17. Cheng Y., Liu H. A short note on the maximal point-biserial correlation under non-normality. *British Journal of Mathematical and Statistical Psychology*. 2016. vol. 69. no. 3. pp. 344–351. DOI: 10.1111/bmsp.12075.
18. Timofeeva A.Y. Data Type Detection for Choosing an Appropriate Correlation Coefficient in the Bivariate Case. *Proceedings of the Second Young Scientist's International Workshop on Trends in Information Processing (YSIP2)*. 2017. vol. 1837. pp. 188–194.
19. Czyz P., Grabowski F., Vogt J., Beerenwinkel N., Marx A. Beyond Normal: On the Evaluation of Mutual Information Estimators. *Advances in Neural Information Processing Systems*. 2023. vol. 36. pp. 1–34. DOI: 10.52202/075280-0742.
20. Mpmi: Mixed-Pair Mutual Information Estimators. Available at: <https://CRAN.R-project.org/package=mpmi> (accessed 20.04.2026).
21. Kraskov A., Stogbauer H., Grassberger P. Estimating mutual information. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics*. 2004. vol. 69. no. 6. 066138 p. DOI: 10.1103/PhysRevE.69.066138.
22. Ross B.C. Mutual Information between Discrete and Continuous Data Sets. *PloS ONE*. 2014. vol. 9. no. 2. e87357 p. DOI: 10.1371/journal.pone.0087357.
23. Knnmi: k-Nearest Neighbor Mutual Information Estimator. Available at: <https://CRAN.R-project.org/package=knnmi> (accessed 20.04.2026).
24. YEAB: Analyze Data from Analysis of Behavior Experiments. Available at: <https://CRAN.R-project.org/package=YEAB> (accessed 20.04.2026).

25. Fan J., Song R. Sure Independence screening in generalized linear models with NP-dimensionality. *Annals of Statistics*. 2010. vol. 38. no. 6. pp. 3567–3604. DOI: 10.1214/10-AOS798.
26. Zyma I., Tukaev S., Seleznev I., et al. Electroencephalograms during Mental Arithmetic Task Performance. *Data*. 2019. vol. 4. no. 1. pp. 14. DOI: 10.3390/data4010014.
27. Timofeeva A.Y., Avdeenko T.V., Alkov S.S. A Study of Robust Partial Correlation between Electroencephalogram Connectivity and Human Arithmetic Abilities. *Pattern Recognition and Image Analysis*. 2025. vol. 35. pp. 189–200. DOI: 10.1134/S1054661825700105.
28. Zakharov I., Tabueva A., Adamovich T., et al. Alpha Band Resting-State EEG Connectivity is Associated with Non-verbal Intelligence. *Frontiers in Human Neuroscience*. 2020. vol. 14. no. 10. DOI: 10.3389/fnhum.2020.00010.
29. Avdeenko T.V., Timofeeva A.Y., Murtazina M.S., Razumnikova O.M. [Identification of stable relationships between EEG connectivity indicators and components of intelligence]. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta – Bulletin of the South Ural State University*. 2022. vol. 11. no. 4. pp. 19–36. (In Russ.).

Timofeeva Anastasia — Ph.D., Associate Professor, Department of Theoretical and Applied Informatics, Novosibirsk State Technical University (NSTU). Research interests: the development of feature engineering methods for building interpretable machine learning models. The number of publications — 100. a.timofeeva@corp.nstu.ru; 20, Karl Marx Ave., 630073, Novosibirsk, Russia; office phone: +7(383)346-0600.

Avdeenko Tatyana — Ph.D., Dr.Sci., Professor, Department of Theoretical and Applied Informatics, Novosibirsk State Technical University (NSTU). Research interests: modeling and identification, mathematical and information methods for decision-making support, knowledge engineering. The number of publications — 250. avdeenko2101@gmail.com; 20, Karl Marx Ave., 630073, Novosibirsk, Russia; office phone: +7(383)346-0600.

Ю.И. СЕНКЕВИЧ, М.А. МИЩЕНКО
**МЕТОД ВЫДЕЛЕНИЯ ПРИЗНАКОВ СВЯЗИ СИГНАЛА
ГЕОАКУСТИЧЕСКОЙ ЭМИССИИ С СИЛЬНЫМИ
ЗЕМЛЕТРЯСЕНИЯМИ НА КАМЧАТКЕ**

Сенкевич Ю.И., Мищенко М.А. Метод выделения признаков связи сигнала геоакустической эмиссии с сильными землетрясениями на Камчатке.

Аннотация. Известна опасность катастрофических последствий сильных землетрясений. Сейсмические процессы, связанные с подготовкой землетрясений в зоне субдукции, вызывают деформацию поверхностных слоев земной коры, включая осадочный слой, в котором наблюдается явление генерации акустических колебаний, известное как геоакустическая эмиссия (ГАЭ). Основной механизм генерации ГАЭ связан с движением дислокаций при взаимодействии фракций и их скоплений в осадочных породах. Деформации в осадочном слое приводят к изменению его физических свойств и, как следствие, к изменению условий генерации акустических колебаний. Установлена связь между степенью напряженно-деформированного состояния осадочных пород и интенсивностью импульсного потока ГАЭ. Сигнал ГАЭ представляет собой поток импульсов, следующих с различной скважностью, амплитудой и формой. В периоды повышенной сейсмической активности в осадочных породах наблюдаются изменения параметров сигнала. Детальное изучение изменений характеристик фрагментов сигнала, фиксируемых в периоды, непосредственно связанные с началом землетрясений на Камчатке, и поиск их отличий от изменений фрагментов сигнала в периоды, не связанные с землетрясениями, составили суть проводимого исследования. В Институте космических исследований и распространения радиоволн (ИКИР ДВО РАН) ведется изучение изменчивости характеристик сигналов ГАЭ в периоды, предшествующие землетрясениям, в интересах разработки методов предсказания повышенной вероятности их возникновения на Камчатке. Приведены факты связи отклонений характеристик сигнала с сильными землетрясениями на Камчатке. Отличительной особенностью исследования является представление динамики характеристик сигнала в виде изображений для разномасштабных периодов наблюдений. Результатом исследования явился разработанный метод выделения признаков связи сигнала геоакустической эмиссии с сильными землетрясениями на Камчатке. Критерием степени связи определено значение сходства изображений динамики характеристик сигнала, связанных с моментами начала землетрясений.

Ключевые слова: приповерхностные осадочные породы, поверхностные волны, геоакустическая эмиссия, аномалии характеристик сигнала, сравнение изображений.

1. Введение. Сильные землетрясения наносят огромный социально-экономический ущерб на всех материках нашей планеты. Поэтому одной из важнейших задач современной геофизики остается поиск методов предсказания сильных землетрясений [1]. Сложность решения задачи поиска связана с необходимостью получения знаний об особенностях изменения физико-механических свойств земной коры и мантии перед землетрясением в каждом конкретном регионе. Деформационные процессы, инициируемые тектоническими

движениями, поддерживают неустойчивое напряженно-деформированное состояние среды в приповерхностных слоях литосферы [2, 3]. Такое состояние среды способно вызывать мгновенную реакцию осадочных пород даже на слабое воздействие поверхностных волн, порождаемых сейсмическими процессами. Реакция проявляется как заметное изменение характеристик акустических шумов, возникающих в породах. Такие шумы известны как геоакустическая эмиссия (ГАЭ) [2 – 4]. С помощью пьезоэлектрических датчиков акустические волны преобразуются в сигнал, который представляет непрерывный поток коротких электрических импульсов различной формы, амплитуды и скважности. Статистические характеристики этих параметров сигнала меняются в зависимости от текущего состояния среды, в которой генерируется акустический сигнал. Контроль параметров получаемого сигнала ГАЭ позволяет осуществлять мониторинг изменений его характеристик и формировать на основании анализа результатов наблюдений маркеры предвестников землетрясений [5 – 7]. Известные представления о природе возникновения ГАЭ в осадочных породах описываются моделью упругих колебаний в среде как результата нелинейной перестройки элементов мелкозернистой структуры сыпучих грунтов, обусловленной деформационными процессами и явлением дилатансии [1 – 5, 8]. Возникающая в осадочных породах ГАЭ опосредованно отражает поведение воздействующих тектонических и сейсмических процессов. Это позволяет, изучая особенности изменений характеристик ГАЭ, опосредованно давать оценку протекания таких геофизических явлений [7, 9]. В определенных случаях наиболее яркие проявления изменений статистических характеристик сигнала ГАЭ удастся связать с конкретными сейсмическими событиями [3, 10, 11].

На практике при анализе изменчивости характеристик ГАЭ следует учитывать факт влияния на эти показатели климатических факторов – атмосферного давления, температуры грунта и его влажности [12].

В ходе многолетних наблюдений, проводимых в лаборатории акустических исследований Института космофизических явлений и распространения радиоволн (ИКИР) ДВО РАН на Камчатке, неоднократно удавалось наблюдать изменчивость характеристик сигнала геоакустической эмиссии в моменты, предшествующие возникновению сильных землетрясений в пределах временных интервалов от 15 суток до единиц часов [8, 11, 12]. Проведенные наблюдения положили начало поиску предвестников землетрясений

и проверке их статистической устойчивости на основе наблюдаемых изменений характеристик сигнала ГАЭ [8 – 14] на различных временных интервалах.

Основной причиной сильных землетрясений на Камчатке является субдукция – тектоническое взаимодействие литосферных плит, при котором происходит погружение океанической Тихоокеанской литосферной плиты под континентальную Охотоморскую плиту Евразии [15]. Нелинейный характер такого взаимодействия составляет серьезную проблему поиска устойчивых признаков подготовки землетрясений [16, 17]. Это явление отражается в изменениях характеристик сигнала ГАЭ и, как следствие, в результатах измерений принимаемого акустического сигнала [18, 19].

Рассматриваемые здесь особенности изменчивости характеристик сигнала ГАЭ далее будут называться аномалиями. Особенности изменчивости характеристик ГАЭ на разных временных масштабах в периоды подготовки землетрясений – это отражение воздействия обширного комплекса нелинейных физических явлений, обусловленных критическими геодеформациями, возникающими по мере приближения напряжений в горных породах к предельным значениям. В результате нелинейного характера взаимодействия явлений аномалии ГАЭ достаточно трудно выделяются из-за большого разнообразия форм их проявления.

Возникает статистическая неустойчивость оценок параметров сигнала ГАЭ на непродолжительных временных интервалах, с одной стороны, и маскирование возможных динамических закономерностей статистических распределений параметров на продолжительных временных интервалах в силу наложений нескольких воздействующих факторов, с другой стороны. В частности, помимо названных причин сложного обнаружения аномалий, присутствует искажение сигнала наложением проявлений климатических и погодных факторов на процесс генерации ГАЭ.

Для детального изучения аномалий ГАЭ в лаборатории акустических исследований ИКИР ДВО РАН был разработан подход к представлению серии последовательных оценок выбранной характеристики сигнала на определенном временном интервале в виде изображения, подготовленного к дальнейшей обработке и анализу с использованием нейронных сетей [20]. В основу подхода заложена гипотеза о возможности сохранения детальной информации о влиянии поверхностных волн на процессы деформации приповерхностных пород. Это влияние отражается в непродолжительных эпизодах изменчивости динамики переходов состояний напряжений осадочного

слоя и, как следствие, на параметрах импульсного потока ГАЭ. Гипотеза базируется на результатах анализа многолетних наблюдений, в ходе которых проводился поиск скрытых закономерностей изменчивости характеристик ГАЭ на различных временных масштабах продолжительностью от нескольких часов до нескольких суток [19, 21]. Рассматриваются не статистические характеристики сигнала или его спектральные преобразования, а конструкция в виде совокупности собранных в матрицу результатов, отображающих изменчивость серии последовательных во времени распределений выбранной характеристики сигнала [22].

Определим целесообразность применения разработанного подхода с позиции отражения влияния изменений напряженных состояний осадочного слоя литосферы под действием деформационных процессов в статистической модели сигнала ГАЭ.

2. Моделирование сигнала геоакустической эмиссии.

На рисунке 1 приведены пятнадцатиминутные фрагменты записи сигнала ГАЭ, каждый из которых представляет собой последовательность акустических импульсов.

По каждому импульсу проводится измерение его параметров (амплитуда, длительность, межимпульсный интервал). На каждом фрагменте можно выделить более тысячи отдельных импульсов (например, 2262 и 3454 импульсов на рисунке 1(а) и 1(б) соответственно).

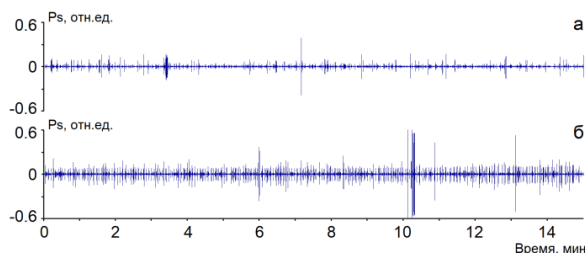


Рис. 1. Фрагменты записи сигнала акустической эмиссии, зарегистрированного в 2025 г.: а) 10 января; б) 25 сентября

Это позволяет получать параметрические оценки с достаточной статистической достоверностью, содержащей необходимую информацию о параметре распределения на каждом таком фрагменте сигнала. Например, за сутки можно получить 96 подобных оценок и проследить изменчивость выбранной характеристики на суточном периоде наблюдения. Для анализа изменчивости характеристик ГАЭ

на суточном периоде наблюдения в ходе проводимого исследования применялся метод сравнения статистических оценок распределений выбранного параметра.

Сигнал ГАЭ на выбранном периоде наблюдения рассматривается как последовательность независимых фрагментов потока акустических импульсов с изменяющимися во времени статистическими характеристиками. Вариации характеристик сигнала обусловлены влиянием изменений свойств среды, в которой генерируется сигнал ГАЭ, регистрируемый датчиками приемной аппаратуры. Под действием тектонических процессов и распространяющихся поверхностных волн изменяется текущее напряжение в приповерхностных слоях и происходит их деформирование [5, 6].

Необходимо сделать важное замечание по поводу искажения характеристик регистрируемых гидрофоном упругих акустических волн геоакустической эмиссии, которые, проходя через преобразующие и фильтрующие нелинейные цепи на выходе регистрирующей аппаратуры, трансформируются в радиоимпульс [23]. В процессе такой трансформации менее всего подвержены искажениям такие характеристики сигнала ГАЭ, как амплитуда, длительность импульса и межимпульсный интервал. Форма же импульса в сигнале существенно зависит от методов фильтрации преобразованного сигнала и частотных характеристик канала усиления приемной аппаратуры [24, 25]. Модуляция радиоимпульсов сигнала ГАЭ является продуктом резонансных свойств акустического приёмника и передаточных характеристик входных цепей регистрирующей аппаратуры [26]. Поэтому для поиска закономерностей изменчивости характеристик сигнала ГАЭ под действием деформации и напряженного состояния грунта предпочтительно проводить анализ характеристик амплитуды (максимальных значений), межимпульсных интервалов и длительности импульсов, а формфактор (огнивающий импульса) и его частотную модуляцию не учитывать для расчетов.

Для лучшего понимания влияния сейсмических воздействий на характеристики ГАЭ опишем условия приема сигнала ГАЭ в пункте комплексных геофизических наблюдений «Карымшина» на основании обзора и обобщения результатов ранее проведенных исследований [16 – 22]. Наблюдаемые на практике сигналы ГАЭ представляют поток релаксационных импульсов. Считая среду распространения упругих акустических волн ГАЭ в радиусе приема однородной и изотропной, в отсутствие внешних воздействий плотность распределения источников импульсов в среде можно

считать равномерной, а активацию источников равновероятной. Тогда сигнал будет представлять собой сумму импульсов, принятых от всех рассеянных по объему активированных точечных источников n за время наблюдения сигнала. В ранее проведенном исследовании [13] было показано, как получить число активированных точечных источников n в объеме полусферы с радиусом уверенного приема сигнала R за время наблюдения сигнала Δt . Схема приема импульсов ГАЭ с заглубленного приемника сигнала показана на рисунке 2.

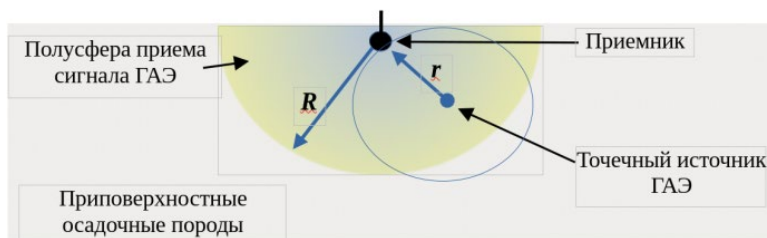


Рис. 2. Схема приема импульсов геоакустической эмиссии в пункте комплексных геофизических наблюдений «Карымшина»

Максимальное число таких источников $n_{\max}$ в объеме полусферы приема, определяющих импульсный поток (1) с учетом равномерной плотности распределения потенциальных источников импульсов ρ , равно (1):

$$n_{\max} = \frac{V}{2} \cdot \rho = \frac{2}{3} \pi R^3 \cdot \rho, \quad (1)$$

где: V – обобщенный коэффициент, зависящий от радиуса приема рассеянных по объему точечных источников акустической эмиссии.

В отсутствие воздействий на среду существует спонтанная или фоновая акустическая эмиссия. Необходимо учитывать снижение энергии импульсов, прямо связанное с затуханием амплитуды с расстоянием от источника до приемника r , линейной величины источников излучения в пористых, зерновых средах $\sim 10^{-4} - 10^{-2}$ м (рассматривается как точечный источник), и частотного диапазона принимающего пьезоэлектрического датчика – 0,1-10 кГц, скорости звука в осадочных породах с $\sim 1500 - 4500$ м/с. В пределах этих значений выполняется условие приема сигнала в зоне Френеля (ближняя зона приема сигнала). Влияние отражения от поверхности сред почва-воздух не учитывается. Тогда параметрическую модель сигнала акустической эмиссии можно представить в виде (2):

$$s(t) = \sum_{i=1}^{n_{max}} p_i \cdot g(A_i, \tau_i, T_i), \quad (2)$$

где: g – функция, описывающая импульс; τ_i – длительность i -го импульса, p_i – средняя вероятность активации i -го источника импульса; A_i – амплитуда i -го источника импульсов; межимпульсный интервал $T_i = t - t_i$, где t_i – время генерации i -го импульса; n_{max} – максимальное число в потенциальных источников импульсов в радиусе чувствительности приемной аппаратуры (1).

Чтобы представить модель сигнала ГАЗ, объясняющую влияние воздействия деформационных процессов на изменение характеристик сигнала, необходимо понять, как основные параметры модели (2): амплитуда A , длительность сигнала τ и межимпульсный интервал T , реагируют на физическое напряжение в поверхностном осадочном слое. Для этого рассмотрим некоторые наиболее известные зависимости формирования сигнала в точке излучения точечного источника акустических колебаний под воздействием напряжений и деформации [27 – 29]. Для расчета числа активированных источников в сфере приема сигнала под действием приложенного напряжения можно ориентироваться на эмпирическую формулу (3):

$$n = \alpha (\sigma - \sigma_0) \cdot m, \quad (3)$$

где: n – кумулятивное (накопленное) число импульсов акустической эмиссии; σ – приложенное напряжение (деформационное сжатие); σ_0 – пороговое напряжение (начало интенсивной генерации акустической эмиссии, соответствует значению предварительного уплотнения); α, m – эмпирические параметры материала, зависящие от плотности, гранулометрии, минералогии, влажности. Параметр m находится в диапазоне 1.5-3.0.

При возрастании напряжений σ интенсивность потока импульсов увеличивается. Можно ввести зависимость периода следования импульсов акустической эмиссии от напряжённого состояния среды [30, 31] (4):

$$\lambda = \lambda_0 \exp(-\alpha \cdot \sigma), \quad (4)$$

где: λ_0 – начальная скорость возникновения импульсов (при отсутствии дополнительной нагрузки), α – чувствительность частоты импульсов к уровню напряжённости.

По данным испытаний сред под нагрузкой зависимость изменений межимпульсных интервалов описывается степенным законом [32, 33], близким к пуассоновскому (5):

$$N(> T) \propto \lambda T^{-\sigma}, \quad (5)$$

где: $N(*)$ – упорядоченное множество распределения межимпульсных интервалов; T – межимпульсный интервал; λ – интенсивность события (частота возникновения импульсов), зависящая от напряжения σ .

На рисунке 3. представлена демонстрация распределения межимпульсных интервалов, полученного для двух фрагментов сигнала ГАЭ, полученных в различные периоды сейсмической активности, закон распределения которых близок к степенному закону.

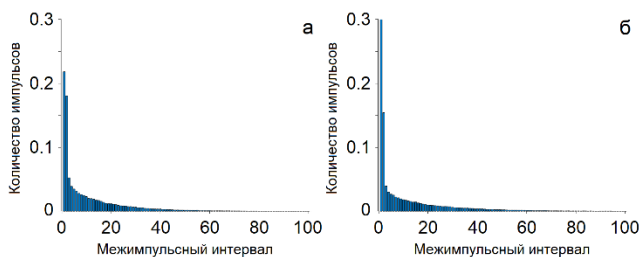


Рис. 3. Демонстрация степенного закона распределения межимпульсных интервалов для 96 15-минутных файлов ГАЭ, полученных в пункте комплексных геофизических наблюдений «Карымшина» в период:

- а) фоновой активности 02.11.2016 г.;
- б) предсейсмической активности 03.07.2018 г.

Функция плотности распределения вероятности межимпульсного интервала может быть представлена в следующем виде (6):

$$P(T \vee \sigma) = \lambda(\sigma) \exp(-\lambda(\sigma) \cdot T), \quad (6)$$

Рассмотрим следующую характеристику сигнала ГАЭ – амплитуду сигнала A и ее зависимость от значений напряжения σ в гравийно-песчаных грунтах. Для описания зависимости часто используют аппроксимацию обратного степенного закона (Закон Омури) [34, 35], описывающего изменение частоты

и амплитуды импульсов акустической эмиссии во времени после резкого изменения напряжения среды под действием деформации (7).

$$N(>A) \propto A^{-\alpha}, \quad (7)$$

где: $N(>A)$ – упорядоченное множество, распределение амплитуд (или энергией) не ниже A ; α – эмпирический коэффициент.

Формальное выражение распределения амплитуд импульсов акустической эмиссии в зависимости от деформации и напряжённого состояния в сыпучих грунтах часто описывается с помощью степенного закона [36, 37]. Для хрупких материалов (бетон, керамика, горные породы) справедливо следующая зависимость функции плотности вероятности распределения амплитуд (8).

$$P(A \vee \sigma) = \frac{b(\sigma)}{A_0} \left(\frac{A}{A_0} \right)^{-b(\sigma)}, \quad (8)$$

где: A – амплитуда сигнала акустической эмиссии; A_0 – минимальная амплитуда, определяемая чувствительностью приемной аппаратуры; параметр b (b -value) характеризует наклон амплитудного распределения и связан с уровнем девиаторного напряжения $q = \sigma_1 - \sigma_3$, где σ_1 – осевое напряжение, σ_3 – радиальное (боковое) напряжение).

На рисунке 4. представлена демонстрация степенного закона распределения амплитуд импульсов ГАЭ для двух фрагментов сигнала.

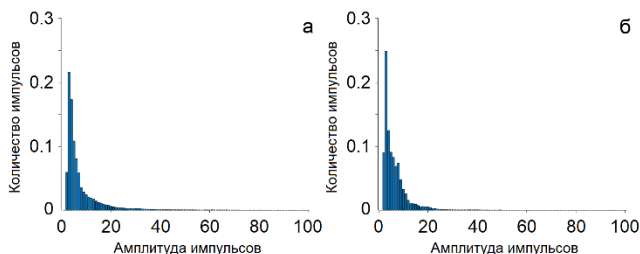


Рис. 4. Демонстрация степенного закона распределения амплитуд импульсов для 96 15-минутных файлов ГАЭ, полученных в пункте комплексных геофизических наблюдений «Карымшина» в период:

- а) фоновой активности 06.10.2016 г.;
- б) сейсмической активности 27.12.2016 г.

В научной литературе степенной закон распределения часто упоминается, как аппроксимация степенного закона распределения частот землетрясений в зависимости от их магнитуды – закон Гутенберга-Рихтера в геофизике [38, 39], который отражает скейлинговые свойства процесса генерации сигнала акустической эмиссии. Исследования [40, 41, 42] показывают значительное сходство статистических характеристик между сжатием пористого материала и напряженностью в различных средах под воздействием деформации, вызванной сейсмическими процессами, что подтверждает их масштабную инвариантность.

Проводимые в акустической лаборатории ИКИР ДВО РАН исследования подтвердили, что с ростом напряжения и интенсивности деформации амплитуда акустической эмиссии в поверхностном осадочном слое увеличивается. Это связано с активизацией микропроцессов в материале – ростом трещин, перестройкой структурных связей, фрикционным взаимодействием частиц [43, 44]. Перенос степенного закона на распределение амплитуд импульсов акустической эмиссии позволяет выразить вероятность акустической эмиссии в зависимости от напряжения среды с учетом изменения формы распределения в зависимости от условий воздействия.

Поэтому функцию распределения вероятности амплитуд часто представляют распределением Вейбулла [45] (9):

$$P(A \vee \sigma) = 1 - \exp\left(-\left(\frac{A(\sigma)-A_0}{\beta}\right)^g\right), \quad (9)$$

где: A – амплитуда сигнала с учетом напряжения среды σ , A_0 – пороговое значение (нижняя граница принимаемых аппаратурой значений амплитуд), β – масштабирующий параметр (определяет диапазон амплитуд), g – параметр формы (характеризует симметричность и склонность распределения к большим или малым амплитудам; определяет форму распределения: $\beta < 1$ – убывающая экспоненциальная форма, $\beta \sim 1$ – асимметричная, $\beta > 1$ – приближающаяся к нормальному).

Далее рассмотрим известные зависимости распределения длительности импульсов акустической эмиссии от напряжения среды и воздействия разрушающей деформации.

Физика изменения длительности импульсов акустической эмиссии под действием напряжений и деформаций объясняется увеличением масштаба порождающего объекта (длина трещин, размер дислокационных ансамблей), а также возрастанием числа взаимодействующих источников [46]. Формально такая зависимость

выражается через изменение длины размера порождающего объекта следующим образом:

$$\tau \sim \frac{L}{c}, \quad (10)$$

где: L – размер порождающего объекта; c – скорость звука в материале.

При нагружении L растёт с напряжением σ по закону $L \sim \sigma^d$, где d зависит от механизма взаимодействия составляющих элементов среды, перераспределения контактных усилий между зёрнами в гранулированной среде. С учетом сказанного приведем формулу к виду (11):

$$\tau \sim \frac{\sigma^d}{c}. \quad (11)$$

Как и для амплитуд, распределение длительностей τ в популяции событий генерации импульсов акустической эмиссии подчиняется степенному закону [46, 37], что отражает масштабную инвариантность процесса возникновения импульсов акустической эмиссии при разрушениях сред. Кумулятивная зависимость распределения длительностей импульсов акустической эмиссии имеет вид упорядоченного множества (12):

$$N(>\tau) \propto C \cdot \tau^{-d}, \quad (12)$$

где: $N(>\tau)$ – количество последовательных событий с длительностью, превышающей τ ; C – нормировочная константа (продуктивность источника); d (d -value) – показатель степени, аналог b -значения для амплитуд. Параметры C и d являются функциями механического состояния: напряжения σ (или девиаторного напряжения $q = \sigma_1 - \sigma_3$, где σ_1 – осевое напряжение, σ_3 – радиальное напряжение).

На рисунке 5 представлена демонстрация степенного закона распределения длительностей импульсов ГАЭ

Функция распределения плотности вероятности длительности импульса τ акустической эмиссии описывается следующим уравнением [34, 25] (13):

$$P(\tau \vee \sigma) = \alpha \cdot \tau \cdot \exp\left(-\beta \frac{\tau}{\sigma}\right), \quad (13)$$

где β – масштабирующий параметр (определяет диапазон значений длительностей импульсов).

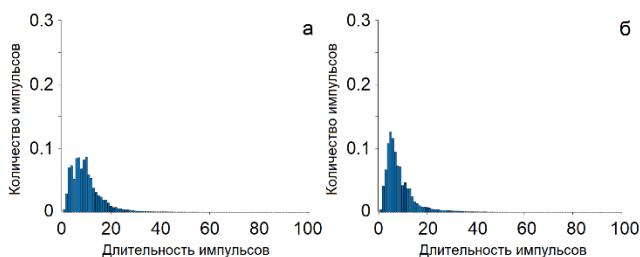


Рис. 5. Демонстрация степенного закона распределения длительностей импульсов для 96 15-минутных файлов ГАЭ, полученных в пункте комплексных геофизических наблюдений «Карымшина» в период:

- а) фоновой активности 13.10.2016 г.;
б) предсейсмической активности 12.09.2018 г.

Приведенные связи (5,7,12) дают возможность переписать модель (2) в виде совокупности зависимых параметров сигнала ГАЭ от величины напряжения поверхностных осадочных пород в полусфере приема (14):

$$s(t) = \sum_{i=1}^{n_{max}} p_i \cdot g(A_i(\sigma), \tau_i(\sigma), T_i(\sigma)), \quad (14)$$

где: p_i – средняя вероятность активации i -го источника импульса; g – функция, описывающая импульс; N – число активированных импульсов, связанное с напряженностью среды.

Модель (14) построена на данных многих практических исследований, указывающих на совокупное влияние напряжения в гранулированных средах на амплитуду, длительность и межимпульсный интервал, определяющие импульсный поток сигнала ГАЭ. Этот результат представления сигнала акустической эмиссии позволяет продемонстрировать физическую связь зависимости параметров сигнала с напряжением среды. Обнаруженное сходство процессов генерации сигнала геоакустической эмиссии в различных материалах при сжатии в лабораторных условиях и генерации сигнала, получаемого в ходе активных сейсмических проявлений [40], согласуются с результатами наблюдений, полученными в нашей лаборатории [47, 48]. Поэтому на базе модели (14) можно строить предположение о текущем состоянии среды, то есть решать обратную задачу.

Однако её решение на основе этой модели ограничивается тем, что оно будет описывать текущее состояние среды в конкретный момент времени и не дает возможности получить информацию о возможных закономерных изменениях, которым подвергалась среда на протяжении длительных наблюдений в предсейсмический период активности. Это существенно ограничивает информативные свойства получаемых данных мониторинга состояния среды и, как следствие, снижает возможности выделения аномалий сигнала, связанных с сильными землетрясениями.

2.1. Обоснование подхода к анализу сигнала ГАЭ. Измерения в Лаборатории акустических исследований ИКИР ДВО РАН, связанные с мониторингом сигналов акустической эмиссии в приповерхностном слое осадочных пород в пункте комплексного геофизического наблюдения «Карымшина» выполняются, начиная 1998 года по настоящее время. Мониторинг показал, что в периоды предсейсмической активности наблюдается заметная активация источников сигнала ГАЭ, которая проявляется на разных временных масштабах от нескольких недель до нескольких часов [49].

Ограничения классического статистического анализа распределений параметров о репрезентативности выборки и робастности предъявляют к выполнению обработки данных необходимое условие большого объема выборки данных, или применительно к сигналу – достаточно длительного времени наблюдения. В наших исследованиях это приводило к маскированию изменений состояния среды, которые происходили за короткие временные интервалы внешних воздействий на нее. Это также приводило к ограничению возможности детального анализа коротких воздействий, которые переносят полезную информацию. В то же время стремление проводить анализ на коротких интервалах времени приводило к потере корректности обобщенной информации. Кроме того, решение задачи анализа на многочисленных коротких интервалах времени приводило к значительным временным потерям и затратам приборных ресурсов.

Чтобы избежать потери информации, предлагается оригинальный подход к подготовке данных, который сможет обойти указанную проблему анализа. Данный подход опирается на представленную модель сигнала ГАЭ (14), которая позволяет рассматривать формирование сигнала, как совокупность отдельных независимых процессов влияния напряжения среды на его параметры. Причем дает возможность наблюдать динамику изменений характеристики каждого из параметров во времени посредством

сопоставления оценок параметра, полученных в последовательные короткие временные интервалы. Основой такого подхода служит разбиение длительного интервала наблюдения сигнала ГАЭ на последовательности равных по длительности коротких интервалов, содержащие достаточное количество импульсов для получения значимой оценки, и анализа полученных оценок для каждого из параметров независимо.

Пусть известно распределение каждого из параметров сигнала на длительном интервале θ : для межимпульсных интервалов $N(> T)$, для амплитуды $N(> A)$, для длительности импульса $N(> \tau)$. Представим это разбиение отдельно для каждого из параметров виде суммы выделенных распределений на равных достаточно коротких интервалах длительностью θ/m (15), зависящих от напряжения среды σ , которое на каждом отдельном таком интервале можно считать неизменным.

$$\begin{aligned} N(T \vee \sigma) &= \sum_{i=1}^m N_i(T \vee \sigma_i), \\ N(A \vee \sigma) &= \sum_{i=1}^m N_i(A \vee \sigma_i), \\ N(\tau \vee \sigma) &= \sum_{i=1}^m N_i(\tau \vee \sigma_i). \end{aligned} \quad (15)$$

Здесь m – число разбиений на короткие равные временные интервалы. Каждое отдельное распределение под знаком суммы в (15) представляет собой упорядоченное множество частоты попадания рассматриваемого параметра в интервалы от 1 до n , охватывающие установленный диапазон значений параметра, и может рассматриваться как вектор значений параметра. Последовательность таких векторов можно представить в виде двумерной матрицы вида (16):

$$D_\theta = \begin{bmatrix} a_{1,n} & a_{2,n} & \dots & a_{m-1,n} & a_{m,n} \\ a_{1,n-1} & a_{2,n-1} & \dots & a_{m-1,n-1} & a_{m,n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{1,2} & a_{2,2} & \dots & a_{m-1,2} & a_{m,2} \\ a_{1,1} & a_{2,1} & \dots & a_{m-1,1} & a_{m,1} \end{bmatrix}, \quad (16)$$

где: $a_{i,j}$ – значение выбранного параметра сигнала; $i = [1, m]$ – порядковый номер вектора распределения; $j = [1, n]$ – порядковый номер интервала распределения.

Матрица D_θ содержит последовательность векторов распределений выбранного параметра ГАЭ на интервале наблюдения θ , тем самым сохраняя информацию о коротких временных фрагментах сигнала, отражающих информацию о деталях изменений напряжений среды при прохождении через нее поверхностных волн, с одной

стороны. При этом в совокупности своих элементов матрица (16) при достаточно больших интервалах наблюдений D_y может содержать информацию о медленных деформационных процессах, порождаемых длительными тектоническими воздействиями, с другой стороны.

В целом матрица $D\theta$ является характеристикой, отражающей динамику изменений напряжений в приповерхностном слое в радиусе приема сигнала ГАЭ за время установленного наблюдения θ , с разрешением по времени – θ/m , и разрешением по значению выбранного параметра – $3 \cdot \sigma/n$, где σ – дисперсия выбираемого параметра. Матрицу (16) в дальнейшем удобно рассматривать как изображение, передающее информативный паттерн динамики распределений выбранного параметра сигнала ГАЭ.

3. Препроцессинг сигналов геоакустической эмиссии. Важный вопрос о выборе временного интервала для обнаружения аномалий характеристик сигнала ГАЭ, основывается на физике происхождения форшоков и афтершоков [50]. Современные исследования, использующие результаты анализа сигналов ГАЭ подтверждают, что с приближением к моменту начала землетрясений активность проявления форшоков, сопровождаемых генерацией АЭ экспоненциально возрастает (Закон Гутенберга-Рихтера), а после катастрофического явления проявление афтершоковой активности экспоненциально снижается (Закон Омори), образуя специфический график общей сейсмической активности [51]. Отсюда следует, что поиск специфических аномалий на краткосрочном периоде наблюдений ГАЭ (от нескольких часов до примерно двадцати суток) с большей вероятностью следует искать вблизи участков форшоковой и афтершоковой активности, связанной с сильными землетрясениями.

Все ниже описываемые эксперименты выполнялись ретроспективно на базе данных сигналов ГАЭ лаборатории акустических исследований, полученных в пункте комплексных геофизических наблюдений «Карымшина» ИКИР ДВО РАН [52]. По результатам бурения приповерхностные породы в данном пункте наблюдения являются осадочными с толщиной слоя более 20 м [56]. Это подтверждается тем, что в 170 м от гидрофона находится буровая скважина № 99-8 Камчатского филиала ФИЦ ЕГС РАН глубиной 19 м. Ее геологическая колонка представлена слоистыми отложениями с песчано-глинистым заполнением (0-5 м), глыбово-щебнистыми отложениями с глинистым заполнителем (5-14 м) и валунно-галечными отложениями с песчаным заполнителем (14-19 м). Таким образом, основная часть источников принимаемых сигналов ГАЭ описывается свойствами фракций геологической колонки,

представляющих приповерхностные осадочные породы сложной слоистой структуры без присутствия твердых монолитных гранитных пород. Это исключает формирование ГАЭ в результате растрескивания пород вследствие деформации, а основным источником ГАЭ в этих слоях будут дилатансия и контракция [53, 54] под действием деформации, вызванной сейсмическими волнами и продолжительным тектоническим напряжением.

Начиная с 2015 года до настоящего времени в пунктах комплексных геофизических наблюдений «Карымшина» и «Мижижа» ведутся непрерывные удаленные наблюдения аномалий сигналов ГАЭ и связанных с ними деформаций, на основе которых формируется специализированная база данных [55, 56]. Из базы данных выбирались участки записей сигналов ГАЭ, которые охватывали разномасштабные по времени фрагменты длительностью от нескольких часов до двадцати суток и, в зависимости от поставленной задачи, либо примыкали непосредственно к моментам начала выбранных из каталога землетрясений и взятые непосредственно после них, либо составляли последовательности, полученные скользящим временным окном в районе от 1 до 20 суток. Для экспериментов в рамках проводимого исследования были отобраны землетрясения, произошедшие в период с сентября 2016 г. по декабрь 2019 г., с энергетическим классом K_S не менее 10.0. В региональном сейсмическом каталоге [57] зарегистрировано 981 такое событие. Из них были отобраны землетрясения, для которых выполнялось следующее условие (17):

$$d/R_D < 1, R_D = 10^{0,43(K_S-4,6)/1,5}, \quad (17)$$

где R_D – радиус зоны Добровольского, d – расстояние от пункта наблюдений до эпицентра землетрясения, K_S – энергетический класс. Условие (17) ограничивает выбор сильных камчатских землетрясений, эпицентры которых находятся в радиусе чувствительности датчиков приемной аппаратуры.

Известно, что принципиально важное значение для качественного анализа приобретает препроцессинг данных, который с появлением нейросетевой обработки данных превратился в огромную индустрию разработки алгоритмов искусственного интеллекта. До последнего десятилетия трудно было представить классификацию более тысячи фрагментов сигналов в практике проведения непрерывных наблюдений в реальном времени. Сравнительный анализ графического представления нечетких картин распределений параметров, какими являются характеристики ГАЭ, визуально практически неразличим

при сравнении даже небольшого числа коротких фрагментов сигнала. Это значительно снижает качество поиска различий при инструментальном сравнении результатов обработки сигналов и их идентификации. Для выхода из создавшейся ситуации была предпринята попытка анализа трехмерных графических изображений, построенных на матрицах вида (18) (Рис. 6).

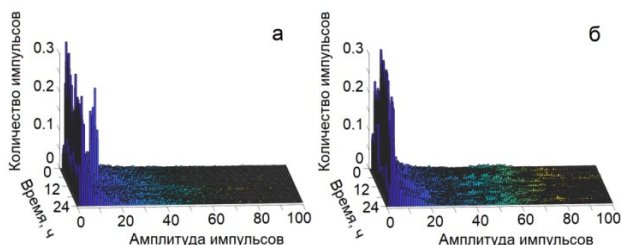


Рис. 6. Пример формирования массива данных динамики распределения амплитуд импульсов сигнала ГАЭ за сутки в 2025 г.: а) 10 января; б) 25 сентября

На рисунке 6 приведены последовательно записанные пятнадцатиминутные фрагменты сигнала ГАЭ в виде поочередно выстроенных нормированных гистограмм распределения амплитуд импульсов. При таком представлении картина изменчивости характеристики сигнала на суточном интервале обработки сигнала ГАЭ более наглядна. Такое графическое представление наилучшим образом отражает динамику амплитудного распределения на интервале обработки. Однако как визуальный, так и инструментальный анализ еще более усложнился. Решение возникшей проблемы с анализом трехмерных изображений было найдено в ходе изучения высококачественной обработки и классификации изображений современными программно-ориентированными архитектурами нейронных сетей [58, 59, 60]. Такие нейронные сети широко используются в различных научных областях для идентификации, кластеризации и классификации изображений.

Поэтому для дальнейших исследований использовались массивы данных динамики распределений параметров импульсов сигнала ГАЭ в виде нормированных и приведенных к цветовой шкале матричных представлений (18).

На рисунке 7(а, б) представлена пара таких изображений, полученных на основе данных, представленных на рисунке 6.

Более детально формирование матриц для изображений и их использование для идентификации аномалий ГЭЭ перед сильными землетрясениями представлено в описании патента [14, 20].

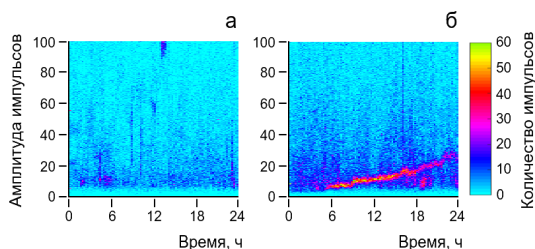


Рис. 7. Графическое представление массивов данных динамики распределения амплитуды импульсов сигнала ГЭЭ (Рис. 3(а, б)), подготовленное для обработки нейронной сетью

4. Обнаружение связи сигнала геоакустической эмиссии с сильными землетрясениями. Задача выявления связи сигнала геоакустической эмиссии с сильными землетрясениями может быть сведена к задаче определения степени сходства паттернов изображений, полученных из фрагментов сигнала ГЭЭ, нарезанных на этапе краткосрочного периода до начала землетрясения $\{D_{(q)}\}$ (период от нескольких часов до двух недель) – отнесенных к связанным с землетрясением изображениям, когда, как было показано в начале статьи, сейсмическая деятельность наиболее ярко отражается на параметрических распределениях сигнала.

Изображения, полученные из фрагментов сигнала ГЭЭ на достаточно значительном временном удалении от момента землетрясения $\{D_{(f)}\}$ принимаются как фоновые изображения. По результатам нарезки фрагментов сигнала ГЭЭ получаем две группы изображений C_q и C_f , соответственно с которыми выполняется следующая последовательность действий. Выполняется объединение множеств $\{D\} = \{D_{(q)}\} \cup \{D_{(f)}\} = \{D_1, D_2, \dots, D_N\}$, где каждое изображение D_i – двумерная матрица размером $h \times w$ ($h, w \in \mathbb{N}$) с элементами из некоторого множества значений $\mathbb{Z}$ дискретных цветовых каналов. Вытягиванием матрицы по столбцам (отдельным распределениям выбранного параметра) для каждого элемента D_i получаем векторы $\vec{D}_i \in \mathbb{Z}^{hw}$. Среди векторов проводится кластеризация с использованием евклидова расстояния в качестве метрики сравнения векторов (18):

$$d(D_i, D_j) = \|\vec{D}_i - \vec{D}_j\| = \sqrt{\sum_{r=1}^h \sum_{c=1}^w (D_i(r, c) - D_j(r, c))^2}. \quad (18)$$

Для заданного числа кластеров $K=2$ находится разбиение, минимизирующее сумму квадратов расстояний от элементов до центров кластеров (19):

$$\min_{C, \{\mu_k\}} \sum_{k=1}^K \sum_{D \in C_k} d^2(D, \mu_k), \quad (19)$$

где μ_k – центр кластера C_k .

Результат – разбиение множества $\{D\}$ на два кластера (группы) C'_q и C'_f , с новым набором изображений в группах $\{D'_{(q)}\}$ и $\{D'_{(f)}\}$, соответственно. Отношение числа совпадающих и несовпадающих изображений (элементов множеств) для каждой пары классов C'_q , C_q , C'_f , C_f позволяет рассчитать показатели качества кластеризации по отношению к изначально заданным группам (20):

$$\begin{aligned} Q_{true} &= \#(D_{(q)} \cap D'_{(q)}) / \# D_{(q)}, \\ Q_{false} &= \#(D_{(q)} \cap D'_{(f)}) / \# D_{(q)}, \\ F_{true} &= \#(D_{(f)} \cap D'_{(f)}) / \# D_{(f)}, \\ F_{false} &= \#(D_{(f)} \cap D'_{(q)}) / \# D_{(f)}. \end{aligned} \quad (20)$$

где показатели: Q_{true} – доля верно отнесенных изображений из кластера, отнесенного к связанным с землетрясениями изображениями $\{D'_{(q)}\}$ по отношению к условно отнесенной группе связанных с землетрясениями изображений $\{D_{(q)}\}$; Q_{false} – доля числа изображений из кластера фоновых $\{D'_{(f)}\}$, ошибочно отнесенных к группе условно связанной с землетрясениями группе изображений $\{D_{(q)}\}$; F_{true} – доля верно отнесенных изображений из кластера фоновых изображений $\{D'_{(f)}\}$ по отношению к числу исходных фоновых изображений $\{D_{(f)}\}$; F_{false} – доля ошибочно отнесенных к связанным с землетрясениями изображений из кластера $\{D'_{(q)}\}$ по отношению к числу исходных фоновых изображений $\{D_{(f)}\}$.

Перечисленные показатели в полной мере характеризуют информативность изображений по отношению к их связи с землетрясениями. Чем выше показатель Q_{true} , отражающий

специфичность связанных с землетрясением изображений, и чем ниже показатель Q_{false} , отражающий отличие от них паттернов фоновых изображений, тем более достоверно утверждение о глубине связи сейсмических событий с динамикой распределений параметров сигналов ГАЭ в период краткосрочных наблюдений перед землетрясениями.

5. Апробация метода. Для проверки разработанного метода обнаружения признаков связи сигнала ГАЭ с землетрясениями на этапе краткосрочного периода подготовки землетрясений был проведен эксперимент с использованием нейросетевых технологий подготовки и кластеризации изображений.

С помощью описанного выше препроцессинга сигналов ГАЭ было подготовлено 31 изображение, связанное с сильными землетрясениями на Камчатке в период с начала сентября 2016 года по конец декабря 2019 года, а также 96 изображений для фрагментов сигнала ГАЭ на удалении от моментов начала землетрясений не менее 20 суток.

Для обработки изображений применялась программа на языке Python с функциями весового выравнивания и кластеризации библиотеки scikit-learn. В таблице 1 представлены сводные результаты обработки и рассчитанные значения показателей искомой связи.

Таблица 1. Результаты обработки изображений

Классы	Число в группе	Число в кластере	Совпало	Показатель
Связанные с землетрясениями	31	36	13	13/31
				8/96
Фоновые	96	91	88	88/96
				8/31

Результаты вычисления показателя верно отнесенных объектов Q_{true} свидетельствуют о том, что в трети случаев наблюдается определенная связь фрагментов сигнала ГАЭ с землетрясениями, поскольку существует совпадение изображений из кластера, отнесенного к связанным с землетрясениями изображениям $\{D'_{(q)}\}$ с изображениями, полученными на суточных фрагментах сигнала ГАЭ накануне 20-суточного интервала перед сильными землетрясениями $\{D_{(q)}\}$. При этом показатель Q_{false} ошибочно отнесенных к группе, условно связанной с землетрясениями группе изображениям $\{D_{(q)}\}$ относительно мал.

Полученные показатели доль, верно отнесенных к фоновым изображениям – F_{true} и F_{false} – доли ошибочно отнесенных к связанным с землетрясениями изображений указывают на высокую степень сходства паттернов фоновых изображений.

6. Заключение. Обнаружение и классификация специфических отклонений характеристик геоакустической эмиссии, связанных с сильными землетрясениями на Камчатке – это многолетнее системное направление научных исследований лаборатории акустических исследований ИКИР ДВО РАН, которое в перспективе направлено на решение задач предсказания катастрофических событий, вызванных сейсмическими явлениями.

В ходе исследования был проведен анализ современных исследований по данным открытых источников, направленных на изучение связи напряжений приповерхностных слоев литосферы, возникающих под действием тектонических движений земной коры. В результате обобщения изложенных фактов, известных и подтвержденных закономерностей связи напряжений поверхностного слоя осадочных пород под действием сейсмических поверхностных волн, была предложена параметрическая модель связи таких явлений, выраженная через основные параметры геоакустической эмиссии. Модель позволяет объяснить опосредованное изменение амплитудных распределений, распределений межимпульсных интервалов и распределений длительностей импульсов ГАЭ от состояния напряженности среды в зоне приема акустического сигнала.

Был предложен и реализован подход представления фрагментов сигнала ГАЭ на разных временных интервалах в виде матриц, составленных из распределений каждого отдельного измеряемого параметра сигнала на заданных разномасштабных временных интервалах. Такая трансформация сигнала ГАЭ позволяет собрать информацию о динамике его последовательных распределений от одного выделенного короткого интервала к другому и представить полученную характеристику в виде матрицы, которая визуально наблюдается как трехмерное изображение.

Совокупность получаемых изображений, собранная в периоды, предшествующие многолетним камчатским землетрясениям, составляет базу данных изображений, ориентированных на использование эффективных алгоритмов искусственного интеллекта с применением нейронных сетей. Применение нейросетевых технологий для анализа данных позволило показать и наглядно оценить сравнимость паттернов

изображений, отражающих динамику характеристик сигнала ГЭЭ, путем применения кластеризации.

Разработанный метод позволил дать основание для утверждения о существовании признаков связи сигнала геоакустической эмиссии с сильными землетрясениями на Камчатке, а также существования признаков значительного сходства характеристик фоновых фрагментов сигнала. Дальнейшее накопление данных наблюдений ГЭЭ, вероятно, позволит составить более убедительную статистику для принятия решений о предупреждении опасности возникновения сильных землетрясений в краткосрочный период перед их началом.

Литература

1. Шебалин П.Н., Гвишиани А.Д., Дзедобоев Б.А., Скоркина А.А. Почему необходимы новые подходы к оценке сейсмической опасности? // Доклады Российской академии наук. Науки о Земле. 2022. Т. 507. №1. С. 91–97. DOI: 10.31857/S2686739722601466.
2. Спивак А.А., Рыбнов Ю.С. Акустические эффекты сильных землетрясений // Физика Земли. 2021. №1. С. 41–50. DOI: 10.31857/S0002333721010099.
3. Сычев В.Н., Богомолов Л.М. Потенциал методов нелинейной динамики для анализа геофизических рядов и сейсмичности // Сб. научн. ст. VI Международной конференции «Солнечно-земные связи и физика предвестников землетрясений». 2013. С. 78.
4. Богомолов Л.М., Сычева Н.А. Прогноз землетрясений в XXI веке: предыстория и концепции // Геосистемы переходных зон. 2022. Т. 6. №3. С. 145–182. DOI: 10.30730/gtr.2022.6.3.145-164.164-182.
5. Салтыков В.А., Кугаенко Ю.А. Развитие приповерхностных зон дилатансии как возможная причина аномалий в параметрах сейсмической эмиссии перед сильными землетрясениями // Тихоокеанская геология. 2012. Т. 31. №1. С. 96–106.
6. Чеботарева И.Я., Володин И.А., Дрягин В.В. Акустические эффекты при деформировании структурно неоднородных сред // Акустический журнал. 2017. Т. 63. №1. С. 84–93. DOI: 10.7868/S0320791916060046.
7. Пономарев А.В., Соболев Г.А. Физика землетрясений и предвестники // М.: Наука. 2003. 270 с.
8. Муратов П.В., Руленко О.П., Марапулец Ю.В., Солодчук А.А. Электрический и акустический отклик приповерхностных осадочных пород на прохождение сейсмических волн от землетрясений // Вестник КРАУНЦ. Физико-математические науки. 2018. №5(25). С. 62–73. DOI: 10.18454/2079-6641-2018-25-5-62-73.
9. Гаврилов В.А. Воздействие переменных электромагнитных полей на геоакустические процессы: эмпирические закономерности и физические механизмы: специальность 25.00.10 «Геофизика, геофизические методы поисков полезных ископаемых»: автореферат диссертации на соискание ученой степени доктора физико-математических наук / Федеральное государственное бюджетное учреждение науки Институт вулканологии и сейсмологии Дальневосточного отделения Российской академии наук (ИВиС ДВО РАН). Петропавловск-Камчатский, 2017. 51 с.
10. Segall P. Earthquake and Volcano Deformation // Princeton: Princeton University Press. 2010. 464 p.
11. Марапулец Ю.В., Шевцов Б.М., Ларионов И.А., Мищенко М.А., Щербина А.О., Солодчук А.А. Отклик геоакустической эмиссии на активизацию

- деформационных процессов при подготовке землетрясений // Тихоокеанская геология. 2012. Т. 31. №6. С. 59–67.
12. Богданов В.В., Бычков В.В., Водинчар Г.М. Динамика физических процессов в активных зонах ближнего космоса и геосфер: избранные работы 2018-2020 гг. / Под ред. Шевцов Б.М. // М.: РАН, 2021. 409 с.
 13. Сенкевич Ю.И., Мищенко М.А. Методика оценки состояния приповерхностных осадочных пород по результатам наблюдений динамических характеристик геоакустической эмиссии // Вестник КРАУНЦ. Физико-математические науки. 2023. Т. 45. №4. С. 109–121. DOI: 10.26117/2079-6641-2023-45-4-109-121.
 14. Сенкевич Ю.И., Марапулец Ю.В., Луковенкова О.О., Солодчук А.А. Методика выделения информативных признаков в сигналах геоакустической эмиссии // Труды СПИИРАН. 2019. Т. 18. №5. С. 1066–1092. DOI: 10.15622/sp.2019.18.5.1066-1092.
 15. Баранов Б.В., Цуканов Н.В., Гедике К., Фрейтаг Р., Дозорова К.А. Морфология и особенности строения зоны субдукции в районе сочленения Курило-Камчатской и Алеутской дуг // Океанология. 2022. Т. 62. №4. С. 528–539. DOI: 10.31857/S0030157422040025.
 16. Авдейко Г.П., Палуева А.А. Сегментация камчатской зоны субдукции и оценка сейсмической активности сегментов // Вестник КРАУНЦ. Науки о земле. 2010. №15(1). С. 69–89.
 17. Сафонов Д.А. Пространственное распределение тектонических напряжений в южной глубокой части Курило-Камчатской зоны субдукции // Геосистемы переходных зон. 2019. Т. 3. №2. С. 175–188. DOI: 10.30730/2541-8912.2019.3.2.175-188.
 18. Гапеев М.И., Марапулец Ю.В. Моделирование зон относительных сдвиговых деформаций перед сильными землетрясениями на Камчатке, произошедшими в период 2018-2021 гг. // Вестник КРАУНЦ. Физико-математические науки. 2021. Т. 37. №4. С. 53–66. DOI: 10.26117/2079-6641-2021-37-4-53-66.
 19. Senkevich Y., Mishchenko M., Solodchuk A. Observation of the State of the Near-surface Lithosphere Layer based on the Analysis of the Dynamics of the Geoaoustic Emission Signal Characteristics // Proceedings of the IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE). 2024. pp. 1140–1143. DOI: 10.1109/PIERE62470.2024.10805026.
 20. Сенкевич Ю.И. Обработка и подготовка данных наблюдений в интересах выделения особенностей динамики характеристик геоакустической эмиссии // Вестник КРАУНЦ. Физико-математические науки. 2024. Т. 47. №2. С. 75–94. DOI: 10.26117/2079-6641-2024-47-2-75-94.
 21. Сенкевич Ю.И., Мищенко М.А., Солодчук А.А., Луковенкова О.О., Гапеев М.И. Стохастическое моделирование сигнала геоакустической эмиссии // Цифровая обработка сигналов. 2023. №1. С. 27–34.
 22. Патент RU 2758582 C1. Способ обнаружения комплексного предвестника землетрясений: №2020138668: заявл. 26.11.2020: опубл. 29.10.2021 / Сенкевич Ю.И., Марапулец Ю.В., Луковенкова О.О., Солодчук А.А., Мищенко М.А., Малкин Е.И., Гапеев М.И.
 23. Буйло С.И. Физико-механические, статистические и аспекты акустико-эмиссионной диагностики: монография // Таганрог: Изд-во Южного федерального университета, 2017. 182 с.
 24. Грешников В.А. Акустическая эмиссия. Применение для испытаний материалов и изделий // М.: Изд-во стандартов, 1976. 272 с.
 25. Семашко Н.А., Шпорт В.И., Марьин Б.Н. Акустическая эмиссия в экспериментальном материаловедении // М.: Машиностроение, 2002. 239 с.
 26. Grosse C.U., Ohtsu M., Aggelis D.G., Shiotani T. Acoustic Emission Testing // Springer Tracts in Civil Engineering. Springer. 2022. 746 p. DOI: 10.1007/978-3-030-67936-1.

27. Березин Ю.А., Сподарева Л.А. Распространение акустических сигналов в грунтах // Прикладная механика и техническая физика. 2001. Т. 42. №4(248). С. 177–183.
28. Ou J., Wang E., Wang X., et al. Acoustic emission characteristics and damage evolution of different rocks under uniaxial compression conditions // Scientific Reports. 2024. vol. 14. pp. 1–10. DOI: 10.1038/s41598-024-54950-9.
29. Lockner D. The Role of Acoustic Emission in the Study of Rock Fracture // International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts. 1993. vol. 30. no. 7. pp. 883–899.
30. Xiangguo K., Enyuan W., Xueqiu H., et. al. Multifractal characteristics and acoustic emission of coal with joints under uniaxial loading // Fractals. 2017. vol. 25. no. 5. pp. 1–11. DOI: 10.1142/S0218348X17500451.
31. Carpinteri A., Lacidogna G., Puzzi Si. From criticality to final collapse: Evolution of the “b-value” from 1.5 to 1.0 // Chaos, Solitons & Fractals. 2009. vol. 41. pp. 843–853 DOI: 10.1016/j.chaos.2008.04.010.
32. Guarino A., Garcimartin A., Ciliberto S. An experimental test of the critical behaviour of fracture precursors // The European Physical Journal B – Condensed Matter and Complex Systems. 1998. vol. 6. pp. 13–24. DOI: 10.1007/s100510050521.
33. Морозова К.Г. Сейсмоакустическая эмиссия, сопровождающая различные режимы скольжения по разломам и трещинам: специальность 1.6.9 «Геофизика»: автореферат диссертации на соискание ученой степени доктора физико-математических наук / Институт динамики геосфер имени академика М.А. Садовского Российской академии наук (ИДГ РАН). Москва, 2023. 144 с.
34. Шереметьева О.В., Шевцов Б.М. Аппроксимация законов распределения времён ожидания форшоков на основе дробной модели деформационной активности // Вестник КРАУНЦ. Физико-математические науки. 2022. Т. 40. №3. С. 137–152. DOI: 10.26117/2079-6641-2022-40-3-137-152.
35. Милешкин М.Б. Библик И.В., Стоев П.И. Анализ амплитудного распределения сигналов акустической эмиссии на основе расчетно-экспериментального метода // Физическая мезомеханика. 2004. Т. 7. №S1-1. С. 206–209.
36. Lacoste M., Ruiz S., Or D. Listening to earthworms burrowing and roots growing – acoustic signatures of soil biological activity // Scientific Reports. 2018. vol. 8. pp. 1–9. DOI: 10.1038/s41598-018-28582-9.
37. Кочарян Г.Г., Кочарян А.А. Акустическая эмиссия при различных режимах межблоковых перемещений // Физико-технические проблемы разработки полезных ископаемых. 2015. №1. С. 3–13.
38. Баранов С.В. Оценка активности афтершокового процесса с помощью закона Гутенберга – Рихтера и etas-модели // Фундаментальные исследования. 2014. №12. С. 751–755.
39. Baro J., Corral A., Illa X., et al. Statistical Similarity between the Compression of a Porous Material and Earthquakes // Physical Review Letters. 2013. vol. 110. 088702 p. DOI: 10.1103/PhysRevLett.110.088702.
40. Sedghizadeh M., Van Den Berghe M., Shcherbakov R. Statistical and clustering analysis of microseismicity from a Saskatchewan potash mine // Frontiers in Applied Mathematics and Statistics. 2023. vol. 9. 1126952 p. DOI: 10.3389/fams.2023.1126952.
41. Турунтаев С.Б. Кинетика техногенной сейсмичности: специальность 25.00.10 «Геофизика, геофизические методы поисков полезных ископаемых»: автореферат диссертации на соискание ученой степени доктора физико-математических наук / Институт физики Земли им. О.Ю. Шмидта РАН (ИФЗ РАН). Москва, 2005. 330 с.
42. Мищенко М.А., Марапулец Ю.В., Ларионов И.А. Анализ временных и частотных характеристик одновременных геодеоформационных сигналов в пункте

- наблюдений «Карымшина» // Вестник КРАУНЦ. Физико-математические науки. 2019. Т. 29. №4. С. 166–172. DOI: 10.26117/2079-6641-2019-29-4-166-172.
43. Marapulets Y., Solodchuk A., Shcherbina A. Changes of geoaoustic emission directivity at «Mikizha» site associated with earthquakes in Kamchatka // Proceedings of the VII International Conference Solar-Terrestrial Relations and Physics of Earthquakes precursors. 2016. 5 p. DOI: 10.1051/e3sconf/20161100014.
44. Feller K.R., Abercrombie R.E., Ekstrom G. A Common Origin for Aftershocks, Foreshocks, and Multiplets // Bulletin of the Seismological Society of America. 2004. vol. 94. no. 1. pp. 88–98. DOI: 10.1785/0120030069.
45. Иванов В.И., Барат В.А. Акустико-эмиссионная диагностика (справочник) // М.: Спектр, 2017. 362 с.
46. Щербина А.О., Солодчук А.А. Локализация источников акустического излучения по данным распределенной системы комбинированных приемников // Вестник КРАУНЦ. Физико-математические науки. 2023. Т. 44. №3. С. 144–156. DOI: 10.26117/2079-6641-2023-44-3-144-156.
47. Senkevich Y., Mishchenko M., Solodchuk A. Observation of the State of the Near-surface Lithosphere Layer based on the Analysis of the Dynamics of the Geoaoustic Emission Signal Characteristics // Proceedings of the IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE). 2024. pp. 1140–1143. DOI: 10.1109/PIERE62470.2024.10805026.
48. Проект РНФ 18-11-00087. Комплексный анализ акустических и электромагнитных сигналов для оценки уровня сейсмической опасности и предупреждения природных и техногенных катастроф. 2021.
49. Татевосян Р. Э. Афтершок // Большая российская энциклопедия: научно-образовательный портал. URL: <https://bigenc.ru/c/aftershok-b80b0f?v=5911142> (дата обращения 25.01.2026)
50. Богомолов Л.М., Родкин М.В., Сычев В.Н. Инстантонное представление форшок-афтершоковых последовательностей // Физика Земли. 2025. №2. С. 43–57. DOI: 10.7868/S3034645225020045.
51. Свидетельство о государственной регистрации базы данных №2017620097 от 19.01.2017. База данных "Геодеформационные наблюдения в долине реки Карымшина (полуостров Камчатка)" / Ларионов И.А., Мищенко М.А.
52. Цирель С.В. Дилатансия в грунтах // Большая российская энциклопедия: научно-образовательный портал. URL: <https://bigenc.ru/c/dilatansia-v-gruntakh-792dde> (дата обращения: 25.01.2026)
53. Мирный А.Ю. Исследования дилатансии в дисперсных грунтах и методы ее количественной оценки // Инженерная геология. 2019. Т. 14. №2. С. 34–43. DOI: 10.25296/1993-5056-2019-14-2-34-43.
54. Свидетельство о государственной регистрации базы данных №2017621061 от 20.09.2017 г. База данных «Аномальные возмущения высокочастотной геоакустической эмиссии в пунктах наблюдений "Микижа" и "Карымшина" (полуостров Камчатка)» / Мищенко М.А., Ларионов И.А.
55. Свидетельство о государственной регистрации базы данных №2018621109 от 19.07.2018 г. База данных «Геоакустические и деформационные аномалии в долине реки Карымшина (полуостров Камчатка)» / Ларионов И.А., Мищенко М.А.
56. Каталог землетрясений Камчатки и Командорских островов // Камчатский филиал ФИЦ ЕГС РАН. URL: <http://sdis.emsd.ru/info/earthquakes/catalogue.php> (дата обращения: 25.01.2026).
57. Мингалев А.В. Метод и алгоритмы обработки тепловизионных изображений на основе сверточных нейронных сетей для автоматического обнаружения заданных типов объектов // Вестник Рязанского Государственного Радиотехнического Университета. 2025. №91. С. 158–167. DOI: 10.21667/1995-4565-2025-91-158-167.

58. Увайсов С.У., Черновская В.В., Данг Н.В., Туан Н.В. Применение искусственной нейронной сети в задаче тепловой диагностики печатного узла устройства контроля разбега самолета // Моделирование, оптимизация и информационные технологии. 2022. Т. 10. №3(38). С. 23–24. DOI: 10.26102/2310-6018/2022.38.3.012.
59. Набатов Н.Н., Ерачина С.М. Использование сверточных нейронных сетей для удаления шума и восстановления изображений в медицинской и телеметрической практике // Международный научно-исследовательский журнал. 2025. №10(160). С. 1–14. DOI: 10.60797/IRJ.2025.160.34.
60. Yang Z., Algesheimer R., Tessone C.J. A Comparative Analysis of Community Detection Algorithms on Artificial Networks // Scientific Reports. 2016. vol. 6. pp. 1–18. DOI: 10.1038/srep30750.

Сенкевич Юрий Игоревич — д-р техн. наук, доцент, ведущий научный сотрудник, лаборатория акустических исследований, Федеральное государственное бюджетное учреждение науки «Институт космофизических исследований и распространения радиоволн Дальневосточного отделения Российской академии наук» (ИКИР ДВО РАН). Область научных интересов: обработка данных и выявление скрытых закономерностей в хранилищах данных, применение методов математики и информатики в геофизических исследованиях, разработка методов и программ обработки и анализа данных с применением нейросетевых технологий. Число научных публикаций — 117. senkevich@ikir.ru; ул. Мирная, 7, 684034, с. Паратунка, Россия; р.т.: +7(415)313-3193.

Мищенко Михаил Александрович — канд. физ.-мат. наук, ведущий научный сотрудник, лаборатория акустических исследований, Федеральное государственное бюджетное учреждение науки «Институт космофизических исследований и распространения радиоволн Дальневосточного отделения Российской академии наук» (ИКИР ДВО РАН). Область научных интересов: подповерхностная геофизика, деформации, геоакустика, атмосферное электрическое поле, ЦОС. Число научных публикаций — 80. micle@ikir.ru; ул. Мирная, 7, 684034, с. Паратунка, Россия; р.т.: +7(415)313-3193.

Поддержка исследований. Исследование проведено в рамках госзадания Института космофизических исследований и распространения радиоволн ДВО РАН (проект № 124012300245-2).

Y. SENKEVICH, M. MISHCHENKO
**SEARCH FOR SPECIFIC CHANGES IN GEOACOUSTIC
EMISSION CHARACTERISTICS ASSOCIATED WITH STRONG
EARTHQUAKES IN KAMCHATKA**

Senkevich Y., Mishchenko M. Search for Specific Changes in Geoacoustic Emission Characteristics Associated with Strong Earthquakes in Kamchatka.

Abstract. The catastrophic consequences of strong earthquakes are a well-recognized hazard. Seismic processes associated with earthquake preparation in subduction zones induce deformation in the surface layers of the Earth's crust, including the sedimentary layer, where the phenomenon of acoustic oscillation generation, known as geoacoustic emission (GAE), is observed. The primary mechanism of GAE generation is associated with the movement of dislocations resulting from the interaction of fractions and their clusters within sedimentary rocks. Deformations in the sedimentary layer lead to changes in its physical properties and, consequently, to alterations in the conditions under which acoustic oscillations are generated. A relationship has been established between the degree of the stress-strain state of sedimentary rocks and the intensity of the pulsed GAE flux. The GAE signal constitutes a sequence of pulses with varying duty cycles, amplitudes, and waveforms. During periods of heightened seismic activity, changes in signal parameters are observed within the sedimentary rocks. A detailed study of the changes in the characteristics of signal fragments recorded during periods directly associated with the onset of earthquakes in Kamchatka, and the search for their distinctions from changes in signal fragments during periods unrelated to earthquakes, constituted the core of this research.

Keywords: sedimentary rocks, surface waves, geoacoustic emission, signal characteristic anomalies, anomaly classification, neural network image processing.

References

1. Shebalin P.N., Gvishiani A.D., Dzeboev B.A., Skorkina A.A. [Why are New Approaches to Seismic Hazard Assessment Required?]. *Doklady Rossiiskoi akademii nauk. Nauki o Zemle – Reports of the Russian Academy of Sciences. Earth Sciences.* 2022. vol. 507. no. 1. pp. 91–97. DOI: 10.31857/S2686739722601466. (In Russ.).
2. Spivak A.A., Rybnov Y.S. [Acoustic Effects of Strong Earthquakes]. *Fizika Zemli – Physics of the Earth.* 2021. no. 1. pp. 41–50. DOI: 10.31857/S0002333721010099. (In Russ.).
3. Sychev V.N., Bogomolov L.M. [The potential of nonlinear dynamics methods for analyzing geophysical series and seismicity.]. *Sb. nauchny'x statej VI Mezhdunarodnoj konferencii «Solnechno-zemny'e svyazi i fizika predvestnikov zemletryasenij».* 2013. pp. 78. (In Russ.).
4. Bogomolov L.M., Sycheva N.A. [Earthquake Predictions in XXI Century: Prehistory and Concepts, Precursors and Problems]. *Geosistemy' perexodny'x zon – Geosystems of Transition Zones.* 2022. vol. 6. no. 3. pp. 145–182. DOI: 10.30730/gtr.2022.6.3.145-164.164-182. (In Russ.).
5. Saltykov V.A., Kugaenko Y.A. [Development of Near-Surface Dilatancy Zones as a Possible Cause for Seismic Emission Anomalies Before Strong Earthquakes]. *Pacific Geology – Tixookenskaya geologiya.* 2012. vol. 31. no. 1. pp. 96–106. (In Russ.).
6. Chebotareva I.Y., Volodin I.A., Dryagin V.V. [Acoustic Effects in the Deformation of Structurally Inhomogeneous Media]. *Akusticheskij zhurnal – Acoustic Journal.* 2017. vol. 63. no. 1. pp. 84–93. DOI: 10.7868/S0320791916060046. (In Russ.).

7. Ponomarev A.V., Sobolev G.A. Fizika zemletryasenij i predvestniki [Physics of Earthquakes and Precursors]. M.: Nauka. 2003. 270 p. (In Russ.).
8. Muratov P.V., Rulenko O.P., Marapulets Yu.V., Solodchuk A.A. [Electrical and acoustic response of near-surface sedimentary rocks to the passage of seismic waves from earthquakes]. Vestnik KRAUNC. Fiziko-matematicheskie nauki – Bulletin of KRAUNTS. Physical and mathematical sciences. 2018. no. 5(25). pp. 62–73. DOI: 10.18454/2079-6641-2018-25-5-62-73. (In Russ.).
9. Gavrilov V.A. Vozdejstvie peremenny'x e'lektromagnitny'x polej na geokusticheskie processy': e'mpiricheskie zakonomernosti i fizicheskie mexanizmy' [The Effect of Variable Electromagnetic Fields on Geoacoustic Processes: Empirical Patterns and Physical Mechanisms]: special'nost' 25.00.10 «Geofizika, geofizicheskie metody' poiskov polezny'x iskopaemy'x»: avtoreferat dissertacii na soiskanie uchenoj stepeni doktora fiziko-matematicheskix nauk. Federal'noe gosudarstvennoe byudzhethoe uchrezhdenie nauki Institut vulkanologii i sejsmologii Dal'nevostochnogo otdeleniya Rossijskoj akademii nauk (IViS DVO RAN). Petropavlovsk-Kamchatskij, 2017. 51 p. (In Russ.).
10. Segall P. Earthquake and Volcano Deformation. Princeton: Princeton University Press. 2010. 464 p.
11. Marapulets Y.V., Shevtsov B.M., Larionov I.A., Mishchenko M.A., Shcherbina A.O., Solodchuk A.A. [Geoacoustic Emission Response to Deformation Processes Activation During Earthquake Preparation]. Tixookeanskaya geologiya – Russian Journal of Pacific Geology. 2012. vol. 31. no. 6. pp. 59–67. (In Russ.).
12. Bogdanov V.V., Bychkov V.V., Vodinchar G.M. Dinamika fizicheskix processov v aktivny'x zonax blizhnego kosmosa i geosfer: izbranny'e raboty' 2018-2020 gg. [Dynamics of Physical Processes in Active Zones of Near Space and Geospheres: Selected Works 2018-2020] (Russ. ed.: Shevtsov B.M.) M.: RAN. 2021. 409 p. (In Russ.).
13. Senkevich Yu.I., Mishchenko M.A. [Methodology for assessing the condition of near-surface sedimentary rocks based on observations of dynamic characteristics of geoacoustic emission]. Vestnik KRAUNC. Fiziko-matematicheskie nauki – Bulletin of KRAUNTS. Physical and mathematical sciences. 2023. vol. 45. no. 4. pp. 109–121. DOI: 10.26117/2079-6641-2023-45-4-109-121. (In Russ.).
14. Senkevich, Y., Marapulets, Y., Lukovenkova, O., Solodchuk, A. [Technique of Informative Features Selection in Geoacoustic Emission Signals]. Trudy' SPIIRAN – SPIIRAS Proceedings. 2019. vol. 18. no. 5. pp. 1066–1092. DOI: 10.15622/sp.2019.18.5.1066-1092. (In Russ.).
15. Baranov B.V., Tsukanov N.V., Gaedicke Ch., Freitag R., Dozorova K.A. [Morphology and Structural Features of the Subduction Zone in the Vicinity of Kuril-Kamchatka-Aleutian Junction]. Okeanologiya – Oceanology. 2022. vol. 62. no. 4. pp. 528–539. DOI: 10.31857/S0030157422040025. (In Russ.).
16. Avdeiko G.P., Palueva A.A. [Historic Analysis of Seismicity and Seismic Hazard of the Kamchatkan Subduction Zone]. Vestnik KRAUNC. Nauki o zemle – Bulletin of KRAUNTS. Earth sciences. 2010. no. 15(1). pp. 69–89. (In Russ.).
17. Safonov D.A. [Spatial Distribution of Tectonic Stress in the Southern Deep Part of the Kuril-Kamchatka Subduction Zone]. Geosistemy' perexodny'x zon – Geosystems of Transition Zones. 2019. vol. 3. no. 2. pp. 175–188. DOI: 10.30730/2541-8912.2019.3.2.175-188. (In Russ.).
18. Gapeev M.I., Marapulets Yu.V. [Modeling of zones of relative shear deformations before strong earthquakes in Kamchatka, which occurred in the period 2018-2021]. Vestnik KRAUNC. Fiziko-matematicheskie nauki – Bulletin of KRAUNTS. Physical and mathematical sciences. 2021. vol. 37. no. 4. pp. 53–66. DOI: 10.26117/2079-6641-2021-37-4-53-66. (In Russ.).

19. Senkevich Y., Mishchenko M., Solodchuk A. Observation of the State of the Near-surface Lithosphere Layer based on the Analysis of the Dynamics of the Geoaoustic Emission Signal Characteristics. *Proceedings of the IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE)*. 2024. pp. 1140–1143. DOI: 10.1109/PIERE62470.2024.10805026.
20. Senkevich Yu.I. [Processing and preparation of observational data in order to identify the features of the dynamics of geoaoustic emission characteristics]. *Vestnik KRAUNC. Fiziko-matematicheskie nauki – Bulletin of KRAUNTS. Physical and mathematical sciences*. 2024. vol. 47. no. 2. pp. 75–94. DOI: 10.26117/2079-6641-2024-47-2-75-94. (In Russ.).
21. Senkevich Yu.I., Mishchenko M.A., Solodchuk A.A., Lukovenkova O.O., Gapeev M.I. [Stochastic modeling of geoaoustic emission signal]. *Cifrovaya obrabotka signalov – Digital signal processing*. 2023. no. 1. pp. 27–34. (In Russ.).
22. Patent RU 2758582 C1. [A method for detecting a complex earthquake precursor]: №2020138668: zayavl. 26.11.2020: opubl. 29.10.2021. Senkevich Yu.I., Marapulets Yu.V., Lukovenkova O.O., Solodchuk A.A., Mishchenko M.A., Malkin E.I., Gapeev M.I. (In Russ.).
23. Builo S.I. *Fiziko-mexanicheskie, statisticheskie i aspekty` akustiko-e`missionnoj diagnostiki: monografiya* [Physical and mechanical, statistical, and aspects of acoustic emission diagnostics: a monograph]. Taganrog: Izd-vo Yuzhnogo federal'nogo universiteta. 2017. 182 p. (In Russ.).
24. Greshnikov V.A. *Akusticheskaya e`missiya. Primenenie dlya ispy`tanj materialov i izdelij* [Physical and mechanical, statistical, and aspects of acoustic emission diagnostics: a monograph]. M.: Izd-vo standartov. 1976. 272 p. (In Russ.).
25. Semashko N.A., Shport V.I., Maryin B.N. *Akusticheskaya e`missiya v e`ksperimental'nom materialovedeni* [Acoustic Emission in Experimental Materials Science]. M.: Mashinostroenie. 2002. 239 p. (In Russ.).
26. Grosse C.U., Ohtsu M., Aggelis D.G., Shiotani T. *Acoustic Emission Testing*. Springer Tracts in Civil Engineering. Springer. 2022. 746 p. DOI: 10.1007/978-3-030-67936-1.
27. Berezina Yu.A., Spodarevab L.A. [Propagation of acoustic signals in soils]. *Prikladnaya Mekhanika i Tekhnicheskaya Fizika – Applied mechanics and technical physics*. 2001. vol. 42. no. 4(248). pp. 177–183. (In Russ.).
28. Ou J., Wang E., Wang X., et al. Acoustic emission characteristics and damage evolution of different rocks under uniaxial compression conditions. *Scientific Reports*. 2024. vol. 14. pp. 1–10. DOI: 10.1038/s41598-024-54950-9.
29. Lockner D. The Role of Acoustic Emission in the Study of Rock Fracture. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*. 1993. vol. 30. no. 7. pp. 883–899.
30. Xiangguo K., Enyuan W., Xueqiu H., et. al. Multifractal characteristics and acoustic emission of coal with joints under uniaxial loading. *Fractals*. 2017. vol. 25. no. 5. pp. 1–11. DOI: 10.1142/S0218348X17500451.
31. Carpinteri A., Lacidogna G., Puzzi Si. From criticality to final collapse: Evolution of the “b-value” from 1.5 to 1.0. *Chaos, Solitons & Fractals*. 2009. vol. 41. pp. 843–853 DOI: 10.1016/j.chaos.2008.04.010.
32. Guarino A., Garcimartin A., Ciliberto S. An experimental test of the critical behaviour of fracture precursors. *The European Physical Journal B – Condensed Matter and Complex Systems*. 1998. vol. 6. pp. 13–24. DOI: 10.1007/s100510050521.
33. Morozova K.G. *Sejsmoakusticheskaya e`missiya, soprovozhdayushhaya razlichny`e rezhimy` skol'zheniya po razlomam i treshhinam* [Seismoacoustic emission accompanying various modes of sliding along faults and fractures]: special'nost' 1.6.9 «Geofizika»: avtoreferat dissertacii na soiskanie uchenoj stepeni doktora

- fiziko-matematicheskix nauk. Institut dinamiki geosfer imeni akademika M.A. Sadovskogo Rossijskoj akademii nauk (IDG RAN). Moskva, 2023. 144 p. (In Russ.).
34. Sheremetyeva O.V., Shevtsov B.M. [Approximation of the laws of the distribution of waiting times for foreshocks based on a fractional model of deformation activity]. Vestnik KRAUNC. Fiziko-matematicheskie nauki – Bulletin of KRAUNTS. Physical and mathematical sciences. 2022. vol. 40. no. 3. pp. 137–152. DOI: 10.26117/2079-6641-2022-40-3-137-152. (In Russ.).
 35. Mileshekin M.B., Biblik I.V., Stoev P.I. [Analysis of the amplitude distribution of acoustic emission signals based on the computational and experimental method]. Fizicheskaya mezomechanika – Physical Mesomechanics. 2004. vol. 7. no. S1-1. pp. 206–209. (In Russ.).
 36. Lacoste M., Ruiz S., Or D. Listening to earthworms burrowing and roots growing – acoustic signatures of soil biological activity. Scientific Reports. 2018. vol. 8. pp. 1–9. DOI: 10.1038/s41598-018-28582-9.
 37. Kocharyan G.G., Ostapchuk A.A. [Acoustic Emission During Different-Type Inter-Block Movements]. Fiziko-technicheskie problemy razrabotki polezny'x iskopaemy'x – Physico-technical problems of mining. 2015. no. 1. pp. 3–13. (In Russ.).
 38. Baranov S.V. [Estimating Aftershock Process Activity by Gutenberg-Richter Law and Etas Model]. Fundamental'ny'e issledovaniya – Fundamental Research. 2014. no. 12. pp. 751–755. (In Russ.).
 39. Baro J., Corral A., Illa X., et al. Statistical Similarity between the Compression of a Porous Material and Earthquakes. Physical Review Letters. 2013. vol. 110. 088702 p. DOI: 10.1103/PhysRevLett.110.088702.
 40. Sedghizadeh M., Van Den Bergh M., Shcherbakov R. Statistical and clustering analysis of microseismicity from a Saskatchewan potash mine. Frontiers in Applied Mathematics and Statistics. 2023. vol. 9. 1126952 p. DOI: 10.3389/fams.2023.1126952.
 41. Turuntaev S.B. Kinetika texnogennoj sejsmichnosti [Kinetics of Anthropogenic Seismicity]: special'nost' 25.00.10 «Geofizika, geofizicheskie metody poiskov polezny'x iskopaemy'x»: avtoreferat dissertacii na soiskanie uchenoj stepeni doktora fiziko-matematicheskix nauk. Institut fiziki Zemli im. O. Yu. Shmidta RAN (IFZ RAN). Moskva, 2005. 330 p. (In Russ.).
 42. Mishchenko M.A., Marapulets Yu.V., Larionov I.A. [Analysis of time and frequency characteristics of simultaneous geodeformation signals at the "Karymshina" observation point]. Vestnik KRAUNC. Fiziko-matematicheskie nauki – Bulletin of KRAUNTS. Physical and mathematical sciences. 2019. vol. 29. no. 4. pp. 166–172. DOI: 10.26117/2079-6641-2019-29-4-166-172. (In Russ.).
 43. Marapulets Y., Solodchuk A., Shcherbina A. Changes of geoacoustic emission directivity at «Mikizha» site associated with earthquakes in Kamchatka. Proceedings of the VII International Conference Solar-Terrestrial Relations and Physics of Earthquakes precursors. 2016. 5 p. DOI: 10.1051/e3sconf/20161100014.
 44. Feller K.R., Abercrombie R.E., Ekstrom G. A Common Origin for Aftershocks, Foreshocks, and Multiplets. Bulletin of the Seismological Society of America. 2004. vol. 94. no. 1. pp. 88–98. DOI: 10.1785/0120030069.
 45. Ivanov V.I., Barat V.A. Akustiko-e'missionnaya diagnostika (spravochnik) [Acoustic Emission Diagnostics (reference book)]. M.: Spektr. 2017. 362 p. (In Russ.).
 46. Shcherbina A.O., Solodchuk A.A. [Localization of Acoustic Radiation Sources According to a Distributed System of Combined Receivers]. Vestnik KRAUNC. Fiziko-matematicheskie nauki – Bulletin of KRAUNTS. Physical and mathematical sciences. 2023. vol. 44. no. 3. pp. 144–156. DOI: 10.26117/2079-6641-2023-44-3-144-156. (In Russ.).
 47. Senkevich Y., Mishchenko M., Solodchuk A. Observation of the State of the Near-surface Lithosphere Layer based on the Analysis of the Dynamics of the Geoacoustic Emission Signal Characteristics. Proceedings of the IEEE 3rd

- International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE). 2024. pp. 1140–1143. DOI: 10.1109/PIERE62470.2024.10805026.
48. Project RNF 18-11-00087. [Comprehensive analysis of acoustic and electromagnetic signals to assess the level of seismic hazard and prevent natural and man-made disasters]. 2021. (In Russ.).
 49. Tatevosyan R.E. Aftershock. The Great Russian Encyclopedia: a scientific and educational portal. Available at: <https://bigenc.ru/c/aftershok-b80b0f/?v=5911142> (accessed 25.01.2026). (In Russ.).
 50. Bogomolov L.M., Rodkin M.V., Sychev V.N. [Instanton Representation of Foreshock-Aftershock Sequences]. *Fizika Zemli – Physics of the Earth*. 2025. no. 2. pp. 43–57. DOI: 10.7868/S3034645225020045. (In Russ.).
 51. Certificate of state registration of the database №2017620097 dated 19.01.2017. [Database "Geodeformation observations in the Karymshina River valley (Kamchatka Peninsula)"]. Larionov I.A., Mishchenko M.A. (In Russ.).
 52. Tsirel S.B. Dilatation in soils. The Great Russian Encyclopedia: a scientific and educational portal. Available at: <https://bigenc.ru/c/dilatansia-v-gruntakh-792dde> (accessed 25.01.2026) (In Russ.).
 53. Mirny A.Yu. [Disperse soils dilatancy studies and the methods of its quantitative assessment]. *Inzhenernaya geologiya – Engineering Geology World*. 2019. vol. 14. no. 2. pp. 34–43. DOI: 10.25296/1993-5056-2019-14-2-34-43. (In Russ.).
 54. Certificate of state registration of the database №2017621061 dated 20.09.2017. ["Anomalous disturbances of high-frequency geoacoustic emission at the observation points "Mikizha" and "Karymshina" (Kamchatka Peninsula)"]. Mishchenko M.A., Larionov I.A. (In Russ.).
 55. Certificate of state registration of the database №2018621109 dated 19.07.2018. [Database "Geoacoustic and deformation anomalies in the Karymshina River Valley (Kamchatka Peninsula)"]. Larionov I.A., Mishchenko M.A. (In Russ.).
 56. Earthquake Catalog of Kamchatka and the Commander Islands. Kamchatka Branch of the Russian Academy of Sciences. Available at: <http://sdis.emsd.ru/info/earthquakes/catalogue.php> (accessed 25.01.2026). (In Russ.).
 57. Mingalev A.V. [Method and algorithms of processing scanner thermal images based on high-accuracy neural networks for automatic detection of specified types of objects]. *Vestnik Ryazanskogo Gosudarstvennogo Radiotekhnicheskogo Universiteta – Vestnik of Ryazan State Radio Engineering University*. 2025. no. 91. pp. 158–167. DOI: 10.21667/1995-4565-2025-91-158-167. (In Russ.).
 58. Uvaysov S.U., Chernoverskaya V.V., Dang N.V., Tuan N.V. [The use of an artificial neural network in thermal diagnostics of the printed node of the on-board take-off control device of an aircraft]. *Modelirovanie, optimizaciya i informacionny'e texnologii – Modeling, optimization and information technology*. 2022. vol. 10. no. 3(38). pp. 23–24. DOI: 10.26102/2310-6018/2022.38.3.012. (In Russ.).
 59. Nabatov N.N., Yerachina S.M. [The use of convolutional neural networks for noise removal and image restoration in medical and telemetrical practice]. *Mezhdunarodny'j nauchno-issledovatel'skij zhurnal – International Research Journal*. 2025. no. 10(160). pp. 1–14. DOI: 10.60797/IRJ.2025.160.34. (In Russ.).
 60. Yang Z., Algesheimer R., Tessone C.J. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific Reports*. 2016. vol. 6. pp. 1–18. DOI: 10.1038/srep30750.

Senkevich Yuriy — Ph.D., Dr.Sci., Associate Professor, Leading Researcher, Laboratory of Acoustic Research, Federal State Budgetary Institution of Science «Institute of Space Physics Research and Radio Wave Propagation of the Far Eastern Branch of the Russian Academy of Sciences» (IKIR FEB RAS). Research interests: data processing and identification of hidden

patterns in data repositories, application of mathematical and computer science methods in geophysical research, development of methods and programs for data processing and analysis using neural network technologies. The number of publications — 117. senkevich@ikir.ru; 7, Mirnaya St., 684034, Paratunka Vill., Russia; office phone: +7(415)313-3193.

Mishchenko Michael — Ph.D., Leading researcher, Laboratory of acoustic research, Federal State Budgetary Institution of Science «Institute of Space Physics Research and Radio Wave Propagation of the Far Eastern Branch of the Russian Academy of Sciences» (IKIR FEB RAS). Research interests: near-surface geophysics, deformations, geoacoustics, atmospheric electric field, digital signal processing. The number of publications — 80. micle@ikir.ru; 7, Mirnaya St., 684034, Paratunka Vill., Russia; office phone: +7(415)313-3193.

Acknowledgements. This study was conducted as part of a government assignment from the Institute of Cosmophysical Research and Radio Wave Propagation, FEB RAS (Project No. 124012300245-2).

Н.Ю. МИРОШНИКОВ, У.В. ТОКАРЕВА, А.Д. ШУЛЬЖЕНКО
**СПОСОБ ОПРЕДЕЛЕНИЯ АИ-ПРЕОБРАЗОВАННЫХ РЕЧЕВЫХ
АУДИОЗАПИСЕЙ НА ОСНОВЕ АКУСТИЧЕСКИХ
И СТАТИСТИЧЕСКИХ ПРИЗНАКОВ**

Мирошников Н.Ю., Токарева У.В., Шутьженко А.Д. Способ определения АИ-преобразованных речевых аудиозаписей на основе акустических и статистических признаков.

Аннотация. В работе предложен способ определения факта преобразования аудиосигнала с использованием интеллектуальных средств голосового синтеза (AI-based Voice Conversion, AI-VC) речевых аудиозаписей (например, с целью подмены автора или смыслового наполнения) на основе анализа акустических и статистических признаков аудиосигнала. Рассмотрены принципы работы современных систем преобразования голоса, сделан акцент на архитектуре преобразования голоса на основе поиска (Retrieval-based Voice Conversion, RVC). Показано, что преобразование представляет собой поэтапный процесс, в ходе которого происходит извлечение признаков целевого голоса из исходной речевой аудиозаписи, что эквивалентно векторизации бихевиаристических свойств говорящего, построение целевого шаблона – модели для синтеза голоса, формирование необходимого смыслового наполнения и синтез новой речевой аудиозаписи, имеющей акустически незначимое различие с целевой. Проведен обзор информативных признаков, чувствительных к такому преобразованию и принятых как артефакты синтеза: мел-частотные и линейно-частотные кепстральные коэффициенты (Mel-Frequency Cepstral Coefficients, MFCC, Linear-Frequency Cepstral Coefficients, LFCC), кепстральные коэффициенты на основе постоянного Q (Constant Q Cepstral Coefficients, CQCC), статистические свойства ошибки предсказания (Linear Predictive Coding, LPC), контур основной частоты и его вариативность (джиттер, шиммер), спектральный центронид и поток, скорость пересечения нуля. Разработана формальная модель бинарной классификации и программная реализация на основе гауссовых смесей (Gaussian Mixture Model, GMM). Экспериментальная проверка предложенного способа на наборе данных ASVspoof2019 показала точность классификации 93,9% (AUC = 0,939, EER = 13,52%). Это говорит о том, что предложенный способ может обеспечить защиту получателей речевых аудиозаписей от целевых вишинговых атак по типу спуфинга (мошенничества с подменой голосовой идентичности). Способ может быть реализован в виде программного модуля и интегрирован в мессенджеры или корпоративные коммуникационные платформы. Актуальность работы обусловлена отсутствием готовых продуктов, определяющих голосовые подделки на основе результатов спектрального анализа, несмотря на то, что подобные разработки ведутся как в нашей, так и в зарубежных странах (и обладают точностью на уровне 91-93%), что подтверждается анализом литературы.

Ключевые слова: противодействие вишингу, искусственный интеллект, дипфейк, преобразование голоса, спуфинг, машинное обучение, акустические признаки, статистические признаки.

1. Введение. Рост доступности средств AI-VC и вычислительных мощностей приводит к возрастанию числа пользователей, в том числе и злоумышленников. Это подтверждается

данными BI.ZONE, которые показывают, что количество фишинговых атак в РФ выросло с 111 тыс. в 2022 году до 350 тыс. в 2024 году [1]. При этом, помимо масштабных атак, разрабатываются целевые атаки на сотрудников организаций, основанные на фишинге с использованием интеллектуальных средств преобразования голоса [2]. При их реализации мошенники выдают себя за высокопоставленных государственных служащих [3, 4], воспользовавшись голосовыми записями последних, находящимися в открытом доступе, для подделки их голоса. Целью таких атак чаще всего является получение идентификационных данных граждан или попрошайничество.

С 2019 года известна новая схема мошенничества с использованием AI-VC, которая представляет собой целевую атаку на предприятие. В ходе атаки сначала отправляется электронное письмо без вложений и ссылок (что затрудняет обнаружение атаки), в котором злоумышленник представляется руководителем адресата, а затем совершается звонок с использованием вышеупомянутых средств обработки аудио и просьбой произвести денежный перевод со счетов предприятия [5]. Именно сочетание этих действий делает мошенника еще более убедительным в глазах жертвы.

Важное замечание. Для удовлетворительного клонирования голоса (оценка MOS > 4.0) средствами RVC достаточно 10 минут (для получения максимально точного результата – 30 минут) чистой речи целевого говорящего. Процесс требует дообучения модели на данных жертвы, которое занимает ~30-60 минут на GPU уровня GTX 1660.

Zero-shot модели могут работать с эталонным аудио длительностью 5-60 секунд, но качество тембральной передачи уступает результату, полученному с применением RVC.

Такой объем данных может быть получен из записей публичных выступлений или голосовых сообщений в открытых чатах мессенджеров или в соцсетях, что подтверждает актуальность разработки детекторов AI-VC.

Итак, статистика говорит об уязвимости людей к целевым атакам (особенно вишингу), что требует разработки эффективных методов детекции синтезированной речи. Технически эта задача состоит в выявлении и оценке артефактов, внесённых процессом синтеза, для принятия многокритериального решения об их значимости. Похожая задача решается при распознавании когнитивных нарушений, где анализируются естественные отклонения в речи [4]. Как отмечено в [4], анализ акустических и лингвистических признаков позволяет выявлять отклонения в речевом сигнале, вызванные патологическими

или искусственными изменениями. В данной работе развивается этот подход применительно к артефактам генеративных моделей.

Цель данного исследования – повышение эффективности защиты от целевых вишинговых атак.

Задачей исследования является разработка способа и программного средства для определения факта использования AI-преобразования речевых аудиозаписей на основе оценки модифицированного набора акустических и статистических признаков.

2. Анализ современных средств преобразования голоса и их артефактов. Современные средства преобразования голоса можно классифицировать по трем группам [5]:

1. Классические методы цифровой обработки сигналов (DSP), включающие в себя pitch / formant shift и вокодеры (MorphVox, Voicemod).

2. Гибридные методы, сочетающие DSP с машинным обучением. К этому разделу относятся нейросетевые вокодеры (YourTTS, ElevenLabs).

3. Глубокие нейросетевые модели, представляющие собой архитектуры на основе вариационных автокодировщиков, диффузионных моделей и retrieval-based подходов (RVC, So-VITS-SVC, Diff-SVC).

При сравнении средств преобразования реалистичность оценивается по шкале MOS. Mean Opinion Score (MOS) – субъективная оценка качества речи по пятибалльной шкале, где 5 – это эквивалент естественной речи. Сравнение средств преобразования голоса по данным источника [5] представлено в таблице 1.

Как видно из таблицы 1, RVC представляет собой оптимальный вариант: высокая оценка по шкале MOS, близкая к уровню диффузионных моделей, а также относительная вычислительная эффективность, так как время инференса в 1.6 раза меньше, чем для VITS и в 6.7 раз ниже, чем для Diff-SVC. Более того, у RVC минимальная задержка в реальном времени составляет всего лишь 20 мс, что является особенно важным фактором, позволяющим использовать RVC в квази-реальном времени для голосовых звонков, а именно этот сценарий представляет наибольшую угрозу при проведении целевых вишинговых атак. Также важно принимать во внимание открытость кода и активное сообщество разработчиков, что делает технологию доступной для злоумышленников.

Таблица 1. Сравнение средств преобразования голоса

Критерий	Классические методы DSP (MorphVox)	Гибридные методы (ElevenLabs)	Глубокие нейросетевые модели		
			VITS	Diff-SVC	RVC
Реалистичность (оценка по шкале MOS)	2.1 ± 0.4	3.8 ± 0.5	4.1 ± 0.3	4.4 ± 0.2	4.3 ± 0.3
Время инференса (3-сек. аудио, CPU i5)	0.8 с	4.2 с	3.1 с	12.7 с	1.9 с
Требования к GPU (минимум для реального времени)	Не требуется	RTX 3060	RTX 3070	RTX 4090	GTX 1660
Доступность (открытый исходный код)	Частично	Нет	Да	Да	Да, активное сообщ.
Распространенность (оценка по GitHub и форумам)	Низкая	Высокая, коммерческая	Средняя	Низкая	Высокая
Уязвимость к детекции (минимальный достигнутый EER на ASVspoof2019)	0.05	0.11	0.13	0.14	0.135
Задержка в реальном времени (мс)	50	200	80	350	20

Архитектура RVC включает 4 модуля, а значит, и 4 этапа обработки аудиозаписи. Первый – это энкодер на базе HuBERT для разделения фонетического и стилистического содержимого, второй – codebook для поиска ближайших стилистических представлений, третий – декодер на основе Flow-модели VITS, четвертый – вокодер HiFi-GAN для синтеза аудиоволны. На каждом этапе обработки сигнала в архитектуре RVC возникают артефакты в синтезированной речи, формальное происхождение которых представлено в данной работе. На первом этапе исходный сигнал $y(t)$ преобразуется в последовательность скрытых представлений $h_t \in R^d$:

$$h_t = \text{HuBERT}(y_{[t-\Delta, t+\Delta]}), \quad (1)$$

здесь h_t – это «сырая» последовательность, содержащая смесь фонетических и стилистических данных аудиозаписи.

HuBERT обучен на дискретизации акустического пространства с потерей информации, что можно представить как проекцию в подпространство меньшей размерности с последующим квантованием:

$$h_t = Q\left(W^T \phi(y_{[t-\Delta, t+\Delta]})\right), \quad (2)$$

где $\phi(\cdot)$ – нелинейное преобразование, $W \in R^{D \times d}$ – матрица проекции ($d \ll D$), $Q(\cdot)$ – оператор векторного квантования. Эта операция необратима и приводит к потере фазовой информации исходного сигнала:

$$\varepsilon_{RVC}[n] = y_{RVC}[n] - \sum_{k=1}^p a_k y_{RVC}[n-k]. \quad (3)$$

Поскольку фазовые компоненты искажены, дисперсия остатка увеличивается:

$$\sigma_{\varepsilon, RVC}^2 = E[\varepsilon_{RVC}^2[n]] > \sigma_{\varepsilon, natural}^2. \quad (4)$$

Далее следует второй этап (codebook). Механизм поиска ближайшего соседа выводит дискретизацию стилистического пространства:

$$c_t^* = \arg \min_{c \in \mathcal{C}} \|h_t^{\text{style}} - c\|_2, \quad (5)$$

где c – конечный набор кодовых векторов; h_t^{style} – стилистическое пространство (тембр, интонация, манера речи), которое затем передаётся в декодер для придания синтезу целевого голоса и отвечает за то, как именно аудио произносится.

Эта операция приводит к квантованию просодических паттернов, что проявляется в снижении вариативности основной частоты F_0 . Для естественной речи стандартное отклонение F_0 составляет $\sigma_{F_0} \approx 35$ Гц, для RVC – $\sigma_{F_0} \approx 22$ Гц (-37%), соответственно, джиттер уменьшается:

$$\text{Jitter}_{RVC} = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| \frac{T_0[i+1] - T_0[i]}{T_0[i]} \right|. \quad (6)$$

На третьем этапе декодер моделирует преобразование через обратимые потоки:

$$z = f_{\theta}(h_{\text{phon}}, c_{\text{style}}^*), \quad (7)$$

где z – латентное представление мел-спектрограммы, h_{phon} – фонетическое (содержательное) пространство (лингвистическая составляющая, фиксирующая, какие именно звуки/слова произносятся), s – вектор кодовых компонент. Однако на практике потоки аппроксимируют истинное распределение речевого сигнала, который формально описывается как случайный процесс с некоторой точностью. При аппроксимации, то есть при попытке математически обобщить исследуемую речь говорящего, возникает неизбежное расхождение между реальным распределением этого случайного процесса и его модельным представлением. Это расхождение характеризует ошибка распределения δ , учет которой выглядит следующим образом:

$$p_{\text{model}}(z) = p_{\text{true}}(z) + \delta(z), \|\delta\|_2 > 0. \quad (8)$$

Ошибка проявляется как сглаживание спектральной огибающей, что количественно измеряется через увеличение дисперсии высокочастотных MFCC и низкочастотных CQCC.

Графики мощности первого мел-кепстрального коэффициента (MFCC-1) представлены на рисунке 1.

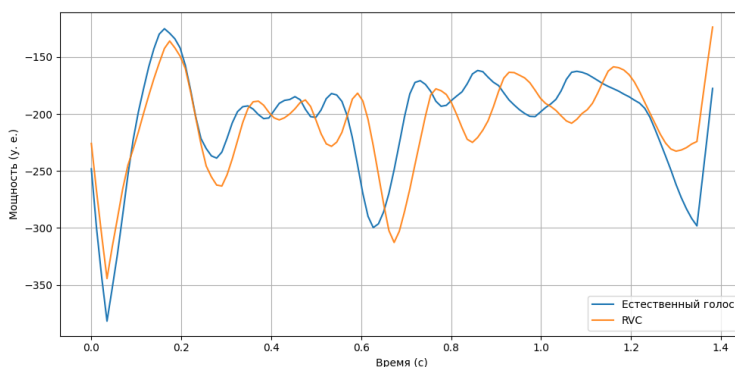


Рис. 1. Графики мощности первого мел-кепстрального коэффициента

Графики мощности второго кохлеарного коэффициента (второго CQCC-коэффициента) показаны на рисунке 2.

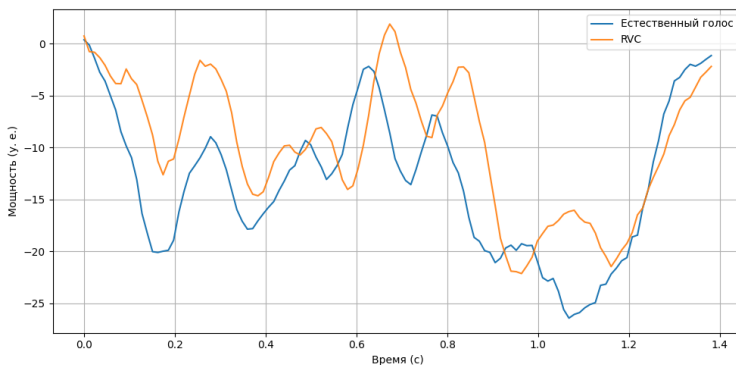


Рис. 2. Графики мощности второго кохлеарного коэффициента

Четвертый этап — это вокодер HiFi-GAN. HiFi-GAN восстанавливает форму волны из мел-спектограммы через генеративную сеть:

$$\hat{y}[n] = G_{\psi}(\mathbf{M}) + \eta_{\text{phase}}[n], E[\eta_{\text{phase}}] \neq 0, \quad (9)$$

где $\eta_{\text{phase}}[n]$ — фазовый шум артефактов.

Это математическое отображение формализует неточное восстановление фазы и базируется на анализе функций потерь нейровокодеров, оптимизирующихся по спектральным метрикам и не содержащих явной фазовой компоненты [12, 13].

Неточное восстановление фазы обосновывает природу искажения на данном этапе. Это приводит к нарушению временной структуры сигнала, что измеряется через скорость пересечения нуля:

$$\text{ZCR}_{\text{RVC}} = \frac{1}{2(N-1)} \sum_{n=1}^{N-1} |\text{sgn}(y[n]) - \text{sgn}(y[n-1])|. \quad (10)$$

ZCR снижается, и это объясняется уменьшением высокочастотных компонент из-за неточного восстановления фазы. График ZCR от естественного и преобразованного с применением RVC сигнала представлен на рисунке 3.

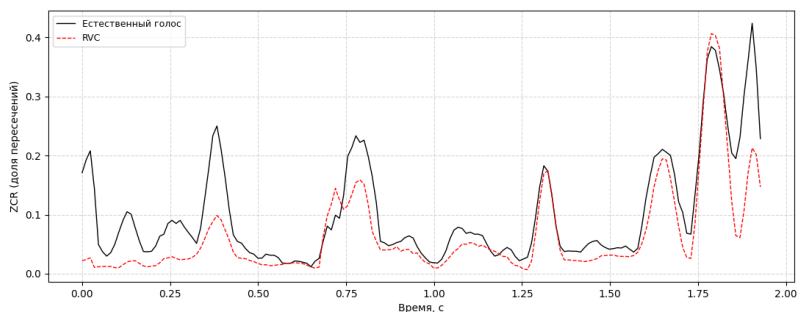


Рис. 3. График ZCR

Примечание: графики 1-3 для иллюстрации различий естественных и сгенерированных аудиозаписей получены экспериментальным путем и подтверждают изложенные в литературе [8, 10, 23, 24] зависимости.

Обобщенное математическое представление искажений при преобразовании голоса через RVC может быть описано следующим образом:

$$y_{\text{RVC}}[n] = \mathcal{T}_{\text{RVC}}(y_{\text{natural}}[n]) + \eta[n], \quad (11)$$

где $\mathcal{T}_{\text{RVC}}$ – нелинейный оператор преобразования, $\eta[n]$ – аддитивный шум артефактов, статистические свойства которого отличаются от естественного шума: $E[\eta^3[n]] \neq 0$ (асимметрия), $E[\eta^4[n]]/\sigma_\eta^4 > 3$ (положительная острота).

Описанное аналитическое обоснование вносимых искажений создает основу для выдвижения базовой гипотезы исследования о возможности определения AI-преобразованных речевых аудиозаписей.

3. Классификация информативных признаков и формальная модель классификации. Для подтверждения и уточнения базовой гипотезы был проведен анализ публикаций, посвященных детекции спуфинга голоса за период с 2019 по 2025 годы. Из всех работ были выбраны 25 наиболее информативных по описанию используемых подходов и оценкам их эффективности. Распределение акустических признаков в исследованиях по детекции AI-преобразованных речевых аудиозаписей представлено в таблице 2.

Таблица 2. Распределение акустических признаков в исследованиях по детекции AI-преобразованных речевых аудиозаписей

Признак	Количество работ	Примеры исследований
MFCC	21	ASVspoof 2019/2021 [22]; Delgado, et al. [8]; Tak, et al. [24]
CQCC	14	ASVspoof 2021 [8]; Sahidullah, et al. [9]; Alzantot, et al. [23]
LFCC	9	ASVspoof 2021 [8]; Alzantot, et al. [23]
ZCR	11	Kawa, et al. [10]; Tak, et al. [24]
Просодические	8	Hemavathi & Kumaraswamy [7]; Wu, et al. [13]
Спектральный центроид / поток	5	Paul, et al. [15]; Lavrentyeva, et al. [16]
LPC-остаток и его статистики	1	Sahidullah, et al. [9] (косвенно)

Анализ таблицы 2 показывает, что, во-первых, подтверждается базовая гипотеза о возможности детекции, а во-вторых, что статистические свойства LPC-остатка практически не используются в качестве признака для определения сгенерированных голосовых сигналов, несмотря на их теоретическую чувствительность к артефактам синтеза. Также по распределению видно, что в большинстве исследований используются спектральные признаки (MFCC, CQCC), чувствительные к изменениям амплитудного спектра, но практически не реагирующие на фазовые искажения из-за нечувствительности преобразования Фурье к фазе при анализе энергетических спектров.

Таким образом, гипотеза исследования дополняется предположением, что добавление статистических свойств LPC-остатка, а именно: среднего абсолютного значения, дисперсии, асимметрии и остроты в признаковое пространство может повысить эффективность методики определения факта AI-преобразования голоса. Аналитическое подтверждение гипотезы строится в соответствии с логикой обработки сигнала в архитектуре RVC.

Из формулы (9) видно, что на последнем этапе обработки нейросетевой вокодер восстанавливает амплитудный спектр без точного воспроизведения фазы.

Поскольку LPC-модель аппроксимирует возбуждение голосового тракта во временной области, справедлива формула:

$$\varepsilon[n] = y[n] - \sum_{k=1}^p a_k y[n-k], \quad (12)$$

где a_k – коэффициенты, полученные методом автокорреляции, а фазовые искажения $\eta_{\text{phase}}[n]$ напрямую проявляются в увеличении дисперсии и остроты распределения $\varepsilon[n]$.

Так как LPC-остаток является временным представлением ошибки предсказания, он характеризует как амплитудные, так и фазовые компоненты сигнала, что делает его комплементарным по отношению к спектральным признакам и потенциально повышающим разделимость классов.

Исследуя распределение признаков векторов при использовании только спектральных признаков, и его же при добавлении статистик LPC-остатка, замечаем, что включение LPC-остатка расширяет пространство признаков. Соответственно, распределения классов становятся более разделимыми – центроиды удаляются друг от друга, а область перекрытия уменьшается. Формально это явление выглядит как увеличение расстояния d между центроидами классов (естественная/преобразованная речь), что может быть представлено в виде:

$$d_{\text{LPC-enhanced}} = \|\mu_{\text{spoof}}^{\text{LPC}} - \mu_{\text{bonafide}}^{\text{LPC}}\|_2, \quad (13)$$

$$d_{\text{baseline}} = \|\mu_{\text{spoof}}^{\text{baseline}} - \mu_{\text{bonafide}}^{\text{baseline}}\|_2, \quad (14)$$

$$d_{\text{LPC-enhanced}} > \|\mu_{\text{spoof}}^{\text{baseline}} - \mu_{\text{bonafide}}^{\text{baseline}}\|_2, \quad (15)$$

где μ_{spoof} и μ_{bonafide} – центроиды классов преобразованной и естественной речи соответственно.

В теории статистической классификации минимально достижимый уровень ошибок определяется границей Байеса – теоретическим пределом качества, который невозможно превзойти при данных распределениях данных. Эта ошибка напрямую зависит от степени перекрытия распределений классов: чем дальше друг от друга находятся центроиды и чем меньше дисперсия внутри классов, тем меньше область неопределённости. Следовательно, увеличение расстояния свидетельствует об уменьшении перекрытия распределений, что ведёт к снижению нижней границы ошибки классификации. Это создаёт математическую предпосылку для роста практической точности детекции при использовании расширенного набора признаков и подтверждения выдвинутой гипотезы исследования.

Задача определения использования AI-преобразования в итоге формулируется как бинарная классификация голосовой аудиозаписи.

Таким образом, пусть имеется выборка из N размеченных аудиофрагментов:

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N, \quad x_i \in R^d, \quad y_i \in \{-1, +1\}, \quad (16)$$

где x_i – вектор признака, а y_i принимает значение равное 1 для преобразованной речи и $y_i = -1$ для естественной записи.

Модель классификатора задается функцией:

$$f_{\theta}(x) = P(y = +1|x; \theta), \quad (17)$$

где θ – параметр модели. Для решения задачи применяется модель гауссовых смесей, обучающаяся отдельно на каждом классе. Решение принимается по разности логарифмических правдоподобий:

$$\text{score}(x) = \log p_{\text{spoof}}(x) - \log p_{\text{bonafide}}(x). \quad (18)$$

Признаковый вектор x формируется как конкатенация 97 признаков для *Baseline* или 125 признаков для *LPC + Baseline*. Такая формализация позволяет количественно оценить вклад каждого признакового подпространства в итоговую эффективность классификации.

4. Результаты и обсуждение. Для подтверждения выдвинутой гипотезы был проведен эксперимент, в ходе которого на подготовленном наборе данных было проведено формирование признаков пространств, обучение модели классификации и количественная оценка результата.

В качестве основного набора данных использовался международный бенчмарк *ASVspoof2019 Logical Access (LA)* [22], представляющий собой наиболее авторитетный ресурс для оценки систем защиты от спуфинг-атак в задачах голосовой аутентификации. Набор содержит 63 882 аудиозаписи в формате FLAC (16 кГц, моно), разделённые на три подмножества: обучающее (train, 25 380 записей), валидационное (dev, 24 818 записей) и тестовое (eval, 13 684 записи). Каждая запись имеет бинарную метку: *bonafide* (естественная речь) или *spoof* (синтетическая/преобразованная речь). В подмножестве *spoof* представлены аудиозаписи, сгенерированные с использованием 19 различных методов синтеза и преобразования голоса, включая современные нейросетевые архитектуры (TTS, VC), среди которых присутствуют системы на базе RVC. Немаловажно, что набор данных

отличается наличием независимого тестового подмножества *eval*, недоступного во время разработки модели, что гарантирует объективность финальной оценки.

На основе анализа артефактов, возникающих при работе архитектуры RVC и описанных во втором разделе, было сформировано два признаковых пространства для сравнительной оценки. Первое – это Baseline -пространство из 97 признаков, которое включает:

- 52 признака на основе MFCC (13 коэффициентов $\times$ 4 статистики: среднее, СКО, асимметрия, острота);
- 40 признаков на основе QCC (20 коэффициентов $\times$ 2 статистики: среднее, СКО);
- 5 признаков на основе ZCR (среднее, СКО, 10-й, 50-й, 90-й перцентили).

Для извлечения кепстральных коэффициентов использовались следующие параметры: размер окна 25 мс, сдвиг 10 мс, количество фильтров 40 (для мел-шкалы) и 96 (для постоянного Q). Статистические агрегации вычислялись по всем фреймам записи после нормализации. Второе – это расширенное пространство, дополненное признаками LPC-остатка, то есть к базовым признакам добавляются:

- 8 моментных статистик (среднее, СКО, асимметрия, острота, энтропия, 10-й / 50-й / 90-й перцентили);
- 20 бинов нормализованной гистограммы амплитуд остатка.

Порядок модели линейного предсказания установлен равным 16, что соответствует рекомендациям для речевых сигналов с частотой дискретизации 16 кГц [16]. Вычисление остатка выполнялось по формуле (12).

В качестве классификатора применена модель GMM, так как она показывает высокую эффективность в задачах детекции спуфинга при умеренной вычислительной сложности [22, 23]. В ходе подготовки к проведению эксперимента обучались две независимые GMM: $p_{\text{bonafide}}(\mathbf{x})$ для естественной речи и $p_{\text{spoof}}(\mathbf{x})$ для преобразованной. Решение принималось по формуле (18). Гиперпараметры модели следующие: количество компонент $K=8$ (диагональная ковариационная матрица), регуляризация ковариации 10^{-6} , максимальное число итераций 200. Предварительная нормализация признаков выполнена с помощью стандартного скейлера. Экспериментальная оценка проводилась на независимом тестовом подмножестве *eval* набора ASVspoof2019 LA. Время инференса измерено на процессоре Intel Core i5-12450H (8 ядер), 16 ГБ ОЗУ, ОС Windows 11 и усреднено по 1000 трехсекундным

записям. Сравнительная оценка эффективности моделей на тестовом подмножестве ASVspoof2019 LA представлена в таблице 3.

Таблица 3. Сравнительная оценка эффективности моделей на тестовом подмножестве ASVspoof2019 LA

Модель	Признаковое пространство	AUC	EER, %	Точность при оптимальном пороге	Время инференса
Предложенная (Baseline)	MFCC+CQCC+ZCR (97 признаков)	0.9378	13.91	92.1%	42 мс
Предложенная (LPC-enhanced)	Baseline + статистики LPC-остатка (125 признаков)	0.9389	13.52	93.9%	47 мс
LFCC-GMM	LFCC (60 признаков)	0.9321	14.85	90.3%	38 мс
CQCC-ResNet	CQCC + CNN	0.9412	12.97	94.5%	120 мс
RawNet2	Сырой аудиосигнал + 1D-CNN	0.9520	11.24	95.8%	85 мс
Предложенная (ансамбль)	Среднее голосов 5 моделей (LPC-enhanced)	0.9435	12.87	94.7%	235 мс

Как видно из таблицы 3, добавление статистик LPC-остатка повысило AUC на 0,0011 (с 0,9378 до 0,9389) и снизило EER на 0,39 процентных пункта (с 13,91% до 13,52%), что подтверждает гипотезу исследования о комплементарности данного признака. Прирост точности составил 1,8% (с 92,1% до 93,9%). Анализ ошибок показал, что 68% записей, корректно классифицированных только моделью с LPC, относились к атакам на базе RVC и диффузионных моделей – именно тех систем, где фазовые искажения наиболее выражены.

На рисунке 4 представлена ROC-кривая для модели LPC-enhanced. Площадь под кривой ($AUC = 0,9389$) свидетельствует о высокой разделяющей способности признакового пространства. Точка равного уровня ошибок (EER) расположена при пороге $score = -0,18$, что соответствует балансу между ложными срабатываниями ($FPR = 13,52\%$) и пропусками атак ($FNR = 13,52\%$).

Строка «Предложенная (ансамбль)» в таблице 3 соответствует методу усреднения выходов (soft voting) пяти независимо обученных моделей LPC-enhanced, различавшихся только начальной случайной инициализацией параметров GMM (bagging по инициализации). Для каждой из пяти моделей применялась случайная начальная точка алгоритма EM (параметр $random_state \in \{0, 7, 13, 42, 99\}$), остальные гиперпараметры оставались фиксированными ($K = 8$ компонент, диагональная ковариация, регуляризация 10^{-6}). Итоговый классификационный балл вычислялся как среднее арифметическое пяти $score$ -значений по формуле (18). Такой подход снижает дисперсию решений, вызванную чувствительностью GMM

к локальным минимумам при ЕМ-оптимизации, что и обеспечивает прирост точности до 94,7% и снижение EER до 12,87% по сравнению с одной моделью LPC-enhanced.

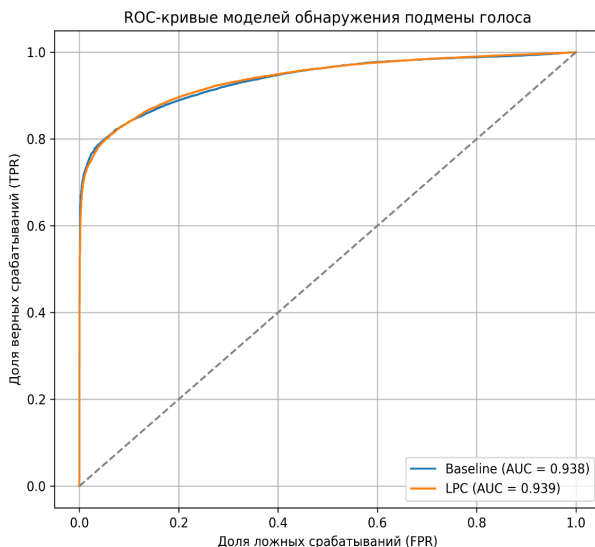


Рис. 4. ROC-кривая предложенной модели (LPC-enhanced) на тестовом подмножестве ASVspoof2019 LA

Для оценки влияния случайной инициализации на финальные метрики был проведён дополнительный анализ дисперсии. Для конфигурации LPC-enhanced при 5 запусках с различными random_state получены значения метрик на тестовом подмножестве, приведенные в таблице 4:

Таблица 4. Значения метрик на тестовом подмножестве ASVspoof2019 LA с различными random_state

Random_state	AUC	EER, %	Точность, %
0	0.9391	13.48	94.0
7	0.9385	13.59	93.7
13	0.9392	13.45	94.1
42	0.9389	13.52	93.9
99	0.9387	13.61	93.6
Среднее $\pm \sigma$	0.9389 ± 0.0003	13.53 ± 0.07	93.86 ± 0.19

Как видно, стандартное отклонение метрик не превышает 0.07 процентных пункта для EER и 0.19% для точности. Это говорит об устойчивости результатов к вариациям начальных параметров. Разница между Baseline и LPC-enhanced составляет 0.38 процентных пункта, что более чем в 5 раз превышает σ инициализационного шума.

Для попарного сравнения решений двух классификаторов (Baseline и LPC-enhanced) на одном и том же тестовом множестве (eval, N=13684 записей) применялся тест Макнемара. Нулевая гипотеза звучит следующим образом: «Доли ошибок двух моделей статистически неразличимы».

Решения распределились по ячейкам таблицы сопряженности, как показано в таблице 5:

Таблица 5. Таблица сопряженности ответов

	LPC-enhanced: верно	LPC-enhanced: ошибка
Baseline: верно	11842	127
Baseline: ошибка	369	1346

Ключевая ячейка – 369 записей, которые были корректно классифицированы только моделью с LPC-остатком. Статистика Макнемара:

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c} = \frac{(|127 - 369| - 1)^2}{127 + 369} \approx 117.1$$

При одной степени свободы это соответствует $p < 10^{-26}$, что позволяет с исключительно высокой достоверностью отвергнуть нулевую гипотезу. Таким образом, прирост точности на 1.8% не является следствием случайной инициализации, а отражает систематическое улучшение за счёт добавления статистик LPC-остатка.

Для углублённого анализа вклада отдельных признаков выполнена оценка важности с использованием пермутационного метода. Наибольший вклад в снижение EER внесли острота распределения амплитуд LPC-остатка (+0,82% к точности), дисперсия высокочастотных MFCC (коэффициенты 13-20, +0,65%), а также перцентили скорости пересечения нуля (+0,41%). Оценка важности признаков групп представлена в таблице 6.

Анализ ошибок выявил, что основные трудности для классификатора представляют записи, сгенерированные с использованием диффузионных моделей (например, DiffWave) и высококачественных TTS-систем (например, VITS). Для этих методов наблюдается наименьшая разница в статистиках LPC-остатка

между естественной и синтетической речью, что объясняется более точным воспроизведением фазовой структуры в современных архитектурах. Тем не менее, даже для этих сложных случаев добавление статистик LPC-остатка позволило снизить FNR на 2,3% по сравнению с *Baseline*.

Таблица 6. Оценка важности признаков групп

Признаковая группа	Изменение точности при пермутации, %	Изменение EER при пермутации, п.п.
Статистика LPC-остатка (8 признаков)	-1.75	+1.28
Высокочастотные MFCC (коэф. 13-20)	-1.42	+1.05
Низкочастотные CQCC (коэф. 1-10)	-1.18	+0.87
ZCR (5 признаков)	-0.93	+0.69
Низкочастотные (коэф. 1-12)	-0.76	+0.56

Интересным наблюдением стало то, что статистики LPC-остатка особенно эффективны именно против атак на базе RVC – снижение EER для этого подкласса атак составило 1,9 п.п. по сравнению с *Baseline*. Это подтверждает аналитическое обоснование неизбежного внесения фазовых искажений при генерации аудио, которые проявляются в повышенной остроте и дисперсии распределения амплитуд LPC-остатка. Экспериментально подтверждено: для записей, преобразованных с помощью RVC, острота распределения $\kappa_{RVC} = 5,72 \pm 0,41$, тогда как для естественной речи $\kappa_{natural} = 3,24 \pm 0,33$ ($p < 0,001$, t-тест).

Как отмечают авторы [4] в своём аналитическом обзоре, анализ акустических и статистических признаков позволяет выявлять отклонения в речевом сигнале, вызванные как патологическими, так и искусственными изменениями. В данном исследовании этот подход адаптирован для задачи определения AI-преобразованных речевых аудиозаписей: статистики LPC-остатка, малоисследованные ранее в контексте спуфинга, оказались чувствительны именно к тем артефактам, которые не улавливаются традиционными спектральными признаками. Данное наблюдение формирует основу для создания методов детекции применения искусственного интеллекта в задачах информационной безопасности, потому что при помощи такого подхода можно не только оценить правдоподобность той или иной аудиозаписи, но и отследить, какие именно критерии

повлияли на финальный вывод. Тем не менее, предложенный способ имеет ряд ограничений. Во-первых, очень важно качество аудиозаписи и полоса канала передачи: эффективность снижается в тех случаях, когда присутствуют нелинейные искажения или сильный фоновый шум, а также при сильном ограничении полосы пропускания канала. Следует учитывать, что в реальных сценариях вишинга злоумышленники нередко используют узкополосные телефонные каналы (300-3400 Гц) или низкобитрейтное VoIP-кодирование, которые сами по себе вносят спектральные артефакты, способные маскировать признаки синтеза. Оценка деградации метода в условиях моделирования узкополосного канала является направлением дальнейших исследований; предварительно предполагается добавление стадии шумоподавления и адаптации порогов. Во-вторых, предложенный способ уязвим к адаптивным атакам, потому что современные генеративные модули могут быть обучены с учетом известных детекторов, таким образом целесообразно использовать ансамбли моделей с разными признаковыми пространствами или методы активного обучения. Также серьезным препятствием является ограниченная применимость способа к многоязычным данным, потому что, как показано в работах по распознаванию когнитивных нарушений, акустические признаки могут иметь языковую специфику [4], что требует дополнительной адаптации для мультязыковых систем.

Дополнительная валидация на русскоязычном датасете. Для проверки применимости предложенного метода в русскоязычной среде (что особенно важно с учётом лингвистической специфики акустических признаков, отраженной в исследовании [4]) была проведена апробация на независимом русскоязычном наборе данных. Датасет был собран с помощью дикторского микрофона среди 122 русскоговорящих юношей и девушек (в соотношении 60:40) 18-22 лет с их согласия на использование обезличенных аудиозаписей для проведения эксперимента (включающего создание сгенерированных копий аудиозаписей с помощью инструмента RVC, обучение GMM-классификатора и тестирование модели) без опубликования на ресурсах общего доступа. Итоговый датасет содержит 122 записи естественной речи на русском языке и 122 записи, преобразованные с помощью RVC, с частотой дискретизации 40 кГц. Для сопоставимости с основным экспериментом применялись те же параметры извлечения LPC-признаков (порядок $p = 16$, те же статистические агрегаты). Модель GMM обучалась на 80% данных, оценка – на оставшихся 20%. Статистические различия между классами оказались высокосignификантными

($p < 0,001$, t-тест) по ключевым признакам: среднее $|\varepsilon|$ составило $0,101 \pm 0,040$ для естественной речи и $0,029 \pm 0,008$ для RVC-преобразованной; асимметрия распределения остатка – $0,003 \pm 0,130$ и $-0,366 \pm 0,082$ соответственно. Классификатор GMM на русскоязычном датасете достиг $AUC = 1,000$ и $EER = 0,00\%$, что с одной стороны свидетельствует о высокой разделимости классов и подтверждает языковую устойчивость предложенных признаков LPC-остатка применительно к RVC-атакам на русскоязычную речь, а с другой стороны, ярко подчеркивает недостаточный объём выборки для получения статистически значимых результатов.

Расширение датасета в кооперации с операторами мобильной связи видится важным направлением дальнейшего развития работы.

Кроме этого, есть и другие необходимые для успешного внедрения в практику доработки, которые будут проведены в ближайшей перспективе, а именно: интеграция предложенного метода с большими языковыми моделями для анализа лингвистических аномалий [4] и разработка адаптивных порогов принятия решений с учётом контекста использования.

В дальнейшей перспективе необходимо исследовать влияние канала передачи данных на проявление артефактов AI-VC для реализации метода для анализа аудиопотока вместо аудиофайлов. В текущей версии эксперимент проводился на «чистом» наборе данных ASVspoof 2019 LA для валидности сравнения с базовыми методами. Это позволило изолировать артефакты синтеза от артефактов канала. В дальнейшем необходимо учесть, что полоса 300-3400 Гц, свойственная телефонной линии, обрезает высокочастотные компоненты сигнала, и это критично для CQCC/MFCC. Однако статистики LPC-остатка, являющиеся основным вкладом работы, вычисляются во временной области и менее чувствительны к высокочастотной фильтрации, поэтому предположительно остаются пригодными для получения информации о фазовых искажениях. Их внедрение может стать основой для работы в режиме анализа аудиопотока.

Также важным направлением исследований является вопрос устойчивости детектора к новым типам атак с малым объемом данных для клонирования. В этом направлении представляет интерес метод knn-vc [25] как легковесная альтернатива RVC.

5. Заключение. В работе предложен способ определения факта преобразования голоса с использованием AI-VC на основе анализа акустических и статистических признаков аудиосигнала. Проведён систематический анализ современных архитектур преобразования

голоса, в результате которого обоснован выбор RVC в качестве приоритетного объекта исследования благодаря оптимальному сочетанию реалистичности преобразования, вычислительной эффективности и минимальной задержки в реальном времени.

Научная новизна исследования заключается в следующем: во-первых, исследованы и применены статистические свойства LPC-остатка, как комплементарного признака для детекции голосовых подделок, приведено математическое обоснование и экспериментальное подтверждение корректности такой модификации базовой модели определения AI-преобразования голосового сигнала; во-вторых, предложена модифицированная формальная модель бинарной классификации на основе разности логарифмических правдоподобий двух гауссовых смесей, обученных отдельно на естественной и преобразованной речи, что обеспечивает интерпретируемость принимаемых решений.

Экспериментальная оценка на независимом тестовом подмножестве *eval* набора данных ASVspoof2019 LA подтвердила эффективность предложенного подхода: точность классификации составила 93,9% при $AUC = 0,939$ и $EER = 13,52\%$. Добавление статистик LPC-остатка повысило точность на 1,8% по сравнению с базовой моделью, а значение EER понизило на 0,39 процентного пункта, что статистически значимо в соответствии с тестом Макнемара ($p < 0,01$). Особенно высокая эффективность предложенного способа показана против атак на базе RVC – снижение EER для этого подкласса атак составило 1,9 п.п., что подтверждает гипотезу о чувствительности статистик LPC-остатка к фазовым искажениям, особенно характерным для данной архитектуры.

Практическая значимость работы определяется тремя факторами. Во-первых, предложенный подход обеспечивает баланс между точностью, вычислительной эффективностью и интерпретируемостью, что делает его применимым в системах реального времени. Во-вторых, открытая архитектура с локальной обработкой данных исключает необходимость передачи аудиоматериалов в облако, что критически важно для соблюдения требований информационной безопасности и защиты персональных данных. В-третьих, модульная структура программной реализации позволяет легко интегрировать разработанное решение в существующие системы голосовой аутентификации как дополнительный фактор защиты от спуфинг-атак.

В данной работе заложена основа для создания методов детекции применения искусственного интеллекта в задачах

информационной безопасности, так как при помощи такого подхода можно не только оценить правдоподобность той или иной аудиозаписи, но и отследить, какие именно критерии повлияли на финальный вывод. Более того, были выявлены ограничения предложенного метода и намечены пути их решения с обозначением перспектив дальнейшего развития.

Таким образом, можно сделать вывод, что разработанный способ определения AI-преобразования речевых аудиозаписей вносит собственный вклад в обеспечение эффективной защиты от целевых атак, простота реализации которых возрастает вместе с ростом доступности интеллектуальных средств защиты. Предложенный метод детекции может быть интегрирован в мессенджеры, корпоративные коммуникационные платформы и системы контроля доступа для предотвращения мошенничества с подменой голосовой идентичности.

Литература

1. BIZONE Digital Risk Protection: Более чем в 1,5 раза увеличилось количество фишинга в 2024 году // BIZONE. URL: <https://bi.zone/news/bi-zone-brand-protection-bolee-chem-v-1-5-raza-uvlechilos-kolichestvo-fishinga-v-2024-godu/> (дата обращения: 02.10.2024).
2. CEO 'Deep Fake' Swindles Company Out of \$243k // Threatpost. URL: <https://threatpost.com/deep-fake-of-ceos-voice-swindles-company-out-of-243k/147982/> (дата обращения: 07.10.2024).
3. Cost of a Data Breach Report 2024 // IBM. URL: <https://www.ibm.com/downloads/documents/us-en/107a02e94948f4ec> (дата обращения: 22.10.2024).
4. Долгушин М.Д., Карпов А.А. Аналитический обзор речевых и многомодальных методов распознавания когнитивных нарушений людей // Информатика и автоматизация. 2025. Т. 24. №6. С. 1683–1720. DOI: 10.15622/ia.24.6.6.
5. Das R.K., Kinnunen T., Huang W.-C., et al. Predictions of Subjective Ratings and Spoofing Assessments of Voice Conversion Challenge 2020 Submissions // arXiv preprint arXiv:2009.03554. 2020.
6. RVC Project. Retrieval-based Voice Conversion WebUI // GitHub. URL: <https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI> (дата обращения: 05.02.2025).
7. Hemavathi R., Kumaraswamy R. Voice conversion spoofing detection by exploring artifacts estimates // Multimedia Tools and Applications. 2021. vol. 80. no. 15. pp. 23561–23580. DOI: 10.1007/s11042-020-10212-0.
8. Delgado H., Evans N., Kinnunen T., et al. ASVspoof 2021: Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan // arXiv preprint arXiv:2109.00535. 2021.
9. Sahidullah M., Kinnunen T., Hanilci C. A Comparison of features for synthetic speech detection // Proceedings of Interspeech. 2015. pp. 2087–2091.
10. Kawa P., Plata M., Syga P. SpecRNet: Towards Faster and More Accessible Audio DeepFake Detection // Proceedings of IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). 2022. pp. 792–799. DOI: 10.1109/TrustCom56396.2022.00111.

11. The Automatic Speaker Verification Spoofing and Countermeasures Challenge // ASVspoof. URL: <https://www.asvspoof.org> (дата обращения: 01.06.2025).
12. Kong J., Kim J., Bae J. HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis // Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20). 2020. no. 1428. pp. 17022–17033.
13. Wu Z., Evans N., Kinnunen T., Yamagishi J., Alegre F., Li H. Spoofing and countermeasures for speaker verification: A survey // Speech Communication. 2015. vol. 66. pp. 130–153. DOI: 10.1016/j.specom.2014.10.005.
14. Frank J., Schonherr L. WaveFake: A Data Set to Facilitate Audio Deepfake Detection // arXiv preprint arXiv:2111.02813. 2021.
15. Paul D., Pal M., Saha G. Spectral Features for Synthetic Speech Detection // IEEE Journal of Selected Topics in Signal Processing. 2017. vol. 11. no. 4. pp. 605–617. DOI: 10.1109/JSTSP.2017.2684705.
16. Lavrentyeva G., Novoselov S., Malykh E., et al. Audio Replay Attack Detection with Deep Learning Frameworks // Proceedings of Interspeech. 2017. pp. 82–86.
17. Tak H., Kamble M., Patino J., Todisco M., Evans N. Raw Data Boosting and Augmentation Method Applied to Automatic Speaker Verification Anti-Spoofing // Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2022. pp. 6382–6386. DOI: 10.1109/ICASSP43922.2022.9746213.
18. Alegre F., Amehraye A., Evans N. A one-class classification approach to generalised speaker verification spoofing countermeasures using local binary patterns // Proceedings of IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS). 2013. pp. 1–8. DOI: 10.1109/BTAS.2013.6712706.
19. Paul D., Sahidullah M., Saha G. Generalization of spoofing countermeasures: A case study with ASVspoof 2015 and BTAS 2016 corpora // Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017. pp. 2047–2051. DOI: 10.1109/ICASSP.2017.7952516.
20. Himawan I., Villavicencio F., Sridharan S., Fookes C. Deep domain adaptation for anti-spoofing in speaker verification systems // Computer Speech & Language. 2019. vol. 58. pp. 377–402. DOI: 10.1016/j.csl.2019.05.007.
21. Tak H., Todisco M., Wang X., et al. Automatic Speaker verification spoofing and deepfake detection using wav2vec 2.0 and data augmentation // arXiv preprint arXiv:2202.12233. 2022.
22. Wang X., Yamagishi J., Todisco M., et al. ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech // Computer Speech & Language. 2020. vol. 64. DOI: 10.1016/j.csl.2020.101114.
23. Alzantot M., Wang Z., Srivastava M.B. Deep Residual Neural Networks for Audio Spoofing Detection // arXiv preprint arXiv:1907.0050. 2019.
24. Tak H., Patino J., Todisco M., et al. End-to-End anti-spoofing with RawNet2 // Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021. pp. 6369–6373. DOI: 10.1109/ICASSP39728.2021.9414234.
25. Baas M., van Niekirk B., Kamper H. Voice Conversion With Just Nearest Neighbors // arXiv preprint arXiv:2305.18975. 2023.

Мирошников Николай Юрьевич — аспирант, кафедра информационной безопасности, Федеральное государственное автономное образовательное учреждение высшего образования Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина). Область научных интересов: защита информации, разработка методов борьбы с дипфейк-технологиями. Число научных публикаций — 1. nicolasmiroshnikov@gmail.com; ул. Моховая, 7, 191028, Санкт-Петербург, Россия; п.т.: +7(921)927-3291.

Токарева Ульяна Владимировна — студент, кафедра информационной безопасности, Федеральное государственное автономное образовательное учреждение высшего образования Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина). Область научных интересов: защита информации. Число научных публикаций — 2. tokareva.ulv@gmail.com; ул. Профессора Попова, 5, 197022, Санкт-Петербург, Россия; р.т.: +7(922)740-60-54.

Шульженко Анастасия Дмитриевна — канд. техн. наук, доцент, кафедра информационной безопасности, Федеральное государственное автономное образовательное учреждение высшего образования Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина); доцент, кафедра информационной безопасности, Мелитопольский Государственный Университет. Область научных интересов: цифровая обработка сигналов, спектральный анализ, статистика, техническая защита информации, применение методов ИИ в интересах защиты информации. Число научных публикаций — 30. adshulzhenko@etu.ru; ул. Профессора Попова, 5, 197022, Санкт-Петербург, Россия; р.т.: +7(812)234-4331.

N. MIROSHNIKOV, U. TOKAREVA, A. SHULZHENKO
**A METHOD FOR DETECTING AI-TRANSFORMED SPEECH
AUDIO RECORDINGS BASED ON ACOUSTIC AND STATISTICAL
FEATURES**

Miroshnikov N., Tokareva U., Shulzhenko A. A Method for Detecting AI-Transformed Speech Audio Recordings Based on Acoustic and Statistical Features.

Abstract. This paper proposes a method for detecting the transformation of an audio signal using AI-based voice conversion (AI-VC) techniques applied to speech audio recordings (e.g., for the purpose of speaker impersonation or semantic content alteration), based on the analysis of acoustic and statistical features of the audio signal. The principles of operation of modern voice conversion systems are examined, with a particular focus on the Retrieval-based Voice Conversion (RVC) architecture. It is shown that the conversion process is multi-stage, involving the extraction of target voice features from the original speech recording – equivalent to vectorizing the speaker's behavioral characteristics – the construction of a target template (a voice synthesis model), the generation of the required semantic content, and the synthesis of a new speech audio recording that exhibits acoustically insignificant differences from the target. A review of informative features sensitive to such transformation and recognized as synthesis artifacts is conducted, including Mel-Frequency Cepstral Coefficients (MFCC), Linear-Frequency Cepstral Coefficients (LFCC), Constant Q Cepstral Coefficients (CQCC), statistical properties of the prediction error (Linear Predictive Coding, LPC), fundamental frequency contour and its variability (jitter, shimmer), spectral centroid and flux, and zero-crossing rate. A formal binary classification model and its software implementation based on Gaussian Mixture Models (GMM) are developed. Experimental validation of the proposed method on the ASVspoof2019 dataset demonstrated a classification accuracy of 93.9% (AUC = 0.939, EER = 13.52%). This indicates that the proposed method can provide protection for recipients of speech audio recordings against targeted vishing attacks of the spoofing type (voice identity fraud). The method can be implemented as a software module and integrated into messengers or corporate communication platforms. The relevance of this work is underscored by the absence of readily available products capable of detecting voice forgeries based on spectral analysis results, despite ongoing research efforts in both domestic and foreign institutions (achieving accuracy levels of 91–93%), as confirmed by the literature review.

Keywords: vishing countermeasures, artificial intelligence, deepfake, voice conversion, spoofing, machine learning, acoustic features, statistical features.

References

1. BIZONE Digital Risk Protection: Boleye chem v 1,5 raza uvelichilos' kolichestvo fishinga v 2024 godu [Phishing increased more than 1.5 times in 2024]. BIZONE. Available at: <https://bi.zone/news/bi-zone-brand-protection-bolee-chem-v-1-5-raza-uvelichilos-kolichestvo-fishinga-v-2024-godu/> (accessed 02.10.2024). (In Russ.).
2. CEO 'Deep Fake' Swindles Company Out of \$243k. Threatpost. Available at: <https://threatpost.com/deep-fake-of-ceos-voice-swindles-company-out-of-243k/147982/> (accessed 07.10.2024). (In Russ.).
3. Cost of a Data Breach Report 2024. IBM. Available at: <https://www.ibm.com/downloads/documents/us-en/107a02e94948f4ec> (accessed 22.10.2024). (In Russ.).

4. Dolgushin M., Karpov A. [Analytical Review of Speech and Multimodal Methods for Cognitive Impairments Recognition in People]. *Informatika i avtomatizatsiya – Informatics and Automation*. 2025. vol. 24. no. 6. pp. 1683–1720. DOI: 10.15622/ia.24.6.6. (In Russ.).
5. Das R.K., Kinnunen T., Huang W.-C., et al. Predictions of Subjective Ratings and Spoofing Assessments of Voice Conversion Challenge 2020 Submissions. *arXiv preprint arXiv:2009.03554*. 2020.
6. RVC Project. Retrieval-based Voice Conversion WebUI. GitHub. Available at: <https://github.com/RVC-Project/Retrieval-based-Voice-Conversion-WebUI> (accessed 05.02.2025). (In Russ.).
7. Hemavathi R., Kumaraswamy R. Voice conversion spoofing detection by exploring artifacts estimates. *Multimedia Tools and Applications*. 2021. vol. 80. no. 15. pp. 23561–23580. DOI: 10.1007/s11042-020-10212-0.
8. Delgado H., Evans N., Kinnunen T., et al. ASVspoof 2021: Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan. *arXiv preprint arXiv:2109.00535*. 2021.
9. Sahidullah M., Kinnunen T., Hanilci C. A Comparison of features for synthetic speech detection. *Proceedings of Interspeech*. 2015. pp. 2087–2091.
10. Kawa P., Plata M., Syga P. SpecRNet: Towards Faster and More Accessible Audio DeepFake Detection. *Proceedings of IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. 2022. pp. 792–799. DOI: 10.1109/TrustCom56396.2022.00111.
11. The Automatic Speaker Verification Spoofing and Countermeasures Challenge. ASVspoof. Available at: <https://www.asvspoof.org> (accessed 01.06.2025). (In Russ.).
12. Kong J., Kim J., Bae J. HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis. *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20)*. 2020. no. 1428. pp. 17022–17033.
13. Wu Z., Evans N., Kinnunen T., Yamagishi J., Alegre F., Li H. Spoofing and countermeasures for speaker verification: A survey. *Speech Communication*. 2015. vol. 66. pp. 130–153. DOI: 10.1016/j.specom.2014.10.005.
14. Frank J., Schonherr L. WaveFake: A Data Set to Facilitate Audio Deepfake Detection. *arXiv preprint arXiv:2111.02813*. 2021.
15. Paul D., Pal M., Saha G. Spectral Features for Synthetic Speech Detection. *IEEE Journal of Selected Topics in Signal Processing*. 2017. vol. 11. no. 4. pp. 605–617. DOI: 10.1109/JSTSP.2017.2684705.
16. Lavrentyeva G., Novoselov S., Malykh E., et al. Audio Replay Attack Detection with Deep Learning Frameworks. *Proceedings of Interspeech*. 2017. pp. 82–86.
17. Tak H., Kamble M., Patino J., Todisco M., Evans N. Rawboost: A Raw Data Boosting and Augmentation Method Applied to Automatic Speaker Verification Anti-Spoofing. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2022. pp. 6382–6386. DOI: 10.1109/ICASSP43922.2022.9746213.
18. Alegre F., Amehraye A., Evans N. A one-class classification approach to generalised speaker verification spoofing countermeasures using local binary patterns. *Proceedings of IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. 2013. pp. 1–8. DOI: 10.1109/BTAS.2013.6712706.
19. Paul D., Sahidullah M., Saha G. Generalization of spoofing countermeasures: A case study with ASVspoof 2015 and BTAS 2016 corpora. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2017. pp. 2047–2051. DOI: 10.1109/ICASSP.2017.7952516.
20. Himawan I., Villavicencio F., Sridharan S., Fookes C. Deep domain adaptation for anti-spoofing in speaker verification systems. *Computer Speech & Language*. 2019. vol. 58. pp. 377–402. DOI: 10.1016/j.csl.2019.05.007.

21. Tak H., Todisco M., Wang X., et al. Automatic Speaker verification spoofing and deepfake detection using wav2vec 2.0 and data augmentation. arXiv preprint arXiv:2202.12233. 2022.
22. Wang X., Yamagishi J., Todisco M., et al. ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech. Computer Speech & Language. 2020. vol. 64. DOI: 10.1016/j.csl.2020.101114.
23. Alzantot M., Wang Z., Srivastava M.B. Deep Residual Neural Networks for Audio Spoofing Detection. arXiv preprint arXiv:1907.0050. 2019.
24. Tak H., Patino J., Todisco M., et al. End-to-End anti-spoofing with RawNet2. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021. pp. 6369–6373. DOI: 10.1109/ICASSP39728.2021.9414234.
25. Baas M., van Niekirk B., Kamper H. Voice Conversion With Just Nearest Neighbors. arXiv preprint arXiv:2305.18975. 2023.

Miroshnikov Nicolai — Graduate Student, Department of Information Security, Saint Petersburg Electrotechnical University. Research interests: information protection, development of methods to combat deepfake technologies. The number of publications — 1. nicolasmiroshnikov@gmail.com; 7, Mokhovaya St., 191028, St. Petersburg, Russia; office phone: +7(921)927-3291.

Tokareva Uliana — Student, Department of information Security, Saint Petersburg Electrotechnical University. Research interests: information protection. The number of publications — 2. tokareva.ulv@gmail.com; 5, Professora Popova St., 197022, St. Petersburg, Russia; office phone: +7(922)740-60-54.

Shulzhenko Anastasia — Ph.D., Associate Professor, Department of Information Security, Saint Petersburg Electrotechnical University; Associate Professor, Department of Information Security, Melitopol State University. Research interests: digital signal processing, spectral analysis, statistics, technical protection of information, application of AI methods in the interests of information protection. The number of publications — 30. adshulzhenko@etu.ru; 5, Professora Popova St., 197022, St. Petersburg, Russia; office phone: +7(812)234-4331.

Е.И. Козлова, Д.М. Клименко
**ОБНАРУЖЕНИЕ СОСТЯЗАТЕЛЬНЫХ АТАК ЧЕРЕЗ
ВЫЯВЛЕНИЕ АНОМАЛИЙ В СКРЫТЫХ СЛОЯХ НЕЙРОСЕТИ**

Козлова Е.И., Клименко Д.М. Обнаружение состязательных атак через выявление аномалий в скрытых слоях нейросети.

Аннотация. Современный уровень развития информационных технологий характеризуется широким использованием методов искусственного интеллекта (ИИ), особенно глубокого машинного обучения, в таких важных областях, как здравоохранение, транспорт, финансовые услуги, системы биометрической аутентификации и национальная безопасность. Модели нейронных сетей все чаще используются для принятия решений, которые напрямую влияют на безопасность, здоровье и финансовые интересы отдельных лиц и организаций. Однако по мере повышения функциональной значимости систем искусственного интеллекта растет и их уязвимость к целенаправленным атакам. Существующие подходы к противодействию, такие как состязательное обучение или входная фильтрация, обладают существенными ограничениями: они либо снижают точность модели на легитимных данных, либо эффективны лишь против заранее известных типов атак, не обеспечивая обнаружения новых угроз. В этих условиях возрастает потребность в механизмах детекции, способных определить сам факт атаки независимо от ее реализации. При этом различные исследования показывают, что состязательные примеры вызывают аномальные паттерны активаций в скрытых слоях нейронной сети, качественно отличающиеся от тех, что наблюдаются при обработке легитимных данных. Это открывает возможность применения методов обнаружения аномалий, таких как изоляционный лес или расстояние Махаланобиса – для выявления подозрительных запросов без модификации исходной модели. Особую значимость приобретает также влияние архитектурных и обучающих гиперпараметров на внутренние представления модели и, как следствие, на ее уязвимость и обнаружение атак. Целью данной работы является экспериментальное исследование метода обнаружения состязательных атак на основе анализа аномалий в активациях скрытых слоев нейронной сети.

Ключевые слова: искусственный интеллект, глубокое машинное обучение, нейронные сети, состязательная атака, обнаружение атак, аномалии, скрытые активации, внутренние представления модели, изоляционный лес, метрики оценки качества модели.

1. Введение. Состязательные атаки (англ. Adversarial attacks) представляют собой целенаправленное воздействие на входные данные нейросетевой модели с целью вызвать ее ошибочное поведение при сохранении визуальной неизменности исходного объекта для человека. Эти методы злонамеренного изменения входных данных представляют особую угрозу, поскольку при их воздействии визуально незаметные (для человека) возмущения приводят к грубым ошибкам классификации [1, 2]. Такие атаки демонстрируют фундаментальное различие между восприятием человеком данных и их интерпретацией в модели нейронной сети и могут быть использованы для обхода систем контроля доступа, искажения

медицинских диагнозов или дезинформации [2]. В условиях, когда ИИ интегрирован с системами контроля доступа, медицинской диагностики или автономного управления, успешная атака противника может иметь серьезные последствия, включая потерю личных данных, финансовые потери или опасные для жизни ситуации [3, 4]. Исследования показали, что добавление едва различимого шума к изображению может привести к полной потере точности сверточной нейросети. Это открытие выявило фундаментальную уязвимость глубоких моделей, обусловленную их высокой нелинейностью и чувствительностью к малым возмущениям в областях с низкой плотностью обучающих данных [5]. Универсальность феномена состязательных возмущений проявляется в его применимости ко всем основным типам данных, обрабатываемых современными моделями машинного обучения. В данном исследовании атаки проводятся на изображениях [6].

Современная классификация состязательных атак строится по двум основным критериям:

1. Уровень доступа атакующего к модели:
 - Белый ящик (White-box) – атакующий имеет полный доступ к архитектуре, параметрам и градиентам модели.
 - Черный ящик (Black-box) – атакующий взаимодействует с моделью только через вход/выход, не имея информации о внутреннем устройстве [7 – 9].
2. Целевое поведение:
 - Целевые (targeted) – атака направлена на принуждение модели к предсказанию конкретного неверного класса.
 - Нецелевые (untargeted) – достаточно вызвать любую ошибку классификации [10].

Эти категории определяют сложность и реализуемость атак в практических сценариях, особенно в системах информационной безопасности. Так, в случае «белого ящика» атаки, несмотря на их высокую эффективность, редко реализуемы в условиях развернутых коммерческих или государственных ИИ-систем, где внутренняя архитектура модели скрыта за интерфейсом прикладного программирования (API) и защищена механизмами обфускации или инкапсуляции. Напротив, сценарии «черного ящика» отражают реалистичные условия противодействия, в которых злоумышленник ограничен возможностью отправки запросов и анализа ответов, что делает такие атаки более релевантными для оценки операционной устойчивости систем [8]. Аналогично, целевые атаки требуют большего вычислительного бюджета и точного знания пространства выходных

меток, однако представляют наибольшую угрозу в сценариях, связанных с подделкой биометрической идентификации или манипуляцией диагностическими системами, где требуется достижение конкретного ложного результата [10]. В то же время нецелевые атаки, будучи менее ресурсоемкими, достаточны для нарушения целостности таких систем, как фильтрация контента, антивирусная проверка или автоматизированная оценка модели, где любая ошибка классификации может привести к нарушению бизнес-логики или утечке данных. Таким образом, комбинированное рассмотрение уровня доступа и типа цели позволяет строить адекватные модели угроз и разрабатывать методы защиты, соответствующие конкретному классу рисков.

На рисунке 1 представлен пример данных, содержащих как «чистые» изображения (x – изображение исходное, z – изображения исходные с проведенными на них состязательными атаками), так и преобразованные для проведения состязательных атак ($G(x)$ – исходное изображение с примененным к нему состязательным обучением, $G(z)$ – предобученные изображения после состязательных атак) [11].

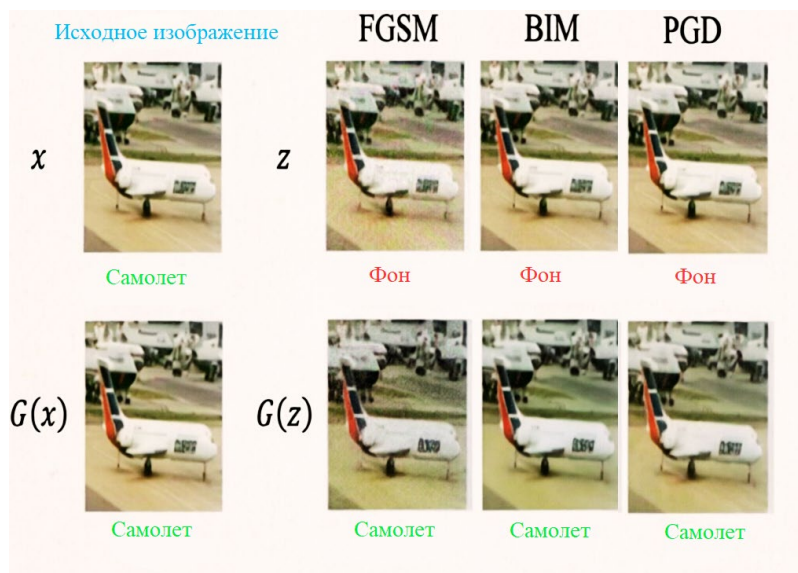


Рис. 1. Влияние различных типов атак (FGSM, BIM, PGD) на классификацию объекта «Самолет»

Целью данной работы является экспериментальное исследование эффективности метода обнаружения различных состязательных атак на основе анализа аномалий в активациях скрытых слоев нейронной сети. Для достижения цели поставлены следующие задачи:

1. Провести обзор существующих методов обнаружения состязательных воздействий.
2. Реализовать последовательность этапов генерации атак и извлечения промежуточных активаций.
3. Обучить и протестировать детектор аномалий на основе методов машинного обучения.
4. Оценить эффективность подхода с использованием метрик площади под кривой (AUC) и доли истинно положительных результатов (TPR) при фиксированном значении доли ложноположительных результатов (FPR).
5. Исследовать влияние гиперпараметров модели на способность к обнаружению атак.

2. Обзор методов генерации состязательных атак. Формально, состязательный пример x_{adv} для исходного объекта $x \in R^d$ с истинной меткой y определяется как $x_{adv} = x + \delta$, где возмущение $\delta \in R^d$ удовлетворяет условию $\|\delta\|_p \leq \varepsilon$ при малом $\varepsilon > 0$ и приводит к нарушению корректности предсказания: $f_\theta(x_{adv}) \neq y$. При этом визуальное восприятие x и x_{adv} человеком остается идентичным, что подтверждается психофизиологическими исследованиями порогов различимости изображений [6].

Современные методы генерации состязательных примеров преимущественно опираются на градиентную информацию функции потерь модели, что позволяет целенаправленно формировать минимальные по норме возмущения, способные нарушить корректность предсказания. Среди множества существующих методов наибольшее распространение получили Метод быстрого градиентного знака (Fast Gradient Sign Method, FGSM), Проецируемый спуск по градиенту (Projected Gradient Descent, PGD) и Базовый итеративный метод (Basic Iterative Method, BIM), отличающиеся по вычислительной сложности, точности аппроксимации и устойчивости к защитным механизмам [12].

Метод быстрого градиентного знака (FGSM), введенный в работе Goodfellow et al. [11], основывается на линейной аппроксимации функции потерь $L(f_\theta(x), y)$ в окрестности входного вектора (исходное изображение) $x \in [0, 1]^d$. Состязательный пример формируется по формуле:

$$x_{adv} = \prod_{[0,1]^d} (x + \varepsilon \cdot \text{sign}(\nabla_x L(f_\theta(x), y))),$$

где $\varepsilon > 0$ – параметр, ограничивающий L_∞ – норму возмущения, а $\prod$ – оператор проекции на допустимый диапазон пикселей. Несмотря на вычислительную эффективность (один прямой и один обратный проход), FGSM часто генерирует возмущения, не являющиеся оптимальными, особенно против моделей с гладкими градиентами, поскольку игнорирует нелинейную структуру функции потерь [7].

Проецируемый спуск по градиенту (PGD) устраняет этот недостаток за счет итеративного уточнения направления возмущения [7]. Формально метод PGD реализует многократный градиентный подъем с проекцией на L_∞ – шар радиуса ε :

$$x_{t+1} = \prod_{B_\varepsilon^\infty(x)} (x_t + \alpha \cdot \text{sign}(\nabla_x L(f_\theta(x_t), y))),$$

где α – шаг итерации, $B_\varepsilon^\infty(x) = \{x' \in R^d: \|x' - x\|_\infty \leq \varepsilon\}$. При малом α и достаточном числе шагов метода PGD аппроксимирует локальный максимум функции потерь и считается «сильнейшей» L_∞ – ограниченной атакой в сценарии белого ящика [7]. Например, на модели ResNet-18, обученной на CIFAR-10, PGD с $\varepsilon = \frac{8}{255}$, $\alpha = \frac{8}{255}$, $T = 20$ снижает точность с 92% до менее 5% [13].

Базовый итеративный метод (BIM) формально совпадает с методом PGD, однако в оригинальной постановке применяется без случайного старта ($x_0 = x$) и часто интерпретируется как итеративная версия метода FGSM. Метод BIM обеспечивает баланс между вычислительной эффективностью и качеством возмущения, демонстрируя более высокий уровень успеха по сравнению с методом FGSM при сопоставимых затратах [14, 15].

3. Построение детектора аномалий. Построение детектора аномалий в настоящем исследовании выполнено в соответствии с парадигмой обучения без учителя [16]. Для выявления состязательных примеров использовалась стратегия построения детекторов аномалий с обучением без учителя, основанных на анализе внутренних представлений, формируемых в промежуточных слоях предварительно обученной сверточной нейронной сети [17]. Обучение детектора проводилось только на выборке чистых (неискаженных) данных, что соответствует сценарию обучения на одном классе [18], при котором нормальное поведение модели считается единственным доступным классом.

В настоящее время существует большое число методов обнаружения аномалий [2]. В эксперименте были реализованы и исследованы четыре из них:

1. Изоляционный лес (Isolation Forest, IF) – ансамблевый метод, основанный на случайном разбиении пространства признаков, где аномалии идентифицируются по меньшей средней длине пути до листа в дереве изоляции. Гиперпараметры: $n_estimators=100$, $contamination=0.1$.

2. Одноклассовый метод опорных векторов (One-Class Support Vector Machine, OCSVM) – метод, максимизирующий зазор между началом координат и гиперплоскостью, ограждающей область сосредоточения нормальных данных в признаковом пространстве с ядром RBF. Гиперпараметры: $\nu=0.1$, $\gamma='scale'$.

3. Анализ главных компонент (Principal Component Analysis, PCA) – линейный метод снижения размерности; аномальность вычисляется как квадрат ошибки реконструкции после проекции на главное подпространство, содержащее 95 % дисперсии. Порог детекции устанавливается как 90-й перцентиль ошибки реконструкции на обучающей (неискаженной) выборке.

4. Расстояние Махаланобиса (Mahalanobis distance) – статистическая мера удаленности до центра класса с учетом ковариационной структуры признакового пространства; теоретически обоснован как наиболее чувствительный к смещениям, вносимым состязательными возмущениями [2].

Для каждого детектора вычисляются оценки аномальности:

- для IF и OCSVM – отрицательная функция принятия решений (чем выше значение, тем выше вероятность аномалии);
- для PCA – евклидова норма разности между исходным вектором активаций и его реконструкцией.

Качество детекции оценивалось с помощью следующих метрик:

- площадь под ROC-кривой (AUC ROC) – интегральный показатель способности метода располагать аномальные образцы выше нормальных в ранжированном списке [19];
- $TPR@FPR = \tau$ (для $\tau \in \{1\%, 5\%, 10\%\}$) – доля истинно положительных срабатываний при заданном уровне ложноположительных срабатываний.

Выбранные для оценки качества детекции метрики в комплексе отражают как общую способность к разделению классов (AUC ROC), так и практическую применимость в условиях строгих требований к доле ложных срабатываний – критически важному аспекту для систем информационной безопасности:

- интегральная метрика бинарного классификатора, инвариантная к порогу AUC ROC (Area Under the Receiver Operating Characteristic Curve) – позволяет сравнивать методы в общем, но не отражает их поведение при реальных операционных ограничениях;

- метрика $TPR@FPR=X\%$, отражающая долю истинно положительных срабатываний при фиксированной доле ложноположительных срабатываний – ключевая для приложений информационной безопасности, где ложные срабатывания (FPR) недопустимы выше заданного уровня (например, 1% – не более 1 ложного тревожного сигнала на 100 легитимных запросов). В исследовании анализировались три уровня: $FPR = 1\%$, 5% , 10% , что охватывает диапазон от режимов с минимально допустимым риском (ядерные системы, финансовый сектор) до режимов с допустимым уровнем ложных срабатываний (внутренние инструменты мониторинга).

Применение такого комплекса метрик позволяет анализировать эффективность детекторов как в общем, так и при строгих ограничениях на долю ложных срабатываний, что особенно актуально для систем информационной безопасности и значимо для практических задач с жесткими требованиями к избирательности [14]. Выбор именно этих методов также обусловлен их различной вычислительной сложностью, требованиями к данным и совместимостью с промышленными практиками разработки и эксплуатации (DevOps и MLOps):

- IF и OCSVM легко конвертируются в формат для хранения/передачи (сериализуются с помощью `joblib/pickle`) и развертываются в виде отдельных программных сервисов;

- PCA может быть реализован на графическом процессоре или даже в виде оператора формата ONNX, что ускоряет этап вывода;

- Расстояние Махаланобиса требует лишь хранения векторов средних и обратной ковариационной матрицы (либо ее диагональной аппроксимации), что обеспечивает минимальные накладные расходы по памяти (< 1 МБ даже для 1024-мерного пространства).

Таким образом, предложенный подход реализует метод обнаружения аномалий в пространстве признаков [20, 21], при котором уязвимость классификатора к состязательным воздействиям компенсируется внешней системой мониторинга, не требующей изменения архитектуры или повторного обучения целевой модели.

4. Исследование эффективности обнаружения состязательных атак. В работе проведены экспериментальные исследования, направленные на анализ устойчивости сверточных нейронных сетей к состязательным атакам и оценку эффективности

четырёх вышеуказанных методов обнаружения аномалий в признаковом пространстве. Эксперименты проводились на четырёх моделях – Model_A, Model_B, Model_A_AT, Model_B_AT, исследованных авторами ранее [22]. Модели построены на архитектурах сверточных нейронных сетей, отличающихся глубиной, размером ядер и стратегией понижения размерности. В качестве признакового пространства использовались векторные активации заданного сверточного слоя: для модели Model_A – выход второго сверточного блока с операцией субдискретизации, для Model_B – выход третьего сверточного блока. Для моделей A и B также были обучены версии (модели Model_A_AT, Model_B_AT) с использованием состязательного обучения (AT), применяющие атаки на основе проекционного градиентного спуска (PGD) на этапе обучения – один из наиболее эффективных в настоящее время подходов к повышению устойчивости. Активации извлекались без вычисления градиентов, нормализация входных данных сохранялась согласно этапу предобработки.

В качестве тестовой среды выбран набор данных Fashion-MNIST – стандартный эталонный набор для задач компьютерного зрения, представляющий собой коллекцию из 70000 изображений одежды 10 классов (28×28 пикселей, в градациях серого). Выбор именно этого набора данных обусловлен следующими факторами:

1. Достаточная сложность по сравнению с набором MNIST при сохранении высокой воспроизводимости результатов.
2. Отсутствие этических и лицензионных ограничений на использование.
3. Широкая распространенность в исследованиях устойчивости к состязательным атакам (например, в работах [8, 10]), что позволяет сопоставлять полученные результаты с данными научной литературы.

Генерация состязательных примеров осуществлялась с использованием библиотеки torchattacks и включала три типа атак, различающихся по механизму и «заметности»: метод быстрого градиентного знака (Fast Gradient Sign Method, FGSM), проекционный градиентный спуск (Projected Gradient Descent, PGD), базовый итеративный метод (Basic Iterative Method, BIM).

Для каждой атаки тестировались три уровня интенсивности возмущения: $\varepsilon = 0.1, 0.15, 0.2$, что соответствует смещению пикселя на 25.5, 38.25 и 51 единицу яркости (при исходном диапазоне [0,255]), и позволяет исследовать зависимость эффективности детекции от силы атаки.

4.1. Влияние типа состязательной атаки на эффективность обнаружения. Одним из ключевых факторов, определяющих уязвимость систем машинного обучения, являются характер и механизм генерации состязательных возмущений. В рамках данного исследования были протестированы три классических типа атак – FGSM, PGD и BIM, – различающиеся по числу итераций, наличию проекции и степени «заметности» вносимых обнаружений. Анализ результатов позволяет сделать вывод о том, что не все атаки равнозначны с точки зрения обнаружения, и выбор тестового набора атак напрямую влияет на оценку надежности системы защиты.

Атаки FGSM (метод быстрого градиентного спуска) и PGD (проекционный градиентный спуск) относятся к классу градиентных методов с контролируемой L -нормой, где величина возмущения ограничена параметром ε . Несмотря на различия в реализации (FGSM – одношаговая атака, PGD – итеративный метод), обе они демонстрируют сходное поведение с точки зрения обнаружения на незащищенных моделях Model_A, Model_B.

Как видно из таблицы 1, для Model_A при $\varepsilon = 0.1$ разница в метриках не превышает статистически значимого уровня ($\Delta TPR < 1\%$ при $N=500$, N – это объем тестовой выборки, использованной в экспериментах: NUM_TEST = 500), что позволяет рассматривать FGSM и PGD как взаимозаменяемые при оценке на незащищенных моделях.

Однако при переходе к моделям с защитным обучением различия становятся значимыми. Как показано в таблице 2, для Model_A_AT при $\varepsilon = 0.2$ PGD обнаруживается хуже FGSM на 5-7 % по $TPR@FPR = 1\%$.

Таблица 1. Сравнение эффективности обнаружения FGSM и PGD на модели Model_A ($\varepsilon = 0.1$)

Детектор	FGSM (AUC)	PGD (AUC)	FGSM (TPR@FPR = 1%)	PGD (TPR@FPR = 1%)
OCVSM	0.9987	0.9987	0.962	0.968
IF	0.9931	0.9931	0.848	0.860
PCA	1.0000	1.0000	0.000	0.000

Таблица 2. Сравнение $TPR@FPR = 1\%$ для FGSM и PGD на модели Model_A_AT ($\varepsilon = 0.2$)

Детектор	FGSM	PGD	Разность (FGSM - PGD)
Mahalanobis	0.914	0.860	+0.054
PCA	0.768	0.696	+0.072

Это объясняется тем, что защитное обучение, выполненное с использованием PGD, формирует устойчивость именно к данному типу возмущения, в то время как FGSM остается относительно незнакомой атакой – что согласуется с выводами работы [10].

В корне иная ситуация наблюдается при использовании BIM (Basic Iterative Method) – итеративной атаки без проекции на L -шар. Несмотря на формальное сходство с PGD, отсутствие ограничения на максимальное изменение пикселя приводит к более «естественному» распределению возмущений.

Как видно из таблицы 3, для Model_A при $\epsilon = 0.2$ обнаруживаемость BIM падает более чем в 200 раз по сравнению с FGSM.

Особый интерес представляет сравнение эффективности обнаружения BIM на разных архитектурах. Как видно из таблицы 4, несмотря на более «сильную» архитектуру, Model_B показывает меньшее значение $TPR@FPR = 1\%$ против BIM.

В то же время Model_B проявляет наивысшую способность к обнаружению «грубых» атак (FGSM, PGD): при $\epsilon = 0.1$ метод PCA обеспечивает $TPR = 0.998$ при $FPR = 1\%$. Это объясняется тем, что глобальное усреднение в Model_B формирует признакововое пространство, в котором реконструкционная ошибка PCA четко отражает аномалии, вносимые крупномасштабными возмущениями.

Таблица 3. Падение эффективности обнаружения при переходе от FGSM к BIM (Model_A, $\epsilon = 0.2$)

Детектор	FGSM ($TPR@FPR = 1\%$)	BIM ($TPR@FPR = 1\%$)	Относительное снижение
OCVSM	0.972	0.004	-99.6 %

Таблица 4. Эффективность обнаружения BIM-атаки ($\epsilon = 0.2$) на базовых моделях

Модель	Лучший детектор	$TPR@FPR = 1\%$
Model_A	Mahalanobis	0.012
Model_B	PCA	0.004

4.2. Влияние силы состязательной атаки на эффективность обнаружения различных моделей. Если тип атаки определяет качественные особенности угрозы, то параметр силы возмущения (ϵ) задает ее количественную интенсивность. В рамках данного исследования, тестирование проводилось при трех значениях ϵ – 0.1, 0.15 и 0.2, – что соответствует максимальному смещению пикселя на 25.5, 38.25 и 51 единицу яркости (в исходном диапазоне [0, 255]). Такой диапазон охватывает как малозаметные ($\epsilon = 0.1$),

так и визуально различимые ($\varepsilon = 0.2$) возмущения, позволяя оценить, как изменяется эффективность детекции при переходе от скрытых (stealth) к явным атакам.

Анализ результатов выявляет три принципиально различных режима поведения, зависящих от сочетания архитектуры модели и наличия состязательного обучения.

Для незащищенных моделей (Model_A, Model_B) рост ε не приводит к значительному улучшению обнаруживаемости – особенно при строгих ограничениях на ложные срабатывания ($\text{FPR} \leq 1\%$). Рассмотрим динамику $\text{TPR@FPR} = 1\%$ для атаки BIM, представленную в таблице 5:

Таблица 5. Значения показателя TPR@FPR при увеличении силы атаки

Модель	$\varepsilon = 0.1$	$\varepsilon = 0.15$	$\varepsilon = 0.2$	$\Delta(0.1 \rightarrow 0.2)$
Model_A (Mahalanobis)	0.010	0.010	0.012	+20 %
Model_B (PCA)	0.002	0.002	0.004	+100 %

Видно, что даже при удвоении ε с 0.1 до 0.2 рост TPR составляет всего 0.002-0.004, что делает детекторы практически бесполезными в реальных условиях, поскольку в большинстве систем информационной безопасности для эффективного обнаружения аномалий требуется значение $\text{TPR} \geq 0.8$ при $\text{FPR} \leq 0.01$.

Причина заключается в том, что базовые модели не обучены различать «тонкие» смещения в признаковом пространстве. Возмущения, даже при $\varepsilon = 0.2$, остаются в пределах «естественной» вариативности данных, и распределение активаций атакованных и чистых примеров сильно перекрывается.

Коренным образом иная картина наблюдается для моделей с состязательным обучением (Model_A_AT, Model_B_AT). Здесь рост ε приводит к резкому, нелинейному увеличению TPR, особенно при использовании метрики расстояния Махаланобиса.

Для Model_B_AT при атаке BIM значения показателя TPR@FPR при увеличении силы атаки представлены в таблице 6:

Таблица 6. Зависимость показателя TPR@FPR предобученной модели от силы атаки BIM

ε	AUC	$\text{TPR@FPR} = 1\%$	$\text{TPR@FPR} = 5\%$	$\text{TPR@FPR} = 10\%$
0.1	0.5806	0.054	0.194	0.240
0.15	0.6235	0.104	0.250	0.292
0.2	0.6758	0.164	0.286	0.312

Из данных таблицы следует, что при увеличении ϵ всего на 0.1 (от 0.1 до 0.2) $\text{TPR@FPR} = 1\%$ растет в 3 раза (с 0.054 до 0.164). Это говорит о том, что состязательное обучение формирует в признаковом пространстве четкие «зоны чувствительности», где даже небольшое превышение порога ϵ вызывает резкое изменение активаций – эффект, описанный в работе [23] как разделение состязательных подпространств (adversarial subspace separation).

Полученные результаты свидетельствуют о взаимном усилении между защитным обучением и применением расстояния Махаланобиса: защитное обучение формирует в признаковом пространстве компактные и четко разделимые классовые кластеры, а расстояние Махаланобиса, учитывающее ковариационную структуру этих кластеров, максимально чувствительно реагирует на отклонения, вносимые состязательными возмущениями.

Особый интерес представляет сравнение моделей Model_A_AT и Model_B_AT – систем, построенных с применением одинаковой схемы защитного обучения, но имеющих различную архитектуру сверточных слоев.

Для атаки BIM при $\epsilon = 0.2$ результаты, представленные в таблице 7, демонстрируют кардинальное различие в эффективности обнаружения:

Таблица 7. Эффективность обнаружения BIM-атаки ($\epsilon = 0.2$) моделями с защитным обучением, оцененная по расстоянию Махаланобиса

Модель	$\text{TPR@FPR} = 1\%$	$\text{TPR@FPR} = 5\%$	$\text{TPR@FPR} = 10\%$
Model A AT	0.010	0.058	0.124
Model B AT	0.164	0.286	0.312

Model_B_AT демонстрирует наиболее высокую обнаруживаемость базовых итеративных методов (BIM-атак) среди всех моделей. Так, при $\epsilon = 0.2$ расстояние Махаланобиса достигает $\text{TPR} = 0.164$ при $\text{FPR} = 1\%$, что в 16.4 раза выше, чем у Model_A_AT (0.010).

Данный эффект объясняется следующим образом:

- большое ядро (5×5) формирует более глобальные структурированные признаки, менее чувствительные к локальному шуму, но четко реагирующие на систематические смещения, вносимые атаками;
- формирование компактного 64-мерного пространства признаков (после операции глобального усреднения) минимизирует влияние «шумовых» измерений, что повышает соотношение сигнал/шум при вычислении расстояния Махаланобиса, поскольку

в низкоразмерных пространствах отклонения от центроида класса обладают большей статистической значимостью;

– оптимальный размер полносвязного слоя (512 нейронов) предотвращает избыточную выразительность модели, которая в противном случае могла бы интегрировать составительные примеры в свое распределение.

Следовательно, гиперпараметрическая оптимизация, изначально направленная на повышение точности классификации, косвенно способствовала лучшему обнаружению составительных атак.

Таким образом, сила атаки (ϵ) является критическим параметром при проектировании систем защиты, но ее влияние нелинейно и зависит от архитектуры и уровня защиты модели. Составительное обучение обеспечивает предсказуемое и управляемое поведение системы защиты: рост силы атаки приводит к стабильному увеличению вероятности ее обнаружения. Кроме того, предварительно проведенная гиперпараметрическая оптимизация нейронной сети, изначально направленная на повышение точности классификации, также способствует лучшему обнаружению составительных атак.

5. Заключение. В ходе исследовательской работы было проведено комплексное экспериментальное исследование, направленное на оценку эффективности методов обнаружения составительных атак на основе анализа аномалий в активациях промежуточных слоев сверточных нейронных сетей. Актуальность работы обусловлена растущими рисками, связанными с внедрением моделей машинного обучения в критически важные информационные системы, где устойчивость к целенаправленным атакам становится неотъемлемой частью требований к информационной безопасности.

Полученные результаты позволяют сделать следующие выводы:

1. Тип атаки определяет выбор метода защиты: FGSM и PGD легко обнаруживаются большинством детекторов ($TPR > 0.96$ при $FPR = 1\%$), в то время как BIM, благодаря своей «незаметной» структуре, снижает TPR до $0.004 - 0.012$ на базовых моделях – что подчеркивает необходимость тестирования на разнообразных атаках, а не только на FGSM.

2. Защитное обучение не только повышает устойчивость, но и усиливает способность к обнаружению атак, особенно в паре с методом обнаружения аномалий на основе расстояния Махаланобиса. Показано взаимное усиление между защитным обучением и статистической оценкой отклонений: первое формирует компактные, четко разделимые классовые кластеры, вторая – максимально чувствительна к их нарушению.

3. Архитектурные гиперпараметры оказывают решающее влияние на эффективность детекции. Модель Model_B, оптимизированная в предыдущем исследовании авторов [22] под критерий точности классификации (ядро 5×5 , глобальное усреднение, 512 нейронов в скрытом слое), продемонстрировала наивысшую способность к обнаружению «грубых» атак ($TPR = 0.998$ при $FPR = 1\%$ с использованием анализа главных компонент), что свидетельствует о совместимости задач повышения точности классификации и обеспечения безопасности информационных систем.

4. Метрика TPR при фиксированном $FPR = 1\%$ является более информативной, чем площадь под ROC-кривой (AUC), при проектировании систем защиты. Например, анализ главных компонент на Model_A показывает $AUC = 1.0$, но при фиксированном уровне ложноположительных срабатываний $FPR = 0.01$ достигается нулевая чувствительность ($TPR = 0$), что свидетельствует об отсутствии способности метода обнаруживать истинные аномалии в режиме, соответствующем практически приемлемой частоте ложных тревог.

Практическая значимость работы заключается в возможности интеграции полученных результатов в процессы обеспечения безопасности при разработке и эксплуатации информационных систем.

Литература

1. Rossolini G., Biondi A., Butazzo G., Attention-based real-time defenses for physical adversarial attacks in vision applications // Proceedings of the ACM/IEEE 15th International Conference on Cyber-Physical Systems (ICCPs). 2024. pp. 23–32. DOI: 10.1109/ICCPs61052.2024.00009.
2. Kotenko I., Saenko I., Lauta O., et al. Analysis of Modern Research on Protection against Adversarial Attacks in Energy Systems // Informatics and Automation. 2025. vol. 24. no. 6. pp. 1751–1809. DOI: 10.15622/ia.24.6.8.
3. Wang H., Wu C., Zheng K. Defense against adversarial attacks based on color space transformation // Neural Networks. 2024. vol. 173. 106176 p. DOI: 10.1016/J.NEUNET.2024.106176.
4. Brau F., Rossolini G., Biondi A., Butazzo G. On the Minimal Adversarial Perturbation for Deep Neural Networks With Provable Estimation Error // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022. vol. 45. pp. 5038–5052. DOI: 10.1109/TPAMI.2022.3195616.
5. Szegedy C., Zaremba W., Sutskever I., et al. Intriguing properties of neural networks // arXiv preprint arXiv:1312.6199. 2013.
6. Hong F., Bouhassira R., Chow J., et al. Comprehensive characterization of human color discrimination thresholds // bioRxiv. 2025. pp. 1–70. DOI: 10.1101/2025.07.16.665219.
7. Goodfellow I.J., Shlens J., Szegedy C. Explaining and harnessing adversarial examples // Proceedings of the 3rd International Conference on Learning Representations (ICLR). 2015. pp. 1–11.
8. Papernot N., McDaniel P., Jha S., et al. The Limitations of Deep Learning in Adversarial Settings // Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P). 2016. pp. 372–387. DOI: 10.1109/EuroSP.2016.36.

9. Zhu Z., Chen H., Wang X., et al. Ge-AdvGAN: Improving the transferability of adversarial samples by gradient editing-based adversarial generative model // Proceedings of the 2024 SIAM International Conference on Data Mining (SDM). 2024. pp. 706–714.
10. Carlini N., Wagner D. Towards Evaluating the Robustness of Neural Networks // Proceedings of the IEEE Symposium on Security and Privacy (SP). 2017. pp. 39–57. DOI: 10.1109/SP.2017.49.
11. Wang Y., Liao X., Cui W., et al. Defending against and generating adversarial examples together with generative adversarial networks // Scientific Reports. 2025. vol. 15. 12994 p. DOI: 10.1038/s41598-024-83444-x.
12. Xu L., Zhai J. Generating adversarial examples with collaborative generative models // International Journal of Information Security. 2024. vol. 23. pp. 1077–1091. DOI: 10.1007/S10207-023-000780-1.
13. Madry A., Makelov A., Schmidt L., et al. Towards deep learning models resistant to adversarial attacks // Proceedings of the International Conference on Learning Representations (ICLR). 2018. pp. 1–28.
14. Kurakin A., Goodfellow I.J., Bengio S. Adversarial Examples in the Physical World // Proceedings of the 5th International Conference on Learning Representations. 2017. pp. 99–112.
15. Gui J., Sun Z., Wen Y., et al. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications // IEEE Transactions on Knowledge and Data Engineering. 2023. vol. 35. pp. 3313–3332. DOI: 10.1109/TKDE.2021.3130191.
16. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey // ACM Computing Surveys (CSUR). 2009. vol. 41. no. 3. pp. 1–58. DOI: 10.1145/1541880.1541882.
17. Li Z., Wang W., Li J., et al. UCG: A Universal Cross-Domain Generator for Transferable Adversarial Examples // IEEE Transactions on Information Forensics and Security. 2024. vol. 19. pp. 3023–3037. DOI: 10.1109/TIFS.2024.3352913.
18. Scholkopf B., Platt J.C., Shawe-Taylor J., et al. Estimating the Support of a High-Dimensional Distribution // Neural Computation. 2001. vol. 13. no. 7. pp. 1443–1471. DOI: 10.1162/089976601750264965.
19. Fawcett T. An introduction to ROC analysis // Pattern Recognition Letters. 2006. vol. 27. no. 8. pp. 861–874. DOI: 10.1016/j.patrec.2005.10.010.
20. Lee K., Lee K., Lee H., Shin J. A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks // Advances in Neural Information Processing Systems. 2018. vol. 31. pp. 1–11.
21. Feng W., Xu N., Zhang T., et al. Robust and Generalized Physical Adversarial Attacks via Meta-Gan // IEEE Transactions on Information Forensics and Security. 2023. vol. 19. pp. 1112–1125. DOI: 10.1109/TIFS.2023.3288426.
22. Клименко Д.М., Козлова Е.И. Исследование влияния гиперпараметров на точность нейросетевого предсказания с использованием набора данных Fashion-MNIST // Цифровая трансформация. 2026. Т. 32. №2. С. 44–52. DOI: 10.35596/1729-7648-2026-32-2-44-52.
23. Ma X., Li B., Wang Y., et al. Characterizing Adversarial Subspaces Using Local Intrinsic Dimensionality // Proceedings of the 6th International Conference on Learning Representations (ICLR). 2018. pp. 1–15.

Козлова Елена Ивановна — канд. физ.-мат. наук, доцент, заведующий кафедрой, кафедра интеллектуальных систем, Белорусский государственный университет (БГУ). Область научных интересов: искусственный интеллект, машинное обучение, кибербезопасность. Число научных публикаций — 80. kozlova@bsu.by; ул. Курчатова, 5, 220047, Минск, Беларусь; р.т.: +375(17)209-5936.

Клименко Даниил Михайлович — магистрант, кафедра интеллектуальных систем, Белорусский государственный университет (БГУ). Область научных интересов: искусственный интеллект, машинное обучение, кибербезопасность. Число научных публикаций — 1. daniil.klimenko2212@yandex.by; ул. Курчатова, 5, 220047, Минск, Беларусь; р.т.: +375(17)209-5936.

Поддержка исследований. НИОКР № 20262086 от 08.07.2026 г. в рамках Задания 75 подпрограммы «Научно-учебное оборудование» государственной научно-технической программы «Научно-методическое и материально-техническое обеспечение образовательного процесса».

E. KOZLOVA, D.M. KLIMENKA

DETECTING ADVERSARIAL ATTACKS THROUGH ANOMALY IDENTIFICATION IN HIDDEN LAYERS OF A NEURAL NETWORK

Kozlova E., Klimenka D.M. Detecting Adversarial Attacks through Anomaly Identification in Hidden Layers of a Neural Network.

Abstract. The current level of information technology development is characterized by the widespread use of artificial intelligence (AI) methods, particularly deep machine learning, in such critical domains as healthcare, transportation, financial services, biometric authentication systems, and national security. Neural network models are increasingly employed to make decisions that directly affect the safety, health, and financial interests of individuals and organizations. However, as the functional significance of artificial intelligence systems increases, so does their vulnerability to targeted attacks. Existing countermeasures, such as adversarial training or input filtering, have significant limitations: they either degrade model accuracy on legitimate data or are effective only against previously known attack types, failing to detect emerging threats. Under these conditions, there is a growing need for detection mechanisms capable of identifying the very fact of an attack regardless of its implementation. Moreover, various studies have shown that adversarial examples induce anomalous activation patterns in the hidden layers of a neural network that differ qualitatively from those observed when processing legitimate data. This opens up the possibility of applying anomaly detection methods, such as isolation forest or Mahalanobis distance, to identify suspicious queries without modifying the original model. Of particular importance is also the influence of architectural and training hyperparameters on the internal representations of the model and, consequently, on its vulnerability and attack detection. The aim of this work is an experimental study of a method for detecting adversarial attacks based on anomaly analysis in the activations of hidden layers of a neural network.

Keywords: artificial intelligence, deep machine learning, neural networks, adversarial attack, attack detection, anomalies, hidden activations, internal representations of the model, isolation forest, model quality assessment metrics.

References

1. Rossolini G., Biondi A., Butazzo G., Attention-based real-time defenses for physical adversarial attacks in vision applications. Proceedings of the ACM/IEEE 15th International Conference on Cyber-Physical Systems (ICCPs). 2024. pp. 23–32. DOI: 10.1109/ICCPs61052.2024.00009.
2. Kotenko I., Saenko I., Lauta O., et al. Analysis of Modern Research on Protection against Adversarial Attacks in Energy Systems. Informatics and Automation. 2025. vol. 24, no. 6. pp. 1751–1809. DOI: 10.15622/ia.24.6.8.
3. Wang H., Wu C., Zheng K. Defense against adversarial attacks based on color space transformation. Neural Networks. 2024. vol. 173. 106176 p. DOI: 10.1016/J.NEUNET.2024.106176.
4. Brau F., Rossolini G., Biondi A., Butazzo G. On the Minimal Adversarial Perturbation for Deep Neural Networks With Provable Estimation Error. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022. vol. 45. pp. 5038–5052. DOI: 10.1109/TPAMI.2022.3195616.
5. Szegedy C., Zaremba W., Sutskever I., et al. Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199. 2013.

6. Hong F., Bouhassira R., Chow J., et al. Comprehensive characterization of human color discrimination thresholds. *bioRxiv*. 2025. pp. 1–70. DOI: 10.1101/2025.07.16.665219.
7. Goodfellow I.J., Shlens J., Szegedy C. Explaining and harnessing adversarial examples. *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. 2015. pp. 1–11.
8. Papernot N., McDaniel P., Jha S., et al. The Limitations of Deep Learning in Adversarial Settings. *Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P)*. 2016. pp. 372–387. DOI: 10.1109/EuroSP.2016.36.
9. Zhu Z., Chen H., Wang X., et al. Ge-AdvGAN: Improving the transferability of adversarial samples by gradient editing-based adversarial generative model. *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)*. 2024. pp. 706–714.
10. Carlini N., Wagner D. Towards Evaluating the Robustness of Neural Networks. *Proceedings of the IEEE Symposium on Security and Privacy (SP)*. 2017. pp. 39–57. DOI: 10.1109/SP.2017.49.
11. Wang Y., Liao X., Cui W., et al. Defending against and generating adversarial examples together with generative adversarial networks. *Scientific Reports*. 2025. vol. 15. 12994 p. DOI: 10.1038/s41598-024-83444-x.
12. Xu L., Zhai J. Generating adversarial examples with collaborative generative models. *International Journal of Information Security*. 2024. vol. 23. pp. 1077–1091. DOI: 10.1007/S10207-023-000780-1.
13. Madry A., Makelov A., Schmidt L., et al. Towards deep learning models resistant to adversarial attacks. *Proceedings of the International Conference on Learning Representations (ICLR)*. 2018. pp. 1–28.
14. Kurakin A., Goodfellow I.J., Bengio S. Adversarial Examples in the Physical World. *Proceedings of the 5th International Conference on Learning Representations*. 2017. pp. 99–112.
15. Gui J., Sun Z., Wen Y., et al. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE Transactions on Knowledge and Data Engineering*. 2023. vol. 35. pp. 3313–3332. DOI: 10.1109/TKDE.2021.3130191.
16. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*. 2009. vol. 41. no. 3. pp. 1–58. DOI: 10.1145/1541880.1541882.
17. Li Z., Wang W., Li J., et al. UCG: A Universal Cross-Domain Generator for Transferable Adversarial Examples. *IEEE Transactions on Information Forensics and Security*. 2024. vol. 19. pp. 3023–3037. DOI: 10.1109/TIFS.2024.3352913.
18. Scholkopf B., Platt J.C., Shawe-Taylor J., et al. Estimating the Support of a High-Dimensional Distribution. *Neural Computation*. 2001. vol. 13. no. 7. pp. 1443–1471. DOI: 10.1162/089976601750264965.
19. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006. vol. 27. no. 8. pp. 861–874. DOI: 10.1016/j.patrec.2005.10.010.
20. Lee K., Lee K., Lee H., Shin J. A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks. *Advances in Neural Information Processing Systems*. 2018. vol. 31. pp. 1–11.
21. Feng W., Xu N., Zhang T., et al. Robust and Generalized Physical Adversarial Attacks via Meta-Gan. *IEEE Transactions on Information Forensics and Security*. 2023. vol. 19. pp. 1112–1125. DOI: 10.1109/TIFS.2023.3288426.
22. Klimenko D., Kozlova E. [Investigation of the Impact of Hyperparameters on the Accuracy of Neural Network Predictions Using the Fashion-MNIST Dataset]. *Cifrovaya transformaciya – Digital Transformation*. 2026. vol. 32. no. 2. pp. 44–52. DOI: 10.35596/1729-7648-2026-32-2-44-52. (In Russ.).

23. Ma X., Li B., Wang Y., et al. Characterizing Adversarial Subspaces Using Local Intrinsic Dimensionality. Proceedings of the 6th International Conference on Learning Representations (ICLR). 2018. pp. 1–15.

Kozlova Elena — Ph.D., Associated Professor, Head of the Department, Department of Intelligent Systems, Belarusian State University (BSU). Research interests: artificial intelligence, machine learning, cybersecurity. The number of publications — 80. kozlova@bsu.by; 5, Kurchatov St., 220047, Minsk, Belarus; office phone: +375(17)209-5936.

Klimenka Daniil — Master's Student, Department of Intelligent Systems, Belarusian State University (BSU). Research interests: artificial intelligence, machine learning, cybersecurity. The number of publications — 1. daniil.klimenko2212@yandex.by; 5, Kurchatov St., 220047, Minsk, Belarus; office phone: +375(17)209-5936.

Acknowledgements. Research and Development Project No. 20262086 dated July 8, 2026, within the framework of Task 75 of the subprogram “Scientific and Educational Equipment” of the state scientific and technical program “Scientific, Methodological and Material-Technical Support of the Educational Process”.

М.Г. ФИРСОВА

**ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ ДЛЯ АНАЛИЗА
ЗАЩИЩЁННОСТИ СИСТЕМ ИНТЕРНЕТА ВЕЩЕЙ***Фирсова М.Г. Онтологическая модель для анализа защищённости систем Интернета вещей.*

Аннотация. В статье рассматривается задача комплексного анализа защищённости систем Интернета вещей. Актуальность работы обусловлена ростом сложности и гетерогенности систем Интернета вещей, а также необходимостью масштабируемой оценки их защищённости с учётом архитектурных и функциональных особенностей. На основе анализа архитектуры систем Интернета вещей разработана многоуровневая онтологическая модель, объединяющая предметную область Интернета вещей и информационной безопасности, дополненная учётом функциональной роли компонентов системы, типов передаваемых сообщений и целей системы, что позволяет отобразить зависимость критичности компонента от выполняемой функции и влияние угрозы на нарушение целевой функции системы. На основе модели предложен механизм вывода угроз, выявляющий каскадный эффект распространения деструктивного воздействия по топологическому вектору и функциональному вектору. В отличие от подходов, рассматривающих угрозу исключительно через факт первичной компрометации элемента, новизна заключается в том, что глубина и последствия угрозы определяются не начальной компрометацией, а структурно-семантическими характеристиками межкомпонентных связей, задающих развитие каскадного эффекта. Онтология интегрирована с внешними репозиториями знаний (CVE, RVD, CAPEC, CWE, БДУ ФСТЭК) посредством ETL-процесса, а вывод скрытых связей автоматизирован с помощью SWRL-правил в среде Protege. Практическая апробация модели выполнена на экспериментальном стенде «Умная лаборатория». Результатом работы является реализованная онтологическая модель, обеспечивающая автоматическое выявление угроз безопасности для систем Интернета вещей. Оценка качества модели по методике OntoMod показала расширение семантического покрытия предметной области в 3 раза и снижение структурной сложности в 1,5-2,4 раза по сравнению с рассмотренными аналогами. Сформулирован вывод о применимости предложенного онтологического подхода для оперативной поддержки принятия решений операторами и проектировщиками распределённых сетей Интернета вещей.

Ключевые слова: онтология, Интернет вещей, угроза, структурно-функциональная характеристика, информационная безопасность.

1. Введение. В настоящее время наблюдается устойчивый рост числа компьютерных атак, нацеленных на системы Интернета вещей. При этом сами устройства Интернета вещей все чаще становятся не только целью, но и инструментом для осуществления этих атак, будучи объединёнными в ботнет-сети. Данная тенденция усугубляется неконтролируемым расширением поверхности атаки, что обусловлено как увеличением количества гетерогенных компонентов, так и множеством технологий, лежащих в основе их реализации [1].

В связи с этим возникает явное противоречие, с которым сталкивается оператор системы Интернета вещей: с одной стороны, наблюдается стремительный рост числа потенциальных угроз,

напрямую связанный с усложнением архитектуры системы [1, 2]. С другой стороны, перед оператором стоит практическая задача: не устранить все угрозы, а идентифицировать и структурировать актуальные угрозы для их последующей минимизации. Ситуация усугубляется спецификой самой парадигмы Интернета вещей. В отличие от традиционных информационных систем, Интернет вещей характеризуется тесной интеграцией вычислительных и физических процессов, использованием ограниченных по ресурсам устройств и многоуровневой архитектурой взаимодействия [2]. Устройства зачастую имеют прямой выход в глобальную сеть, располагаются в физически неконтролируемой зоне и используют беспроводные каналы связи, которые по своей природе являются уязвимыми. Компрометация отдельных компонентов может приводить к каскадным последствиям и оказывать влияние на функционирование физического объекта управления. Более того, попытка устранить угрозы, связанные, например, с особенностями беспроводной передачи данных, может повлечь неоправданно высокие затраты, ставящие под вопрос экономическую целесообразность всей системы.

Существующие подходы к обеспечению информационной безопасности Интернета вещей, как правило, базируются либо на применении обобщённых классификаций уязвимостей и атак, либо на экспертных методах [3 – 5]. При этом в большинстве работ отсутствует формализованное представление взаимосвязей между архитектурными компонентами системы, уязвимостями, атаками и возникающими угрозами, что затрудняет воспроизводимую и масштабируемую идентификацию угроз.

Таким образом, актуальной становится задача получения объективной картины угроз. Ключевая потребность заключается в определении того, на какие именно компоненты системы эти угрозы могут повлиять, какова степень критичности этих компонентов в контексте работы всей системы. При этом определить каким образом эта критичность зависит от конкретной области применения и сферы функционирования Интернета вещей – домена. Решение этой задачи требует формализации не столько самих угроз, сколько связей между архитектурными компонентами системы, их уязвимостями и сценариями атак с учетом контекста функционирования. Для такой формализации предлагается использовать онтологический подход, позволяющий построить модель формализованного представления знаний о предметной области (онтологию), объединяющую предметные области Интернета вещей и информационной безопасности. Именно такой подход способен отразить киберфизическую природу Интернета

вещей, где компрометация отдельного датчика или может привести к каскадным последствиям в физическом мире.

Целью работы является разработка онтологической модели, объединяющей предметную область Интернета вещей и информационной безопасности, обеспечивающей автоматизированную идентификацию и детальный анализ угроз с учётом архитектурных особенностей, межкомпонентного взаимодействия и специфики домена применения систем.

Объектом исследования является система Интернета вещей, предметом исследования – модели, семантические связи и логические правила формализации знаний предметной области, определяющие процессы автоматизированного выявления угроз информационной безопасности систем Интернета вещей на основе онтологического инжиниринга.

Для достижения поставленной цели необходимо решить следующие задачи:

- анализ архитектурных особенностей систем Интернета вещей, определяющих их уязвимость к компьютерным атакам и специфику каскадного распространения угроз;
- систематический обзор и сравнительный анализ существующих онтологических моделей и подходов к моделированию и оценке угроз информационной безопасности в сфере Интернета вещей;
- разработка структуры концептуальной и формальной онтологической модели, объединяющей концепты Интернета вещей и информационной безопасности;
- реализация разработанной онтологической модели в среде Protege и наполнение базы знаний экземплярами классов на основе интеграции гетерогенных внешних источников (CVE, RVD, CAPEC, CWE, БДУ ФСТЭК) с применением методов ETL;
- разработка системы логических правил на языке SWRL для автоматизации вывода скрытых зависимостей и сквозной декомпозиции векторов атак, угроз и рисков безопасности;
- экспериментальное исследование разработанной онтологии на натурном стенде «Умная лаборатория» и оценка её структурно-поведенческого качества с использованием метрик OntoMod в сравнении с существующими решениями.

Научная новизна работы заключается в разработке метода автоматического выявления каскадных угроз в системах Интернета вещей, основанного на логическом выводе с использованием онтологической модели, которая интегрирует предметные области Интернета вещей и информационной безопасности, формализует межкомпонентные связи

через семантически нагруженные объектные свойства и обеспечивает возможность многопараметрической оценки угроз для повышения защищенности систем Интернета вещей. Многопараметрическая оценка угрозы определяется совместным учётом трёх разнородных параметров: самого факта воздействия на компонент, количества связанных компонентов и семантического класса передаваемых сообщений. В отличие от подходов, оценивающих угрозу исключительно по факту первичной компрометации элемента, предложенный метод создает производные угрозы и распространяет их по функциональным связям, что позволяет выявить каскадный эффект, определяемый не простой связностью компонентов, а семантикой их взаимодействия. На основе выведенных производных угроз и связи компонентов с целями системы метод устанавливает факт нарушения целевых функций, делая вывод о компрометации цели не напрямую из факта взлома, а как результат цепочки логически обоснованных шагов, что обеспечивает выявление скрытых путей нарушения целей функций системы.

2. Архитектурные особенности систем Интернета вещей.

Типовая архитектура систем Интернета вещей включает совокупность гетерогенных устройств, сетевых компонентов, программного обеспечения и сервисов обработки данных, которые являются структурно-функциональными характеристиками (СФХ) системы и взаимодействуют в рамках распределённой среды. В общем случае в структуре Интернета вещей выделяют сенсорный уровень (сенсоры и актуаторы), сетевой уровень для передачи данных и сетевого взаимодействия, а также уровень контроллеров, реализуемый с использованием локальных, туманных или облачных вычислений для управления и обработки данных [6] (Рис. 1).

Оператор инициализирует работу системы, передавая управляющие команды («1» стрелка). Далее сигнал через сеть направляется к контроллеру, который непосредственно отвечает за приведение в действие актуатора. В современных реализациях промежуточным звеном может выступать облачная платформа [5], выполняющая роль аналитического центра, или же уровень туманных вычислений, обеспечивающий локальную обработку и снижение сетевых задержек [7]. Результатом прохождения команды является физическое изменение состояния объекта управления.

Потоки данных иницируются либо запросом от оператора, либо автоматически – сенсорами. При запросе данных («5» стрелка) контроллер, получив сигнал через сеть, опрашивает соответствующий сенсор и возвращает информацию оператору («4» стрелка). В фоновом режиме сенсоры непрерывно генерируют измерительную информацию

о параметрах внешней среды («2» стрелка). Эта информация передается на контроллер для анализа. Для повышения эффективности анализа могут привлекаться технологии искусственного интеллекта или машинного обучения, при этом вычислительная нагрузка может распределяться между облаком и туманными узлами. Результаты мониторинга выводятся на интерфейс оператора, а в случае обнаружения отклонений или по заданному алгоритму контроллер инициирует управляющее воздействие, передавая команду актуатору («3» стрелка), как того требует принцип обратной связи [8].



Рис. 1. Схема взаимодействия компонентов системы Интернета вещей

Задача идентификации угроз, как части процесса оценки защищённости, состоит из определения множества актуальных угроз для заданной системы Интернета вещей на основе анализа её архитектуры и известных уязвимостей. Угроза считается актуальной, если она может быть реализована в системе Интернета вещей и впоследствии приводит к негативным последствиям для самой системы, окружающей среды или человека.

Для решения поставленной задачи предлагается использовать онтологический подход, в рамках которого онтология выступает не только как средство формализованного описания предметных областей Интернета вещей и информационной безопасности, но и как инструмент логического вывода. Онтологическая модель позволяет явно задать отношения между компонентами системы, уязвимостями, атаками и угрозами, а система логических правил обеспечивает автоматическое порождение новых знаний.

3. Обзор существующих онтологий и подходов к идентификации и оценке угроз. Исследования в области информационной безопасности систем Интернета вещей охватывают широкий спектр подходов, включая анализ архитектурных уязвимостей, разработку моделей угроз и применение онтологий для представления знаний. Целью данного обзора является анализ существующих исследований, посвященных моделированию и идентификации угроз систем Интернета вещей с фокусом на работы, использующие онтологический подход. Критериями анализа выступают: (1) полнота охвата архитектуры Интернета вещей, (2) концепты онтологии, (3) интеграция с внешними базами знаний, (4) наличие формализованного механизма вывода скрытых логических связей для идентификации угроз, (5) учёт домена функционирования системы, целей системы, (6) наличие взаимосвязей между компонентами системы.

Проведенный анализ позволяет выделить несколько направлений в развитии подходов к анализу защищённости на основе онтологий и графов атак. Основы применения онтологического подхода в области информационной безопасности были заложены в [9]. Была предложена публично доступная онтология, моделирующая активы, угрозы, уязвимости и контрмеры. Ключевой особенностью данной онтологии является введение отношения, связывающего контрмеру с активом, целью безопасности и стратегией защиты. Авторы впервые продемонстрировали возможность использования логического вывода для классификации угроз и контрмер по целям безопасности, что позволяет отвечать на сложные запросы к модели без ручного перечисления, а также служит основой для будущих алгоритмов ранжирования контрмер против конкретных угроз. Развивая концептуальные основы, в [10] была обоснована необходимость модульного построения онтологий информационной безопасности с выделением верхнего, среднего и доменного уровней, а также подчеркнута важность интеграции с существующими стандартами, такими как CVE (Common Vulnerabilities and Exposures), CAPEC (Common Attack Pattern Enumeration and Classification) и MAEC (Malware Attribute Enumeration and Characterization).

Применение семантических технологий для анализа компьютерных атак получило развитие в работе [11], предложившей онтологию графа атак на основе RDF (Resource Description Framework)/OWL (Web Ontology Language) с транзитивным свойством, позволяющим выполнять логический вывод о возможных векторах атак. Развивая этот подход, в работе [12] была реализована онтология

графа атак в среде Protege, показав преимущества SPARQL-запросов для извлечения информации о векторах атак и привилегиях в корпоративных сетях. Однако данные работы не рассматривали подробно СФХ системы и не позволяли выстроить причинно-следственные связи между компонентами системы и их угрозами. Более поздние исследования интегрируют онтологии с графами атак для динамической оценки и реагирования. В [13] было предложено комплексное решение, в котором логический граф атак обогащается данными из онтологии уязвимостей VDO (Vulnerability Description Ontology), а выбор контрмер осуществляется путем сопоставления с базой знаний D3FEND (Detection, Denial, and Disruption Framework Empowering Network Defense) и многокритериальной оптимизации реагирования для киберфизических систем (CPS).

С ростом популярности парадигмы Интернета вещей появились работы, адаптирующие онтологические подходы к специфике этой предметной области. Масштабный систематический обзор [14] показал, что основной целью применения онтологий в Интернете вещей является обеспечение совместимости (31,30% работ), в то время как вопросам безопасности и приватности уделяется существенно меньше внимания (13,04% работ). Отмечается, что лишь 27,83% исследований повторно используют существующие онтологические наработки, что свидетельствует о фрагментарности и отсутствии единой стандартизированной модели. В работе [1] описан подход, который позволяет рассчитывать интегральные метрики безопасности для устройств Интернета вещей на основе анализа уязвимостей, разрешений приложений и онтологического представления политик конфиденциальности. Другой подход предложен в работе [15], где разработана базовая онтология безопасности для промышленного Интернета вещей (IIoT), включающая классы Asset, Attack, Fault, SecurityGoal, Vulnerability, Risk, Capability, Criticality. Модель учитывает асимметричные связи между возможностями атакующего и атаками, а также связь риска с критичностью. Однако она не интегрируется с внешними базами данных уязвимостей и не содержит формализованных правил логического вывода для идентификации угроз (лишь декларируется возможность использования SWRL в будущем). Практическая реализация онтологического подхода для обеспечения безопасности устройств Интернета вещей представлена в [16], где разработана онтология, связывающая уязвимости устройств с лучшими практиками безопасности. На примере промышленного стенда авторы продемонстрировали использование онтологии и данных из базы CVE

для анализа реализованных мер защиты, однако предложенный подход не включает механизмы логического вывода с целью вывода скрытых связей о состоянии защищённости системы. Близкий по духу подход, но с фокусом на SIEM, представлен в [17, 18], где формализованы понятия сообщения и события, введены явные и неявные связи между устройствами, однако отсутствует интеграция с внешними базами знаний об уязвимостях и нет реализации онтологии на OWL.

Важным направлением исследований является разработка контекстно-зависимых онтологий безопасности, учитывающих специфику окружения и роли участников в системе Интернета вещей. В работе [19] предложена онтология контекстно-зависимой безопасности для Интернета вещей, построенная на базе STAC (Security Toolbox: Attacks & Countermeasures) и реализованная с использованием SWRL-правил в Protege. Авторы демонстрируют, как на основе контекста (тип пользователя, архитектура, протокол) можно выполнять логический вывод для рекомендации соответствующих контрмер. Данный подход учитывает доменную специфику на примере здравоохранения, однако не интегрируется с открытыми базами данных.

Наряду с академическими исследованиями, важную роль в стандартизации играют официальные спецификации, разработанные международными организациями. Эталонная онтология SAREF (Smart Applications REference) [20], разработанная Европейским институтом телекоммуникационных стандартов (ETSI), представляет собой пример успешной стандартизации в области Интернета вещей, обеспечивая семантическую совместимость за счёт модульной структуры и согласования с существующими стандартами, такими как W3C® Semantic Sensor Network (SSN) и oneM2M. Архитектурный принцип расширяемости SAREF позволяет создавать доменно-специфичные модули, что концептуально близко к вводимому в настоящей работе понятию «домена функционирования». Однако SAREF не охватывает аспекты информационной безопасности. Другой важной инициативой является онтология безопасности Web of Things (WoT) Security Ontology [21], представляющая собой стандартизированный словарь для описания механизмов аутентификации и авторизации, включающий классы для различных схем безопасности (OAuth2, Basic, Bearer, PSK и др.). Вместе с тем она ограничена описанием требований к безопасности и не охватывает моделирование угроз, уязвимостей или рисков.

Параллельно с развитием онтологических моделей формировалось направление исследований, ориентированное на анализ и оценку защищённости с использованием открытых баз знаний. В работе [22] предложена иерархия статических и динамических показателей, объединяющая оценки рисков нарушения CIA, рисков на основе анализа политик конфиденциальности и рисков на основе анализа APK-файлов. Ключевой особенностью является разработка оригинальных алгоритмов вычисления интегральной оценки с нелинейным логарифмическим масштабированием, что позволяет не «усреднять» показатели, а сохранять вклад наиболее критичного из них. При этом авторы для анализа защищённости используют данные только об APK-файлах, не учитывая полную архитектуру системы. В [23] предлагается онтологическая модель, автоматизирующая расчет интегральных метрик защищённости в реальном времени и динамически адаптирующаяся к домену через корреляцию событий и подключение баз знаний, включая матрицы MITRE ATT&CK и ATT&CK for ICS. В работе [24] предложен подход к оцениванию защищённости гетерогенных инфраструктур на основе графов атак, формируемых с использованием баз NVD и MITRE ATT&CK, и реализован расчёт вероятностей реализации атак с применением байесовского вывода. В [25] был представлен масштабный обзор угроз безопасности Интернета вещей, систематизировав атаки по уровням архитектуры и детально проанализировав анатомию вредоносного программного обеспечения, что сформировало требования к полноте разрабатываемых моделей для систем Интернета вещей. Важность учёта контекста применения системы при анализе защищённости подчеркивается в работе [26], в которой при обзоре проблем безопасности робототехники явно выделены различные домены применения (медицина, промышленность, энергетика и т.д.) и показано, что последствия реализации атак существенно различаются в зависимости от домена. Однако предложенный авторами анализ носит описательный характер и не предлагает формального способа учёта домена.

Анализ существующих онтологических решений в области Интернета вещей и информационной безопасности показал, что большинство рассмотренных онтологий ориентированы на решение частных задач и не обеспечивают комплексную поддержку анализа защищённости систем Интернета вещей. Существенная часть онтологий предназначена для описания устройств, протоколов взаимодействия и сервисов Интернета вещей и используется преимущественно для обеспечения совместимости и интеграции компонентов. В таких моделях аспекты информационной безопасности либо отсутствуют, либо представлены без формализованных связей

с архитектурой системы. Отдельную группу составляют онтологии, направленные на представление знаний об уязвимостях и атаках. Эти решения позволяют структурировать сведения о типах уязвимостей и сценариях атак, однако, как правило, не учитывают архитектурные особенности Интернета вещей и не связывают угрозы с компонентами объекта. В результате идентификация угроз осуществляется вне контекста реальной конфигурации системы, что ограничивает применимость таких онтологий для анализа защищённости.

В таблице 1 представлено сравнение работ, в которых онтологическая модель объединяет понятия Интернета вещей и информационной безопасности.

Таблица 1. Сравнительный анализ работ

Работа	Охват архитектуры	Концепты модели	Интеграция с базами знаний	Механизм вывода	Учёт домена и целей системы	Наличие межкомпонентного взаимодействия
[1]	Нет	Static CIA score, APK based score, Privacy policy score, Readability score, Dynamic CIA/privacy score	NVD, CPE, CVE	Нет	Косвенно	Нет
[15]	Да	Asset, Attack, Fault, SecurityGoal, SecurityMechanism, Vulnerability, Capability, Criticality, Risk	IICSF, INSSA	Потенциально	Нет	Да
[16]	Да	Vulnerability, Best Security Practice, System Component, Stakeholder	CVE, IEEE Sensors Registry	SPARQL-запросы	Косвенно	Да
[17]	Да	Message, Event, Security Threat, Device, Network Parameter	Нет	Да	Нет	Нет
[19]	Да	TypeOfContext, Threat, SecurityAsset, SecurityProperty, SecurityMechanism, ArchitectureType, Protocol	STAC, LOV4IoT	SWRL	Да	Да
[20]	Да	Device, Function, Command, State, Property, Service, Measurement, Task, Commodity, Profile, FeatureOfInterest	SSN, oneM2M, Time Ontology, OM	OWL (подразумевается)	Да	Да
[21]	Да	No, Auto, Basic, Digest, APIKey, Bearer, PSK, OAuth2, Combo	OpenAPI, IETF RFC	RDF/SPARQL	Косвенно	Да
[22]	Да	Confidentiality, Integrity, Availability	CVE, CPE, NVD	Нет	Косвенно	Да

Продолжение таблицы 1. Сравнительный анализ работ

[25]	Да	Уровни Интернета вещей, Угрозы, Уязвимости, Атаки, Malware, Ботнеты, Контрмеры	NVD (CVE), CERT	Не рассматривается	Упоминается, но не используется	Да
[26]	Да	Домены, Уязвимости, Атаки, Риски, Контрмеры	NVD (CVE)	Не рассматривается	Да	Да

В существующих исследованиях отсутствует комплексная семантическая модель анализа защищенности систем Интернета вещей, которая одновременно учитывала бы (1) топологическую и функциональную связность компонентов, (2) распространение угроз по этим связям, включая явные и неявные логические зависимости, (3) функциональные цели системы и то, как их нарушение связано с компрометацией компонентов. Кроме того, отсутствует целостная онтологическая цепочка «компонент – его уязвимость – атака на компонент – распространение угрозы на связанные компоненты – риск». Без такой цепочки невозможно формализовать каскадные эффекты компрометации и, как следствие, оценить, как реализация атаки на отдельном компоненте приводит к нарушению работы системы как целого. Необходимость построения новой онтологической модели обусловлена потребностью преодоления этого разрыва.

4. Разработка онтологической модели информационной безопасности систем Интернета вещей. Предлагаемая онтологическая модель формализует знания в области безопасности Интернета вещей, объединяя архитектурные характеристики систем, сведения об уязвимостях и атаках из общедоступных баз данных, а также контекстную информацию о домене функционирования объекта, целях системы, зависимостях между компонентами и потоках сообщений. Ключевое отличие модели – ориентация на киберфизическую природу Интернета вещей с учётом динамики информационного обмена и каскадных эффектов.

На основе анализа архитектуры систем Интернета вещей [2, 6] выделены ключевые понятия предметной области. Рисунок 2 представляет концептуальную модель, которая задает каркас для построения онтологической модели, детально описанной в разделах 4.1-4.5.

Данная концептуальная модель отражает предметную область Интернета вещей и информационной безопасности, взаимосвязи между ними, а также сохраняет преемственность по отношению к базовой архитектуре системы (Рис. 1). Модель наглядно показывает, как деструктивные воздействия влияют на каждый уровень системы, а также описывает цепочку «уязвимость – атака – угроза – риск – ущерб», интегрируя технические аспекты безопасности с их конечными

последствиями. Основными концептами модели являются «Объект» (совокупность уровней системы), «СФХ» (компоненты каждого уровня), «Уязвимость», «Атака», «Риск», «Угроза». В рамках данной работы под объектом понимается система Интернета вещей.

В рамках предлагаемой концептуальной модели именно угроза является центральным, системообразующим понятием, которое придает смысл и контекст всем остальным элементам. Угрозо-ориентированный подход не исключает анализ уязвимостей и атак, а интегрирует их, обеспечивая причинно-следственную связь. Его ключевое преимущество заключается в том, что угроза представляет собой комплексное понятие, включающее не только технические аспекты (атаки и уязвимости), но и контекст реализации, потенциальные последствия и особенности защищаемого объекта. Именно эта способность к интеграции делает угрозу центральным концептом модели, преобразующим набор разрозненных данных (перечень уязвимостей, классификация атак) в целостную картину безопасности системы Интернета вещей.

Определение основ и структуры концептуальной модели Интернета вещей с точки зрения безопасности позволяет перейти к построению онтологической модели, объединяющей предметные области Интернета вещей и информационной безопасности (Рис. 3).

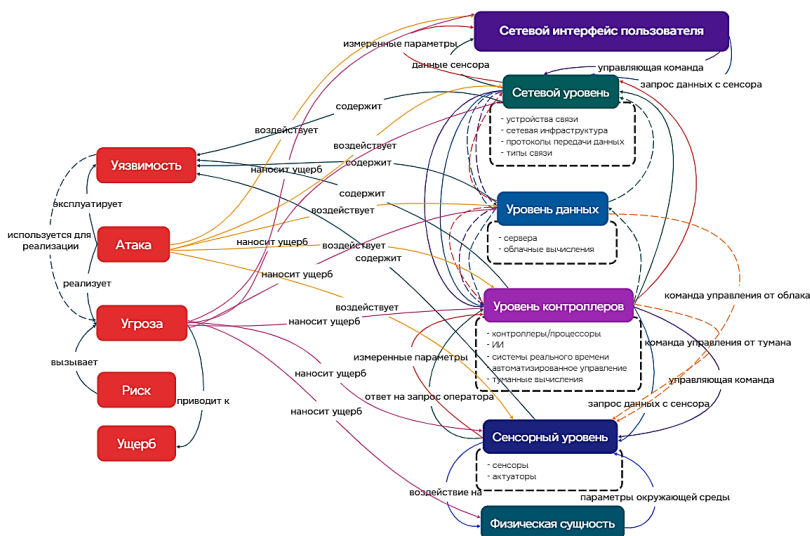


Рис. 2. Концептуальная модель системы Интернета вещей с точки зрения безопасности

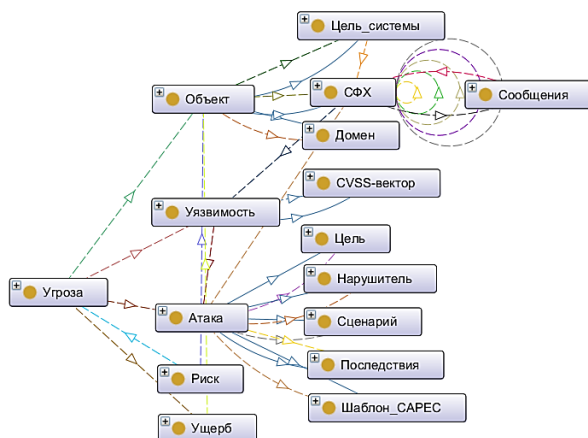


Рис. 3. Ключевые концепты онтологической модели, объединяющей предметную область Интернета вещей и информационной безопасности

Для формального представления структуры разрабатываемой онтологической модели определим её как кортеж, задающий сигнатуру знаний:

$$\Sigma = \langle C, R, D, A \rangle, \quad (1)$$

где C – конечное множество классов (концептов) предметной области, R – множество объектных свойств (бинарных отношений), связывающих экземпляры, D – множество свойств-данных, связывающих экземпляры с литеральными значениями, A – множество аксиом, определяющих семантику классов и свойств.

В рамках определённого таким образом базиса Σ содержательное наполнение онтологии задаётся путём спецификации подмножеств классов, свойств и аксиом, описывающих ключевые концепты предметной области. В частности, вводятся следующие онтологические соглашения, связывающие понятия Интернета вещей и информационной безопасности.

Система Интернета вещей представлена классом объект (Object). Каждый экземпляр этого класса характеризуется областью функционирования, – с этой целью в онтологию введён класс «домен» (Domain), связанный с объектом отношением hasDomain. Любой объект состоит из множества компонентов – экземпляров класса СФХ (SFC), – соединённых с объектом отношением hasSFC.

Система Интернета вещей всегда предназначена для достижения некоторого конечного результата, что формализовано введением класса «цель системы» (SystemGoal). Объект связан с одной или несколькими целями через отношение hasGoal, а цель, в свою очередь, реализуется конкретными СФХ (отношение realizedBySFC).

Для моделирования информационного обмена между компонентами введён класс «сообщение» (Message). В отличие от прямых объектных связей между структурно-функциональными характеристиками, которые фиксируют исключительно факт и направление взаимодействия, сообщения как самостоятельные сущности позволяют учесть семантику передаваемой информации, а именно – тип данных и целевое назначение передаваемого потока. Каждое сообщение характеризуется отправителем и получателем (объектные свойства hasSrc и hasDst соответственно), а также атрибутами, определяющими тип и роль сообщения в функционировании системы. Выделение сообщений в качестве отдельных элементов онтологической модели имеет принципиальное значение для анализа распространения угроз. Компрометация компонента-отправителя не означает автоматической компрометации всех связанных с ним компонентов; вместо этого характер деструктивного воздействия, передаваемого по цепочке взаимодействия, определяется типом сообщения. Так, подмена данных мониторинга и искажение управляющей команды представляют собой качественно различные виды воздействия, приводящие к различным последствиям для целевых функций системы. Таким образом, сообщения выступают связующим звеном, через которое первичная угроза преобразуется в производные угрозы, классифицированные в соответствии с семантикой взаимодействия.

Кроме того, между СФХ определены прямые объектные отношения, отражающие физические и логические связи в системе:

- connectsTo – обобщённая связь (родительское свойство);
- sendDataTo – передача данных;
- controls – управляющее воздействие;
- receivesFrom – обратное свойство к sendDataTo.

Данные отношения задают базовую архитектуру системы, а их совместное использование с сообщениями обеспечивает возможность моделирования каскадного распространения угроз не на основе простого отношения связности, а с учётом типа взаимодействия и содержания передаваемой информации. Компрометация одного компонента порождает производные угрозы для связанных с ним элементов, при этом характер порождаемой угрозы

определяется семантикой связи и типом сообщения, что позволяет дифференцировать последствия каскадного эффекта в зависимости от функциональной роли затронутых компонентов.

Каждая СФХ может содержать слабые места, для их представления служит класс «уязвимость» (Vulnerability), связанный с классом SFC отношением hasVulnerability. Каждая уязвимость может быть проэксплуатирована атакой – этот факт фиксируется при помощи класса «атака» (Attack) и отношения exploitedBy между атакой и уязвимостью. Успешная реализация атаки приводит к возникновению угрозы (Threat), которая связывается с атакой отношением realizedBy. Наконец, риск (Risk) является вероятностью реализации угрозы (свойство causedByThreat) с последующим ущербом (свойство leadsToDamage).

Область исследования ограничена угрозами от целенаправленных атак нарушителя, что обосновано спецификой Интернета вещей, где такие атаки представляют наихудший сценарий с потенциальным физическим ущербом. В контексте разрабатываемой онтологической модели нарушитель рассматривается как неотъемлемый атрибут атаки, что позволяет сохранить иерархическую стройность модели без потери семантической полноты. Исключение случайных событий (ошибок, отказов) необходимо для избежания громоздкости моделей и обеспечения последовательности угроз-ориентированного подхода.

Далее для компактной формальной записи структуры классов и связей в онтологии используется нотация, приближенная к синтаксису языка описания онтологий *Ontology Web Language* (OWL). В данной нотации символ $\subseteq$ означает «подкласс» (subClassOf), $\exists$ – экзистенциальное ограничение «существует хотя бы один», $\cap$ – пересечение классов (логическое «И»), а запись вида $\exists R.C$ соответствует наличию отношения R хотя бы с одним экземпляром класса C . Объектные свойства обозначаются именем свойства (например, $\exists \text{hasSFC.SFC}$). Данные обозначения используются исключительно для формального описания онтологии.

Онтологическая модель расположена на веб-сервисе GitHub [27].

4.1. Объект. Угроза направлена на объект, который характеризуется областью функционирования, определяющей предметную область его применения, целью, к достижению которой стремится система, а также совокупностью СФХ, необходимых для достижения целей системы и выступающих первичным фактором уязвимости, архитектуры и функциональных возможностей, и информационными потоками (сообщениями) между СФХ, включая связи управления и передачи данных, что позволяет моделировать

каскадное распространение угроз. Формально «Объект» можно представить следующим образом:

$$Object \subseteq (= hasDomain.Domain) \cap (\exists hasSFC.SFC) \cap (\exists hasGoal.SystemGoal), \quad (2)$$

где *Object* – класс системы Интернета вещей, *Domain* – класс домена (области функционирования), *SFC* – класс структурно-функциональных характеристик, *SystemGoal* – класс цели системы; *hasDomain*, *hasSFC*, *hasGoal* – объектные свойства, связывающие объект с соответствующими сущностями. Данное ограничение утверждает, что каждый экземпляр объекта обязан иметь один домен, хотя бы одну СФХ и хотя бы одну цель.

Объекту сопоставляется свойство «уровень критичности», которое отображает процент нарушения работоспособности объекта, функционирующего в определенной области, в случае отказа СФХ, реализующих определенную цель объекта:

$$S_{Object} = [S_D, S_G, S_{SFC}], \quad (3)$$

где S_{Object} – уровень критичности объекта, S_D – уровень критичности домена, S_G – уровень критичности цели объекта, S_{SFC} – максимальная опасность от каждой узвимой СФХ.

На рисунке 4 представлена модель «Объекта».

Цели системы определены на основе анализа стандартов в области Интернета вещей [2, 6, 28 – 31]. Домены выделены в соответствии со сферами функционирования субъектов критической информационной инфраструктуры Российской Федерации согласно [32], что придает модели практическую значимость. Для определения СФХ задействованы параметры из стандарта ETSI TS 103 264 [33] и онтологии SAREF [20]. Использование стандарта ETSI TS 103 264 позволяет привести модель в соответствие с глобальными архитектурными подходами к Интернету вещей, а применение онтологии SAREF обеспечивает совместимость с существующими решениями в этой области.

Как было сказано ранее, между СФХ были определены отношения, отражающие физические и логические связи в системе (Рис. 5).

Представленная схема отражает транзитивный характер связей между компонентами, что позволит в последующем выявлять каскадные угрозы. Например, компрометация контроллера влечет автоматическое распространение угрозы на все связанные актуаторы.



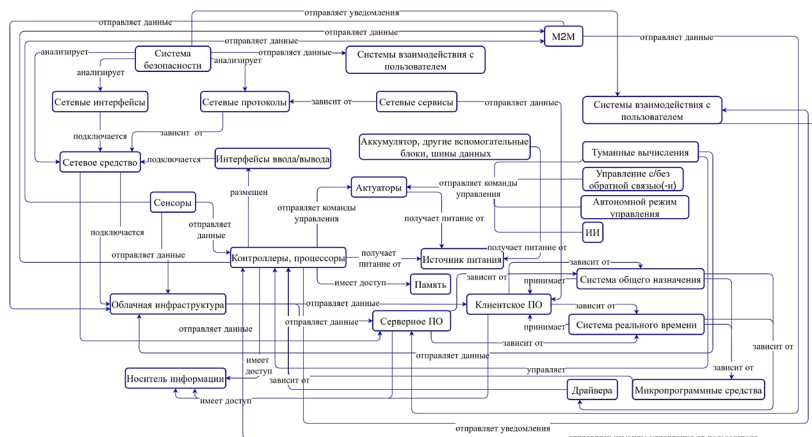


Рис. 5. Схема связей между СФХ

4.2. Уязвимость. Уязвимость рассматривается как недостаток или слабость реализации СФХ объекта, эксплуатация которой возможна в процессе атаки и приводит к формированию угрозы. Формально «Уязвимость» определяется как:

$$Vu l n e r a b i l i t y \subseteq (\exists h a s V u l n e r a b i l i t y ^{-}. S F C) \cap (\exists e x p l o i t e d B y . A t t a c k), \quad (4)$$

где $h a s V u l n e r a b i l i t y ^{-}$ – инверсное свойство к $h a s V u l n e r a b i l i t y$, связывающее уязвимость с компонентом, $e x p l o i t e d B y$ – «эксплуатируется атакой», $S F C$ – класс СФХ; $A t t a c k$ – класс атак.

Источниками данных об уязвимостях выбраны базы CVE [34] и RVD [35]. Обращение к репозиторию CVE обеспечивает методологическую совместимость результатов с устоявшимися практиками в области информационной безопасности и создает предпосылки для корректной интеграции разработанной модели с иными отраслевыми стандартами. Включение базы RVD позволяет охватить спектр специфичных для систем Интернета вещей уязвимостей, которые могут не отражаться в общих реестрах. Существенным преимуществом выступает наличие механизма соотнесения записей RVD с идентификаторами CVE, что гарантирует непротиворечивость и сравнительную сопоставимость данных в ходе аналитической обработки.

Описание уязвимостей осуществляется с использованием векторов оценки CVSS [36] или RVSS, позволяющих формализовать их характеристики. Данный выбор обусловлен тремя ключевыми

факторами: (1) наличием формализованных, машиночитаемых форматов и API, что критически важно для автоматизации последующих процессов ETL (Extract, Transform, Load); (2) широкой распространённостью и признанием в академическом и индустриальном сообществе, что обеспечивает воспроизводимость и сравнимость результатов с мировыми аналогами; (3) прямыми семантическими связями между базами (CVE – CWE – CAPEC), что позволяет строить формальные цепочки логического вывода. В соответствии с данными базы CVE уязвимости сопоставляется свойство уровень опасности, отражающий степень потенциального воздействия на объект:

$$S_{Vulnerability} = \{m_E, m_I, m_S\}, \quad (5)$$

где $S_{Vulnerability}$ – уровень опасности уязвимости, m_E, m_I, m_S – базовые метрики уязвимости в соответствии с CVSS.

Наряду с международными источниками, разработанная онтологическая модель ориентирована на совместимость с нормативной базой Российской Федерации. В частности, для оценки уровня критичности уязвимостей может быть использована «Методика оценки уровня критичности уязвимостей программных, программно-аппаратных средств» [37]. Данная методика устанавливает порядок оценки и обязательна к применению операторами информационных систем, в том числе объектов критической информационной инфраструктуры. Она определяет уровень критичности уязвимости на основе базовой оценки по шкале CVSS, дополненной контекстными метриками. Таким образом, предлагаемая модель не противопоставляет международный и отечественный подходы, а обеспечивает возможность работы с обоими.

4.3. Атака. Реализация угрозы осуществляется посредством атаки, эксплуатирующей уязвимость. Атака описывается через набор ключевых атрибутов, включающих субъект реализации (нарушитель), целевой компонент (цель), последовательность действий (сценарий) и потенциальные результаты (последствия). Формально «Атака» определяется в следующем виде:

$$Attack \subseteq (\exists exploitedBy^-.Vulnerability) \cap (\exists hasTarget.SFC) \cap (\exists realizedBy^-.Threat), \quad (6)$$

где $exploitedBy^-$ – «эксплуатирует уязвимость», $hasTarget$ – «нацелена на СФХ», $realizedBy^-$ – обратное к $realizedBy$, означающее, что атака реализует угрозу.

В качестве источника данных об атаках выбран каталог шаблонов атак CAPEC [38], который имеет семантические связи с базой уязвимостей CVE, что позволяет проследивать корреляцию между конкретными уязвимостями и потенциальными векторами их эксплуатации. Кроме того, CAPEC совместим с каталогом CWE, что повышает точность кросс-анализа данных.

В соответствии с каталогом CAPEC атака характеризуется вероятностью реализации, определяющей успешность её выполнения нарушителем, а также уровнем влияния, отражающим масштаб последствий при успешной реализации, которые в совокупности формируют свойство «уровень критичность атаки»:

$$S_{Attack} = [L_{Attack}, I_{Attack}], \quad (7)$$

где S_{Attack} – уровень критичности атаки, L_{Attack} – вероятность реализации атаки (заимствованный параметр «Likelihood of Attack» из CAPEC), I_{Attack} – уровень влияния атаки (заимствованный параметр «Typical Severity» из CAPEC).

На рисунке 6 представлена модель «Атаки».

В качестве целей атак выбраны ключевые параметры безопасности Интернета вещей, выделенные в эталонной архитектуре [2]. Типология нарушителей в разработанной модели соответствует подходам, зафиксированным в Методике оценки угроз безопасности информации ФСТЭК России от 5 февраля 2021 г. [39]. За основу последствий реализации атак взята классификация «ConsequencesType» из каталога CAPEC. Однако предложенная модель не ограничивается исключительно стандартизированными категориями: она дополнена перечнем потенциальных последствий, выявленных в ходе анализа исследований [31, 40 – 43]. Такое расширение таксономии позволяет оценивать воздействие атак не только на технологическую составляющую системы, но и на конечных пользователей.

Хотя разработанная онтологическая модель фокусируется на технических аспектах атак, для обеспечения совместимости с существующими методологиями в нее заложена базовая типизация нарушителя. В текущей реализации атрибуты нарушителя не участвуют в правилах логического вывода для идентификации угроз, однако класс «Нарушитель» введен в модель. Типология нарушителей соответствует подходам, зафиксированным в Методике оценки угроз безопасности информации ФСТЭК России. На текущем этапе исследования, посвященном идентификации наличия угроз, учет нарушителя не является критическим, так как наличие задокументированного

шаблона атаки (CAPEC) уже свидетельствует о принципиальной возможности ее реализации.



Рис. 6. Модель «Атаки»

4.4. Угроза. Угроза определяется как потенциальное событие, при котором нарушитель эксплуатирует уязвимости СФХ объекта, функционирующего в определенной области, путём реализации атаки, что приводит к негативным последствиям как для объекта, так и для окружающей среды и человека. Формально «Угроза» определяется в следующем виде:

$$Threat \subseteq (\exists affectObject. Object) \cap (\exists realizedBy. Attack) \cap (\exists causedByVuInerability. VuInerability), \quad (8)$$

где *affectObject* – «воздействует на объект», *realizedBy* – «реализуется атакой», *causedByVulnerability* – «обусловлена уязвимостью».

В данной работе СФХ являются параметрами безопасности. Угроза существует для объекта, если СФХ имеет уязвимости и возможность проэксплуатировать эти уязвимости путём реализации атаки [44]. При этом следует учитывать, что отсутствие известного вектора атаки в текущий момент не гарантирует защищённости в долгосрочной перспективе: уязвимость может быть потенциально опасной, если для нее теоретически возможна эксплуатация, даже если соответствующий шаблон атаки еще не зафиксирован в общедоступных базах знаний. В разработанной онтологической модели данный аспект учитывается путем классификации угроз по статусу актуальности. Угроза считается актуальной, если для неё существует известная атака, и потенциальной – при наличии уязвимости в отсутствие известного эксплойта.

Формализация процедуры идентификации угроз базируется на анализе онтологических взаимосвязей, установленных между СФХ объекта, уязвимостями и атаками. В качестве основы для идентификации угроз используется банк данных угроз безопасности информации ФСТЭК России (БДУ ФСТЭК) [45], где для каждой угрозы определены объекты воздействия. В контексте настоящего исследования такими объектами воздействия выступают СФХ.

Угроза имеет свойство «уровень опасности», которое указывает на потенциальную возможность нарушения безопасности и нанесения ущерба объекту, что определяется вероятностью эксплуатации слабых мест СФХ системы нарушителем путём реализации атак [46]:

$$S_{Threat} = [P, SFC], \quad (9)$$

где S_{Threat} – уровень опасности угрозы, $P = P(Vulnerability, Attack)$ – вероятность эксплуатации слабых мест атак.

4.5. Риск. Риск интерпретируется как сочетание вероятности реализации угрозы и ожидаемого ущерба от её реализации [46]. Формально «Риск» определяется в следующем виде:

$$Risk \subseteq (\exists causedByThreat.Threat) \cap (\exists leadsToDamage.Damage), \quad (10)$$

где *causedByThreat* – «вызван угрозой», *leadsToDamage* – «приводит к ущербу».

Риск имеет свойство «уровень критичности», отражающее степень негативного воздействия угрозы и масштаба потенциального ущерба для объекта:

$$S_{Risk} = [S_{Threat}, Damage], \quad (11)$$

где S_{Risk} – уровень критичности риска.

Ущерб (*Damage*) введён в онтологическую модель для обеспечения целостного и завершённого представления предметной области информационной безопасности систем Интернета вещей. Ущерб определяется как мера потенциального негативного воздействия от реализации угрозы. Для формализации ущерба присваивается свойство «уровень критичности», характеризующее процент тяжести негативных последствий для объекта в контексте его области функционирования. Контекстно-зависимый характер риска является критически важным требованием для гетерогенных сред Интернета вещей, где семантика одного и того же события варьируется в зависимости от сценария использования.

Ввиду отсутствия в открытых источниках готовых классификаций рисков для систем Интернета вещей, в работе проведен анализ авторитетных источников: отчеты ENISA [47], публикации IoT Security Foundation [48], база угроз Лаборатории Касперского [49], а также исследования [31, 40 – 44]. На основе этого анализа выделены классы рисков безопасности, релевантные для систем Интернета вещей (Рис. 7).

Модель представляет собой конечную точку пути от конкретной СФХ до итогового риска. В результате формируется не абстрактный перечень угроз, а множество последовательностей деструктивных воздействий, каждая из которых завершается негативным событием. Перечисленные риски характерны именно для систем Интернета вещей и отличаются от рисков, присущих чисто программным информационным системам, что критически важно при анализе защищённости гетерогенных систем Интернета вещей.

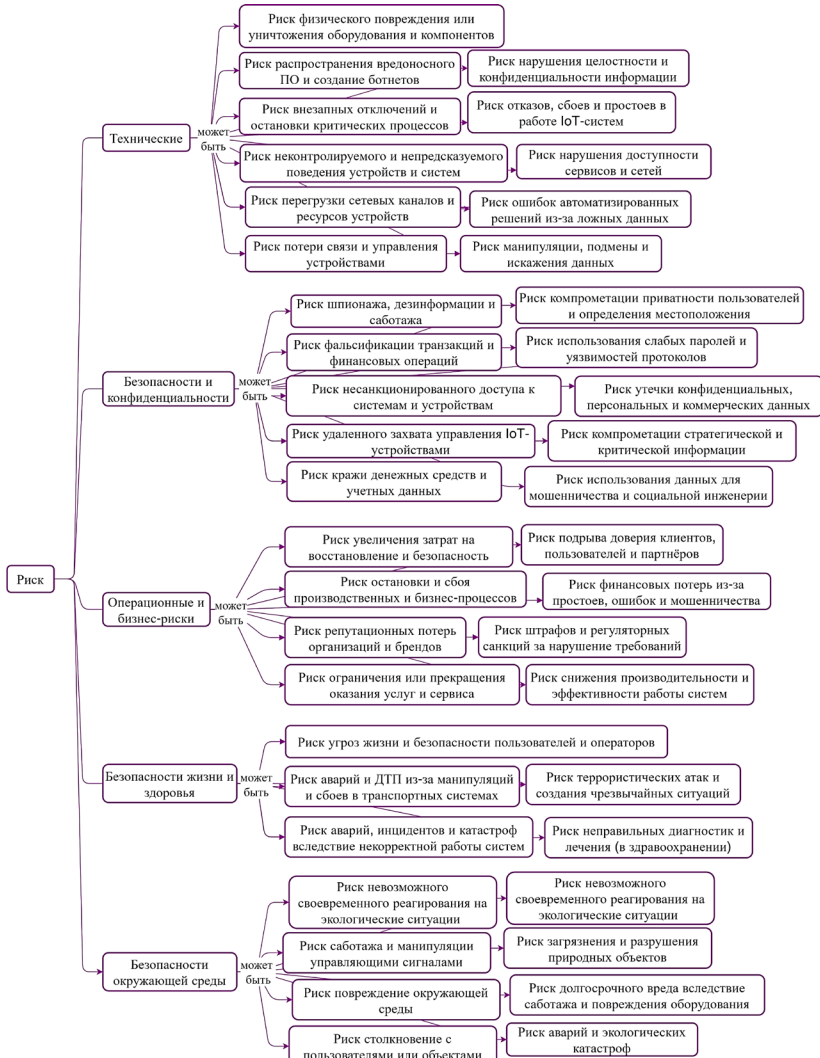


Рис. 7. Модель «Риска»

4.6. Реализация онтологической модели. Использование онтологического подхода обеспечивает не только структурирование информации, но и формализацию семантических связей между элементами предметной области. Это создаёт предпосылки для применения логического вывода и автоматизированных

механизмов анализа защищённости информационной безопасности систем Интернета вещей.

Для реализации разработанной онтологической модели выбрано программное обеспечение Protege 5.6.5 [50]. Для представления знаний используется язык OWL, обеспечивающий формальную семантику и поддержку автоматического логического вывода. На начальном этапе были последовательно созданы классы и подклассы в соответствии с иерархиями, представленными на рисунках 4-6 (Рис. 8). Для описания отношений между классами выделены основные объектные свойства (Object Properties), фрагментарно представленные на рисунке 9, которые формализуют их взаимосвязи. Например:

- hasSFC (Объект $\rightarrow$ СФХ): определяет состав системы;
- hasVulnerability (СФХ $\rightarrow$ Уязвимость): идентифицирует слабые места компонентов;
- exploits (Атака $\rightarrow$ Уязвимость): указывает, какая атака эксплуатирует уязвимость;
- causedByThreat (Риск $\rightarrow$ Угроза): устанавливает причину возникновения риска.

Назначены дата свойства для описания атрибутов классов (Data Properties), которыми являются показатели защищённости каждого класса, позволяющие в будущем перейти от качественного описания к количественной оценке (Рис. 10). Такими свойствами являются уровень критичности (для «Объекта», «Атаки», «Риска») и уровень опасности (для «СФХ», «Уязвимости», «Угрозы»). Все свойства нормализованы в диапазоне [0,2; 1], что обеспечит согласованность будущих расчетов.

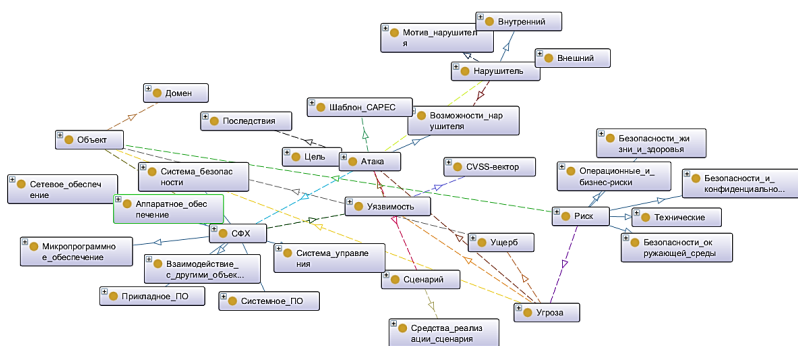


Рис. 8. Фрагмент графического представления классов

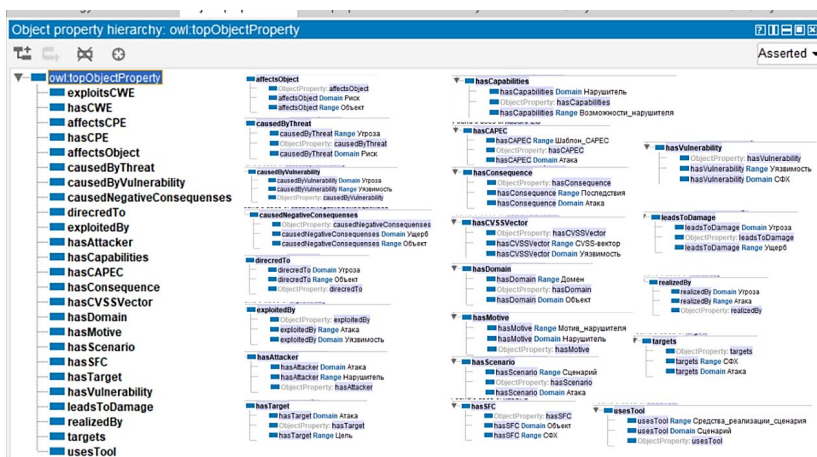


Рис. 9. Назначенные Object Properties

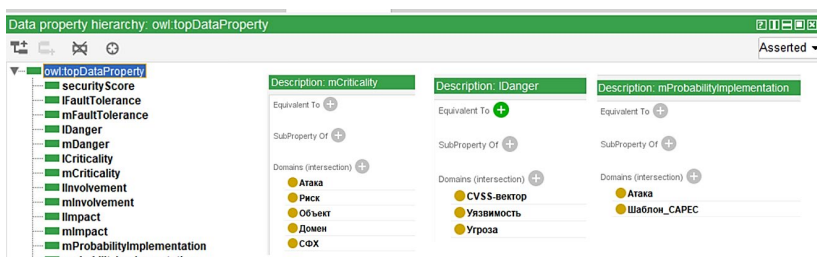


Рис. 10. Назначенные Data Properties

Для отражения ресурсных ограничений компонентов системы Интернета вещей введены дата свойства, которые также характеризуют и объект исследования. Например, для класса СФХ «Актuatorы» введены дата свойства `hasCurrentConsumptionMax`, отражающий напряжение питания, `hasSupplyVoltageMax`, отражающий потребляемый ток, и `hasControlType`, отражающий тип управления (Рис. 11). Данные свойства показывают, что могут использоваться только простые электрические приводы и реле.

Наполнение онтологии экземплярами классов производилось как вручную – с помощью вкладки «Individuals» редактора Protege, так и полуавтоматически – с использованием модуля «Cellfie» для импорта предварительно преобразованных данных. Для наполнения использовались внешние базы знаний: БДУ ФСТЭК, CVE, RVD, CWE, CPE (Common Platform Enumeration) [51], CAPEC, NVD [52]. Поскольку перечисленные источники используют различные

форматы данных, для их интеграции в единое непротиворечивое хранилище применяется метод ETL (Extract, Transform, Load). Ключевым этапом, обеспечивающим совместимость и целостность знаний, является Transform, в рамках которого выполняется установление соответствия (маппинг) исходных данных на единый формат, заданный онтологией. Единый формат описания классов (Рис. 12), выработанный на основе анализа всех используемых источников, гарантирует, что каждый элемент будет представлен в базе знаний с унифицированным набором атрибутов и связей. Это создаёт основу для последующего логического вывода.

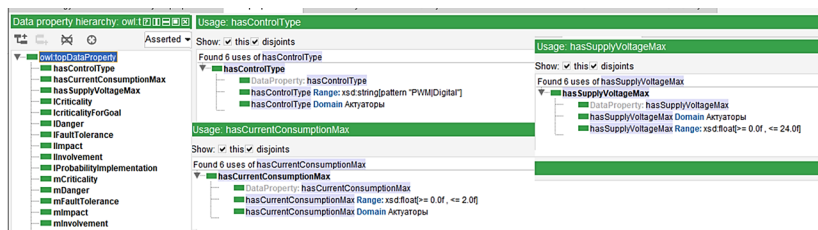


Рис. 11. Назначенные Data Properties для актуаторов

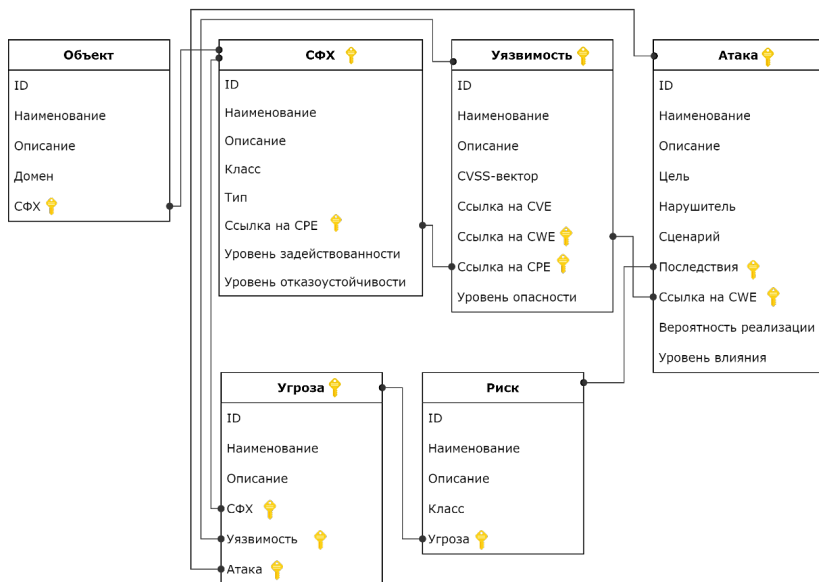


Рис. 12. Единый формат описания классов в базе знаний

Для отражения взаимодействия между СФХ и последующего вывода каскадных угроз созданы объектные отношения, описанные в разделе 4. Пример таких отношений для класса «Контроллеры, процессоры» показан на рисунке 13.

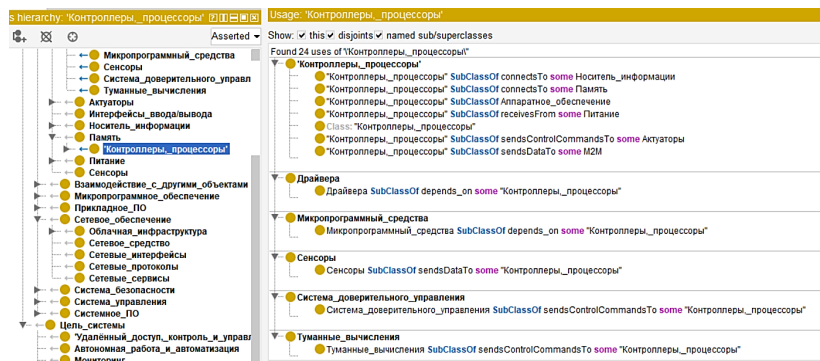


Рис. 13. Назначенные Object Properties для отражения взаимодействия между СФХ на примере класса «Контроллеры, процессоры»

Для демонстрации ETL-процесса ниже приведен пример интеграции записи о программном обеспечении датчика для мониторинга показателей электроэнергии Panasonic KW Watcher, который может быть частью системы Интернета вещей (таблица 2).

Для автоматического вывода связей между сущностями используется язык логических правил Semantic Web Rule Language (SWRL). Правила SWRL позволяют выразить логические зависимости между свойствами и классами онтологии. Правило состоит из двух частей: антецедента (условие) – конъюнкция атомарных утверждений (атомов), и консеквента (следствие) – также конъюнкция атомарных утверждений. Если все атомы в антецеденте истинны (выполняются), то все атомы в консеквенте также должны быть истинны. Логические правила для вывода информации о связанных с СФХ уязвимостях, атаках, угрозах и рисках для объекта приведены ниже.

Правило 1. Вывод связанных с СФХ уязвимостей:
 $IoT:hasCPE(?sfc, ?cpe) \wedge IoT:hasCPE(?vuln, ?cpe) \rightarrow IoT:hasVulnerability(?sfc, ?vuln)$ если СФХ имеет идентификатор уязвимого продукта (CPE), и уязвимость также влияет на этот же самый идентификатор продукта (CPE), то можно сделать вывод, что данная СФХ обладает этой уязвимостью.

Таблица 2. Процессы ETL для наполнения базы знаний

Этап	Действие	Пример для на примере СФХ Panasonic KW Watcher
Extract (Извлечение)	Идентификация и сбор сырых данных из приведённых источников	Извлечение записи для <code>сре:/a:panasonic:control_fwwin_pro:6.414</code> из базы данных CPE
Transform (Преобразование)	Нормализация и установление соответствий с базой знаний. Данные из каждого источника преобразуются к единому формату, определяемому классами и свойствами онтологии	Вход: Запись <code>сре:/a:panasonic:control_fwwin_pro:6.414</code> из CPE Процесс: 1. Создается новый экземпляр класса «СФХ» 2. Заполнение свойств: - Идентификатор = «CPE-ID» - Наименование = «Panasonic KW Watcher» - Описание = «Description» (текст из CPE) - Класс = «Прикладное ПО» - Тип = «Клиентское ПО» - Ссылка на CPE = <code>«сре:/a:panasonic:kw_watcher:-»</code> - Уровень задействованности = «0,2» - Уровень отказоустойчивости = «0,5» 3. Устанавливается связь <code>hasVulnerability</code> с соответствующими экземплярами класса
Load (Загрузка)	Сохранение в базу знаний. Преобразованные и связанные данные загружаются в базу знаний	Структурированный экземпляр уязвимости CVE-2023-3472 со всеми свойствами и связями сохраняется в базе знаний

Правило 2. Вывод связанных с уязвимостью атак: $IoT:hasCWE(?vuln, ?cwe) \wedge IoT:exploitsCWE(?attack, ?cwe) \rightarrow IoT:exploits(?attack, ?vuln)$ если уязвимость имеет идентификатор слабого места (CWE), и атака также имеет этот же самый идентификатор слабого места (CWE), то можно сделать вывод, что данная атака эксплуатирует данную уязвимость.

Правило 3. Вывод угроз: $IoT:hasSFC(?obj, ?sfc) \wedge IoT:hasVulnerability(?sfc, ?vuln) \wedge IoT:exploits(?attack, ?vuln) \rightarrow IoT:hasThreat(?obj, ?threat) \wedge IoT:comprises(?threat, ?sfc, ?vuln, ?attack)$ если объект содержит СФХ, И эта СФХ имеет уязвимость, и существует атака, которая эксплуатирует данную уязвимость, то можно сделать вывод, что для объекта возникает угроза, которая представляет собой

комбинацию уязвимого компонента, конкретной уязвимости и реализующей её атаки.

Правило 4. Вывод каскадных угроз через связи СФХ: $IoT:comprises(?threat, ?sfc1, ?vuln, ?attack) \wedge IoT:connectsTo(?sfc1, ?sfc2) \rightarrow IoT:comprises(?threat, ?sfc2, ?vuln, ?attack)$ если угроза уже связана с компонентом СФХ1, и этот компонент имеет связь connectsTo с компонентом СФХ2, то угроза распространяется и на СФХ2.

Правило 5. Вывод каскадного распространения угроз через потоки сообщений: $IoT:comprises(?threat, ?sfc1, ?vuln, ?attack) \wedge IoT:hasSrc(msg, ?sfc1) \wedge IoT:hasDst(msg, ?sfc2) \rightarrow IoT:threatensMessage(?threat, msg) \wedge IoT:comprises(?threat, ?sfc2, ?vuln, ?attack)$ если скомпрометирован СФХ, который является отправителем сообщения, то под угрозой оказывается само сообщение, и эта же угроза каскадно проецируется на СФХ-получатель.

Правило 6. Связь угрозы компонента с целью системы: $IoT:comprises(?threat, ?sfc, ?vuln, ?attack) \wedge IoT:realizedBySFC(?goal, ?sfc) \rightarrow IoT:threatensGoal(?threat, ?goal)$ если угроза ассоциирована с конкретной СФХ, и данная СФХ задействована в реализации целевой функции системы, то данная угроза качественно классифицируется как компрометирующая цель системы.

Правило 7. Вывод рисков: $IoT:hasSFC(?obj, ?sfc) \wedge IoT:hasThreatForComponent(?sfc, ?threat) \wedge IoT:hasProbabilistic(?threat, ?risk) \rightarrow IoT:hasRisk(?obj, ?risk)$ если объект содержит СФХ, и для СФХ существует угроза (связка из правила 3), и для этой угрозы есть связанный с ней риск, то можно сделать вывод, что для объекта существует данный риск.

Рассмотрим синтаксис на примере Правила 1: $IoT:hasCPE(?sfc, ?cpe) \wedge IoT:hasCPE(?vuln, ?cpe) \rightarrow IoT:hasVulnerability(?sfc, ?vuln)$. Здесь ?sfc, ?cpe, ?vuln – переменные, связывающие экземпляры. Часть до стрелки ($\rightarrow$) – антецедент (условие «если»), а после – консеквент (вывод «то»). Правило срабатывает для всех экземпляров онтологии, для которых истинен антецедент.

Выполнение правил инициируется запуском механизма логического вывода резонера, такого как Pellet, непосредственно в среде Protege. Резонер в автоматическом режиме проверяет аксиомы онтологии и правила, добавляя выведенные отношения между экземплярами, например, утверждение о наличии уязвимости у конкретной СФХ. Последовательность применения правил определяется самим резонером на основе имеющихся фактов, что и обеспечивает автоматизированный вывод.

Представленные правила формализуют цепочки логических выводов, позволяя на основе базовых фактов о конфигурации системы и общеизвестных баз знаний автоматизированно выводиться скрытые логические связи о потенциальных угрозах и рисках.

Разработанная в настоящей работе онтологическая модель отличается от существующих решений тем, что ориентирована не только на структурирование знаний, но и на их использование в процессе анализа защищённости систем Интернета вещей. Ключевой особенностью модели является явное представление архитектуры системы Интернета вещей через СФХ и выделение домена для определения контекста функционирования объекта, а также выделение рисков, специфичных для Интернета вещей, и установление формализованных связей между компонентами, уязвимостями, атаками, угрозами и рисками. Такой подход позволяет учитывать архитектурно-зависимый характер угроз и обеспечивает идентификацию только тех угроз, которые актуальны для конкретной конфигурации системы.

Важным свойством разработанной модели является возможность её расширения за счёт добавления новых классов, атрибутов и отношений без нарушения целостности уже сформированного представления знаний. Это позволяет адаптировать онтологию к различным предметным областям применения Интернета вещей, а также учитывать появление новых типов устройств, протоколов взаимодействия и угроз информационной безопасности. Кроме того, модульная структура онтологии упрощает её интеграцию с внешними источниками знаний и средствами анализа безопасности.

4.7. Место онтологической модели в процессе анализа защищённости. Разработанная онтологическая модель не является самостоятельным средством количественной оценки защищённости. Она выполняет роль семантического каркаса, который формализует и связывает в единую структуру следующие элементы:

- архитектуру системы Интернета вещей (через классы Object, SFC, Domain и отношения hasSFC, hasDomain);
- известные уязвимости (класс Vulnerability с данными из CVE, RVD);
- шаблоны компьютерных атак (класс Attack с данными из CAPEC);
- возникающие угрозы безопасности информации (класс Threat);
- потенциальный ущерб и риски (классы Damage, Risk).

Модель выводит связи между этими элементами с помощью SWRL-правил. Полученный в результате граф знаний, в котором каждый компонент системы (SFC) явно связан с актуальными для него

уязвимостями, а те, в свою очередь, – с конкретными атаками и итоговыми угрозами.

Этот граф знаний служит формализованным входом для последующих этапов анализа защищённости, в частности:

- количественной оценки защищённости: на основе выведенных связей могут быть рассчитаны вероятности реализации угроз (например, с использованием байесовских сетей или марковских процессов), а также интегральные показатели защищённости с учётом критичности компонентов;

- расчёта рисков: классы Risk и Damage в модели предоставляют необходимую структуру для применения стандартных методик оценки рисков (в том числе совместимых с подходами ФСТЭК России);

- обоснованного выбора контрмер: зная, какие угрозы связаны с какими компонентами, оператор может целенаправленно выбирать средства защиты, минимизируя затраты.

Важно подчеркнуть, что сами количественные расчёты (вычисление численных значений вероятностей, ущерба, интегральных показателей) выходят за рамки данной работы. Их реализация требует привлечения дополнительных математических методов и является отдельной задачей. Однако без предложенной онтологии любые такие расчёты будут опираться на неформализованные или неполные данные о связях между компонентами, уязвимостями и атаками.

Следует подчеркнуть, что представленные в разделах 4.1-4.5 математические выражения формируют теоретический и алгоритмический базис для аналитического программного модуля. В рамках настоящего этапа работы разработанная онтологическая модель решает критически важную задачу верхнего уровня – агрегацию гетерогенных данных и построение базы знаний, которая выступает в качестве структурированного источника входных данных для последующих автоматизированных количественных расчётов

Таким образом, модель не подменяет полный цикл анализа защищённости, а создаёт его фундамент – формализованное, воспроизводимое и расширяемое представление знаний об угрозах. Это позволяет отделить логическую идентификацию угроз (что делает онтология) от количественной оценки и принятия решений (что остаётся за рамками, но может быть надстроено на основе результатов работы модели).

5. Оценка качества разработанной онтологии. Для оценки качества разработанной онтологической модели применены метрики методологии OntoMod [53], включающие показатели комплексной структурной сложности (обобщенный показатель, охватывающий

связанность и сложность онтологии), инкапсуляции (степень скрытия внутренней структуры модулей онтологии при сохранении необходимых связей между ними), покрытия (полнота представления ключевых понятий предметной области) и глубины иерархии (уровень детализации классов). Онтология разделена на ключевые тематические модули: «Угроза», «Объект», «Уязвимость», «Атака» и «Риск». Исходные параметры для оценки (таблица 3) определялись на основе онтологических метрик из Protege и результатов SPARQL-запросов, которые включали сведения о классах и их модулях, объектных свойствах, отношениях подклассов и модульной структуре онтологии.

Для расчета комплексной структурной сложности применяется формула:

$$ComplexOnto\ score = 0,25 * (LD + LC + LR + CC_{norm}), \quad (12)$$

где *ComplexOnto score* – комплексная структурная сложность, $LD = \frac{2 \cdot (op+dp)}{n(n-1)}$ – плотность связей, где *op* – количество объектных свойств, *dp* – количество дата-свойств, *n* – общее число концептов в онтологии, $LC = \frac{op+dp}{n}$ – среднее количество связей на концепт, $LR = \frac{op+dp}{op+dp+sc}$ – насыщенность связей, где *sc* – количество подкласс-аксиом, $CC_{norm} = \frac{op+dp+sc-n+2}{n}$ – нормализованная цикломатическая сложность.

На основе данных из таблицы 3 рассчитаем комплексную структурную сложность по формуле (12). Результаты представлены в таблице 4.

Чтобы определить инкапсуляцию для всей онтологии необходимо определить вес (вклад) каждого модуля:

$$\omega_i = \frac{NCMO(m_i) + NRolesMO(m_i) + N_iRolesMO(m_i)}{\sum_{j=1}^k NCMO(m_j) + NRolesMO(m_j) + N_jRolesMO(m_i)}, \quad (13)$$

где ω_i – инкапсуляция каждого модуля, m_i – выделенный модуль онтологии, $NCMO(m_i)$ – количество концептов (классов) в модуле m_i , $NRolesMO(m_i)$ – количество связей (объектных свойств) в модуле m_i , $N_iRolesMO(m_i)$ – количество связей (объектных свойств) с другими модулями, $\sum_{j=1}^k NCMO(m_j) + NRolesMO(m_j) + N_jRolesMO(m_i)$ – общий размер онтологии с учётом всех её модулей.

Таблица 3. Количественные данные об онтологии

Метрика	Угроза	Объект	Уязвимость	Атака	Риск	Общая онтология
Количество концептов (n)	1	49	5	20	7	88
Связи внутри модуля	1	7	2	6	1	
Связи между модулями	10	7	3	3	2	
Объектные свойства (op)						35
Дата-свойства (dp)						27
Подкласс-аксиомы (sc)						111
Экземпляры						209

Таблица 4. Расчёт ComplexOnto score

Показатель	Расчёт
Плотность связей	$LD = (35 + 27)/(88 * (88 - 1)/2) = 0,02$
Среднее количество связей на концепт	$LC = (35 + 27)/88 = 0,7$
Насыщенность связей	$LR = (35 + 27)/(35 + 27 + 111) = 0,4$
Цикломатическая сложность	$CC_{norm} = (35 + 27 + 111 - 88 + 2)/88 = 0,9$
Комплексная структурная сложность	$ComplexOnto = 0,25 * (0,02 + 0,7 + 0,4 + 0,9) = 0,5$

Для расчёта общей инкапсуляции используем формулу:

$$knowEncap(O) = \sum_i \omega_i * knowEncap(m_i), \quad (14)$$

где $knowEncap(O)$ – общая инкапсуляция онтологии.

Для расчётов используем данные из таблицы 3. Результаты расчётов представлены в таблице 5.

Для расчета покрытия применяется формула:

$$Cov(O) = \alpha * ConCov * RelCov, \quad (15)$$

где $Cov(O)$ – покрытие онтологии, α – эмпирический коэффициент для масштабирования или коррекции оценки покрытия онтологии, $ConCov = \frac{1}{k} \sum_{j=1}^k \sum_{i=1}^n f(c_i, t_j)$ – покрытие концепций онтологии, где k – эмпирический коэффициент корректировки, который служит для учета различных факторов, $f(c_i, t_j)$ – функция соответствия

метки концепта c_i термину поиска t_j , $RelCov = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n g(c_i, c_j)$ – покрытие отношений онтологии, где n – число всех релевантных отношений (свойств, связей), которые характеризуют предметную область; $g(c_i, c_j)$ – функция наличия отношений между парами концепций c_i, c_j .

Таблица 5. Расчёт knowEncap

Показатель	Расчёт
ω_i для угрозы	$\omega_i = (1 + 1 + 10)/150 = 0,08$
ω_i для объекта	$\omega_i = (49 + 7 + 7)/150 = 0,42$
ω_i для уязвимости	$\omega_i = (5 + 2 + 3)/150 = 0,06$
ω_i для атаки	$\omega_i = (20 + 6 + 3)/150 = 0,2$
ω_i для риска	$\omega_i = (7 + 1 + 2)/150 = 0,06$
Общая инкапсуляция онтологии	$knowEncap(O) = 0,56$

Для оценки покрытия онтологии необходимо установить соответствие между метками концептов и поисковыми запросами. Под меткой каждого концепта понимается его собственное наименование. В качестве поисковых терминов задействован следующий перечень: «IoT», «Интернет вещей», «Объект», «Уязвимость», «Угроза», «Атака», «Риск», «Ущерб», «Сенсор», «Актуатор», «Контроллер». Функция $f(c_i, t_j)$ для $ConCov$ представляется в виде матрицы для определения соответствия метки концепта термину поиска. Матрица заполняется по правилу $f(c_i, t_j) = \begin{cases} 1, & \text{если метка концепта } c_i \text{ совпадает с термином } t_i \\ 0, & \text{иначе} \end{cases}$. Учтем, что термин «IoT» совпадает с меткой «Объект». В качестве коэффициента k примем количество выделенных терминов. Ввиду значительного объема полученных данных матрица не приводится в тексте статьи. Функцией $g(c_i, c_j)$ для $RelCov$ является количество объектных свойств в онтологии в соответствии с таблицей 3. Для расчёта $Cov(O)$ коэффициент α вычисляется как соотношение числа экземпляров и концептов, что отражает как структурную полноту, так и качество наполнения онтологии экземплярами.

Результаты расчета покрытия по формуле (15) приведены в таблице 6.

Глубина иерархии определяется по формуле:

$$DIT = \max(\sum D|c_i|), \quad (16)$$

где DIT – глубина иерархии онтологии, $D|c_i|$ – длина пути от корневого класса Thing до листового концепта c_i .

Таблица 6. Расчёт $Cov(O)$

Показатель	Расчёт
Покрытие концепций	$k = 11, f(c_i, t_j) = 6$ $ConCov = 6/11 = 0,5$
Покрытие связей	$n = 10, g(c_i, c_j) = 35$ $RelCov = (2 * 35)/(10 * 9) = 0,7$
Покрытие онтологии	$\alpha = 209/88 = 2,4$ $Cov(O) = 2,4 * 0,5 * 0,7 = 0,84$

В результате получим $DIT = 3$, так как максимальное количество ребер по иерархии классов во всех случаях равно 3.

Итоговым показателем качества онтологии служит *Behavioral Quality*. Это оценка онтологии с точки зрения её богатства и иерархической структуры, которая определяется по формуле:

$$Behavioral\ Quality = \omega_1 * knowEncap + \omega_2 * Cov + \omega_3 * DIT, \quad (17)$$

где *Behavioral Quality* – поведенческое качество онтологии, ω_i – веса, определяющие важность каждого аспекта (по умолчанию могут быть равными).

Так как ранее все параметры были вычислены, то получим $Behavioral\ Quality = 0,3 * 0,56 + 0,5 * 0,84 + 0,2 * 3 = 1,2$.

Для сравнения взяты онтологии из работ [15, 19, 21]. Для обеспечения объективности сравнительного анализа все метрики были рассчитаны по единой методологии OntoMod на основе анализа данных из [19] и OWL-файлов онтологий, полученных из публичных репозиториях, указанных в [21, 15]. Для каждой из сравниваемых и разработанной онтологий был произведен подсчет количества классов, объектных свойств, свойств-данных и аксиом подклассов, после чего по формулам (12-17) были вычислены итоговые показатели. Таким образом, все онтологии были оценены в одинаковых условиях на основе их формальных описаний безотносительно к конкретным наборам экземпляров. Результаты оценки представлены в таблице 7.

Проведённый сравнительный анализ позволяет сделать следующие выводы. Разработанная онтология обладает наилучшей структурной организацией среди всех аналогов, что выражается в снижении комплексной структурной сложности в 1,5-2,4 раза. По полноте покрытия предметной области разработанная онтология

превосходит рассмотренные решения в 3,36 раза, что является наиболее существенным преимуществом. Глубина иерархии находится на максимальном уровне среди сравнимых онтологий. Несколько более низкое значение инкапсуляции (снижение в 1,8 раза) является следствием модульной архитектуры и рассматривается не как недостаток, а как архитектурный компромисс, обеспечивающий гибкость.

Таблица 7. Расчёт метрик онтологических моделей
для сравнительного анализа

Показатель	SOtoTS [15]	Security of heathcare [19]	WoT [21]	Разработанная онтология
Плотность связей (<i>LD</i>)	0,5	0,07	0,33	0,02
Среднее количество связей на концепт (<i>LC</i>)	2	1,07	1,5	0,7
Насыщенность связей (<i>LR</i>)	1	0,44	0,63	0,4
Цикломатическая сложность (<i>CC_{norm}</i>)	1,2	1,47	1,6	0,9
Комплексная структурная сложность (<i>ComplexOnto</i>)	1,2	0,76	1,015	0,5
Инкапсуляция знаний (<i>knowEncap</i>)	1	1	1	0,56
Покрывтие онтологии (<i>Cov</i>)	0,25	0,25	0,25	0,84
Глубина онтологии (<i>DIT</i>)	1	3	2	3
Поведенческое качество (<i>Behavioral Quality</i>)	0,74	1,4	1,07	1,2

Таким образом, с точки зрения практического применения разработанная онтология обеспечивает: высокую сопровождаемость и лёгкость расширения (благодаря минимальной структурной сложности); максимальную полноту покрытия предметной области (благодаря высокому значению покрытия); хорошую выразительную силу (благодаря максимальной глубине иерархии); готовность к использованию в модульных приложениях (благодаря модульной архитектуре).

Рекомендуемым направлением дальнейшего улучшения является повышение инкапсуляции знаний, в частности, оптимизация интерфейсных связей модуля «Угроза», что позволит дополнительно повысить интегральное поведенческое качество.

6. Экспериментальный анализ результатов. Ключевым преимуществом разработанной онтологической модели является возможность автоматизированного вывода скрытых логических связей, отражающих знания об угрозах для конкретной системы Интернета вещей. В отличие от существующих подходов, где идентификация угроз выполняется вручную экспертом или основывается на статических списках,

предложенная модель позволяет формализовать причинно-следственные связи между архитектурой системы и потенциальными угрозами.

Для проведения эксперимента использована натурная модель «Умная лаборатория» – программно-аппаратный комплекс на базе концепции Интернета вещей, разработанный в Южном федеральном университете (Рис. 14). В состав натурной модели, функционирующей в домене «Наука», входят следующие компоненты, реализующие типовые СФХ системы Интернета вещей, которые представлены в таблице 8.

Репрезентативность стенда обусловлена тем, что он включает три ключевых уровня архитектуры Интернета вещей (аппаратный, сетевой, программный), использует протоколы передачи данных и физические устройства. Это позволяет проследить полную цепочку реализации угрозы: от атаки до физического воздействия (например, изменений рабочих функций одноплатного компьютера), что полностью соответствует киберфизической природе систем, для анализа которых предназначена модель.

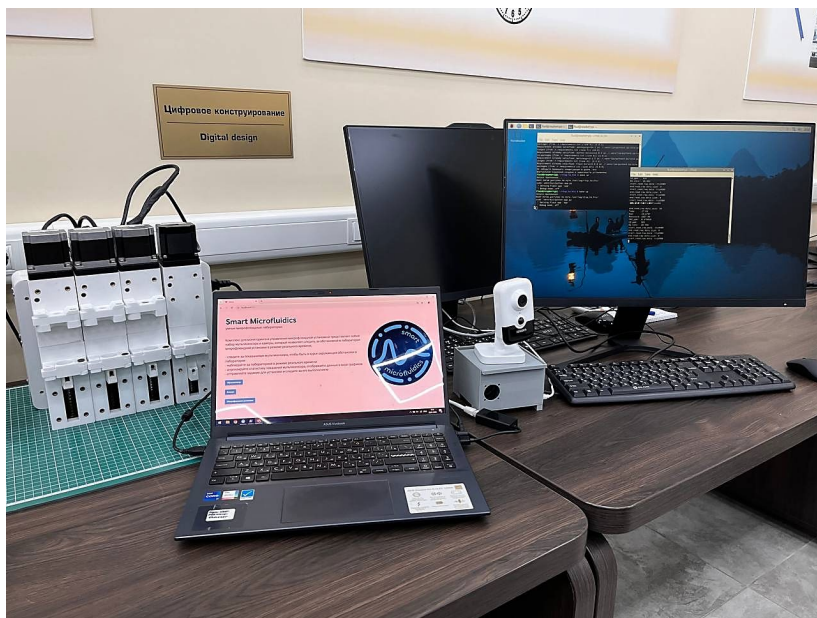


Рис. 14. Натурная модель «Умной лаборатории»

Таблица 8. Компоненты натурной модели «Умная лаборатория»

Класс СФХ	Тип СФХ	Реализация в стенде
Аппаратное обеспечение	Сенсор	Сенсор углекислого газа MH-Z19
		Сенсор угарного газа MQ-7
		Сенсор влажности, состава атмосферы, температуры BME-280
		Сенсор концентрация кислорода AO-02
	Контроллеры, процессоры	Микроконтроллер ESP32
	Интерфейсы ввода/вывода	Одноплатный компьютер Raspberry Pi
Сетевое обеспечение	Сетевые протоколы	IP-камера
		Протокол MQTT
		Протокол Modbus
		Протокол RTSP
Прикладное ПО	Серверное ПО	Rabbit MQ
Взаимодействие с другими объектами	Системы взаимодействия с пользователем	PostgreSQL
		Веб-приложение

В онтологию были загружены экземпляры классов, соответствующие СФХ лаборатории. Процесс идентификации угроз был автоматизирован с помощью набора SWRL-правил, описанных в разделе 4.6. Правила последовательно выводят связи между сущностями, наполняя базу знаний новыми фактам. На рисунке 15 представлен фрагмент графа онтологии, визуализирующий выведенные связи в Protege. Он наглядно демонстрирует, как онтология связывает конкретный компонент системы (микроконтроллер ESP32) с уязвимостями, векторами атак, итоговыми угрозами и рисками. Также на схеме показаны каскадные эффекты от угроз, переходящие на компоненты, связанные с микроконтроллером.

На представленной схеме стрелкой 1 обозначена взаимосвязь между микроконтроллером и уязвимостями; стрелкой 2 – взаимосвязь между уязвимостями и атаками; стрелкой 3 – взаимосвязь между атаками и угрозами. Группа стрелок 4 иллюстрирует каскадное распространение угроз посредством связей СФХ, а стрелки 5 – каскадное распространение угроз через потоки сообщений. Связи между угрозами и целевой системой (целью системы) на данной схеме не показаны. Стрелкой 7 отражена связь между рисками и угрозами.

На рисунке 16 представлены количественные результаты автоматического логического вывода, выполненного резонером Pellet на полном наборе SWRL-правил (Правила 1-3, 7) и экземпляров, описывающих стенд «Умная лаборатория». Правила были запущены последовательно в одном сеансе вывода. По оси ординат указано

абсолютное количество выведенных экземпляров соответствующих классов и свойств.

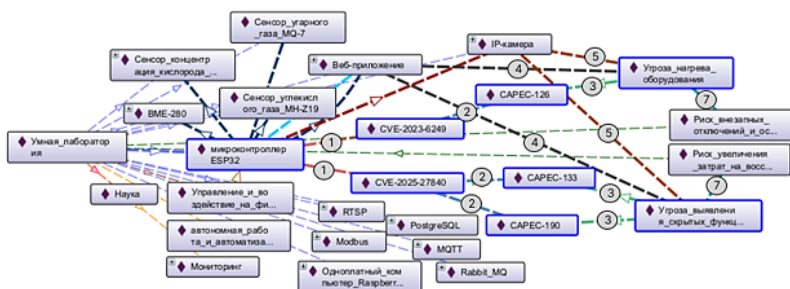


Рис. 15. Фрагмент графа онтологии, демонстрирующий выведенные связи «СФХ-Уязвимость-Атака-Угроза»

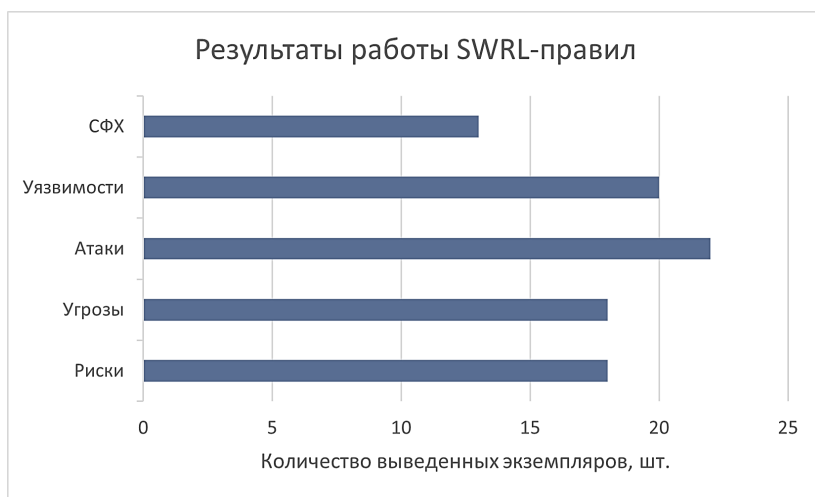


Рис. 16. Результаты работы SWRL-правил

Для иллюстрации процесса вывода (трассировки) рассмотрим получение одной из угроз. Исходный факт: экземпляр микроконтроллер ESP32 (СФХ) связан свойством hasCPE с cpe:/h:espressif:esp32:-. Резонер, проверив Правило 1, находит экземпляр CVE-2025-27840 (Уязвимость) с тем же CPE и генерирует новый факт: микроконтроллер ESP32 hasVulnerability CVE-2025-27840. Далее, согласно Правилу 2, у уязвимости CVE-2025-27840 есть CWE-идентификатор CWE-912,

и у атаки CAPEC-190 есть такой же CWE-идентификатор. Соответственно создается связь CAPEC-190 exploits CVE-2025-27840. Наконец, Правило 3, имея все три факта, создает экземпляр класса Threat, связывая его с объектом натурной модели hasThreat Угроза выявления скрытых функций, а также связывая вместе уязвимый компонент, атаку и угрозу.

Для оценки полноты идентификации угроз был проведён сравнительный эксперимент. В качестве инструмента «ручного» (регламентированного) построения исходного перечня угроз использовался автоматизированный конструктор угроз на сайте Банка данных угроз ФСТЭК России [54], который формирует агрегированный перечень угроз для заданной архитектуры системы.

Для натурной модели «Умная лаборатория» конструктор выделил 2 системных сценария: «Фальсификация данных эксперимента» и «Компрометация ядра системы». Разработанная онтология с использованием SWRL-правил автоматически выявила 18 конкретных угроз, каждая из которых привязана к конкретным компонентам (СФХ), уязвимостям (CVE/RVD) и шаблонам атак (CAPEC). Все 2 сценария, сформированные конструктором ФСТЭК, полностью покрываются набором выведенных онтологий угроз.

Таким образом, онтология обеспечивает как минимум ту же полноту на агрегированном уровне, но дополнительно предоставляет детализированное техническое описание угроз, что невозможно получить напрямую из конструктора ФСТЭК. Это позволяет автоматизировать последующий этап анализа – выбор контрмер и оценку рисков.

7. Заключение. Проведенное исследование направлено на решение ключевого противоречия в области безопасности Интернета вещей: между стремительным ростом числа угроз, обусловленным гетерогенностью и усложнением архитектур, и потребностью операторов в объективном анализе защищённости таких систем. Основной научный результат работы заключается в разработке онтологической модели, которая не просто объединяет предметные области Интернета вещей и информационной безопасности, но и впервые предлагает формальный механизм выявления каскадного распространения угроз, учитывающий не только топологическую связность компонентов, но и семантику передаваемых сообщений, а также функциональную роль каждого элемента в достижении целевых функций системы. Ключевое отличие предложенного подхода от аналогов состоит в том, что глубина и последствия угрозы определяются не фактом первичной компрометации, а структурно-семантическими характеристиками

межкомпонентных связей, что позволяет выявлять скрытые пути нарушения работы системы. Практическая значимость работы раскрывается через возможность автоматизации ключевого этапа анализа защищённости – идентификации угроз. Разработанная онтология совместно с SWRL-правилами позволяет получить исчерпывающий и воспроизводимый перечень конкретных угроз, привязанных к реальным компонентам системы, уязвимостям и шаблонам атак. Апробация на экспериментальном стенде «Умная лаборатория» продемонстрировала, что модель выявляет детализированные технические угрозы, привязанные к конкретным компонентам, уязвимостям и шаблонам атак, обеспечивая полноту не ниже агрегированных сценариев из БДУ ФСТЭК. Практическая применимость модели не ограничивается лабораторным стендом. Её модульная архитектура обеспечивает масштабирование на реальные системы за счёт интеграции с авторитетными базами знаний (CVE, CAPEC), которые непрерывно пополняются. Оценка качества по методологии OntoMod показала расширение семантического покрытия предметной области в 3 раза и снижение структурной сложности в 1,5-2,4 раза по сравнению с рассмотренными аналогами, что свидетельствует о высокой сопровождаемости и пригодности модели для практического использования операторами и проектировщиками распределённых сетей Интернета вещей.

Вместе с тем исследование имеет ряд ограничений. Во-первых, ограничения модели связаны с границами её онтологического базиса: идентифицируются только те угрозы, которые реализуются через известные и задокументированные уязвимости (CVE/RVD) и шаблоны атак (CAPEC). Угрозы класса «нулевого дня» и атаки, специфичные для уникальных аппаратных решений, не имеющих CPE-идентификаторов, остаются за рамками автоматического анализа. Важно отметить, что ограничения на типы охватываемых систем и угроз задаются не самой моделью, а составом эталонного классификатора СФХ. Разработанный классификатор покрывает широкий спектр типовых компонентов (сенсоры, контроллеры, ПО, протоколы), что позволяет описывать большинство распространённых архитектур. Для интеграции узкоспециализированной системы достаточно расширить таксономию СФХ, не изменяя логические правила, заложенные в онтологию. Во-вторых, область исследования ограничена целенаправленными атаками нарушителя, исключая случайные события, ошибки и отказы, что, хотя и обосновано спецификой Интернета вещей, может не охватывать все сценарии нарушения функционирования. В-третьих, актуальность выявленных угроз напрямую зависит

от полноты и своевременности обновления внешних баз знаний, что накладывает требования к эксплуатационному сопровождению модели. Несмотря на указанные ограничения, разработанная онтологическая модель представляет собой законченное и практически применимое решение, создающее формализованный фундамент для последующего этапа количественного анализа защищённости и обоснованного выбора контрмер в системах Интернета вещей.

Литература

1. Doynikova E., Novikova E., Murenin I., et. al. Security Measuring System for IoT Devices // *Lecture Notes in Computer Science*. Springer. 2022. vol. 13106. pp. 256–275.
2. ISO/IEC 30141:2024. Internet of Things (IoT) – Reference architecture: international standard // Geneva: ISO. 2024. 61 p.
3. Parsons E., Panaousis E., Loukas G., Sakellari G. Survey on Cyber Risk Management for the Internet of Things // *Applied Sciences*. 2023. vol. 13. no. 15. pp. 1–44.
4. Salayma M. Risk and threat mitigation techniques in internet of things (IoT) environments: a survey // *Frontiers in the Internet of Things*. 2023. vol. 2. pp. 1–22.
5. Dione D., Seck B., Diop I., et. al. Hardware Security for IoT in the Quantum Era: Survey and Challenges // *Journal of Information Security*. 2023. vol. 14. pp. 227–249.
6. ГОСТ Р 70924-2023. Информационные технологии. Интернет вещей. Типовая архитектура. Издание официальное // М.: Российский институт стандартизации. 2024.
7. Gupta H., Bharti A. Fog Computing & IoT: Overview, Architecture and Applications // *International Journal of Advanced Research in Computer and Communication Engineering*. 2018. vol. 7. pp. 30–34.
8. Zahoor S., Sofi S. Internet of Things: A Review of Functionality, Applications, Architectures and Challenges // *Journal of the University of Shanghai for Science and Technology*. 2021. vol. 23. no. 9. pp. 1216–1224.
9. Herzog A., Shahmehri N., Duma C. An Ontology of Information Security // *International Journal of Information Security and Privacy*. 2007. vol. 1. no. 4. pp. 1–23.
10. Moller D.P.F. Cybersecurity Ontology // *Cybersecurity in Digital Transformation*. Springer. 2020. pp. 79–96.
11. Lee J., Moon D., Kim I., Lee Y. A semantic approach to improving machine readability of a large-scale attack graph // *The Journal of Supercomputing*. 2019. vol. 75. pp. 3028–3045.
12. Falodiya K., Das M.L. Security Vulnerability Analysis using Ontology-based Attack Graphs // *Proceedings of the IEEE India Council International Conference*. 2017. pp. 1–5.
13. Saint-Hilaire K., Neal C., Cuppens F., et. al. A real-time automated attack-defense graph generation approach // *Journal of Information Security and Applications*. 2025. vol. 94. pp. 1–22.
14. Qaswar F., Rahmah M., Raza M., et. al. Applications of Ontology in the Internet of Things: A Systematic Analysis // *Electronics*. 2023. vol. 12. no. 1. pp. 1–30.
15. Jarwar M., Watson J., Ani U., Chalmers S. Industrial Internet of Things Security Modelling using Ontological Methods // *Proceedings of the ACM Workshop on Secure and Trustworthy Cyber-Physical Systems*. 2022. pp. 163–170.
16. Bhole M., Kastner W., Sauter T. Towards Unveiling Vulnerabilities and Securing IoT Devices: An Ontology-Based Approach // *Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation*. 2024. pp. 1–8.
17. Lavrova D.S., Vasilev Y.S. An ontological model of the domain of applications for the Internet of Things in analyzing information security // *Automatic Control and Computer Sciences*. 2017. vol. 51. no. 8. pp. 817–823.

18. Zegzhda D.P., Lavrova D.S. Approach to Internet of Things detection of security incidents using SIEM technology // Интеллектуальные технологии на транспорте. 2017. №1. pp. 35–41.
19. Nazir A., Sholla S., Bashir A. An Ontology based Approach for Context-Aware Security in the Internet of Things (IoT) // International Journal of Wireless and Microwave Technologies (IJWMT). 2021. vol. 11. no. 1. pp. 28–46.
20. SAREF: Smart Applications REference ontology: v. 3.1.1 // European Telecommunications Standards Institute. URL: <https://saref.etsi.org/core/v3.1.1/> (дата обращения: 05.02.2026).
21. Web of Things (WoT) Security Ontology // World Wide Web Consortium (W3C). URL: <https://www.w3.org/2019/wot/security> (дата обращения: 05.02.2026).
22. Федорченко Е.В., Новикова Е.С., Котенко И.В., Гайфулина Д.А., Тушканова О.Н. Левшун Д.С., Мелешко А.В., Муренин И.Н., Коломеец М.В. Система измерения защищенности информации и персональных данных для устройств интернета вещей // Вопросы кибербезопасности. 2022. №5. С. 28–46.
23. Kotenko I., Fedorchenko A., Doynikova E., Chechulin A. An Ontology-based Storage of Security Information // Information Technology and Control. 2018. vol. 47. no. 4. pp. 1–13.
24. Fedorchenko E.V., Kryukov R.O., Kotenko I.V., et. al. Security assessment based on attack graphs using NVD and MITRE ATT & CK database for heterogeneous infrastructures // Information and Control Systems. 2024. no. 2(129). pp. 39–50.
25. Makhdoom I., Abolhasan M., et. al. Anatomy of Threats to the Internet of Things // IEEE Communications Surveys & Tutorials. 2019. vol. 21. no. 2. pp. 1636–1675.
26. Yaacoub J.P.A., Noura H.N., Salman O., Chehab A. Robotics cyber security: vulnerabilities, attacks, countermeasures, and recommendations // International Journal of Information Security. 2022. vol. 21. pp. 115–158.
27. Mishura M. CyberSecOnto-IoT: онтология кибербезопасности Интернета вещей // GitHub. URL: <https://github.com/MashaMishura/CyberSecOnto-IoT.git> (дата обращения 05.02.2026).
28. Приказ Министерства цифрового развития, связи и массовых коммуникаций РФ от 29 марта 2019 г. №113 «Об утверждении Концепции построения и развития узкополосных беспроводных сетей связи «Интернета вещей» на территории Российской Федерации». URL: <https://docs.cntd.ru/document/554066760> (дата обращения 05.02.2026).
29. ISO/IEC 27400:2022. Cybersecurity – IoT security and privacy – Guidelines: international standard // Geneva: ISO. 2022.
30. NISTIR 8228. Considerations for Managing Internet of Things (IoT) Cybersecurity and Privacy Risks // NIST. 2019.
31. NISTIR 8259A. IoT Device Cybersecurity Capability Core Baseline / National Institute of Standards and Technology // NIST. 2020.
32. Федеральный закон Российской Федерации от 26 июля 2017 г. №187 «О безопасности критической информационной инфраструктуры Российской Федерации». <https://docs.cntd.ru/document/436752114> (дата обращения 05.02.2026).
33. ETSI TS 103 264 V3.1.1 (2020-02). SmartM2M; Smart Applications; Reference Ontology and oneM2M Mapping: Technical Specification // European Telecommunications Standards Institute. Sophia Antipolis: ETSI. 2020. 78 p.
34. CVE: Common Vulnerabilities and Exposures // The MITRE Corporation. URL: <https://cve.mitre.org/> (дата обращения: 05.02.2026).
35. RVD: Robotics Vulnerability Database // Alias Robotics. URL: <https://github.com/aliasrobotics/RVD> (дата обращения: 05.02.2026).
36. CVSS v3.0 specification document // FIRST. URL: <https://www.first.org/cvss/specification-document> (дата обращения: 05.02.2026).

37. Методика оценки уровня критичности уязвимостей программных, программно-аппаратных средств: методический документ: утвержден ФСТЭК России 30 июня 2025 г. URL: <https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/metodicheskij-dokument-ot-30-iyunya-2025-g> (дата обращения: 05.02.2026).
38. CAPEC – Common Attack Pattern Enumeration and Classification // The MITRE Corporation. URL: <https://capec.mitre.org/> (дата обращения: 05.02.2026).
39. Методика оценки угроз безопасности информации: методический документ: утвержден ФСТЭК России 5 февраля 2021 г. URL: <https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/metodicheskij-dokument-ot-5-fevralya-2021-g> (дата обращения: 05.02.2026).
40. Wu C.K. Architectures and security measures // Internet of Things security. Springer. 2021. pp. 13–26.
41. Dubey A., Yadav S. Basics of Internet of Things // International Journal of Scientific Research in Engineering and Management (IJSREM). 2024. vol. 8. no. 10. pp. 1–6.
42. Blinowski G., Piotrowski P. CVE based classification of vulnerable IoT systems // arXiv preprint arXiv:2006.16640. 2020.
43. Afroze M., Reena J. A Survey Study on IoT Application and its Attacks // International Research Journal on Advanced Engineering Hub (IRJAEH). vol. 2. no. 02. pp. 80–86.
44. Alam M., Khan I. Cyber Physical Attacks and IoT // Intelligent Cyber-Physical Systems Security for Industry 4.0. 2022. pp. 79–104.
45. Банк данных угроз безопасности информации. URL: <https://bdu.fstec.ru/threat> (дата обращения: 07.02.2026).
46. Дойникова Е.В., Котенко И.В. Оценивание защищённости и выбор контрмер для управления кибербезопасностью // М.: Российская академия наук, 2021. 184 с.
47. ENISA Threat Landscape 2024 // European Union Agency for Cybersecurity. Athens: ENISA. 2024. 112 p.
48. Best Practice Guidelines // IoT Security Foundation. URL: <https://iotsecurityfoundation.org/best-practice-guidelines/> (дата обращения: 07.02.2026).
49. Kaspersky IoT Threat Data Feed. URL: <https://www.kaspersky.ru/iot-threat-data-feed> (дата обращения: 07.02.2026).
50. Protege. URL: <http://protege.stanford.edu/> (дата обращения: 07.02.2026).
51. Common Platform Enumeration (CPE) // National Institute of Standards and Technology. URL: <https://nvd.nist.gov/products/cpe> (дата обращения: 07.02.2026).
52. National Vulnerability Database (NVD) // National Institute of Standards and Technology. URL: <https://nvd.nist.gov> (дата обращения: 07.02.2026).
53. Kumar S., Baliyan N. Semantic Web-Based Systems // SpringerBriefs in Computer Science. Springer. 2018. pp. 19–50.
54. Банк данных угроз ФСТЭК России. Формирование перечня угроз. URL: <https://bdu.fstec.ru/threat-section/shaper-threats> (дата обращения: 26.05.2026).

Фирсова Мария Геннадьевна — старший преподаватель, кафедры безопасности информационных технологий им. О.Б. Макаревича института компьютерных технологий и информационной безопасности, Федеральное государственное автономное образовательное учреждение высшего образования «Южный федеральный университет» (ФГАОУ ВО «ЮФУ»). Область научных интересов: анализ информационной безопасности киберфизических систем и систем Интернета вещей, онтологический инжиниринг. Число научных публикаций — 25. mshulika@sfedu.ru; ул. Чехова, 2, 347922, Таганрог, Россия; р.т.: +7(951)502-1484.

Поддержка исследований. Работа выполнена при поддержке Фонда поддержки молодых ученых имени Геннадия Комиссарова в рамках конкурса на получение стипендии имени Евгения Касперского.

M. FIRSOVA

ONTOLOGY MODEL FOR SECURITY ANALYSIS OF INTERNET OF THINGS SYSTEMS

Firsova M. Ontology model for security analysis of Internet of Things systems.

Abstract. The article addresses the problem of comprehensive security analysis of Internet of Things systems. The relevance of the work is driven by the growing complexity and heterogeneity of Internet of Things systems, as well as the need for scalable assessment of their security while taking into account architectural and functional characteristics. Based on an analysis of the architecture of Internet of Things systems, a multi-level ontological model has been developed that integrates the domains of the Internet of Things and information security, supplemented by consideration of the functional role of system components, the types of transmitted messages, and the goals of the system, which makes it possible to represent the dependence of a component's criticality on the function it performs and the influence of a threat on disruption of the system's target function. Based on the model, a threat inference mechanism is proposed that identifies the cascading effect of the propagation of destructive impact along the topological vector and the functional vector. In contrast to approaches that consider a threat exclusively through the fact of primary compromise of an element, the novelty lies in the fact that the depth and consequences of a threat are determined not by the initial compromise, but by the structural-semantic characteristics of inter-component connections that determine the development of the cascading effect. The ontology is integrated with external knowledge repositories (CVE, RVD, CAPEC, CWE, FSTEC BDU) through an ETL process, and the inference of hidden connections is automated using SWRL rules in the Protege environment. Practical validation of the model was carried out on the experimental testbed "Smart Laboratory." The result of the work is an implemented ontological model that provides automatic identification of security threats for Internet of Things systems. Evaluation of the model's quality using the OntoMod methodology showed a threefold expansion of the semantic coverage of the domain and a reduction in structural complexity by 1.5-2.4 times compared with the considered analogues. A conclusion is formulated about the applicability of the proposed ontological approach for prompt decision support for operators and designers of distributed Internet of Things networks.

Keywords: ontology, Internet of Things, threat, structural and functional characteristics, information security.

References

1. Doynikova E., Novikova E., Murenin I., et al. Security Measuring System for IoT Devices. Lecture Notes in Computer Science. 2022. vol. 13106. pp. 256–275.
2. ISO/IEC 30141:2024. Internet of Things (IoT) – Reference architecture: international standard. Geneva: ISO. 2024. 61 p.
3. Parsons E., Panaousis E., Loukas G., Sakellari G. Survey on Cyber Risk Management for the Internet of Things. Applied Sciences. 2023. vol. 13. no. 15. pp. 1–44.
4. Salayma M. Risk and threat mitigation techniques in internet of things (IoT) environments: a survey. Frontiers in the Internet of Things. 2023. vol. 2. pp. 1–22.
5. Dione D., Seck B., Diop I., et al. Hardware Security for IoT in the Quantum Era: Survey and Challenges. Journal of Information Security. 2023. vol. 14. pp. 227–249.
6. GOST R 70924-2023. Informatsionnye tekhnologii. Internet veshchei. Tipovaya arkhitektura. Izdanie ofitsialnoe [Information Technology. Internet of Things. Reference Architecture. Official Edition]. M.: Rossiiskii institut standartizatsii. 2024. (In Russ.).

7. Gupta H., Bharti A. Fog Computing & IoT: Overview, Architecture and Applications. *International Journal of Advanced Research in Computer and Communication Engineering*. 2018. vol. 7. pp. 30–34.
8. Zahoor S., Sofi S. Internet of Things: A Review of Functionality, Applications, Architectures and Challenges. *Journal of the University of Shanghai for Science and Technology*. 2021. vol. 23. no. 9. pp. 1216–1224.
9. Herzog A., Shahmehri N., Duma C. An Ontology of Information Security. *International Journal of Information Security and Privacy*. 2007. vol. 1. no. 4. pp. 1–23.
10. Moller D.P.F. Cybersecurity Ontology. *Cybersecurity in Digital Transformation*. Springer. 2020. pp. 79–96.
11. Lee J., Moon D., Kim I., Lee Y. A semantic approach to improving machine readability of a large-scale attack graph. *The Journal of Supercomputing*. 2019. vol. 75. pp. 3028–3045.
12. Falodiya K., Das M.L. Security Vulnerability Analysis using Ontology-based Attack Graphs. *Proceedings of the IEEE India Council International Conference*. 2017. pp. 1–5.
13. Saint-Hilaire K., Neal C., Cuppens F., et al. A real-time automated attack-defense graph generation approach. *Journal of Information Security and Applications*. 2025. vol. 94. pp. 1–22.
14. Qaswar F., Rahmah M., Raza M., et al. Applications of Ontology in the Internet of Things: A Systematic Analysis. *Electronics*. 2023. vol. 12. no. 1. pp. 1–30.
15. Jarwar M., Watson J., Ani U., Chalmers S. Industrial Internet of Things Security Modelling using Ontological Methods. *Proceedings of the ACM Workshop on Secure and Trustworthy Cyber-Physical Systems*. 2022. pp. 163–170.
16. Bhole M., Kastner W., Sauter T. Towards Unveiling Vulnerabilities and Securing IoT Devices: An Ontology-Based Approach. *Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation*. 2024. pp. 1–8.
17. Lavrova D.S., Vasilev Y.S. An ontological model of the domain of applications for the Internet of Things in analyzing information security. *Automatic Control and Computer Sciences*. 2017. vol. 51. no. 8. pp. 817–823.
18. Zegzhda D.P., Lavrova D.S. Approach to Internet of Things detection of security incidents using SIEM technology. *Intellectualnye tekhnologii na transporte – Intellectual Technologies on Transport*. 2017. no. 1. pp. 35–41. (In Russ.).
19. Nazir A., Sholla S., Bashir A. An Ontology based Approach for Context-Aware Security in the Internet of Things (IoT). *International Journal of Wireless and Microwave Technologies (IJWMT)*. 2021. vol. 11. no. 1. pp. 28–46.
20. SAREF: Smart Applications REference ontology: v. 3.1.1. European Telecommunications Standards Institute. Available at: <https://saref.etsi.org/core/v3.1.1/> (accessed 05.02.2026). (In Russ.).
21. Web of Things (WoT) Security Ontology. World Wide Web Consortium (W3C). Available at: <https://www.w3.org/2019/wot/security> (accessed 05.02.2026). (In Russ.).
22. Fedorchenko E.V., Novikova E.S., Kotenko I.V., Gajfulina D.A., Tushkanova O.N., Levshun D.S., Meleshko A.V., Murenin I.N., Kolomeecz M.V. [Security Measurement System for Information and Personal Data of Internet of Things Devices]. *Voprosy kiberbezopasnosti – Cybersecurity Issues*. 2022. no. 5. pp. 28–46. (In Russ.).
23. Kotenko I., Fedorchenko A., Doynikova E., Chechulin A. An Ontology-based Storage of Security Information. *Information Technology and Control*. 2018. vol. 47. no. 4. pp. 1–13.
24. Fedorchenko E.V., Kryukov R.O., Kotenko I.V., et al. Security assessment based on attack graphs using NVD and MITRE ATT&CK database for heterogeneous infrastructures. *Information and Control Systems*. 2024. no. 2(129). pp. 39–50.
25. Makhdoom I., Abolhasan M., et al. Anatomy of Threats to the Internet of Things. *IEEE Communications Surveys & Tutorials*. 2019. vol. 21. no. 2. pp. 1636–1675.

26. Yaacoub J.P.A., Noura H.N., Salman O., Chehab A. Robotics cyber security: vulnerabilities, attacks, countermeasures, and recommendations. *International Journal of Information Security*. 2022. vol. 21. pp. 115–158.
27. Mishura M. CyberSecOnto-IoT: ontologiya kiberbezopasnosti Interneta veshhej [CyberSecOnto-IoT: Cybersecurity Ontology of the Internet of Things]. GitHub. Available at: <https://github.com/MashaMishura/CyberSecOnto-IoT.git> (accessed 05.02.2026). (In Russ.).
28. Prikaz Ministerstva cifrovogo razvitiya, svyazi i massovy'x kommunikacij RF ot 29 marta 2019 g. №113 «Ob utverzhdenii Konceptii postroeniya i razvitiya uzkopolosny'x besprovodny'x setej svyazi «Interneta veshhej» na territorii Rossijskoj Federacii» [Order of the Ministry of Digital Development, Communications and Mass Media of the Russian Federation No. 113 dated March 29, 2019 «On Approval of the Concept for the Construction and Development of Narrowband Wireless Communication Networks Internet of Things in the Russian Federation»]. Available at: <https://docs.cntd.ru/document/554066760> (accessed 05.02.2026). (In Russ.).
29. ISO/IEC 27400:2022. Cybersecurity – IoT security and privacy – Guidelines: international standard. Geneva: ISO. 2022.
30. NISTIR 8228. Considerations for Managing Internet of Things (IoT) Cybersecurity and Privacy Risks. NIST. 2019.
31. NISTIR 8259A. IoT Device Cybersecurity Capability Core Baseline. National Institute of Standards and Technology. NIST. 2020.
32. Federal'ny'j zakon Rossijskoj Federacii ot 26 iyulya 2017 g. №187 «O bezopasnosti kriticheskoy informacionnoj infrastruktury» Rossijskoj Federacii». [Federal Law of the Russian Federation No. 187 dated July 26, 2017 «On the Security of the Critical Information Infrastructure of the Russian Federation»]. Available at: <https://docs.cntd.ru/document/436752114> (accessed 05.02.2026). (In Russ.).
33. ETSI TS 103 264 V3.1.1 (2020-02). SmartM2M; Smart Applications; Reference Ontology and oneM2M Mapping: Technical Specification. European Telecommunications Standards Institute. Sophia Antipolis: ETSI. 2020. 78 p.
34. CVE: Common Vulnerabilities and Exposures. The MITRE Corporation. Available at: <https://cve.mitre.org/> (accessed 05.02.2026). (In Russ.).
35. RVD: Robotics Vulnerability Database. Alias Robotics. Available at: <https://github.com/aliasrobotics/RVD> (accessed 05.02.2026). (In Russ.).
36. CVSS v3.0 specification document. FIRST. Available at: <https://www.first.org/cvss/specification-document> (accessed 05.02.2026). (In Russ.).
37. Metodika ocenki urovnya kritichnosti uyazvimostej programmny'x, programmno-apparatny'x sredstv: metodicheskij dokument: utverzhden FSTE'K Rossii 30 iyunya 2025 g. [Methodology for Assessing the Criticality Level of Vulnerabilities in Software and Hardware: Methodological Document: Approved by FSTEC of Russia on June 30, 2025]. Available at: <https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/metodicheskij-dokument-ot-30-iyunya-2025-g> (accessed 05.02.2026). (In Russ.).
38. CAPEC – Common Attack Pattern Enumeration and Classification. The MITRE Corporation. Available at: <https://capec.mitre.org/> (accessed 05.02.2026). (In Russ.).
39. Metodika ocenki ugroz bezopasnosti informacii: metodicheskij dokument: utverzhden FSTE'K Rossii 5 fevralya 2021 g. [Methodology for Assessing Information Security Threats: Methodological Document: Approved by FSTEC of Russia on February 5, 2021]. Available at: <https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/metodicheskij-dokument-ot-5-fevralya-2021-g> (accessed 05.02.2026). (In Russ.).
40. Wu C.K. Architectures and security measures. *Internet of Things security*. Springer. 2021. pp. 13–26.

41. Dubey A., Yadav S. Basics of Internet of Things. International Journal of Scientific Research in Engineering and Management (IJSREM). 2024. vol. 8. no. 10. pp. 1–6.
42. Blinowski G., Piotrowski P. CVE based classification of vulnerable IoT systems. arXiv preprint arXiv:2006.16640. 2020.
43. Afroze M., Reena J. A Survey Study on IoT Application and its Attacks. International Research Journal on Advanced Engineering Hub (IRJAEH). vol. 2. no. 02. pp. 80–86.
44. Alam M., Khan I. Cyber Physical Attacks and IoT. Intelligent Cyber-Physical Systems Security for Industry 4.0. 2022. pp. 79–104.
45. Bank dannykh ugroz bezopasnosti informacii [Data Bank of Information Security Threats]. Available at: <https://bdu.fstec.ru/threat> (accessed 07.02.2026). (In Russ.).
46. Dojnikova E.V., Kotenko I.V. Ocenivanie zashhishhyonnosti i vy`bor kontrmer dlya upravleniya kiberbezopasnost`yu [Security Assessment and Countermeasure Selection for Cybersecurity Management]. Moscow: Rossiyskaya akademiya nauk. 2021. 184 p. (In Russ.).
47. ENISA Threat Landscape 2024. European Union Agency for Cybersecurity. Athens: ENISA. 2024. 112 p.
48. Best Practice Guidelines. IoT Security Foundation. Available at: <https://iotsecurityfoundation.org/best-practice-guidelines/> (accessed 07.02.2026). (In Russ.).
49. Kaspersky IoT Threat Data Feed. Available at: <https://www.kaspersky.ru/iot-threat-data-feed> (accessed 07.02.2026). (In Russ.).
50. Protege. Available at: <http://protege.stanford.edu/> (accessed 07.02.2026). (In Russ.).
51. Common Platform Enumeration (CPE). National Institute of Standards and Technology. Available at: <https://nvd.nist.gov/products/cpe> (accessed 07.02.2026). (In Russ.).
52. National Vulnerability Database (NVD). National Institute of Standards and Technology. Available at: <https://nvd.nist.gov> (accessed 07.02.2026). (In Russ.).
53. Kumar S., Baliyan N. Semantic Web-Based Systems. SpringerBriefs in Computer Science. Springer. 2018. pp. 19–50.
54. Bank dannykh ugroz FSTEK Rossii. Formirovanie perechnya ugroz [FSTEC of Russia Threat Data Bank. Formation of the Threat List]. Available at: <https://bdu.fstec.ru/threat-section/shaper-threats> (accessed 26.05.2026). (In Russ.).

Firsova Maria — Senior Lecture, Department of Information Technology Security Named After O.B. Makarevich, Institute of Computer Technologies and Information Security, Federal State Autonomous Educational Institution of Higher Education “Southern Federal University” (FGAOU VO “SFU”). Research interests: analysis of information security of cyber-physical systems and Internet of Things systems, ontological engineering. The number of publications — 25. mshulika@sfsedu.ru; 2, Chekhov St., 347922, Taganrog, Russia; office phone: +7(951)502-1484.

Acknowledgements. This work is supported by the Gennady Komissarov Foundation for Young Scientists as part of the Eugene Kaspersky Scholarship Competition.

Д.В. ЕФАНОВ, Е.И. ЕЛИНА

СИНТЕЗ САМОПРОВЕРЯЕМЫХ ЦИФРОВЫХ УСТРОЙСТВ С УЛУЧШЕННЫМИ ПОКАЗАТЕЛЯМИ КОНТРОЛЕПРИГОДНОСТИ

Ефанов Д.В., Елина Е.И. Синтез самопроверяемых цифровых устройств с улучшенными показателями контролепригодности.

Аннотация. Предложен метод синтеза полностью самопроверяемых цифровых устройств с улучшенными показателями контролепригодности. Решение основано на использовании принципа логической коррекции сигналов и свойств композиции равновесных кодов «1 из 4» и «3 из 4». Описаны особенности организации схемы встроенного контроля по двум диагностическим признакам: принадлежности формируемых кодовых слов композиции равновесных кодов «1 из 4» и «3 из 4», а также каждой функции, описывающей разряды кодовых слов в схеме встроенного контроля классу самодвойственных булевых функций. Приведена структура самопроверяемого цифрового устройства с контролем вычислений по двум диагностическим признакам. Разработаны алгоритмы синтеза схемы встроенного контроля на основе логической коррекции сигналов и композиции равновесных кодов «1 из 4» и «3 из 4», учитывающие требования по тестируемости ее компонентов в процессе эксплуатации. Предложен метод выделения подмножеств контролепригодных четверок выходов на множестве выходов объекта диагностирования, позволяющий гарантированно покрывать ошибки с помощью выбранной композиции равновесных кодов с учетом ее диагностических свойств. Даны некоторые результаты экспериментов с тестовыми схемами, демонстрирующие преимущества разработанного авторами метода синтеза самопроверяемых цифровых устройств по сравнению с известными: дублированием с контролем самодвойственности и с использованием кода «2 из 4» с дополнительным контролем самодвойственности. Разработанный метод может эффективно применяться при синтезе самопроверяемых цифровых устройств как на современной микроэлектронной базе, так и может рассматриваться в будущем при реализации устройств на новых физических принципах.

Ключевые слова: самопроверяемое цифровое устройство, логическая коррекция сигналов, схема встроенного контроля, контроль вычислений на выходах устройств, композиция равновесных кодов «1 из 4» и «3 из 4», обнаружение неисправностей.

1. Введение. Классические методы синтеза самопроверяемых цифровых устройств основаны на использовании свойств помехозащищенных и помехоустойчивых кодов, применяемых при кодировании состояний, входов и выходов, а также при проектировании схем встроенного контроля (СВК) [1 – 3]. При использовании кодирования состояний, входов и выходов цифровых устройств на этапе их синтеза для контроля неисправностей и ошибок в вычислениях уже в процессе эксплуатации используют самопроверяемые тестеры выбранных кодов. Вся избыточность за исключением схем тестеров и компараторов для получения единственной пары контрольных сигналов вносится в структуру самого цифрового устройства на этапе его проектирования. При синтезе СВК,

наоборот, структура исходного цифрового устройства (будем дальше говорить о нем, как об *объекте диагностирования*, поскольку в этом случае в процессе эксплуатации реализуется как раз рабочее диагностирование [4]) никак не изменяется, и вся избыточность вносится в них. Таким образом, в обоих подходах к реализации самопроверяемых цифровых устройств классические методы подразумевают дополнение некоторых информационных сигналов, реализуемых на входах и выходах исходных устройств, рядом контрольных сигналов. Их особенности и свойства напрямую определяются использованными помехозащищенными или помехоустойчивыми кодами [5].

Среди классических методов выделяются ставшие типовыми метод дублирования со сравнением сигналов (далее – просто дублирование) и метод контроля вычислений по паритету (далее – просто метод паритета). Метод дублирования при построении самопроверяемых цифровых устройств позволяет сделать процесс проектирования максимально простым за счет использования типовых устройств. Он основан на применении кодов с повторением. За счет возможностей покрытия любых сочетаний искажений на выходах объекта диагностирования дублирование широко распространено в практике синтеза высоконадежных и безопасных систем управления как на аппаратном [6], так и на программном уровнях реализации [7]. Исследования применения дублирования при различных способах реализации цифровых устройств актуальны и до сих пор [8]. Недостаток дублирования состоит в невозможности учета особенностей структур объектов диагностирования, что приводит практически в любых реализациях к внесению чрезмерной избыточности для решения задачи рабочего диагностирования. Диаметрально противоположный метод паритета также широко распространен и основан на использовании функции свертки по модулю $M = 2$ всех сигналов с контролируемых выходов. СВК получается максимально простой, но при этом не все сочетания искажений на контролируемых выходах могут быть обнаружены. Решается задача обнаружения полного множества ошибок путем выделения подмножеств независимых выходов на множестве выходов объекта диагностирования, а сам метод, позволяющий это делать, хорошо известен из [9].

Известны подходы к синтезу СВК, учитывающие свойства реализуемых на выходах объекта диагностирования булевых функций. Например, могут быть реализованы устройства, в которых постоянные сигналы логического 0 и логической 1 представляются последовательностями 0101...01 и 1010...10 соответственно,

а сама их работа осуществляется при постоянной смене этих сигналов. Такой метод функционирования цифровых устройств некоторые исследователи называют методом инвертирования данных [10]. Он позволяет использовать свойства самодвойственных булевых функций для организации контроля вычислений и изучен в классических подходах к синтезу СВК в [11, 12].

Исследования последних лет в области синтеза самопроверяемых цифровых устройств сосредоточены в нескольких направлениях. Например, в [13] предложен метод синтеза СВК, принципиально отличающийся от известных тем, что связан с построением двоичного кода, исходя из свойств структуры объекта диагностирования и заданной модели неисправности. Такой код «наследует» свойства структур объектов диагностирования и применяется исключительно для организации СВК для них. Для каждого объекта диагностирования с новой структурой он строится вновь. Авторы [14] использовали этот метод в несколько иной интерпретации – для построения структуры с обнаружением неисправностей с определенной вероятностью, но не 100 %. В [15] предлагается использовать не полное множество кодовых слов некоторого помехозащищенного кода, а его подмножество для организации СВК, что также отличается от классических методов и требует модификаций контрольных схем тестеров, но также, как и затронутый выше ориентируется на свойства структуры объекта диагностирования. В [16] предложено при синтезе самопроверяемых цифровых устройств использовать сразу же два диагностических признака – контроль принадлежности кодовых слов выбранному помехозащищенному коду, а также контроль самодвойственности каждой функции, описывающей разряды кодовых слов, формируемых в СВК. Он исследовался в применении с различными кодами. И здесь можно отметить фундаментальные работы [17, 18] по применению классических кодов Хэмминга и их модификаций для организации СВК таким способом. При синтезе СВК с контролем по двум диагностическим признакам существенно повышаются показатели контролепригодности цифровых устройств в части наблюдаемости ошибок при рабочем диагностировании.

Особые структуры с контролем вычислений по двум диагностическим признакам получаются при использовании принципа логической коррекции сигналов (ЛКС), предложенного в [19, 20]. Принцип ЛКС, в отличие от классического подхода, ориентированного на дополнение информационных сигналов контрольными в СВК, подразумевает преобразование всех или части сигналов с выходов объекта диагностирования в СВК при подаче на его входы каждого

из наборов значений аргументов для придания формируемому кодовому вектору особых диагностических свойств. Исследования авторов настоящей статьи показывают, что принцип ЛКС можно эффективно использовать в сочетании с композициями хорошо известных равновесных кодов для построения самопроверяемых цифровых устройств с улучшенными показателями наблюдаемости. При этом особыми свойствами обладают структуры, синтезируемые с использованием композиции равновесных кодов «1 из 4» и «3 из 4» [21]. Изложению особенностей новейшего метода синтеза самопроверяемых цифровых устройств на основе принципа ЛКС и композиции равновесных кодов «1 из 4» и «3 из 4» и посвящена настоящая статья.

2. Особенности организации контроля вычислений на основе принципа логической коррекции сигналов и композиции равновесных кодов «1 из 4» и «3 из 4». Как отмечалось выше, при синтезе самопроверяемых цифровых устройств контроль может быть организован любым сочетанием следующих множеств точек контроля: входов объекта диагностирования, выходов элементов памяти и выходов объекта диагностирования. Как правило, комбинационные составляющие цифровых устройств выделяют среди всех схемных решений в устройстве, и контролируют вычисления на их выходах [22]. Для элементов памяти используют особые методы тестирования, например, основанные на разрыве цепей обратных связей и прогоне тестовых комбинаций (здесь речь уже идет о тестовом диагностировании устройств) [23].

Используем следующие исходные предпосылки. Положим, что $F(X)$ – некое цифровое устройство комбинационного типа, где $\langle X \rangle = \langle x_i, x_{i-1} \dots x_2 x_1 \rangle$ – вектор наборов значений аргументов (или входных воздействий). Так как речь идет о современных цифровых устройствах, то реализуются они в бинарной логике с применением компонентов, где базовой единицей является транзистор [24]. Его используют в цифровых схемах чаще всего в ключевом режиме и в режиме отсечки. Отсюда становится понятным, что устройство $F(X)$ должно реализовывать некоторые булевы функции $f_1(X), f_2(X), \dots, f_{n-1}(X), f_n(X)$, а вектор наборов значений аргументов также является булевым. Работа устройства может быть определена на полном множестве наборов значений аргументов, либо же на его подмножестве. Далее положим, что работа определена на полном множестве наборов значений аргументов, то есть осуществляется в режиме исчерпывающего тестирования. На суть излагаемых принципов, однако, это не повлияет.

Устройство $F(X)$ функционирует в импульсном режиме при подаче на входы не стационарных наборов значений аргументов с фиксированными значениями 0 и 1, а при подаче последовательностей, соответствующих сигналам логического 0 и логической 1. Сигналу логического 0 соответствует последовательность 0101...01, а сигналу логической 1 – последовательность 1010...10. Сам по себе данный режим не организуется постоянной сменой сигналов на входах, а сигналы от входов $x_1, x_2, \dots, x_{t-1}, x_t$ трансформируются в последовательности путем использования специальной схемы преобразования. Она реализуется на t двухвходовых элементах сложения по модулю $M=2$ (XOR). На один из входов каждого XOR подается сигнал с соответствующего входа $x_1, x_2, \dots, x_{t-1}, x_t$, а на второй – сигнал a с генератора G прямоугольных импульсов со скважностью 2, что позволяет осуществлять преобразования по правилу:

$$x_i^a = x_i \oplus a, \quad i = \overline{1, t}. \quad (1)$$

Поскольку сигнал a представляет собой последовательность нулей и единиц, стационарные сигналы логических 0 и 1 трансформируются в последовательности импульсов, что следует из формулы (1). Поэтому на входы устройства $F(X)$ наборы значений аргументов поступают парами. Наборы из данной пары всегда ортогональны по всем разрядам. При подаче первого набора значений аргументов формируется информационное кодовое слово, а при подаче второго – контрольное. Таким образом, меняется принцип организации контроля вычислений относительно классических подходов – контрольные сигналы следуют сразу же за информационными, а соответствие их проверяется в одном такте работы. Иными словами, требуется использование временной избыточности для реализации работы устройства. Это, однако, и позволяет организовать контроль вычислений сразу же по двум диагностическим признакам.

На структуру устройства $F(X)$ не накладывается никаких ограничений, кроме обозначенных ранее свойств. Для контроля возникающих в его структуре неисправностей устраивается специализированная СВК (Рис. 1). Она содержит блок $G(X)$ вычисления значений функций коррекции, сам блок коррекции сигналов (БКС), а также модуль контроля вычислений по двум диагностическим признакам (МКВ).

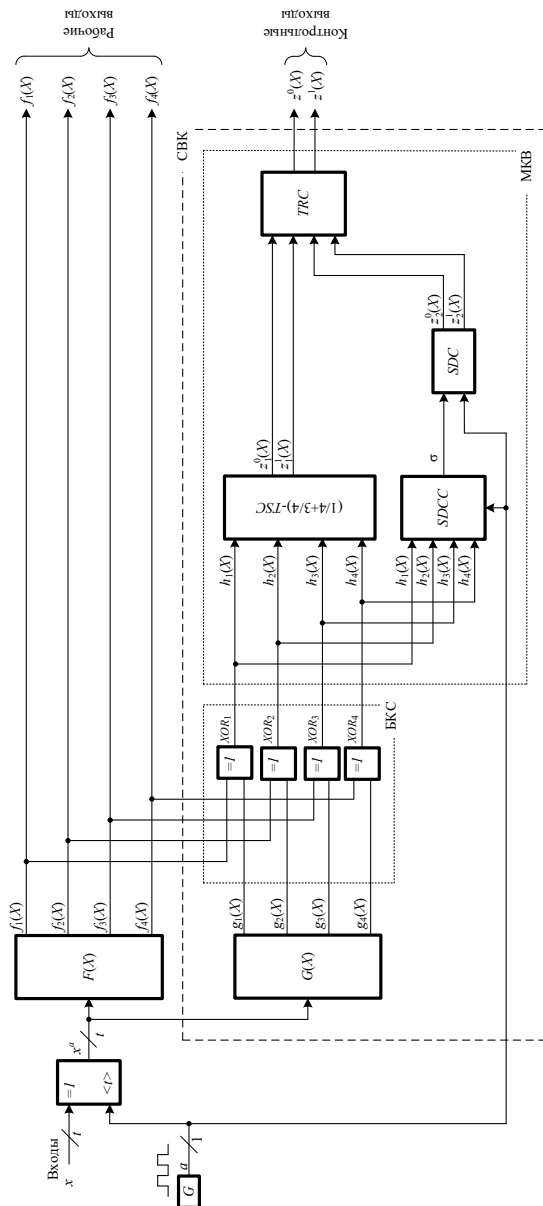
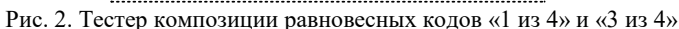


Рис. 1. Структура самопроверяемого цифрового устройства с контролем вычислений по двум диагностическим признакам

В [17, 18] показано, что для организации контроля вычислений по двум диагностическим признакам требуется на каждой паре наборов значений аргументов, поступающих на входы объекта диагностирования, обеспечить формирование ортогональных по всем разрядам кодовых слов. Это возможно не для любого двоичного кода. Например, использование классического кода с суммированием (кода Бергера) [25] или его модификаций при суммировании в кольце вычетов по модулю $M=4$ или $M=8$ (кодов Боуза – Лина) [26], а также любых равновесных кодов, отличных от вида « r из $2r$ », где r – вес кодового слова, не отвечает данным требованиям. Тот же код «1 из 4» содержит следующие кодовые слова: {0001, 0010, 0100, 1000}. Как следует из анализа данного множества, в него не входят пары ортогональных по всем разрядам кодовых слов. Добиться этого свойства можно объединением множеств кодовых слов кодов «1 из 4» и «3 из 4», что даст следующие пары ортогональных по всем разрядам кодовых слов: {0001, 1110}, {0010, 1101}, {0100, 1011}, {1000, 0111}. Конечно же, увеличение мощности множества кодовых слов влечет за собой и ухудшение обнаруживающих свойств получаемого кода, однако, как будет показано далее, это ухудшение «парировается» на схемотехническом уровне реализации СВК путем использования особого метода выделения контролепригодных по заданному коду выходов.

Еще одним крайне важным фактором выбора именно композиции равновесных кодов «1 из 4» и «3 из 4» является простота ее полностью самопроверяемого тестера и малая длина проверяющего теста для него. Его структура разработана авторами настоящей статьи в [21] и изображена на рисунке 2.



Структура тестера абсолютно проста и по сложности соизмерима с простейшим тестером для равновесного кода «2 из 4». Для ее полной проверки достаточно на входы подать следующие комбинации, принадлежащие заданной композиции:

$$T_{II} = T_{II_1} \cup T_{II_2} = \{1000 \cap 0010 \cap 1101 \cap 0111\} \{0001 \cap 1011 \cap 0100 \cap 1110\}. \quad (2)$$

В (2) представлены все проверяющие тесты T_{II_1} и T_{II_2} , длина каждого из которых равна 4. Это минимально возможная длина проверяющего теста в принципе, для самопроверяемых тестеров равновесных кодов [27].

Наличие простейшего тестера с минимальной длиной проверяющего теста открывает хорошие перспективы использования данной композиции равновесных кодов при синтезе цифровых устройств со свойствами контролепригодности, самопроверяемости и отказоустойчивости.

Итак, СВК строится не для всех выходов объекта диагностирования, а для подмножества из четырех его выходов. Обозначим их следующим образом: $\{f_1(X), f_2(X), f_3(X), f_4(X)\}$.

БКС представляет собой каскад параллельно расположенных двухвходовых XOR . На один из входов каждого элемента XOR подается сигнал с j -го выхода объекта диагностирования, а на второй – с j -го выхода блока $G(X)$. Это позволяет выполнить коррекцию вида:

$$h_j(X) = f_j(X) \oplus g_j(X), \quad j = \overline{1, 4}. \quad (3)$$

В соответствии с (3) для коррекции используются именно элементы XOR , поскольку они позволяют любой из сигналов 0 (1) преобразовать в любой из сигналов 0 (1). Аналогичным свойством обладает еще одна элементарная булева функция – функция эквивалентности, реализуемая элементом $XNOR$. Он может использоваться как альтернативный. Иные элементарные булевы функции могут быть использованы только в ряде частных случаев структур объектов диагностирования.

Коррекция сигналов $f_j(X)$, $j = \overline{1, 4}$, осуществляется таким образом, чтобы выполнялись следующие условия:

1. Кодовые слова $\langle h_4(X) h_3(X) h_2(X) h_1(X) \rangle$, формируемые на выходах БКС, должны принадлежать композиции равновесных кодов «1 из 4» и «3 из 4» при поступлении каждого из наборов значений аргументов.

2. Каждая из функций $h_j(X)$, $j = \overline{1, 4}$, должна принадлежать классу самодвойственных булевых функций.

Корректность выполнения обозначенных условий устанавливается на этапе проектирования СВК и проверяется в процессе ее эксплуатации МКВ. В данном модуле выделено две подсхемы контроля: по принадлежности к кодовым словам композиции равновесных кодов «1 из 4» и «3 из 4» и по принадлежности функций к классу самодвойственных. Первая подсхема образована тестером данной композиции – устройством $(1/4+3/4)$ -TSC (totally self-checking checker) (Рис. 2). Данное устройство имеет два выхода $z_1^0(X)$ и $z_1^1(X)$. Именно два выхода требуется потому, что данное устройство является «последним сторожем» в подсхеме контроля. Неисправности элемента, связанного с единственным выходом, не обнаруживались бы. Вторая подсхема содержит два устройства: схему сжатия самодвойственных сигналов (*SDCC* – self-dual compression circuit) [11] и тестер самодвойственных сигналов (*SDC* – self-dual checker) [16 – 18]. Схема сжатия строится на трехвыходовых *XOR* по формуле:

$$\sigma = [h_1(X) \oplus h_2(X) \oplus h_3(X)] \oplus h_4(X) \oplus a]. \quad (4)$$

Вложенные скобки в (4) вида [...] обозначают сложение на первом *XOR*, а внешние – на втором *XOR*.

Самодвойственность сигнала σ контролируется с помощью специального тестера самодвойственного сигнала. Его структура приведена на рисунке 3. При помощи линии задержки, равной одному такту импульсной последовательности a , сигнал σ преобразуется в двухфазный сигнал $\langle v_1 \ v_2 \rangle$. При отсутствии ошибок в импульсной последовательности в контрольных тактах в точках v_1 и v_2 формируются сигналы с противоположными значениями. В рабочих тактах могут формироваться сигналы как с одинаковыми, так и с противоположными значениями. При наличии ошибки в сигнале в некотором периоде в контрольном такте этого периода в точках v_1 и v_2 всегда формируются сигналы с одинаковыми значениями. Отсюда следует, что контроль самодвойственного сигнала можно осуществлять только в контрольном такте. Схема *SDC* преобразует самодвойственный сигнал σ в пространственный парафазный сигнал $\langle z_2^1(X) \ z_2^0(X) \rangle$. При этом во время рабочего такта, когда $a = 0$, значения сигналов на выходах $z_2^1(X)$ и $z_2^0(X)$ не зависят от сигналов в точках v_1 и v_2 , а на них всегда формируется парафазный сигнал $\langle 10 \rangle$. Во время контрольного такта, когда $a = 1$, устанавливаются равенства $z_2^1(X) = v_2$ и $z_2^0(X) = v_1$,

а следовательно, осуществляется контроль самодвойственного сигнала в текущем периоде.

Выходы отдельных подсистем контроля по разным диагностическим признакам подключаются к входам одного модуля сжатия парафазных сигналов *TRC* (two-rail checker) [28] для получения одной пары наблюдаемых контрольных выходов $z^0(X)$ и $z^1(X)$.

Важной схемотехнической особенностью СВК является то, что БКС должен быть построен таким образом, чтобы корректировались сигналы со всех четырех выходов выделяемого подмножества выходов объекта диагностирования.

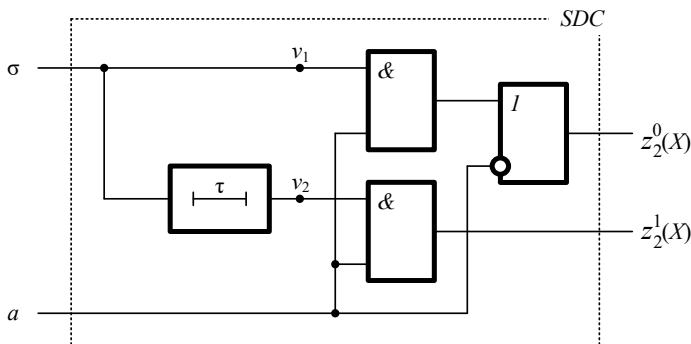


Рис. 3. Тестер самодвойственного сигнала

Теорема 1. Обеспечить условия функционирования СВК с контролем вычислений по признаку принадлежности композиции равновесных кодов «1 из 4» и «3 из 4» и классу самодвойственных булевых функций для любых устройств $F(X)$ можно тогда и только тогда, когда корректируются все четыре сигнала от объекта диагностирования.

Доказательство. Предположим, что условие теоремы неверно. Тогда можно использовать меньшее количество элементов коррекции в СВК. Пусть оно равно трем. Тогда, к примеру, $h_4(X) = f_4(X)$, а для остальных выходов используется коррекция вида (3).

Скорректировать любое кодовое слово $\langle f_4(X) f_3(X) f_2(X) f_1(X) \rangle$ в кодовое слово $\langle h_4(X) h_3(X) h_2(X) h_1(X) \rangle$, принадлежащее композиции равновесных кодов «1 из 4» и «3 из 4», в этом случае можно всегда, так как при всех значениях $f_4(X)$ можно во множестве кодовых слов заданной композиции найти требуемый вариант: для случая $\langle 0 h_3(X) h_2(X) h_1(X) \rangle$ – это слова из множества $\{0001, 0010, 0100, 0111\}$,

для случая $\langle 1 h_3(X) h_2(X) h_1(X) \rangle$ – это слова из множества $\{1000, 1110, 1101, 1011\}$.

Второе условие говорит о том, что каждая функция $h_j(X)$ должна быть самодвойственной. Положим, что значение функции $h_4(X)$ контролироваться не будет вовсе и на нее ограничений не накладывается (это потенциально возможно). Определим, можно ли в таком случае обеспечить самодвойственность оставшихся функций при установлении значений векторов $\langle h_4(X) h_3(X) h_2(X) h_1(X) \rangle$ на этапе проектирования СВК. Самодвойственная функция на ортогональных наборах значений аргументов принимает противоположные значения, поэтому нужно обеспечить это свойство для каждой функции $h_j(X)$. Однако если на таких наборах $f_4(X)$ не меняет своего значения, то имеется два случая кодовых слов: $\langle 0 h_3(X) h_2(X) h_1(X) \rangle$, $\langle 0 h_3(X) h_2(X) h_1(X) \rangle$ и $\langle 1 h_3(X) h_2(X) h_1(X) \rangle$, $\langle 1 h_3(X) h_2(X) h_1(X) \rangle$. Анализ множества слов композиции равновесных кодов «1 из 4» и «3 из 4» показывает, что для каждой из указанных пар в выбранной композиции равновесных кодов не существует требуемых кодовых слов. При увеличении числа не корректируемых сигналов указанная особенность сохраняется.

Таким образом, предположение неверно и теорема доказана. ■

Можно обратить внимание на то, что коррекция всех четырех сигналов от объекта диагностирования соответствует по избыточности дублированию. Однако за счет возможностей оптимизации структуры блока $G(X)$ можно добиться снижения показателей структурной избыточности СВК. Структуры всех контрольных блоков СВК при этом стандартны и просты. Тут также следует подчеркнуть, что при описанной организации СВК при незначительном ее усложнении достигается существенный эффект в части обнаружения неисправностей. Исследования [17, 18] показывают, что достигается увеличение показателя наблюдаемости неисправностей до 20-25 % в зависимости от структуры и конфигурации связей с входами объекта диагностирования, элементами его внутренней структуры и выходами объекта диагностирования. Аналогично может быть организовано и дублирование в импульсном режиме функционирования для повышения показателей наблюдаемости неисправностей. Именно с данной структурой целесообразно дальнейшее сравнение предложенной авторами структуры.

3. Синтез полностью самопроверяемой структуры схемы встроенного контроля. В структуре рисунка 1 все элементы являются типовыми, кроме блока $G(X)$ вычисления значений функций коррекции. Он строится на этапе проектирования СВК с учетом

особенностей структуры объекта диагностирования $F(X)$: на каждом наборе значений аргументов следует установить требуемые значения на выходах блока $G(X)$ с учетом значений на выходах блока $F(X)$.

Теорема 2. Число способов построения СВК без учета тестируемости элементов коррекции сигналов равно:

$$N = 2^{3 \cdot 2^{t-1} - 2}. \quad (5)$$

Доказательство. Всего существует 2^t наборов значений аргументов. Поскольку речь идет о самодвойственных функциях $h_j(X)$, $j = \overline{1, 4}$, то их значения требуется установить на половине наборов – на 2^{t-1} наборах значений аргументов, соответствующих десятичным числам $0 \dots 2^{t-1} - 1$. Если брать полное множество кодовых слов, а не только подмножества T_{Π_1} или T_{Π_2} из (2), то потребуется обеспечить хотя бы единожды формирование каждого из наборов, входящих в T_{Π_1} (T_{Π_2}) при формировании полных тестов для элементов XOR . Обратим внимание, что в формуле (2) в каждом из подмножеств T_{Π_1} (T_{Π_2}) присутствует ровно по две пары ортогональных по всем разрядам кодовых слов. Это означает, что достаточно выбрать одно из двух подмножеств и хотя бы единожды сформировать одно из слов в паре на первой половине наборов значений аргументов $0 \dots 2^{t-1} - 1$. Выбор первого слова может быть осуществлен восемью способами для любого из наборов значений аргументов $0 \dots 2^{t-1} - 1$, второго для любого из оставшихся – уже двумя. Автоматически будут сформированы и кодовые слова, ортогональные выбранным по всем аргументам на соответствующих наборах из второй половины $2^{t-1} \dots 2^t - 1$. На оставшихся $2^{t-1} - 2$ наборах значений аргументов из первой их половины можно использовать любые из кодовых слов полного их множества. Число способов установления значений равно восьми для каждого набора. Отсюда становится ясным следующее:

$$\begin{aligned} N &= 8 \cdot 2 \cdot 8^{2^{t-1}-2} = 2 \cdot 8^{2^{t-1}-1} = 2 \cdot (2^3)^{2^{t-1}-1} = \\ &= 2^{3 \cdot 2^{t-1} - 3 + 1} = 2^{3 \cdot 2^{t-1} - 2}. \end{aligned} \quad (6)$$

Выражение (6) приводит к формуле (5), что и требовалось доказать. ■

Ясно, что число способов построения СВК в соответствии с (5) растет с ростом t . Поэтому требуется использовать какие-либо зафиксированные последовательности действий при синтезе СВК.

Одним из таких алгоритмов является следующий, основанный на использовании всех восьми кодовых слов композиции равновесных

кодов «1 из 4» и «3 из 4» при равномерном их использовании. Такой алгоритм позволяет упростить процедуру полной проверки $(1/4+3/4)\text{-}TSC$, однако требует проверки тестируемости каждого из элементов коррекции, схемы сжатия и SDC . Уже при небольших значениях $t \geq 4$ тестируемость обеспечить просто, а с ростом t – вероятность автоматической генерации тестов возрастает. При этом возможности обеспечения тестируемости компонентов СВК зависят от распределения возможных кодовых слов на выходах объекта диагностирования – они должны быть достаточно разнообразны.

Алгоритм 1. Правила синтеза блока $G(X)$:

1. Определяется так называемый коэффициент заполняемости μ , характеризующий то, какое количество раз будет повторяться каждое кодовое слово композиции равновесных кодов «1 из 4» и «3 из 4»:

$$\mu = \frac{2^t}{8} = 2^{t-3}. \quad (7)$$

2. Наборы значений аргументов упорядочиваются. Им в соответствие ставятся десятичные эквиваленты $0 \dots 2^t - 1$.

3. На основании (7) определяются наборы значений аргументов, на которых повторяются кодовые слова из множества кодовых слов кода «1 из 4».

3.1. На наборах значений аргументов, соответствующих числам $0 \dots 2^{t-3} - 1$, устанавливают значения разрядов вектора $\langle h_4(X) \ h_3(X) \ h_2(X) \ h_1(X) \rangle$ до кодового слова $\langle 0001 \rangle$.

3.2. На наборах значений аргументов, соответствующих числам $2^{t-3} \dots 2 \cdot 2^{t-3} - 1$, устанавливают значения разрядов вектора $\langle h_4(X) \ h_3(X) \ h_2(X) \ h_1(X) \rangle$ до кодового слова $\langle 0010 \rangle$.

3.3. На наборах значений аргументов, соответствующих числам $2 \cdot 2^{t-3} \dots 3 \cdot 2^{t-3} - 1$, устанавливают значения разрядов вектора $\langle h_4(X) \ h_3(X) \ h_2(X) \ h_1(X) \rangle$ до кодового слова $\langle 0100 \rangle$.

3.4. На наборах значений аргументов, соответствующих числам $3 \cdot 2^{t-3} \dots 2^{t-1} - 1$, устанавливают значения разрядов вектора $\langle h_4(X) \ h_3(X) \ h_2(X) \ h_1(X) \rangle$ до кодового слова $\langle 1000 \rangle$.

4. Кодовые слова из множества кодовых слов кода «3 из 4» получают путем инвертирования разрядов кодовых слов на взаимно инверсных наборах значений аргументов из первой их половины.

5. Проверяются условия тестируемости элементов коррекции, схемы сжатия SDC (для схемы сжатия используются тесты элементов XOR (в самом общем случае – в канонической

форме реализации), элемент SDC тестируется на одной паре инверсных комбинаций при подаче всего двух циклов импульсов и пауз с генератора G , что видно по структуре рисунка 3). Если проверяющие тесты на заданных наборах значений аргументов не формируются для какого-либо устройства, то осуществляется изменение части установленных кодовых слов без нарушений условий формирования кодовых слов заданной композиции равновесных кодов «1 из 4» и «3 из 4» и самодвойственности каждой функции $h_j(X)$, $j = \overline{1, 4}$. Изменения делаются выборочно аналогично тому, как это предложено в [29].

6. Рассчитываются значения функций коррекции:

$$g_j(X) = f_j(X) \oplus h_j(X), \quad j = \overline{1, 4}. \quad (8)$$

Формула (8) непосредственно следует из (3).

7. Получаются выражения, описывающие функции блока $G(X)$.

8. При необходимости функции оптимизируются.

9. Блок $G(X)$ реализуется в выбранном элементном базисе.

Приведем пример использования алгоритма 1.

Пример 1. Для устройства, заданного в первых девяти столбцах таблицы 1, привести описание работы блока $G(X)$.

Решение. Пользуясь непосредственно алгоритмом 1, определим значение $\mu = 2^{4-3} = 2$. Далее на наборах $0 \dots 2^{4-3} - 1 = 1$ устанавливаем значения разрядов вектора $\langle h_4(X) \ h_3(X) \ h_2(X) \ h_1(X) \rangle$ до кодового слова $\langle 0001 \rangle$; на наборах $2^{4-3} = 2 \dots 2 \cdot 2^{4-3} - 1 = 3 - \langle 0010 \rangle$; на наборах $2 \cdot 2^{4-3} = 4 \dots 3 \cdot 2^{4-3} - 1 = 5 - \langle 0100 \rangle$; на наборах $3 \cdot 2^{4-3} = 6 \dots 2^{4-1} - 1 = 7 - \langle 1000 \rangle$. Далее получают значения функций $h_j(X)$, $j = \overline{1, 4}$, на наборах $2^{4-1} = 8 \dots 2^4 - 1 = 15$. Проверяем формирование полного множества тестовых комбинаций для элементов преобразования (они записаны в последних четырех столбцах таблицы 1).

Анализ формулы (2) показывает, что алгоритм 1 может быть модифицирован под использование не полного множества кодовых слов композиции равновесных кодов «1 из 4» и «3 из 4», а под использование только тех кодовых слов, которые входят отдельно в T_{Π_1} и T_{Π_2} . Это следует из того, что проверяющие тесты содержат ровно по две пары ортогональных по всем разрядам кодовых слов. Тогда можно предложить следующий алгоритм синтеза блока $G(X)$ с учетом использования неполного множества кодовых слов композиции равновесных кодов «1 из 4» и «3 из 4».

Алгоритм 2. Правила синтеза блока $G(X)$ при условии формирования проверяющего теста T_{Π_1} и использовании только входящих в него кодовых слов:

1. Определяется коэффициент заполняемости μ , характеризующий то, какое количество раз будет повторяться каждое кодовое слово из T_{Π_1} :

$$\mu = \frac{2^t}{4} = 2^{t-2}. \quad (9)$$

2. Наборы значений аргументов упорядочиваются. Им в соответствие ставятся десятичные эквиваленты $0 \dots 2^t - 1$.

3. Кодовые слова из проверяющего теста разбиваются на пары $\{1000, 0111\}$ и $\{0010, 1101\}$.

3.1 На наборах значений аргументов, соответствующих числам $0 \dots 2^{t-2} - 1$, что следует из (9), устанавливают значения разрядов вектора $\langle h_4(X) \ h_3(X) \ h_2(X) \ h_1(X) \rangle$ до кодового слова $\langle 1000 \rangle$.

3.2 На наборах значений аргументов, соответствующих числам $2^{t-2} \dots 2^{t-1} - 1$, что также следует из (9), устанавливают значения разрядов вектора $\langle h_4(X) \ h_3(X) \ h_2(X) \ h_1(X) \rangle$ до кодового слова $\langle 0010 \rangle$.

4. Кодовые слова $\langle 0111 \rangle$ и $\langle 1101 \rangle$ получают путем инвертирования разрядов кодовых слов на взаимно инверсных наборах значений аргументов из первой их половины.

5. Выполняются п. 5-9 алгоритма 1.

Аналогично будет работать алгоритм, использующий кодовые слова из T_{Π_2} .

Приведем пример использования алгоритма 2.

Пример 2. Для устройства из примера 1 привести описание работы блока $G(X)$ по алгоритму 2.

Решение. Аналогично решению в примере 1. Результаты работы алгоритма 2 приведены в таблице 2.

Алгоритмы 1 и 2 подразумевают работу с подмножеством из четырех выходов объекта диагностирования. Для устройства с числом выходов $n > 4$ можно использовать следующий алгоритм.

Таблица 1. Описание работы самопроверяемого устройства
для данных примера 1

№	Значения входов				Значения выходов $F(X)$				Значения на выходах БКС				Значения выходов $G(X)$				Тестовые комбинации			
	x_4	x_3	x_2	x_1	$f_4(X)$	$f_5(X)$	$f_2(X)$	$f_1(X)$	$h_4(X)$	$h_3(X)$	$h_2(X)$	$h_1(X)$	$g_4(X)$	$g_3(X)$	$g_2(X)$	$g_1(X)$	XOR_4	XOR_3	XOR_2	XOR_1
0	0	0	0	0	0	1	1	0	0	0	0	1	0	1	1	1	00	11	11	01
1	0	0	0	1	1	0	0	1	0	0	0	1	1	0	0	0	11	00	00	10
2	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	1	00	00	01	11
3	0	0	1	1	1	1	1	0	0	0	1	0	1	1	0	0	11	11	10	00
4	0	1	0	0	0	1	1	1	0	1	0	0	0	0	1	1	00	10	11	11
5	0	1	0	1	0	0	1	1	0	1	0	0	0	1	1	1	00	01	11	11
6	0	1	1	0	0	0	0	0	1	0	0	0	1	0	0	0	01	00	00	00
7	0	1	1	1	0	1	1	1	1	0	0	0	1	1	1	1	01	11	11	11
8	1	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	11	01	01	01
9	1	0	0	1	1	1	0	0	0	1	1	1	1	0	1	1	11	10	01	01
10	1	0	1	0	0	1	0	0	1	0	1	1	1	1	1	1	01	11	01	01
11	1	0	1	1	1	0	1	1	1	0	1	1	0	0	0	0	10	00	10	10
12	1	1	0	0	1	0	1	1	1	1	0	1	0	1	1	0	10	01	11	10
13	1	1	0	1	0	0	1	1	1	1	0	1	1	1	1	0	01	01	11	10
14	1	1	1	0	0	1	0	0	1	1	1	0	1	0	1	0	01	10	01	00
15	1	1	1	1	1	0	0	0	1	1	1	0	0	1	1	0	10	01	01	00

Таблица 2. Описание работы самопроверяемого устройства
для данных примера 2

№	Значения входов				Значения выходов $F(X)$				Значения на выходах БКС				Значения выходов $G(X)$				Тестовые комбинации			
	x_4	x_3	x_2	x_1	$f_4(X)$	$f_5(X)$	$f_2(X)$	$f_1(X)$	$h_4(X)$	$h_3(X)$	$h_2(X)$	$h_1(X)$	$g_4(X)$	$g_3(X)$	$g_2(X)$	$g_1(X)$	XOR_4	XOR_3	XOR_2	XOR_1
0	0	0	0	0	0	1	1	0	1	0	0	0	1	1	1	0	01	11	11	00
1	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	1	10	00	00	11
2	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	1	01	00	00	11
3	0	0	1	1	1	1	1	0	1	0	0	0	0	1	1	0	10	11	11	00
4	0	1	0	0	0	1	1	1	0	0	1	0	0	1	0	1	00	11	10	11
5	0	1	0	1	0	0	1	1	0	0	1	0	0	0	0	1	00	00	10	11
6	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	00	00	01	00
7	0	1	1	1	0	1	1	1	0	0	1	0	0	1	0	1	00	11	10	11
8	1	0	0	0	1	0	0	0	1	1	0	1	0	1	0	1	10	01	00	01
9	1	0	0	1	1	1	0	0	1	1	0	1	0	0	0	1	10	10	00	01
10	1	0	1	0	0	1	0	0	1	1	0	1	1	0	0	1	01	10	00	01
11	1	0	1	1	1	0	1	1	1	1	0	1	0	1	1	0	10	01	11	10
12	1	1	0	0	1	0	1	1	0	1	1	1	1	1	0	0	11	01	10	10
13	1	1	0	1	0	0	1	1	0	1	1	1	0	1	0	0	00	01	10	10
14	1	1	1	0	0	1	0	0	0	1	1	1	0	0	1	1	00	10	01	01
15	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	11	01	01	01

Алгоритм 3. Правила синтеза СВК для устройств с $n > 4$:

1. Множество выходов $W = \{f_n(X), f_{n-1}(X), \dots, f_2(X), f_1(X)\}$ исходного устройства разбивается на подмножества $W_1, W_2, \dots, W_{q-1}, W_q$ мощностью 4, $q = \left\lfloor \frac{n}{4} \right\rfloor$, такие, что $W = W_1 \cup W_2 \cup \dots \cup W_{q-1} \cup W_q$. Причем, если $n(\bmod 4) = 0$, то:

$$W_i = \{f_{4(i-1)+1}, f_{4(i-1)+2}, f_{4(i-1)+3}, f_{4(i-1)+4}\}, \overline{i = 1, q}, \quad (10)$$

если $n(\bmod 4) \neq 0$, то:

$$\begin{aligned} W_i &= \{f_{4(i-1)+1}, f_{4(i-1)+2}, f_{4(i-1)+3}, f_{4(i-1)+4}\}, \overline{i = 1, q-1}, \\ W_q &= \{f_{n-3}, f_{n-2}, f_{n-1}, f_n\}. \end{aligned} \quad (11)$$

2. Для каждого полученного подмножества $W_1, W_2, \dots, W_{q-1}, W_q$ в соответствии с формулами (10) и (11) строится СВК на основе структуры рисунка 2.

3. Устанавливаются q тестеров композиции равновесных кодов «1 из 4» и «3 из 4» в каждой СВК для каждого подмножества $W_1, W_2, \dots, W_{q-1}, W_q$.

4. Выходы каждого (1/4+3/4)-TSC подключаются к входам самопроверяемого компаратора, состоящего из $q-1$ самопроверяемых модулей сжатия парафазных сигналов TRC .

Работа алгоритма 3 видится тривиальной и, по мнению авторов, не требует примера.

4. Метод выделения подмножеств контролепригодных для композиции равновесных кодов «1 из 4» и «3 из 4» выходов на множестве выходов объекта диагностирования. Для построения полностью самопроверяемой структуры СВК требуется не только обеспечить тестируемость ее компонентов, но и покрыть все возможные ошибки, вызываемые неисправностями из заданной модели на выходах объекта диагностирования $F(X)$. Это можно сделать путем выделения подмножеств контролепригодных выходов относительно каждой неисправности из заданного класса на полном множестве выходов объекта диагностирования $F(X)$.

Пусть имеется объект диагностирования $F(X)$, снабженный t входами и n выходами, а также зафиксирована модель неисправностей – например, модель одиночных константных неисправностей выходов внутренних элементов структуры (stuck-at faults) [30, 31]. Покажем, каким образом следует выделять подмножества выходов объекта

диагностирования для фиксации любых ошибок, вызываемых неисправностями на его выходах.

Алгоритм 4. Правила выделения подмножеств контролепригодных выходов объекта диагностирования:

1. Рассматривается структура объекта диагностирования $F(X)$. Среди множества логических элементов Θ выделяются только те элементы, которые связаны путями с двумя и более выходами устройства: $\Theta^2 \subset \Theta$. Все элементы, связанные путями с одним выходом, исключаются из рассмотрения, поскольку композиция равновесных кодов «1 из 4» и «3 из 4» гарантирует покрытие одиночных искажений в кодовых словах.

2. Строится таблица истинности устройства, работающего без неисправностей, – формируются эталонные кодовые слова $\langle f_n(X) f_{n-1}(X) \dots f_2(X) f_1(X) \rangle$.

3. Выполняется произвольное преобразование каждой функции $f_j(X)$ в самодвойственную функцию $h_j(X)$ с применением фиктивных функций $g_j(X)$. Можно использовать следующую зависимость:

$$h_j(X) = x_i f_j(X) \vee \overline{x_i f_j(\overline{X})}, \quad j = \overline{1, n}. \quad (12)$$

После получения значений $h_j(X)$ по формуле (12) вычисляются значения $g_j(X)$ на каждом наборе значений аргументов:

$$g_j(X) = f_j(X) \oplus h_j(X), \quad j = \overline{1, n}. \quad (13)$$

Значения $g_j(X)$, полученные по формуле (13), фиксируются и не изменяются далее.

4. Моделируются одиночные константные неисправности тех логических элементов, которые входят во множество Θ^2 . Неисправности нумеруются и ранжируются: 1, 2, ..., d . Оценивается их влияние на выходы, а также определяются возможности покрытия их при контроле самодвойственности функций $h_j(X)$. Для этого вычисляются две функции:

$$\psi_j^d(X) = f_j(X) \oplus f_j'(X), \quad (14)$$

$$\psi_j^g(X) = h_j'(X) \oplus h_j'(\overline{X}), \quad (15)$$

где $f'_j(X)$ – значение функции, реализуемой на выходе $f_j(X)$, при внесении неисправности d в объект диагностирования; $h'_j(X)$ – значение функции, реализуемой на выходе $h_j(X)$ БКС в СВК, при внесении неисправности d в объект диагностирования.

Функция (14) представляет собой функцию ошибки: $\psi_j^d(X) = 1$, если ошибка повлияла на выходы устройства $F(X)$. Функция (15) – функция фиксации ошибки при контроле самодвойственности: $\psi_j^g(X) = 1$, если ошибка обнаружена при контроле самодвойственности вычислений.

Дальнейшие действия алгоритма основаны на следующих важнейших положениях.

Теорема 3. Ошибка при контроле самодвойственности вычислений функции $h_j(X)$ будет обнаружена тогда и только тогда, когда:

$$\chi_j = \psi_j^d(X)\psi_j^g(X) = 1. \quad (16)$$

Доказательство. Функция (16) является логическим произведением функций (14) и (15). При решении уравнения $\psi_j^d(X) = 1$ получаются те наборы значений аргументов, на которых неисправность проявляется в виде ошибки выхода $f_j(X)$. При решении уравнения $\psi_j^g(X) = 1$ получаются те наборы значений аргументов, на которых фактически нарушается самодвойственность функции $h_j(X)$. Ясно, что значение $h_j(X)$ напрямую зависит от значений $f_j(X)$ и $g_j(X)$. Из (13) имеем $h_j(X) = f_j(X) \oplus g_j(X)$, $j = \overline{1, n}$. Но у нас зафиксированы значения $g_j(X)$ и меняются значения $f_j(X)$ при внесении неисправностей. Это приводит к искажениям значений $h_j(X)$. Таким образом, во множестве наборов значений аргументов, на которых $\psi_j^g(X) = 1$, неизбежно будут и наборы значений аргументов, на которых $\psi_j^d(X) = 1$. Логически перемножая их, мы отсекаем те наборы, на которых ошибка не фиксируется при контроле самодвойственности вычислений, но выделяем те, которые позволяют зафиксировать ошибку. Выражение (16) требуется проверять именно из-за того, что в структуре рисунка 1 наборы значений аргументов подаются парами, и ошибки фиксируются фактически дважды – когда на первой половине таблицы истинности наборы первой половины таблицы являются рабочими, а второй – контрольными, и наоборот. ■

Теорема 4. Ошибка при контроле самодвойственности вычислений функции $h_j(X)$ не будет обнаружена тогда и только тогда, когда:

$$\varsigma_j = \psi_j^d(X) \overline{\psi_j^\sigma(X)} = 1. \quad (17)$$

Доказательство. Справедливость (17) вытекает непосредственно из теоремы 3 и обосновывается тем, что функция $\overline{\psi_j^\sigma(X)} = 1$ тогда и только тогда, когда ошибка не повлияла на свойство самодвойственности функции. Она либо не исказила функцию $h_j(X)$, либо исказила ее на обоих наборах комбинации. Логически умножая функцию $\overline{\psi_j^\sigma(X)}$ на $\psi_j^d(X)$, мы выделяем те наборы значений аргументов, на которых ошибку зафиксировать не удалось. ■

5. Все ошибки, удовлетворяющие (16), необходимо исключить из рассмотрения. Далее рассматриваются только те ошибки, которые удовлетворяют (17).

6. Далее нужно исключить все одиночные и трехкратные ошибки. Они обнаруживаются композицией равновесных кодов «1 из 4» и «3 из 4». Для этого можно посчитать арифметическую сумму:

$$\varsigma_{max} = \sum_{j=1}^n \varsigma_j. \quad (18)$$

Если при расчетах по выражению (18) получены значения $\varsigma_{max} = 1$ или $\varsigma_{max} = 3$, то ошибку следует исключить из рассмотрения.

7. Все выходы, для которых $\varsigma_j = 1$ при условии $\varsigma_{max} \neq 1$ и $\varsigma_{max} \neq 3$, требуется выполнить разбиение на различные подмножества и включить их в разные четверки выходов при выделении контролепригодных по композиции равновесных кодов «1 из 4» и «3 из 4» подмножеств выходов.

Пример 3. Выполнить контролепригодное разделение выходов объекта диагностирования $F(X)$, структура которого приведена на рисунке 4, а описание дано в таблице 3, в соответствии с алгоритмом 4.

Решение. В соответствии с алгоритмом 4 выполняем расчеты. В рассматриваемом случае требуется проанализировать влияние константных неисправностей элементов G^* , G^{**} и G^{***} на выходы объекта диагностирования. Чтобы это сделать, сформируем самодвойственные функции $h_j(X)$ в соответствии с п. 2 и 3 алгоритма 4. Далее последовательно вносим одиночные константные неисправности элементов G^* , G^{**} и G^{***} в схему устройства и моделируем его работу. В качестве примера приведем работу алгоритма и формул (14) и (15) при внесении неисправности типа «константа 1» (stuck-at-1) на выходе элемента G^{**} , показано в таблице 4.

Расчеты показали, что нужно разделить выходы $f_1(X)$ и $f_3(X)$, а также выходы $f_3(X)$ и $f_4(X)$.

Отсюда становится ясным итоговое разбиение выходов объекта диагностирования на подмножества из четырех выходов: $\{f_1(X), f_2(X), f_4(X), f_5(X)\}$ и $\{f_2(X), f_4(X), f_5(X), f_6(X)\}$. Для них далее строится СВК по описанному выше методу с контролем вычислений по нескольким диагностическим признакам.

В таблице 4 выделены те строки, которые соответствуют необнаруживаемым ошибкам при контроле самодвойственности функций $h_j(X)$. Это как раз показывает необходимость контроля вычислений еще по одному диагностическому признаку.

Представленный в данном разделе метод позволяет одновременно учесть и свойства контроля неисправностей по признаку самодвойственности вычисляемых функций, и свойства контроля неисправностей по композиции равновесных кодов «1 из 4» и «3 из 4». Тут следует отметить, что при построении самопроверяемого цифрового устройства можно учитывать не все ошибки, возникающие на выходах объекта диагностирования, а фиксировать каждую из неисправностей хотя бы по одной из ошибок, как это предложено в [32] при использовании классического кода паритета. Однако данное развитие алгоритма 4 оставим за пределами настоящей работы.

Таблица 3. Описание работы объекта диагностирования и СВК, к примеру 3

N_0	x_4	x_3	x_2	x_1	$f_6(X)$	$f_5(X)$	$f_4(X)$	$f_3(X)$	$f_2(X)$	$f_1(X)$	$h_6(X)$	$h_5(X)$	$h_4(X)$	$h_3(X)$	$h_2(X)$	$h_1(X)$	$g_6(X)$	$g_5(X)$	$g_4(X)$	$g_3(X)$	$g_2(X)$	$g_1(X)$
0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	1	1	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
5	0	1	0	1	1	1	1	0	1	0	1	1	1	0	1	0	0	0	0	0	0	0
6	0	1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0
7	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0
8	1	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	1	0	0	1	1	1
9	1	0	0	1	1	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	1
10	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1
11	1	0	1	1	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	1	0	1
12	1	1	0	0	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0
13	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	0	1
14	1	1	1	0	0	1	1	1	1	0	0	0	0	1	1	1	0	1	1	0	0	1
15	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0

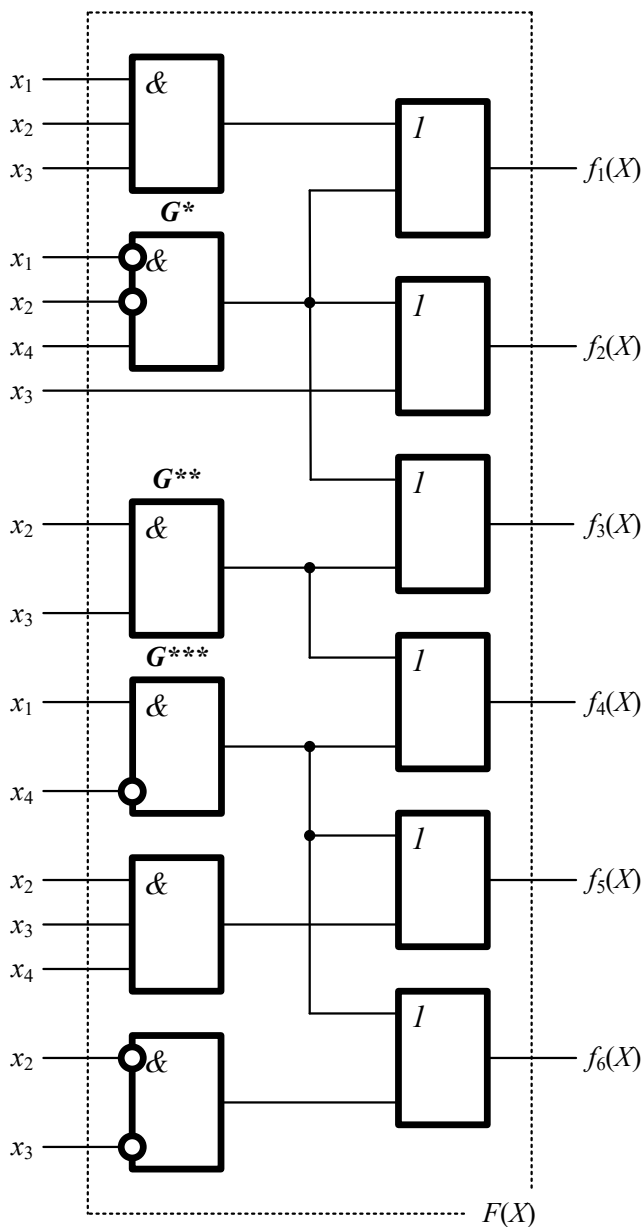


Рис. 4. Объект диагностирования, к примеру 3

Таблица 4. Пример фиксации необнаруживаемых при контроле самодвойственности ошибок

№	x_4	x_3	x_2	x_1	$f_6(X)$	$f_5(X)$	$f_4(X)$	$f_3(X)$	$f_2(X)$	$f_1(X)$	$h_6(X)$	$h_5(X)$	$h_4(X)$	$h_3(X)$	$h_2(X)$	$h_1(X)$	$g_6(X)$	$g_5(X)$	$g_4(X)$	$g_3(X)$	$g_2(X)$	$g_1(X)$	φ^6
0	0	0	0	0	1	0	1	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
3	0	0	1	1	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0
5	0	1	0	1	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0	0
6	0	1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0
7	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
8	1	0	0	0	1	0	1	1	1	1	0	0	1	0	0	0	1	0	0	1	1	1	0
9	1	0	0	1	1	0	1	1	0	0	1	1	1	1	0	1	0	1	0	0	0	1	0
10	1	0	1	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0
11	1	0	1	1	0	0	1	1	0	0	1	1	0	0	0	1	1	1	1	1	0	1	0
12	1	1	0	0	0	0	1	1	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0
13	1	1	0	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	0	1	0
14	1	1	1	0	0	1	1	1	1	0	0	0	0	1	1	1	0	1	1	0	0	1	0
15	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0	0	0
№	φ^5	φ^4	φ^3	φ^2	φ^1	φ^6	φ^5	φ^4	φ^3	φ^2	φ^1	χ_6	χ_5	χ_4	χ_3	χ_2	χ_1	ζ_6	ζ_5	ζ_4	ζ_3	ζ_2	ζ_1
0	0	1	1	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0
3	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0
5	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
6	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9	0	1	1	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0
10	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0
11	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0
12	0	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0
14	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

5. Некоторые результаты экспериментов с тестовыми цифровыми устройствами. В ходе исследований особенностей разработанного метода синтеза самопроверяемых цифровых устройств были проведены эксперименты по оценке показателей структурной избыточности реализуемых устройств для тестовых комбинационных

схем из набора MCNC Benchmarks [33]. В качестве показателя структурной избыточности была использована метрика площади, занимаемой устройством на кристалле, которую можно получить с применением известного интерпретатора SIS [34]. В исследованиях использована стандартная библиотека функциональных элементов stdcell2_2.genlib.

В ходе исследований было произведено сравнение разработанного авторами метода с двумя методами: основанным на дублировании с импульсным режимом работы и разработанным ранее методом, основанного на ЛКС и применении равновесного кода «2 из 4» из [16].

Для выбранных тестовых схем, заданных в матричной (интервальной) форме в формате *.pla, были получены файлы-описания блоков $G(X)$ в соответствии с алгоритмами 1 и 3. Сами файлы-описания получались путем конкатенации векторов $\langle h_4(X) h_3(X) h_2(X) h_1(X) \rangle$ отдельных подсхем контроля. Затем структуры блоков $G(X)$ оптимизировались с использованием SIS при выполнении процедуры full_simplify и осуществлялась процедура map -s для реализации устройства в заданной библиотеке элементов и определения показателей площади. Результаты, полученные в ходе исследований, занесены в таблицу 5.

В таблице 5 столбец $L_{F(X)}$ характеризует площадь объекта диагностирования, а столбец $L_{G(X)}$ – площадь блока вычисления значений функций коррекции для разработанного авторами метода. В столбце $L_{1/4+3/4}$ представлено суммарное значение площади для самопроверяемой реализации устройства. Для сравнения в столбцах L_{D+SD} и $L_{2/4}$ приведены данные по применению дублирования с самодвойственным контролем и равновесного кода «2 из 4».

Расчеты для структур, подразумевающих использование равновесного кода «2 из 4» и композиции равновесных кодов «1 из 4» и «3 из 4», осуществлялись по формуле:

$$L_{CED} = L_{F(X)} + L_{G(X)} + 4L_{XOR} + L_{TSC} + L_{SDCC} + L_{SDC} + L_{TRC}, \quad (19)$$

где нижние индексы указывают на конкретные элементы в СВК (Рис. 1).

Данные по стандартным элементам в СВК определены с использованием показателей stdcell2_2.genlib: $L_{XOR} = 40$ (усл. ед.), $L_{SDCC} = 160$ (усл. ед.), $L_{SDC} = 96$ (усл. ед.), $L_{TRC} = 12$ (усл. ед.), $L_{(1/4+3/4)-TSC} = 96$ (усл. ед.), $L_{2/4-TSC} = 192$ (усл. ед.).

В расчетах не учитывалась входная цепь преобразования и генератор G . Во всех рассматриваемых самопроверяемых устройствах данные цепи одинаковы.

Таблица 5. Результаты экспериментов с тестовыми схемами
MCNC Benchmarks

Номер	Тестовая схема	n	q	$L_{xor}^{(n)}$ усл. ед.	$L_{and}^{(n)}$ усл. ед.	$L_{or}^{(n)}$ усл. ед.	$L_{nand}^{(n)}$ усл. ед.	$L_{nor}^{(n)}$ усл. ед.	$\delta_{1,4+3,4}$ %	$\delta_{2,4}$ %	$\epsilon_{1,4+3,4}$ %	$\epsilon_{2,4}$ %	χ %
1	rd84	4	1	4912	2680	8296	10304	12472	66,517	82,617	33,483	17,383	80,512
2	sqrt8	4	1	1160	1200	3064	3560	4848	63,201	73,432	36,799	26,568	86,067
3	sao2	4	1	2992	3584	7280	7960	9448	77,053	84,251	22,947	15,749	91,457
4	dist	5	2	6968	9016	17584	19168	19384	90,714	98,886	9,286	1,114	91,736
5	newcwp	5	2	440	776	2816	2896	2680	105,075	108,06	-5,075	-8,06	97,238
6	root	5	2	3496	2968	8064	9048	10360	77,838	87,336	22,162	12,664	89,125
7	max512	6	2	9632	7824	19056	20264	25680	74,206	78,91	25,794	21,09	94,039
8	dc1	7	2	976	832	3408	3480	4544	75	76,585	25	23,415	97,931
9	dekoder	7	2	736	776	3112	3344	4176	74,521	80,077	25,479	19,923	93,062
10	wim	7	2	712	976	3288	3256	4152	79,191	78,42	20,809	21,58	100,983
11	newapla2	7	2	600	1008	3208	3400	4056	79,093	83,826	20,907	16,174	94,353
12	dc2	7	2	2424	2736	6760	7984	8816	76,679	90,563	33,321	9,437	84,669
13	newbyte	8	2	592	840	3032	3288	4280	70,841	76,822	29,159	23,178	92,214
14	mlp4	8	2	7224	8040	16864	18368	20352	82,862	90,252	17,138	9,748	91,812
15	f51m	8	2	2272	5448	9320	12560	7480	124,599	167,914	-24,599	-67,914	74,204
16	inc	9	3	2376	2864	7736	8408	9064	85,349	92,763	14,651	7,237	92,008
17	dk27	9	3	528	2360	5384	7856	4608	116,84	170,486	-16,84	-70,486	68,534
18	newepla2	10	3	1896	3632	8024	9048	8512	94,267	106,297	5,733	-6,297	88,683
19	sq6	12	3	2648	2968	8112	8608	10192	79,592	84,458	20,408	15,542	94,238
20	m1	12	3	3064	2032	7592	7888	11112	68,323	70,986	31,677	29,014	96,247
21	p82	14	4	2368	2296	8056	8520	10384	77,581	82,049	22,419	17,951	94,554
22	m2	16	4	10096	4616	18104	20576	29136	62,136	70,621	37,864	29,379	87,986
23	m3	16	4	13464	6208	23064	24640	36040	63,996	68,368	36,004	31,632	93,604
24	m4	16	4	18704	8248	30344	31616	49064	61,846	64,438	38,154	35,562	95,977
25	tms	16	4	6784	5368	15544	16656	21664	71,75	76,883	28,25	23,117	93,324
26	newepla1	16	4	2520	5144	11056	14232	12080	91,523	117,815	8,477	-17,815	77,684
27	newxcpla1	23	6	3760	6096	15040	21048	17704	84,953	118,888	15,047	-18,888	71,456
28	max128	24	6	20192	9872	35248	39792	52408	67,257	75,927	32,743	24,073	88,581

Расчет для структуры дублирования с самодвойственным контролем выполнялся по формуле:

$$L_{D+SD} = 2L_{F(X)} + nL_{NOT} + (n-1)L_{TRC} + L_{G(X)} + nL_{XOR} + \left\lfloor \frac{n}{2} \right\rfloor L_{SDCC} + L_{SDC} + L_{TRC}, \quad (20)$$

где нижние индексы указывают на конкретные элементы в СВК, синтезированной по методу дублирования. Следует пояснить, что $L_{NOT} = 16$ (усл. ед.), а схема сжатия является преобразователем n сигналов в 1 самодвойственный – для ее реализации требуется $\left\lfloor \frac{n}{2} \right\rfloor$ трехходовых элементов XOR .

Для удобства сравнения данных рассчитаны следующие относительные показатели:

$$\delta_{1/4+3/4} = \frac{L_{1/4+3/4}}{L_{D+SD}} \cdot 100 \%, \quad (21)$$

$$\delta_{2/4} = \frac{L_{2/4}}{L_{D+SD}} \cdot 100 \%, \quad (22)$$

$$\varepsilon_{1/4+3/4} = \frac{L_{D+SD} - L_{1/4+3/4}}{L_{D+SD}} \cdot 100 \%, \quad (23)$$

$$\varepsilon_{2/4} = \frac{L_{D+SD} - L_{2/4}}{L_{D+SD}} \cdot 100 \%, \quad (24)$$

$$\chi = \frac{L_{1/4+3/4}}{L_{2/4}} \cdot 100 \%. \quad (25)$$

Результаты расчетов, произведенных по формулам (20)-(25), занесены в таблицу 4.

Показатели $\delta_{1/4+3/4}$ и $\delta_{2/4}$ характеризуют долю площади, занимаемой самопроверяемым цифровым устройством, синтезированного с СВК на основе равновесных кодов и их композиции, от площади системы дублирования. Показатели $\varepsilon_{1/4+3/4}$ и $\varepsilon_{2/4}$ характеризуют выигрыш в применении методов, основанных на равновесных кодах, по сравнению с дублированием с самодвойственным контролем. Показатель χ характеризует преимущество применения композиции равновесных кодов «1 из 4» и «3 из 4» перед отдельным кодом «2 из 4».

Из анализа таблицы 5 можно сделать следующие выводы. В подавляющем большинстве случаев структуры, основанные на равновесных кодах, оказались более эффективными по показателю структурной избыточности, нежели структура дублирования с самодвойственным контролем. Здесь следует отдельно подчеркнуть, что рассматривалась именно структура с контролем вычислений по двум диагностическим признакам, чтобы можно было одинаково характеризовать эффект от импульсного режима функционирования. Естественно, ее обнаруживающие способности без схемотехнических приемов относительно выходов блока $F(X)$ гораздо выше структуры рисунка 1.

При повышенных показателях наблюдаемости неисправностей в случае организации контроля вычислений по двум диагностическим признакам разработанный авторами метод дал более простые структуры самопроверяемых устройств и в сравнении с классическим дублированием без дополнительного контроля самодвойственности для 8 тестовых схем (примерно 30 % всех схем). При этом для ряда схем достигнуто улучшение в 20-25 %, что также свидетельствует

о возможностях применения разработанного метода на практике. Для еще большего уменьшения показателей структурной избыточности могут использоваться специализированные схемы сжатия сигналов от объекта диагностирования, которые, однако, не должны допускать потери диагностических свойств, что требует выбора подмножеств сжимаемых сигналов с учетом свойств структуры самой схемы сжатия [35]. Этот вопрос остался за рамками настоящего исследования.

Применение композиции равновесных кодов «1 из 4» и «3 из 4» практически во всех случаях оказалось выгоднее использования равновесного кода «2 из 4». Для основной массы тестовых схем значение коэффициента χ лежит в диапазоне 85-95 %.

Экспериментальные результаты свидетельствуют о том, что разработанный авторами метод имеет некоторые преимущества перед известными, а сами полученные результаты могут оказаться полезными при разработке самопроверяемых цифровых устройств. Дальнейшие эксперименты с использованием представленного авторами метода будут направлены на исследования особенностей повышения показателей наблюдаемости неисправностей на контрольных выходах СВК при реализации самопроверяемых тестовых схем в различных функциональных базах при различных методах реализации как при исчерпывающем тестировании, так и при подаче заданного подмножества пар наборов значений аргументов.

6. Заключение. Уникальные свойства композиции равновесных кодов «1 из 4» и «3 из 4» совместно с использованием принципа ЛКС позволяют синтезировать самопроверяемые цифровые устройства с контролем вычислений по нескольким диагностическим признакам. При этом удастся обеспечивать более высокие показатели наблюдаемости неисправностей за счет постоянной смены значений на выходах объекта диагностирования и контроля вычислений по двум диагностическим признакам, что приводит к увеличению числа наборов значений аргументов, на которых неисправности проявляются в виде установления непарафазных значений на контрольных выходах СВК.

В качестве достоинства метода можно отметить полный учет свойств композиции равновесных кодов «1 из 4» и «3 из 4» при синтезе СВК, а также учет свойств структуры объекта диагностирования при заданной модели неисправностей. Это исключает излишнюю избыточность при наделении устройства свойством самопроверяемости. Несмотря на то, что устройство $G(X)$ имеет такое же количество выходов, что и блок $F(X)$, в ряде случаев удастся синтезировать самопроверяемые цифровые устройства с меньшей площадью, чем по известным методам. Это возможно из-за потенциально различной сложности булевых

функций, вычисляемых блоками $F(X)$ и $G(X)$ [36]. В качестве недостатков разработанного метода выделим внесение большей избыточности, чем при контроле только по одному из диагностических признаков, а также требуемую временную избыточность, что снижает быстродействие структур вдвое. Для современных устройств микроэлектроники, однако, это не является существенным ввиду высокой частоты работы. Кроме того, потенциально можно уменьшить период подачи второго набора в каждой паре, на котором формируются контрольные сигналы. Это делается путем изменения скважности генератора, однако определение, до какой величины это допустимо – задача дальнейших исследований. К ним можно также отнести и исследования различных алгоритмов формирования значений функций на выходах блока $G(X)$ с учетом их совместной оптимизации, а также совершенствование алгоритма выделения подмножеств контролепригодных выходов на множестве выходов объекта диагностирования с учетом функциональных зависимостей на его выходах и выходах внутренних элементов, а не только структурных особенностей объекта диагностирования, а также исследования в области рабочего диагностирования неисправностей из разных моделей, включая, например, модель временной задержки сигнала (delay fault) [37, 38]. Такие неисправности вызывают неправильную работу схем относительно тактовой частоты и тестируются при подаче на входы пар наборов значений аргументов (как и в предлагаемом методе). Известно [12], что при определенных реализациях цифровых устройств проверяющий тест для неисправностей типа «временная задержка» содержится во множестве пар наборов, ортогональных по всем значениям аргументов.

Разработанный метод интересен в практических приложениях в любых отраслях экономики, где используются системы автоматического управления. Однако наибольший интерес он может представлять для использования в таких системах, данные в которых меняются не столь часто (к ним относится большое число систем управления в промышленности, на транспорте, в военной и космической отраслях [39 – 41]). В таких системах следует исключать потенциальные «скрытые» неисправности [42, 43], что частично обеспечивается импульсным режимом работы представленной авторами структуры и контролем вычислений по двум диагностическим признакам.

Литература

1. Согомонян Е.С., Слабаков Е.В. Самопроверяемые устройства и отказоустойчивые системы // М.: Радио и связь. 1989. 208 с.
2. Goessel M., Ocheretny V., Sogomonyan E., Marienfeld D. New Methods of Concurrent Checking // Frontiers in Electronic Testing. Springer. 2008. vol. 42. 185 p.

3. Sahana A.R., Chiraag V., Suresh G., et al. Application of Error Detection and Correction Techniques to Self-Checking VLSI Systems: An Overview // Proceedings of the IEEE Guwahati Subsection Conference (GCON). 2023. pp. 1–6. DOI: 10.1109/GCON58516.2023.10183449.
4. Дрозд А.В. Нетрадиционный взгляд на рабочее диагностирование вычислительных устройств // Проблемы управления. 2008. №2. С. 48–56.
5. Ефанов Д.В. Методы синтеза самопроверяемых дискретных устройств. М.: ЛЕНАНД. 2025. 268 с.
6. Borecky J., Kohlik M., Kubatova H. Parity driven reconfigurable duplex system // Microprocessors and Microsystems. 2017. vol. 52. pp. 251–260. DOI: 10.1016/j.micpro.2017.06.015.
7. Paschalis A.M., Halatsis C., Philokyprou G. Concurrently totally self-checking microprogram control unit with duplication of microprogram sequencer // Microprocessing and Microprogramming. 1987. vol. 20. no. 4-5. pp. 271–281. DOI: 10.1016/0165-6074(87)90216-X.
8. Hahanov V., Litvinova E., Chumachenko S., et al. Vector Logic Modeling Self-Tested Circuits // Proceedings of the 21st IEEE East-West Design & Test Symposium (EWDTS). 2025. pp. 1–4. DOI: 10.1109/EWDTS67441.2025.11303706.
9. Согомонян Е.С. Построение самопроверяемых схем встроенного контроля для комбинационных устройств // Автоматика и телемеханика. 1974. №2. С.121–133.
10. Аксёнова Г.П. Восстановление в дублированных устройствах методом инвертирования данных // Автоматика и телемеханика. 1987. №10. С. 144–153.
11. Dmitriev A., Sapozhnikov V., Sapozhnikov V.I., et al. New Self-Dual Circuits for Error Detection and Testing // VLSI Design. 1999. vol. 11. no. 1. pp. 1–21. DOI: 10.1155/2000/84720.
12. Сапожников В.В., Сапожников Вл.В., Гессель М. Самодвойственные дискретные устройства // СПб: Энергоатомиздат. 2001. 330 с.
13. Efanov D.V. The equal-length redundant code development for the self-checking combinational devices synthesis based on data on their structures // Electronic Modeling. 2022. vol. 44. no. 1. pp. 43–52. DOI: 10.15407/emodel.44.01.043.
14. Стемковский А.Л., Тельпухов Д.В., Жукова Т.Д. Разработка схем функционального контроля на основе автоматизированной генерации помехоустойчивого кода // Проблемы разработки перспективных микро- и нанoeлектронных систем (МЭС). 2022. №4. С. 2–8. DOI: 10.31114/2078-7707-2022-4-2-8.
15. Ефанов Д.В., Елина Е.И. Синтез самопроверяемых дискретных устройств на основе логической коррекции сигналов с применением взвешенных кодов с суммированием в кольце вычетов по заданному модулю // Автоматика и телемеханика. 2026. №5. С. 108–130.
16. Efanov D., Sapozhnikov V., Sapozhnikov V.I., et al. Self-Dual Complement Method up to Constant-Weight Codes for Arrangement of Combinational Logical Circuits Concurrent Error-Detection Systems // Proceedings of the 17th IEEE East-West Design & Test Symposium (EWDTS). 2019. pp. 1–8. DOI: 10.1109/EWDTS.2019.8884398.
17. Ефанов Д.В., Погодина Т.С. Исследование свойств самодвойственных комбинационных устройств с контролем вычислений на основе кодов Хэмминга // Информатика и автоматизация. 2023. Т. 22. №2. С. 349–392. DOI: 10.15622/ia.22.2.5.
18. Ефанов Д.В., Погодина Т.С. Самодвойственные цифровые устройства с контролем вычислений по кодам Сяо // Вестник Томского государственного университета. 2023. №63. С. 118–136. DOI: 10.17223/19988605/63/14.
19. Morozov M., Sapozhnikov V., Sapozhnikov V.I., Goessel M. New self-checking circuits by use of berger-codes // Proceedings of the 6th IEEE International On-Line Testing Workshop. 2000. pp. 171–176. DOI: 10.1109/OLT.2000.856626.

20. Гессель М., Морозов А.В., Сапожников В.В., Сапожников Вл.В. Логическое дополнение – новый метод контроля комбинационных схем // Автоматика и телемеханика. 2003. №1. С. 167–176.
21. Efanov D.V., Yelina Y.I. On the Compositions of Constant-Weight Codes for the Synthesis of Self-Checking and Controllable Digital Devices // Proceedings of the 21st IEEE East-West Design & Test Symposium (EWDTS). 2025. pp. 1–10. DOI: 10.1109/EWDTS67441.2025.11303696.
22. Chioktour V., Kakarountas A. Adaptive BIST for Concurrent On-Line Testing on Combinational Circuits // Electronics. 2022. vol. 11. no. 11. pp. 1–20. DOI: 10.3390/electronics11193193.
23. Juracy L.R., Moreira M.T., Kuentzer F.A., de Morais Amory A. Optimized Design of an LSSD Scan Cell // IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2017. vol. 25. no. 2. pp. 765–768. DOI: 10.1109/TVLSI.2016.2589548.
24. Melikyan V., Grigoryan G., Zheng X., et al. Aging-Aware Design Method for Reliable Analog Integrated Circuits Combining Reliability-Aware Circuit Architectures, Transistor's Threshold Voltage and Operating Point-Dependent Degradation // Proceedings of the IEEE East-West Design & Test Symposium (EWDTS). 2024. pp. 1–4. DOI: 10.1109/EWDTS63723.2024.10873644.
25. Acharya G.P., Rani M.A. Berger code based concurrent online self-testing of embedded processors // Journal of Semiconductors. 2018. vol. 39. no. 11. 115001 p. DOI: 10.1088/1674-4926/39/11/115001.
26. Das D., Touba N.A. Synthesis of Circuits with Low-Cost Concurrent Error Detection Based on Bose – Lin Codes // Journal of Electronic Testing. 1999. vol. 15. pp. 145–155. DOI: 10.1023/A:1008344603814.
27. Сапожников В.В., Сапожников Вл.В. Самопроверяемые дискретные устройства // СПб: Энергоатомиздат. 1992. 224 с.
28. Lala P.K. Self-Checking and Fault-Tolerant Digital Design // San Francisco: Morgan Kaufmann. 2001. 216 p.
29. Ефанов Д.В. Синтез самопроверяемых комбинационных устройств на основе метода логической коррекции сигналов с применением кодов Боуза–Лина // Информационные технологии. 2023. Т. 29. №10. С. 503–511. DOI: 10.17587/it.29.503-511.
30. Higami Y., Yamauchi T., Inamoto T., et al. Machine Learning Based Fault Diagnosis for Stuck-at Faults and Bridging Faults // Proceedings of the 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). 2022. pp. 477–480. DOI: 10.1109/ITC-CSCC55581.2022.9894966.
31. Higami Y., Inamoto T., Wang S., et al. Diagnosis of Double Faults Consisting of a Stuck-at Fault and a Transition Fault // Proceedings of the International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC). 2024. DOI: 10.1109/ITC-CSCC62988.2024.1062879.
32. Sogomonyan E.S., Gossel M. Design of Self-Testing and On-Line Fault Detection Combinational Circuits with Weakly Independent Outputs // Journal of Electronic Testing. 1993. vol. 4. pp. 267–281. DOI:10.1007/BF00971975.
33. McElvain K. IWLS'93 Benchmark Set: Version 4.0. // Proceedings of the International Workshop on Logic Synthesis. 1993.
34. Sentovich E.M., Singh K.J., Moon C., et al. Sequential circuit design using synthesis and optimization // Proceedings of the IEEE International Conference on Computer Design: VLSI in Computers & Processors. 1992. pp. 328–333. DOI: 10.1109/ICCD.1992.276282.
35. Efanov D.V., Sapozhnikov V., Sapozhnikov Vl. Organization of Testing of Combinational Devices Based on Boolean Complement to Constant-Weight “1-out-of-4” Code with Signal Compression // Automatic Control and Computer Sciences. 2021. vol. 55. no. 2. pp. 113–124. DOI: 10.3103/S014641162102005X.

36. Коршунов А.Д. Сложность вычислений булевых функций // Успехи математических наук. 2012. Т. 67. №1(403). С. 97–168. DOI: 10.4213/mn9459.
37. Матросова А.Ю., Чернышов С.В., Ким О.Х., Николаева Е.А. Построение последовательности, обнаруживающей робастно тестируемые неисправности задержек путей в схемах с памятью // Автоматика и телемеханика. 2021. №11. С. 148–168. DOI: 10.31857/S0005231021110106.
38. Manikandan P. Delay Testing in Integrated Circuits: Methodologies for Path Delay Fault Detection and Hardware Security // International Journal of Network Security & Its Applications (IJNSA). 2025. vol. 17. no. 2. pp. 21–41. DOI: 10.5121/ijnsa.2025.17202.
39. Wang T., Sun J., Yao W., et al. Comparison Study on Parametric Fault Diagnosis Using BPNN, SVM and SDAE for DC-DC Converters in Aircraft // Proceedings of the 25th European Conference on Power Electronics and Applications (EPE ECCE Europe). 2023. pp. 1–9. DOI: 10.23919/EPE23ECCEEurope58414.2023.10264256.
40. Irrera I., Zentai A., Cunha J.C., Madeira H. Validating a Safety Critical Railway Application Using Fault Injection // Certifications of Critical Systems – The CECRIS Experience. 2022. pp. 227–246. DOI: 10.1201/9781003337485-11.
41. Garg K., Dawson C., Xu K., et al. Model-Free Neural Fault Detection and Isolation for Safe Control // IEEE Control Systems Letters. 2023. vol. 7. pp. 3169–3174. DOI: 10.1109/LCSYS.2023.3302768.
42. Drozd O., Romankevich V., Martynyuk O., et al. Problems of Matrix Structures in the Components of Modern Computer Systems // Proceedings of the IEEE 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). 2022. pp. 222–227. DOI: 10.1109/TCSET55632.2022.9766954.
43. Tejaswini P., Ganiger P.A., Reshma B., Sanjana B.C. Transfer learning approach for Automatic Test Pattern Generation // Proceedings of the IEEE 6th India Council International Subsections Conference (INDISCON). 2025. pp. 1–4. DOI: 10.1109/INDISCON66021.2025.11253948.

Ефанов Дмитрий Викторович — д-р техн. наук, директор, Федеральное государственное бюджетное учреждение науки «Институт проблем транспорта им. Н.С. Соломенко Российской академии наук» (ИПТ РАН); профессор, высшая школа транспорта института машиностроения, материалов и транспорта и высшая школа управления кибер-физическими системами института компьютерных наук и кибербезопасности, Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого» (ФГАОУ ВО СПбПУ). Область научных интересов: дискретная математика, надежность, безопасность и техническая диагностика дискретных систем, методы непрерывного мониторинга систем автоматического управления и сложных инженерных конструкций и сооружений. Число научных публикаций — 650. TrES-4b@yandex.ru; 12-я линия В.О., 13, 199178, Санкт-Петербург, Россия; р.т.: +7(911)709-2164.

Елина Есения Игоревна — научный сотрудник, Федеральное государственное бюджетное учреждение науки «Институт проблем транспорта им. Н.С. Соломенко Российской академии наук» (ИПТ РАН); ассистент, высшая школа транспорта института машиностроения, материалов и транспорта и высшая школа управления кибер-физическими системами института компьютерных наук и кибербезопасности, Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский политехнический университет Петра Великого» (ФГАОУ ВО СПбПУ). Область научных интересов: самопроверяемые цифровые устройства, методы синтеза схем, теория кодирования, дискретная математика. Число научных публикаций — 70. eseniya-elina@mail.ru; 12-я линия В.О., 13, 199178, Санкт-Петербург, Россия; р.т.: +7(911)935-1039.

D. EFANOV, Y. YELINA
**DESIGN OF SELF-CHECKING DIGITAL DEVICES
WITH IMPROVED TESTABILITY METRICS**

Efanov D., Yelina Y. **Design of Self-Checking Digital Devices with Improved Testability Metrics.**

Abstract. A method is proposed for the designing of fully self-checking digital devices with improved testability metrics. The solution is based on the principle of Boolean signal correction and the properties of a composition of constant-weight codes «1-out-of-4» and «3-out-of-4». The organizational features of a the concurrent error-detection circuit based on two diagnostic criteria are described: membership of the generated codewords in the composition of constant-weight codes «1-out-of-4» and «3-out-of-4», as well as the membership of each function describing the codeword concurrent error-detection circuit in the class of self-dual Boolean functions. The structure of a self-checking digital device with computation checking based on two diagnostic criteria is presented. Algorithms for designing the concurrent error-detection circuit based on Boolean signal correction and the composition of constant-weight codes «1-out-of-4» and «3-out-of-4» are developed, taking into account the requirements for testability of its components during operation. A method is proposed for identifying subsets of testable quadruples of outputs on the set of outputs of the diagnostic object, which guarantees error coverage using the selected composition of constant-weight codes codes, considering its diagnostic properties. Some experimental results with benchmarks are presented, demonstrating the advantages of the proposed self-checking digital device design method compared to known ones: duplication with self-duality checking and the use of the «2-out-of-4» code with additional self-duality checking. The developed method can be effectively applied in the design of self-checking digital devices on modern microelectronic components, and can also be considered in the future for implementing devices based on new physical principles.

Keywords: self-checking digital device, Boolean signal correction, concurrent error-detection circuit, computation control at the outputs of devices, composition of constant-weight codes «1-out-of-4» and «3-out-of-4» codes, fault detection.

References

1. Soghomonyan E.S., Slabakov E.V. Samoproveryaemy'e ustroystva i otkazoustojchivye sistemy' [Self-testable devices and fault-tolerant systems]. M.: Radio i svyaz' Publ. 1989. 208 p. (In Russ.).
2. Goessel M., Ocheretny V., Sogomonyan E., Marienfeld D. New Methods of Concurrent Checking. *Frontiers in Electronic Testing*. Springer. 2008. vol. 42. 185 p.
3. Sahana A.R., Chiraag V., Suresh G., et al. Application of Error Detection and Correction Techniques to Self-Checking VLSI Systems: An Overview. *Proceedings of the IEEE Guwahati Subsection Conference (GCON)*. 2023. pp. 1–6. DOI: 10.1109/GCON58516.2023.10183449.
4. Drozd A.V. [An unconventional view on operational diagnostics of computing devices]. *Problemy' upravleniya – Control Sciences*. 2008. no. 2. pp. 48–56. (In Russ.).
5. Efanov D.V. Metody' sinteza samoproveryaemy'x diskretny'x ustroystv [Methods of synthesis of self-testable discrete devices]. M.: LENAND. 2025. 268 p. (In Russ.).
6. Borecky J., Kohlik M., Kubatova H. Parity driven reconfigurable duplex system. *Microprocessors and Microsystems*. 2017. vol. 52. pp. 251–260. DOI: 10.1016/j.micpro.2017.06.015.

7. Paschalis A.M., Halatsis C., Philokyrou G. Concurrently totally self-checking microprogram control unit with duplication of microprogram sequencer. *Microprocessing and Microprogramming*. 1987. vol. 20. no. 4-5. pp. 271–281. DOI: 10.1016/0165-6074(87)90216-X.
8. Hahanov V., Litvinova E., Chumachenko S., et al. Vector Logic Modeling Self-Tested Circuits. *Proceedings of the 21st IEEE East-West Design & Test Symposium (EWDTS)*. 2025. pp. 1–4. DOI: 10.1109/EWDTS67441.2025.11303706.
9. Sogomonyan E.S. [Construction of self-testing in-built monitoring circuits for combinational devices]. *Avtomatika i telemekhanika – Automation and Telemechanics*. 1974. no. 2. pp.121–133. (In Russ.).
10. Aksenova G.P. [Restoration in duplicated units by the method of data inversion]. *Avtomatika i telemekhanika – Automation and Telemechanics*. 1987. no. 10. pp. 144–153. (In Russ.).
11. Dmitriev A., Sapozhnikov V., Sapozhnikov V.I., et al. New Self-Dual Circuits for Error Detection and Testing. *VLSI Design*. 1999. vol. 11. no. 1. pp. 1–21. DOI: 10.1155/2000/84720.
12. Sapozhnikov V., Sapozhnikov V.I., Gessel M. *Samodvoystvenny'e diskretny'e ustroystva* [Self-dual discrete devices]. SPb: E'nergoatomizdat. 2001. 330 p. (In Russ.).
13. Efanov D.V. The equal-length redundant code development for the self-checking combinational devices synthesis based on data on their structures. *Electronic Modeling*. 2022. vol. 44. no. 1. pp. 43–52. DOI: 10.15407/emodel.44.01.043.
14. Stempkovsky A.L., Telpukhov D.V., Zhukova T.D. [Development of Concurrent Error Detection Circuit Based on Automated Generation of Error-Correcting Code]. *Problemy' razrabotki perspektivny'x mikro- i nanoe'lektronny'x sistem (ME'S) – Problems of Advanced Micro- and Nanoelectronic Systems Development (MES)*. 2022. no. 4. pp. 2–8. DOI: 10.31114/2078-7707-2022-4-2-8. (In Russ.).
15. Efanov D.V., Yelina Y.I. [Design of Self-Checking Discrete Devices Based on Boolean Signal Correction with Weighted Sum Codes in the Residue Ring of a Given Modulus]. *Avtomatika i telemekhanika – Automation and Telemechanics*. 2026. no. 5. pp. 108–130. (In Russ.).
16. Efanov D., Sapozhnikov V., Sapozhnikov V.I., et al. Self-Dual Complement Method up to Constant-Weight Codes for Arrangement of Combinational Logical Circuits Concurrent Error-Detection Systems. *Proceedings of the 17th IEEE East-West Design & Test Symposium (EWDTS)*. 2019. pp. 1–8. DOI: 10.1109/EWDTS.2019.8884398.
17. Efanov D.V., Pogodina T.S. [Properties Investigation of Self-Dual Combinational Devices with Calculation Control Based on Hamming Codes]. *Informatika i avtomatizatsiya – Informatics and Automation*. 2023. vol. 22. no. 2. pp. 349–392. DOI: 10.15622/ia.22.2.5. (In Russ.).
18. Efanov D.V., Pogodina T.S. [Self-dual digital devices with calculations testing by Hsiao codes]. *Vestnik Tomskogo gosudarstvennogo universiteta – Bulletin of Tomsk State University*. 2023. no. 63. pp. 118–136. DOI: 10.17223/19988605/63/14. (In Russ.).
19. Morozov M., Sapozhnikov V., Sapozhnikov V.I., Goessel M. New self-checking circuits by use of berger-codes. *Proceedings of the 6th IEEE International On-Line Testing Workshop*. 2000. pp. 171–176. DOI: 10.1109/OLT.2000.856626.
20. Gessel M., Morozov A.V., Sapozhnikov V., Sapozhnikov V.I. [Logical Complementation – A New Method for Testing Combinational Circuits]. *Avtomatika i telemekhanika – Automation and Telemechanics*. 2003. no. 1. pp. 167–176. (In Russ.).
21. Efanov D.V., Yelina Y.I. On the Compositions of Constant-Weight Codes for the Synthesis of Self-Checking and Controllable Digital Devices. *Proceedings of the 21st IEEE East-West Design & Test Symposium (EWDTS)*. 2025. pp. 1–10. DOI: 10.1109/EWDTS67441.2025.11303696.

22. Chioktour V., Kakarountas A. Adaptive BIST for Concurrent On-Line Testing on Combinational Circuits. *Electronics*. 2022. vol. 11. no. 11. pp. 1–20. DOI: 10.3390/electronics11193193.
23. Juracy L.R., Moreira M.T., Kuentzer F.A., de Moraes Amory A. Optimized Design of an LSSD Scan Cell. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 2017. vol. 25. no. 2. pp. 765–768. DOI: 10.1109/TVLSI.2016.2589548.
24. Melikyan V., Grigoryan G., Zheng X., et al. Aging-Aware Design Method for Reliable Analog Integrated Circuits Combining Reliability-Aware Circuit Architectures, Transistor's Threshold Voltage and Operating Point-Dependent Degradation. *Proceedings of the IEEE East-West Design & Test Symposium (EWDTS)*. 2024. pp. 1–4. DOI: 10.1109/EWDTS63723.2024.10873644.
25. Acharya G.P., Rani M.A. Berger code based concurrent online self-testing of embedded processors. *Journal of Semiconductors*. 2018. vol. 39. no. 11. 115001 p. DOI: 10.1088/1674-4926/39/11/115001.
26. Das D., Toubia N.A. Synthesis of Circuits with Low-Cost Concurrent Error Detection Based on Bose – Lin Codes. *Journal of Electronic Testing*. 1999. vol. 15. pp. 145–155. DOI: 10.1023/A:1008344603814.
27. Sapozhnikov V., Sapozhnikov V.I. *Samoproveryaemy'e diskretnye ustroystva* [Self-testable discrete devices]. SPb: E'nergoatomizdat. 1992. 224 p. (In Russ.).
28. Lala P.K. *Self-Checking and Fault-Tolerant Digital Design*. San Francisco: Morgan Kaufmann. 2001. 216 p.
29. Efanov D. V. [Self-Checking Combinational Devices Synthesis Based on the Boolean Signal Correction Method Using Bose–Lin Codes]. *Information Technologies*. 2023. vol. 29. no. 10. pp. 503–511. DOI: 10.17587/it.29.503-511. (In Russ.).
30. Higami Y., Yamauchi T., Inamoto T., et al. Machine Learning Based Fault Diagnosis for Stuck-at Faults and Bridging Faults. *Proceedings of the 37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*. 2022. pp. 477–480. DOI: 10.1109/ITC-CSCC55581.2022.9894966.
31. Higami Y., Inamoto T., Wang S., et al. Diagnosis of Double Faults Consisting of a Stuck-at Fault and a Transition Fault. *Proceedings of the International Technical Conference on Circuits/Systems, Computers, and Communications (ITC-CSCC)*. 2024. DOI: 10.1109/ITC-CSCC62988.2024.10628279.
32. Sogomonyan E.S., Gossel M. Design of Self-Testing and On-Line Fault Detection Combinational Circuits with Weakly Independent Outputs. *Journal of Electronic Testing*. 1993. vol. 4. pp. 267–281. DOI:10.1007/BF00971975.
33. McElvain K. IWLS'93 Benchmark Set: Version 4.0. *Proceedings of the International Workshop on Logic Synthesis*. 1993.
34. Sentovich E.M., Singh K.J., Moon C., et al. Sequential circuit design using synthesis and optimization. *Proceedings of the IEEE International Conference on Computer Design: VLSI in Computers & Processors*. 1992. pp. 328–333. DOI: 10.1109/ICCD.1992.276282.
35. Efanov D.V., Sapozhnikov V., Sapozhnikov V.I. Organization of Testing of Combinational Devices Based on Boolean Complement to Constant-Weight “1-out-of-4” Code with Signal Compression. *Automatic Control and Computer Sciences*. 2021. vol. 55. no. 2. pp. 113–124. DOI: 10.3103/S014641162102005X.
36. Korshunov A.D. [Computational complexity of Boolean functions]. *Uspehi matematicheskikh nauk – Russian Mathematical Surveys*. 2012. vol. 67. no. 1(403). pp. 97–168. DOI: 10.4213/rm9459. (In Russ.).
37. Matrosova A.Yu., Cherhyshov S.V., Kim O.Kh., Nikolaeva E.A. [Constructing a sequence detecting robustly testable path delay faults in sequential circuits]. *Avtomatika i Telemekhanika – Automation and Telemechanics*. 2021. no. 11. pp. 148–168. DOI: 10.31857/S0005231021110106. (In Russ.).

38. Manikandan P. Delay Testing in Integrated Circuits: Methodologies for Path Delay Fault Detection and Hardware Security. *International Journal of Network Security & Its Applications (IJNSA)*. 2025. vol. 17. no. 2. pp. 21–41. DOI: 10.5121/ijnsa.2025.17202.
39. Wang T., Sun J., Yao W., et al. Comparison Study on Parametric Fault Diagnosis Using BPNN, SVM and SDAE for DC-DC Converters in Aircraft. *Proceedings of the 25th European Conference on Power Electronics and Applications (EPE ECCE Europe)*. 2023. pp. 1–9. DOI: 10.23919/EPE23ECCEEurope58414.2023.10264256.
40. Irrera I., Zentai A., Cunha J.C., Madeira H. Validating a Safety Critical Railway Application Using Fault Injection. *Certifications of Critical Systems – The CECRIS Experience*. 2022. pp. 227–246. DOI: 10.1201/9781003337485-11.
41. Garg K., Dawson C., Xu K., et al. Model-Free Neural Fault Detection and Isolation for Safe Control. *IEEE Control Systems Letters*. 2023. vol. 7. pp. 3169–3174. DOI: 10.1109/LCSYS.2023.3302768.
42. Drozd O., Romankevich V., Martynyuk O., et al. Problems of Matrix Structures in the Components of Modern Computer Systems. *Proceedings of the IEEE 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*. 2022. pp. 222–227. DOI: 10.1109/TCSET55632.2022.9766954.
43. Tejaswini P., Ganiger P.A., Reshma B., Sanjana B.C. Transfer learning approach for Automatic Test Pattern Generation. *Proceedings of the IEEE 6th India Council International Subsections Conference (INDISCON)*. 2025. pp. 1–4. DOI: 10.1109/INDISCON66021.2025.11253948.

Efanov Dmitry — Ph.D., Dr.Sci., Director, Solomenko Institute of Transport Problems of the Russian Academy of Sciences (ITP RAS); Professor, Higher School of Transport at the Institute of Machinery, Materials, and Transport, and Higher School of Cyber-Physical Systems & Control, Peter the Great St.Petersburg Polytechnic University (SPbPU). Research interests: discrete mathematics, reliability, safety and technical diagnostics of discrete systems, methods of continuous monitoring of automatic control systems and complex engineering structures. The number of publications — 650. TrES-4b@yandex.ru; 13, 12th line V.O., 199178, St. Petersburg, Russia; office phone: +7(911)709-2164.

Yelina Yeseniya — Research Employee, Solomenko Institute of Transport Problems of the Russian Academy of Sciences (ITP RAS); Assistant, Higher School of Transport at the Institute of Machinery, Materials, and Transport, and Higher School of Cyber-Physical Systems & Control, Peter the Great St.Petersburg Polytechnic University (SPbPU). Research interests: self-testable digital devices, circuit synthesis methods, coding theory, discrete mathematics. The number of publications — 70. eseniya-elina@mail.ru; 13, 12th line V.O., 199178, St. Petersburg, Russia; office phone: +7(911)935-1039.

М.К. До, В.А. БОГАТЫРЕВ
**ПОЛУМАРКОВСКАЯ МОДЕЛЬ НАДЕЖНОСТИ
КОМПЬЮТЕРНОЙ СИСТЕМЫ С КОНТЕЙНЕРНОЙ
ВИРТУАЛИЗАЦИЕЙ**

До М.К., Богатырев В.А. Полумарковская модель надежности компьютерной системы с контейнерной виртуализацией.

Аннотация. В современных облачных инфраструктурах и микросервисных архитектурах критически важной задачей является обеспечение высокой надежности, отказоустойчивости и непрерывности функционирования приложений. Основным фактором, снижающим эффективность таких систем, выступает феномен программного старения – постепенная деградация системных ресурсов (утечки памяти, фрагментация данных) в процессе длительной эксплуатации, приводящая к нелинейному росту интенсивности отказов. Сложность прогнозирования надежности в таких условиях обусловлена иерархической зависимостью между общим программным обеспечением хоста и изолированными контейнерами, а также динамическим характером процессов восстановления. В данной статье рассматривается актуальная проблема обеспечения надежности программного обеспечения (ПО) в компьютерной системе с контейнерной виртуализацией, подверженной явлению программного старения. Научная новизна заключается в разработке комплексной математической модели на основе полумарковских процессов, которая учитывает иерархическую структуру системы (общее ПО хоста и изолированные контейнеры) и позволяет оптимизировать стратегии проактивного восстановления. Для описания процесса старения использовано распределение Вейбулла, что отражает возрастающую интенсивность отказов. Предложена полумарковская модель состояний и переходов, включающая механизмы проактивного перезапуска и миграции контейнеров. Для снижения вычислительной сложности разработана упрощенная модель, базирующаяся на допущении о пренебрежимо малой вероятности вторичных отказов в короткие интервалы восстановления. Расчеты стационарных вероятностей и коэффициента готовности реализованы в среде Mathcad. Проведенные численные эксперименты показали, что реализация стратегии проактивного восстановления в диапазоне интервалов 1200-2400 часов существенно повышает коэффициент готовности системы. Установлено, что погрешность упрощенной модели пренебрежимо мала по сравнению со строгой моделью. Предложенная модель позволяет определить максимальное значение коэффициента готовности в зависимости от числа развернутых контейнеров и интервалов проактивного перезапуска контейнеров. Сравнение разработанного подхода с традиционными реактивными моделями восстановления подтверждает преимущество проактивных методов в минимизации простоев, вызванных накоплением ошибок (утечки памяти, фрагментация ресурсов). Предложенная модель может быть применена при проектировании высоконадежных облачных инфраструктур и DevOps-конвейеров.

Ключевые слова: надежность, коэффициент готовности, программное обеспечение, контейнерная виртуализация, полумарковская модель, проактивное восстановление, перезапуск контейнеров.

1. Введение. При разработке современных инфокоммуникационных систем первостепенное внимание уделяется требованиям по обеспечению

высокой готовности, надёжности и отказоустойчивости [1 – 3]. Актуальность данных требований обусловлена критической значимостью непрерывности бизнес-процессов и необходимостью предотвращения потенциальных потерь данных в условиях цифровой трансформации.

Достижение заданных показателей качества обслуживания в современных децентрализованных компьютерных системах и распределенных сетях [4] неразрывно связано со сложной инженерной задачей – необходимостью поддержания непрерывности вычислений и минимизации задержек при обработке и передаче гетерогенного трафика. В условиях экспоненциального роста объемов обрабатываемой информации данная задача успешно решается за счет глубокой консолидации вычислительных ресурсов в рамках кластерных архитектур. Такие архитектуры поддерживают интеллектуальные механизмы динамической реконфигурации, базирующиеся на технологиях репликации данных, оперативной миграции виртуальных машин [5, 6] и легких контейнеров [7]. Эти механизмы активируются не только в случаях возникновения аппаратных отказов, но и в процессах планового технического обслуживания, а также при реализации процедур аварийного восстановления кластерной системы. Современные кластерные решения и платформы оркестрации обеспечивают интеллектуальное автоматическое перераспределение рабочей нагрузки между физическими и виртуальными узлами в режиме реального времени. Это позволяет не только минимизировать деструктивное влияние отказов отдельных аппаратных или программных компонентов на общую производительность и пропускную способность системы [8, 9], но и гарантировать пользователям бесперебойный доступ к сервисам независимо от текущего состояния базовой инфраструктуры.

Обоснование проектных решений в распределенных кластерных системах должно опираться на построение моделей, ориентированных на оптимизацию процессов обслуживания потоков функциональных задач с учетом обеспечения требований надежности [10 – 12], своевременности и непрерывности вычислений. Для современных распределенных систем заслуживают внимания модели, учитывающие особенности систем граничных, туманных и облачных вычислений [13, 14]. Для обоснования организации современных инфокоммуникационных систем представляют интерес модели надежности кластерных компьютерных систем, отражающие процесс накопления отказов и различных дисциплин восстановления после отказов при различных дисциплинах деградации и восстановления [15 – 17], а также модели надёжности, отражающие процесс контроля и мониторинга состояний компьютерных систем.

Модели надежности кластеров должны отражать особенности процессов их физического и информационного восстановления при альтернативах разделения и совмещения этих процессов [15, 18], в том числе при миграции виртуальных машин [19, 20]. Модели двухузловых кластерных систем с учетом миграции виртуальных контейнеров при физических отказах узлов предложены в [21]. Для оценки надежности многоузловых кластерных систем из структурно сложных узлов в [22] предложен метод декомпозиции на основе построения марковской модели процессов отказов и восстановлений в одном узле. Однако следует заметить, что рассмотренные модели надежности кластерных систем ориентированы на отказы аппаратуры и не учитывают особенности миграции виртуальных контейнеров при отказах общего системного программного обеспечения узлов и программного обеспечения поддержки функционирования виртуальных контейнеров.

Современные тенденции развития информационных технологий характеризуются массовым переходом к микросервисной архитектуре и использованию контейнеризации (Docker, Kubernetes) [23]. По данным исследований последних лет, более 80% крупных IT-компаний внедрили контейнерные технологии в свою производственную инфраструктуру, а темпы роста рынка контейнеризации превышают 30% в год [24]. Глобальный рынок контейнерных технологий оценивается экспертами в миллиарды долларов и продолжает стремительно расти [25]. Контейнеризация обеспечивает ряд существенных преимуществ: быстрое развертывание приложений, эффективное использование ресурсов, портативность между различными средами исполнения и упрощение процессов непрерывной интеграции и доставки (CI/CD) [26]. Несмотря на гибкость и масштабируемость, такие системы сталкиваются с проблемой программного старения – постепенного снижения производительности и роста вероятности отказов из-за накопления ошибок в процессе непрерывной эксплуатации [27, 28]. Исследования показывают, что программное старение является одной из основных причин деградации облачных сервисов, приводящее к снижению качества обслуживания и нарушению соглашений об уровне сервиса [29]. В долгосрочной перспективе даже хорошо спроектированные системы подвержены эффектам старения, которые проявляются через увеличение времени отклика, рост потребления памяти и снижение пропускной способности [30].

В работе [31] вопросы надежности часто рассматриваются либо для физического оборудования, либо для монолитного программного

обеспечения без учета специфики контейнерной виртуализации и распределенных микросервисных архитектур. Традиционные подходы к восстановлению носят реактивный характер, т.е. активируются только после наступления отказа, что в критических системах ведет к значительным финансовым и репутационным потерям, снижению удовлетворённости клиентов и потере конкурентных преимуществ. Реактивные стратегии не способны предотвратить каскадные отказы в распределенных микросервисных архитектурах, где отказ одного компонента может привести к лавинообразному эффекту во всей системе из-за наличия сложных зависимостей между сервисами [32].

Анализ инцидентов в облачных инфраструктурах ведущих провайдеров [33] свидетельствует о том, что значительная доля отказов – от 40% до 60% по различным оценкам – вызвана не только дефектами кода, но и явлением программного старения. Данный феномен проявляется в постепенной деградации системных ресурсов – утечках оперативной памяти, фрагментации файловой системы, переполнении буферов, накоплении устаревших записей в таблицах маршрутизации и базах данных, что неизбежно приводит к снижению производительности и зависанию контейнеров. Исследования подтверждают, что даже при отсутствии логических ошибок в коде приложения, длительная эксплуатация контейнеров приводит к накоплению временных файлов, росту числа открытых дескрипторов файлов и сетевых соединений, увеличению фрагментации памяти и деградации производительности кэшей [8].

Применение традиционных моделей надежности программного обеспечения (ПО) к компьютерной системе с контейнерной виртуализацией сталкивается с серьезными ограничениями и зачастую не позволяет получить адекватные оценки характеристик надежности. Модель Джелинского-Моранды (JM) основывается на предположении, что интенсивность отказов пропорциональна числу оставшихся ошибок в программном коде, и после каждого исправления дефекта надежность системы мгновенно возрастает [34]. Однако в компьютерной системе с контейнерной виртуализацией отказы часто вызваны исчерпанием системных ресурсов в процессе длительной работы, даже при полном отсутствии логических ошибок в коде приложения. Исправление программных багов не останавливает процесс накопления «мусора» и деградации ресурсов операционной системы, что делает модель JM принципиально неприменимой для описания долгоживущих облачных сессий и непрерывно работающих сервисов [35].

Модели Мусы и Шумана рассматривают программную систему как единый независимый монолит с фиксированной архитектурой [36]. Контейнеры же, в отличие от виртуальных машин с полной изоляцией, используют общее ядро хостовой операционной системы и разделяют системные ресурсы. Это создает проблему «шумных соседей» и недостаточной изоляции, когда отказ или утечка ресурсов в одном контейнере оказывает существенный побочный эффект на другие сервисы и хост-систему в целом, приводя к снижению их производительности и увеличению вероятности отказов. Традиционные модели не учитывают эту корреляцию отказов и взаимное влияние компонентов в мультитенантных средах с разделяемыми ресурсами. Экспериментальные исследования показывают, что взаимное влияние может приводить к снижению производительности до 40% [37].

Модель Миллса (Mills model, модель с посевом ошибок) предполагает статичное состояние среды выполнения и постоянную конфигурацию системы [38]. Контейнерные оркестраторы (например, Kubernetes, Docker Swarm, Apache Mesos) характеризуются высокой динамикой: экземпляры приложений постоянно масштабируются в соответствии с нагрузкой, перезапускаются при обнаружении отказов и мигрируют между узлами кластера для балансировки нагрузки и оптимизации использования ресурсов [39]. Использование фиксированной интенсивности отказов в таких динамических условиях приводит к значительным погрешностям прогнозирования, которые могут достигать порядка величины. Современные облачные системы требуют моделей, способных адаптироваться к изменяющимся условиям нагрузки, конфигурации инфраструктуры и топологии развертывания [40].

В связи с вышеизложенным, для моделирования надежности сервисов на базе контейнеров, подверженных эффектам старения, наиболее целесообразным признано использование математического аппарата полумарковских процессов (semi-Markov processes, SMP) [41]. В отличие от классических марковских цепей, строго ограничивающих время пребывания системы в каждом состоянии экспоненциальным законом распределения, SMP позволяют использовать произвольные функции распределения и описывать переходы между качественно различными состояниями системы: нормальное функционирование, постепенное старение, критическое состояние, отказ и процесс омоложения (rejuvenation) [42]. Это критически важно, так как вероятность отказа из-за накопленного старения увеличивается со временем непрерывной работы, что свидетельствует об отсутствии

свойства отсутствия памяти, характерного для экспоненциального распределения [43].

Для математического описания процесса деградации в данной работе применяется распределение Вейбулла [44]. Оно позволяет моделировать возрастающую интенсивность отказов, характерную для фазы износа ресурсов хостовой операционной системы и среды исполнения контейнеров, что обеспечивает высокую точность прогнозирования моментов наступления критических состояний и планирования профилактических мероприятий [45]. Параметр формы распределения Вейбулла позволяет гибко настраивать модель для различных сценариев деградации системы – от постепенного до резкого старения, что делает модель универсальной для широкого класса приложений [46].

Методологическим прототипом применения аппарата SMP к оценке надёжности ПО является работа [47], в которой предложено строить граф-схему полумарковского процесса по блок-схеме алгоритма вычислительного процесса, формировать матрицу вероятностей переходов и решать систему алгебраических уравнений в среде MathCAD для нахождения среднего времени до отказа. В работе [47] ПО рассматривается как единое целое без разделения на общее ПО и ПО виртуальных контейнеров, при этом не отражено возможное возрастание интенсивности отказов при программном старении. Исследование стратегий проактивного восстановления, предполагающего плановый перезапуск контейнеров или их миграцию до наступления отказа, в [47] не проводилось. Соответственно, не рассматривалась задача оптимизации числа контейнеров и интервалов их перезапуска для обеспечения максимального коэффициента готовности кластерной системы с контейнерной виртуализацией.

Цель данной работы заключается в разработке и исследовании математической модели надежности компьютерной системы с контейнерной виртуализацией, учитывающей эффекты программного старения, для обоснования стратегий проактивного восстановления и проектирования оптимальной конфигурации системы по количеству используемых контейнеров.

Научная новизна данной работы заключается в создании многоуровневой математической модели надежности, которая разделяет процессы старения общего (системного) программного обеспечения и прикладных контейнеров, учитывая их взаимное влияние. В отличие от существующих решений [47 – 49], предлагаемый подход интегрирует механизмы проактивного перезапуска и миграции контейнеров в единую полумарковскую

модель с возможностью оптимизации параметров обслуживания. Преимуществом разработанного метода является возможность аналитического определения оптимальных значений числа развернутых контейнеров и интервалов их проактивного перезапуска, обеспечивающих баланс между временными затратами на профилактику и коэффициентом готовности системы. Это позволяет осуществить переход от традиционной стратегии восстановления по факту отказа к стратегии превентивного оздоровления системы (proactive rejuvenation), что существенно повышает общую надежность и доступность сервисов.

2. Процесс старения программного обеспечения. Старение программного обеспечения (ПО) – это явление постепенного снижения производительности или увеличения вероятности отказа программной системы в процессе её непрерывной работы. В отличие от физического старения, старение ПО обусловлено не механическим износом, а истощением системных ресурсов, утечками памяти, фрагментацией данных и накоплением ошибок округления.

В теории надежности данное явление характеризуется функцией интенсивности отказов, возрастающей со временем. Это фундаментально отличает ПО от электронных компонентов, которые часто подчиняются экспоненциальному закону распределения (где $\lambda = \text{const}$). Многочисленные эмпирические исследования показывают, что распределение Вейбулла является наиболее подходящей моделью для описания этого этапа [44]. В частности, при параметре формы $\beta > 1$ интенсивность отказов монотонно возрастает, что отражает свойство старения системы.

Принимается допущение, что аппаратное обеспечение является абсолютно надежным. В отказоустойчивых кластерах, использующих технологию контейнерной виртуализации, надежность ПО рассматривается как композиция двух составляющих: общего ПО и прикладного ПО контейнеров. Отказ общего ПО приводит к полной потере работоспособности всех запущенных контейнеров. Рассмотрим программную систему, состоящую из общего ПО и одного изолированного контейнера. Процессы возникновения отказов общего ПО и контейнера описываются функциями распределения Вейбулла $F_1(t)$ и $F_2(t)$ соответственно:

$$F_1(t) = 1 - e^{-\left(\frac{t}{\eta_1}\right)^{\beta_1}}, \quad F_2(t) = 1 - e^{-\left(\frac{t}{\eta_2}\right)^{\beta_2}}, \\ R_1(t) = 1 - F_1(t), \quad R_2(t) = 1 - F_2(t).$$

$R_1(t)$ и $R_2(t)$ представляют собой вероятности безотказной работы общего ПО и контейнера соответственно. Параметр масштаба распределения Вейбулла η представляет собой наработку, в течение которой в среднем отказывает 63,2% наблюдаемых объектов. Диаграмма состояний и переходов ПО, построенная на основе полумарковской модели, представлена на рисунке 1. В рассматриваемой модели состояние S_0 соответствует работоспособности общего ПО и контейнера. В процессе непрерывной работы программная среда контейнера подвергается старению, что со временем приводит к возникновению отказов. При отказе контейнера система переходит в состояние S_1 (восстановления контейнера) и S_2 (миграции контейнера). В случае отказа общего ПО фиксируется переход в состояние S_3 , характеризующееся необходимостью полного восстановления как хостовой платформы, так и запущенного в ней контейнера.

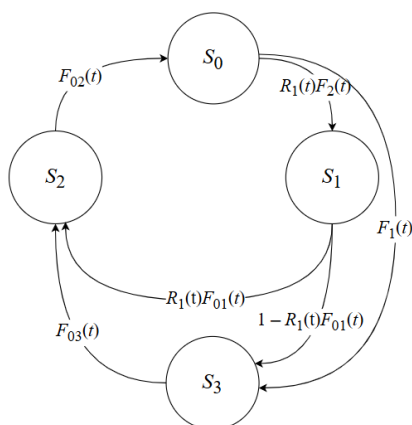


Рис. 1. Диаграмма состояний и переходов общего ПО и контейнера в условиях естественного старения

Предполагается, что процессы восстановления в состояниях S_1 , S_2 и S_3 описываются экспоненциальными законами распределения $F_{01}(t)$, $F_{02}(t)$ и $F_{03}(t)$ соответственно:

$$F_{01}(t) = 1 - e^{-\mu_1 t}, F_{02}(t) = 1 - e^{-\mu_2 t}, F_{03}(t) = 1 - e^{-\mu_3 t},$$

где μ_1 , μ_2 , μ_3 – интенсивности восстановления контейнера, миграции и полного восстановления хостовой платформы соответственно. Для расчета надежности системы строится матрица вероятностей

переходов полумарковского процесса $M_1(t)$ и вектор вероятностей состояний $v(t) = (v_0(t), v_1(t), v_2(t), v_3(t))$. Согласно графу состояний, матрица имеет вид:

$$M_1(t) = \begin{pmatrix} 1 - F_1(t) - R_1(t)F_2(t) & R_1(t)F_2(t) & 0 & F_1(t) \\ 0 & 0 & R_1(t)F_{01}(t) & 1 - R_1(t)F_{01}(t) \\ F_{02}(t) & 0 & 1 - F_{02}(t) & 0 \\ 0 & 0 & F_{03}(t) & 1 - F_{03}(t) \end{pmatrix}.$$

Вектор $v(t) = (v_0(t), v_1(t), v_2(t), v_3(t))$ является единственным решением системы линейных алгебраических уравнений:

$$\begin{cases} v(t) = v(t)M_1(t), \\ \sum_{i=0}^3 v_i(t) = 1. \end{cases}$$

Расчет вектора стационарных вероятностей был выполнен в программной среде Mathcad. Полученные значения компонентов вектора вероятностей состояний представлены на рисунке 2(а). На основе этих данных коэффициент готовности системы рассчитывается по следующей формуле:

$$K_1(\tau) = \frac{\int_0^{\tau} e^{-\left(\frac{t}{\eta_1}\right)^{\beta_1}} e^{-\left(\frac{t}{\eta_2}\right)^{\beta_2}} dt}{\int_0^{\tau} e^{-\left(\frac{t}{\eta_1}\right)^{\beta_1}} e^{-\left(\frac{t}{\eta_2}\right)^{\beta_2}} dt + v_1(\tau)t_1 + v_2(\tau)t_2 + v_3(\tau)t_3}.$$

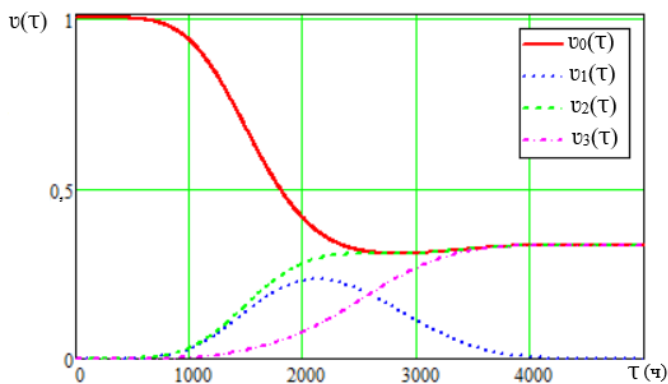
В данной формуле использование определенного интеграла в числителе обусловлено процессом программного старения, поскольку вероятность безотказной работы общего ПО $R_1(t)$ и контейнера $R_2(t)$ непрерывно снижается, а интеграл позволяет рассчитать математическое ожидание времени полезной работы системы. Здесь t – текущая наработка (внутренняя переменная интегрирования), а τ – заданная наработка системы до момента оценки её состояния (интервал наблюдения). Параметры t_1 , t_2 и t_3 представляют собой средние времена восстановления, соответствующие вероятностям пребывания системы в состояниях

S_1, S_2 и S_3 соответственно. Таким образом, знаменатель отражает полную среднюю продолжительность цикла, включая как работу, так и суммарное время восстановления. Математически коэффициент готовности представляет собой отношение среднего времени пребывания системы в работоспособном состоянии к суммарному среднему времени полного цикла «работа – отказ – восстановление».

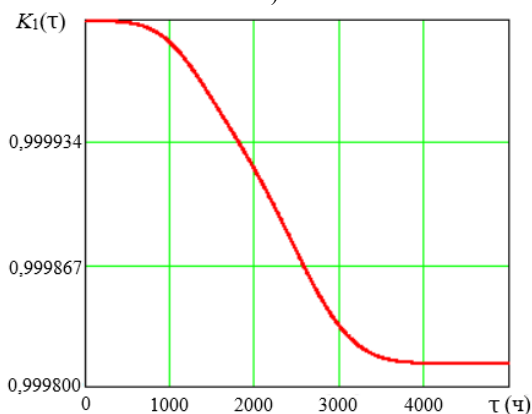
При проведении численных экспериментов были приняты следующие значения параметров. Значения параметров формы $\beta_1 = \beta_2 = 5$ выбраны на основе следующих соображений: согласно теории надёжности, значения $\beta > 4$ характерны для систем, демонстрирующих ярко выраженный износ ресурсов – режим, при котором интенсивность отказов стремительно возрастает в конце жизненного цикла [45]. Применительно к контейнерной виртуализации данный режим соответствует фазе накопленной деградации: утечкам памяти, росту числа открытых дескрипторов и истощению буферов – явлениям, экспериментально подтверждённым для платформ Docker и Kubernetes в работах [30]. Равенство $\beta_1 = \beta_2$ обусловлено тем, что общее ПО и контейнер функционируют на базе единого общего ядра и подвержены схожим механизмам деградации. Значения параметра масштаба $\eta_1 = 3000$ ч для общего ПО и $\eta_2 = 2000$ ч для контейнера отражают типичное для производственных облачных систем соотношение сроков службы. Общее ПО хоста, как правило, функционирует дольше без перезапуска, тогда как контейнер чаще подвергается накоплению ошибок вследствие интенсивных рабочих нагрузок. Конкретные числовые значения выбраны в качестве репрезентативных для демонстрации метода; для реальных систем их следует определять методом максимального правдоподобия по статистике времён до отказа, получаемой из систем мониторинга (например, из открытых наборов данных Google Cluster Data или Alibaba Cluster Data).

Значения временных параметров приняты на основе экспериментальных данных для типичных контейнеризированных систем. Время восстановления контейнера после отказа $t_1 = 0,25$ ч включает диагностику и освобождение ресурсов, что соответствует оценкам, полученным при экспериментальном исследовании программного старения в контейнерной виртуализации [30]. Время миграции контейнера $t_2 = 1/120$ ч описывает не перемещение между физическими серверами, а процесс внутриузлового переразвертывания и миграции контекста, характерный для современных платформ контейнеризации [8]. Этот процесс включает удаление старых пространств имен, которые могли быть загрязнены эффектами старения,

и инициализацию полностью нового контейнера из эталонного образа для обеспечения чистого состояния. Время восстановления общего ПО $t_3 = 1$ ч соответствует полному перезапуску хостовой ОС и среды исполнения контейнера при критическом отказе, вызванном накоплением эффектов программного старения [28, 30]. Результаты расчета коэффициента готовности системы в зависимости от длительности интервала наблюдения τ представлены на рисунке 2(б).



а)



б)

Рис. 2. Зависимость компонентов вектора вероятностей состояний $v(\tau)$ и коэффициента готовности системы $K_1(\tau)$ от длительности интервала наблюдения τ : а) компоненты вектора $v(\tau)$; б) коэффициент готовности $K_1(\tau)$

Анализ временной динамики коэффициента готовности позволяет выделить три характерные фазы. В начальный момент

времени значение K_1 равно 1, что объясняется идеальным состоянием системы, при котором все программные компоненты полностью работоспособны и ресурсы не исчерпаны. В последующий период наблюдается постепенное снижение показателя, вызванное кумулятивным эффектом старения ПО. Согласно закону Вейбулла, с течением времени интенсивность отказов возрастает, что приводит к более частому переходу системы в состояния восстановления S_1 , S_2 и S_3 . Наконец, при стремлении времени к бесконечности график выходит на плато, стремясь к своему стационарному значению. Эта стабилизация объясняется достижением динамического равновесия между процессами возникновения отказов и процессами восстановления в полумарковской модели. Данное стационарное значение является ключевым индикатором долгосрочной надежности, определяющим эффективность системы в реальных условиях эксплуатации.

Для практического применения модели в конкретных реальных системах параметры распределения Вейбулла должны определяться на основе статистических данных мониторинга (например, из открытых наборов данных, таких как Google Cluster Data или Alibaba Cluster Data). Наиболее достоверным методом оценки является метод максимального правдоподобия [45]. Если имеется выборка времен до отказа $\tau_1, \tau_2, \dots, \tau_m$ – функция правдоподобия для распределения Вейбулла имеет вид:

$$F(\eta, \beta) = \prod_{i=1}^m \frac{\beta}{\eta} \left(\frac{\tau_i}{\eta} \right)^{\beta-1} e^{-\left(\frac{\tau_i}{\eta} \right)^{\beta}}.$$

Процедура нахождения точечных оценок параметров η и β для конкретного типа контейнера или общего ПО заключается в максимизации логарифмической функции правдоподобия. Для двухпараметрического распределения Вейбулла логарифм функции правдоподобия имеет вид:

$$\ln F(\eta, \beta) = m \ln \beta - m \beta \ln \eta + (\beta - 1) \sum_{i=1}^m \ln \tau_i - \sum_{i=1}^m \left(\frac{\tau_i}{\eta} \right)^{\beta}.$$

Приравнявая к нулю частные производные, получают систему уравнений метода максимального правдоподобия:

$$\begin{cases} \frac{\partial \ln F(\eta, \beta)}{\partial \beta} = \frac{m}{\beta} - m \ln \eta + \sum_{i=1}^m \ln \tau_i - \sum_{i=1}^m \left(\frac{\tau_i}{\eta} \right)^\beta \ln \left(\frac{\tau_i}{\eta} \right) = 0, \\ \frac{\partial \ln F(\eta, \beta)}{\partial \eta} = -\frac{m\beta}{\eta} + \frac{\beta}{\eta^{\beta+1}} \sum_{i=1}^m \tau_i^\beta = 0, \end{cases}$$

$$\Leftrightarrow \begin{cases} \frac{\sum_{i=1}^m \tau_i^\beta \ln \tau_i}{\sum_{i=1}^m \tau_i^\beta} - \frac{1}{\beta} = \frac{1}{m} \sum_{i=1}^m \ln \tau_i, \\ \eta = \left(\frac{1}{m} \sum_{i=1}^m \tau_i^\beta \right)^{\frac{1}{\beta}}. \end{cases}$$

Моделирование подтвердило, что в условиях естественного старения и при отсутствии профилактических мер коэффициент готовности системы неуклонно снижается, особенно при высоких значениях параметра формы распределения Вейбулла β . Это свидетельствует о быстром накоплении критических ошибок и утечек ресурсов. Установлено, что общая надежность инфраструктуры в значительной степени лимитируется временем безотказной работы общего ПО, отказ которого приводит к полной потере работоспособности всех запущенных контейнеров. Выявленная динамика деградации показателей обосновывает необходимость перехода от пассивного ожидания отказов к разработке стратегий проактивного восстановления, которые будут подробно рассмотрены в последующих разделах работы.

3. Механизмы противодействия старению ПО на основе перезапуска контейнеров. Для повышения рассчитанного ранее коэффициента готовности и минимизации негативного влияния программного старения необходимо внедрение механизмов проактивного восстановления. Одной из наиболее эффективных стратегий является проактивная перезагрузка и обновление контейнера. Суть данной стратегии заключается в плановом превентивном перезапуске контейнера до момента наступления критического отказа. Такой подход позволяет сбросить накопленные в процессе эксплуатации ошибки, устранить утечки памяти и высвободить системные ресурсы, возвращая программную среду в исходное чистое состояние. На рисунке 3 приведена полумарковская модель состояний системы при реализации предложенной стратегии

проактивного восстановления. В отличие от базовой модели, здесь учитывается переход в состояние активного восстановления до момента возникновения отказа.

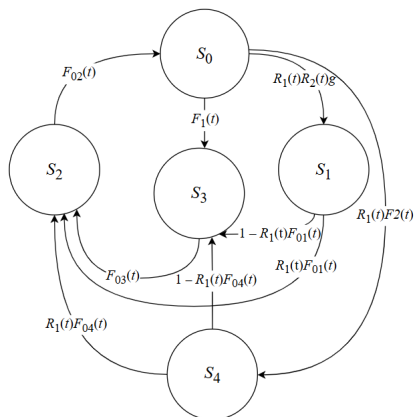


Рис. 3. Диаграмма состояний и переходов системы с проактивным восстановлением

В предложенной модели выделяются следующие состояния системы: S_0 – работоспособность при возможности нормального функционирования общего ПО и контейнера; S_1 – проактивный перезапуск контейнера; S_2 – миграция контейнера; S_3 – восстановление контейнера после отказа, S_4 – отказ системного (общего) ПО. При этом процедуры проактивного обновления и восстановления контейнера реализуются только при условии работоспособности общего ПО. В случае отказа общего ПО (состояние S_4) требуется восстановление системной среды, перезапуск контейнера и его последующая миграция. Процессы восстановления в состояниях S_1 , S_2 , S_3 и S_4 описываются экспоненциальными законами распределения $F_{01}(t)$, $F_{02}(t)$, $F_{03}(t)$ и $F_{04}(t)$ соответственно.

Для описания стратегии проактивного перезапуска контейнера вводится управляемый параметр $g \in [0, 1]$, определяющий вероятность принятия решения о проведении проактивного восстановления при достижении детерминированного интервала проактивного перезапуска контейнера t_0 . Параметр g представляет собой управляемую характеристику стратегии проактивного перезапуска контейнера и определяет вероятность того, что при достижении наработки $t = t_0$ система перейдет в состояние проактивного восстановления S_1 , а не продолжит работу в состоянии S_0 .

Параметр g определяет режим функционирования системы. При $g = 0$ реализуется полностью реактивная стратегия восстановления, при которой переход в состояния восстановления S_3 или S_4 осуществляется исключительно по факту отказа общего ПО или контейнера соответственно. При $g = 1$ система гарантированно переходит в состояние проактивного восстановления S_1 при достижении интервала проактивного перезапуска контейнера t_0 . При промежуточных значениях $0 < g < 1$ реализуется смешанная стратегия проактивного перезапуска контейнера, при которой с вероятностью g осуществляется проактивное восстановление в момент $t = t_0$, а с вероятностью $(1 - g)$ система продолжает эксплуатацию до наступления отказа. На основе представленной (Рис. 3) диаграммы состояний и переходов матрица вероятностей переходов полумарковского процесса $M_2(t, g)$ записывается в виде:

$$M_2(t, g) = \begin{pmatrix} R_1(t)R_2(t)(1-g) & R_1(t)R_2(t)g & 0 & F_1(t) & R_1(t)F_2(t) \\ 0 & 0 & R_1(t)F_{01}(t) & 1-R_1(t)F_{01}(t) & 0 \\ F_{02}(t) & 0 & 1-F_{02}(t) & 0 & 0 \\ 0 & 0 & F_{03}(t) & 1-F_{03}(t) & 0 \\ 0 & 0 & R_1(t)F_{04}(t) & 1-R_1(t)F_{04}(t) & 0 \end{pmatrix}.$$

Произведение вероятностей $R_1(t)R_2(t)$ определяет вероятность безотказной работы системы к моменту наработки t , т.е. вероятность того, что ни общее ПО, ни контейнер не отказали в течение интервала $[0, t]$. При достижении детерминированного интервала проактивного перезапуска контейнера $t = t_0$ и выполнении условия безотказной работы $R_1(t_0)R_2(t_0)$ с вероятностью g принимается решение о проведении проактивного перезапуска контейнера (переход в состояние S_1), а с вероятностью $(1 - g)$ система продолжает функционировать в работоспособном состоянии S_0 . Стационарное распределение вероятностей $v(t, g) = (v_0(t, g), v_1(t, g), v_2(t, g), v_3(t, g), v_4(t, g))$ находится из решения системы линейных уравнений:

$$\begin{cases} v(t, g) = v(t, g)M_2(t, g), \\ \sum_{i=0}^4 v_i(t, g) = 1. \end{cases}$$

Решение данной системы уравнений позволило получить динамику компонентов вектора $v(t, g)$ в зависимости от наработки

и параметра g . На рисунке 4(а) представлена динамика составляющих вектора $v(t, g)$ в зависимости от наработки при фиксированном значении $g = 0,5$, а также на рисунке 4(б) приведено семейство кривых для компонента $v_3(t, g)$, отражающего вероятность пребывания системы в состоянии отказа S_3 , при различных значениях g .

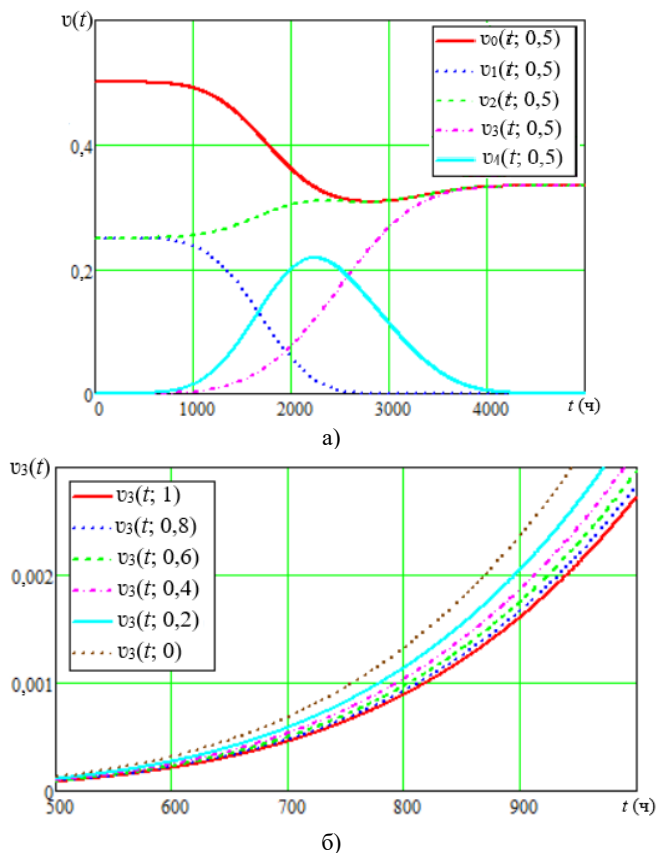


Рис. 4. Результаты компонентов вектора вероятностей состояний $v(t)$ при $g = 0,5$ и семейство кривых $v_3(t)$ при различных значениях вероятности проактивного восстановления g : а) компоненты вектора $v(t)$ при $g = 0,5$; б) семейство кривых $v_3(t)$ при различных значениях g

Анализ полученных зависимостей показывает, что увеличение вероятности проактивного восстановления g приводит к сокращению среднего времени пребывания системы в состоянии отказа общего ПО.

Физический смысл данного явления заключается в том, что реализация стратегии омоложения позволяет превентивно устранять накопленные программные дефекты и утечки ресурсов до момента их перехода в критическую стадию. Таким образом, при максимальной вероятности реализации стратегии ($g = 1$) система чаще переходит в режим контролируемого планового проактивного восстановления, что минимизирует длительность нахождения в состояниях реактивного восстановления. Это подтверждает, что активное управление жизненным циклом системы эффективно подавляет негативные последствия программного старения, предотвращая длительные простои инфраструктуры.

С учетом найденного распределения вероятностей $v(t, g)$ итоговая формула для коэффициента готовности принимает вид:

$$K_2(t_0, g) = \frac{\int_0^{t_0} e^{-\left(\frac{t}{\eta_1}\right)^{\beta_1}} e^{-\left(\frac{t}{\eta_2}\right)^{\beta_2}} dt}{\int_0^{t_0} e^{-\left(\frac{t}{\eta_1}\right)^{\beta_1}} e^{-\left(\frac{t}{\eta_2}\right)^{\beta_2}} dt + v_1(t_0, g)t_1 + v_2(t_0, g)t_2 + v_3(t_0, g)t_3 + v_4(t_0, g)t_4}. \quad (1)$$

Для проведения численного анализа и верификации модели были использованы следующие исходные данные: время перезагрузки контейнера $t_1 = 1/30$ ч, время миграции контейнера $t_2 = 1/120$ ч, время восстановления общего ПО $t_3 = 1$ ч и время восстановления контейнера $t_4 = 0,25$ ч. Значения остальных параметров приняты идентичными использованным выше. Графические зависимости коэффициента готовности от интервала проактивного перезапуска контейнера t_0 при различных значениях g приведены на рисунке 5.

Для оценки устойчивости модели был проведен анализ чувствительности коэффициента готовности к вариации параметров распределения Вейбулла. Результаты расчетов представлены на рисунке 6. Увеличение параметра формы $\beta_1 = \beta_2 = \beta$ (кривые 2, 1, 3 соответствуют $\beta = 3; 5; 7$) при $\eta_1 = 3000$ ч и $\eta_2 = 2000$ ч приводит к росту максимального значения коэффициента готовности и смещению оптимального интервала проактивного восстановления t_0 вправо. Это объясняется тем, что при высоких значениях β программное старение проявляется позже, но более интенсивно: система дольше остается стабильной в начале цикла, что позволяет отложить перезапуск контейнера без риска отказа. Чем выше β , тем дольше система сохраняет высокую надежность в начале цикла,

что позволяет увеличить интервал между перезапусками контейнера без риска возникновения отказа. При фиксированном значении $\beta_1 = \beta_2 = \beta = 5$ изменение параметра масштаба η (кривые 4 и 5) демонстрирует практически прямо пропорциональную зависимость. Снижение показателей надежности ($\eta_1 = 2400$ ч; $\eta_2 = 1600$ ч – кривая 4) сокращает безопасный период работы ПО, что требует более раннего вмешательства и смещает t_0 влево. Напротив, при увеличении запаса стабильности ПО ($\eta_1 = 3600$ ч; $\eta_2 = 2400$ ч – кривая 5) время до наступления фазы старения увеличивается, вследствие чего оптимальная точка t_0 смещается вправо.

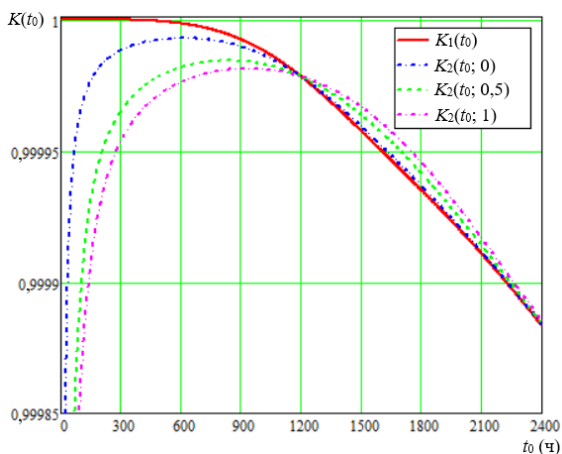


Рис. 5. Зависимость коэффициента готовности от интервала проактивного перезапуска контейнера t_0 при различных значениях g

В результате исследования установлено, что интервал проактивного перезапуска контейнера t_0 является ключевым фактором управления надежностью системы. Сравнительный анализ показал, что реализация стратегии проактивного восстановления в диапазоне от 1200 до 2400 часов обеспечивает существенный прирост коэффициента готовности, достигая максимума при полной вероятности проактивного восстановления ($g = 1$). Выявлено, что чрезмерно частое проведение проактивного восстановления ведет к неоправданному простоям и снижению производительности. В то же время увеличение интервала t_0 нивелирует эффект профилактики, переводя систему в режим реактивного восстановления после критических отказов. Полученные результаты позволяют

оптимизировать стратегию проактивного перезапуска контейнера, обеспечивая своевременное устранение накопленных ошибок, высвобождение избыточных ресурсов и предотвращение критических отказов, требующих значительных временных и финансовых затрат на восстановление.

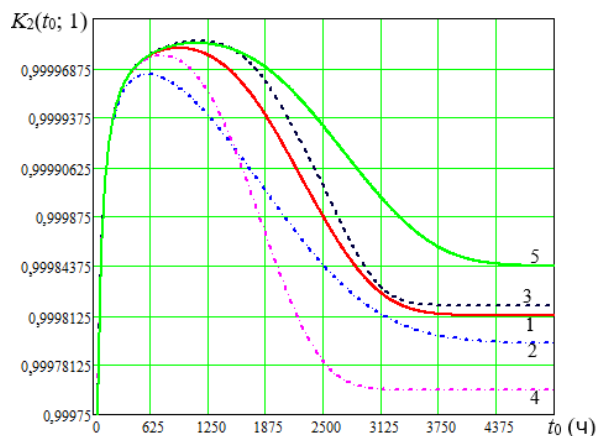


Рис. 6. Анализ чувствительности коэффициента готовности к изменению параметров распределения Вейбулла (η и β)

4. Упрощенная модель. Использование предложенной полумарковской модели для оптимизации в реальном времени может быть затруднено из-за её высокой вычислительной сложности. В связи с этим представлена упрощенная математическая модель процесса обслуживания ПО в компьютерной системе с контейнерной виртуализацией. Переход к упрощенной форме обоснован тем, что процессы восстановления системного ПО и контейнеров описываются экспоненциальными законами распределения, при которых вероятность завершения восстановления быстро стремится к единице.

Поскольку в реальных сценариях оптимальный интервал проактивного перезапуска контейнера t_0 значительно превышает время восстановления, вероятность возникновения вторичного отказа непосредственно в моменты перезапуска или миграции становится пренебрежимо малой. Это допущение позволяет существенно снизить вычислительную сложность модели, сохраняя при этом её высокую точность и физическую адекватность.

В рамках упрощенной модели рассматривается случай полной реализации стратегии проактивного восстановления ($g = 1$). В этом случае матрица вероятностей переходов полумарковского процесса $M_3(t)$ принимает следующий вид:

$$M_3(t) = \begin{pmatrix} 0 & R_1(t)R_2(t) & 0 & F_1(t) & R_1(t)F_2(t) \\ 0 & 0 & R_1(t) & 1-R_1(t) & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & R_1(t) & 1-R_1(t) & 0 \end{pmatrix}.$$

На основе матрицы $M_3(t)$ путем решения соответствующей системы линейных алгебраических уравнений выведены аналитические выражения для компонентов вектора вероятностей состояний $v(t)$ в явном виде.

$$v_0(t) = v_2(t) = \frac{1}{3 + R_1(t) - R_1^2(t)}; v_1(t) = \frac{R_1(t)R_2(t)}{3 + R_1(t) - R_1^2(t)};$$

$$v_3(t) = \frac{1 - R_1^2(t)}{3 + R_1(t) - R_1^2(t)}; v_4(t) = \frac{R_1(t) - R_1(t)R_2(t)}{3 + R_1(t) - R_1^2(t)}.$$

Это позволяет выполнять быстрые расчеты показателей надежности без необходимости проведения итерационных вычислений. Расчет коэффициента готовности произведен по формуле (1). Для подтверждения допустимости предложенных упрощений проведена серия сравнительных расчетов в программной среде MathCad. Основной целью верификации являлось определение относительной погрешности коэффициента готовности, возникающей при использовании упрощенной модели по сравнению со строгой полумарковской моделью. Результат сравнительного анализа представлен на рисунке 7.

Проведенный сравнительный анализ показал, что результаты, полученные с помощью упрощенной модели, имеют пренебрежимо малую погрешность по сравнению со строгой полумарковской моделью. Упрощенная модель подтверждает гипотезу о том, что при оптимальных интервалах проактивного перезапуска контейнера t_0 , значительно превышающих время восстановления, экспоненциальный характер процессов позволяет пренебречь

вероятностью возникновения отказов в фазах перезапуска и миграции. Таким образом, разработанная упрощенная модель эффективна для решения задач оперативной оптимизации стратегий проактивного восстановления. Она обеспечивает необходимый баланс между математической строгостью и вычислительной эффективностью. Полученные аналитические выражения позволяют интегрировать данную модель в системы автоматизированного управления облачными ресурсами для динамической настройки параметров обслуживания в зависимости от текущей интенсивности старения программного обеспечения.

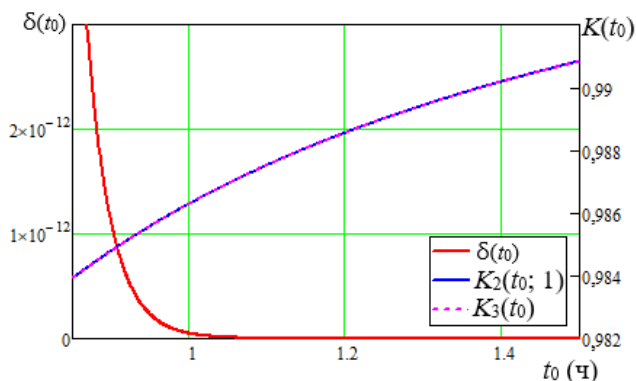


Рис. 7. Относительная погрешность коэффициента готовности $\delta(t_0)$ упрощенной модели

5. Моделирование надежности компьютерной системы с n контейнерами. На данном этапе исследования осуществляется логическое расширение разработанной ранее математической модели на случай сложной кластерной инфраструктуры. Рассматриваемая система состоит из общего системного ПО и n изолированных функциональных контейнеров. Будем считать, что для сохранения работоспособности всей системы в целом достаточно функционирования хотя бы одного из n контейнеров. Исходя из этого, вероятность работоспособного состояния кластерной системы рассчитывается как произведение надежности общего ПО на вероятность того, что хотя бы один из n контейнеров находится в рабочем состоянии:

$$L(t, n) = R_1(t) \left[1 - \prod_{i=1}^n (1 - R_{c,i}(t)) \right],$$

где $R_1(t)$ – функция вероятности безотказной работы общего ПО, а $R_{c,i}(t)$ – функция вероятности безотказной работы отдельных контейнеров. Введение множителя $R_1(t)$ в данную формулу является критически важным, так как отказ хост-платформы является точкой единого отказа, приводящей к неработоспособности всего кластера независимо от количества резервных контейнеров.

Для упрощения математической формализации и последующего анализа кластерной архитектуры в данной работе принимается допущение, что все n контейнеров являются идентичными и характеризуются одинаковой функцией надежности $R_{c,i}(t) = R_2(t)$ для всех $i = 1, \dots, n$. Такое допущение обосновано практикой развертывания современных облачных микросервисов, где контейнеры создаются на основе единого эталонного образа и функционируют в однородной программной среде. В этом случае вероятность работоспособного состояния контейнеров определяется через биномиальное распределение, и общая формула вероятности функционирования системы принимает более компактный вид:

$$L(t, n) = R_1(t) \left[1 - (1 - R_2(t))^n \right].$$

Использование идентичных функций $R_2(t)$ позволяет сфокусировать исследование на изучении влияния масштабируемости (числа развернутых контейнеров n) и стратегий проактивного восстановления на итоговый коэффициент готовности, исключив влияние стохастической неоднородности отдельных узлов. Это обеспечивает высокую прозрачность моделирования и позволяет получить аналитические зависимости для определения оптимальных параметров технического обслуживания всей кластерной группы как единого функционального объекта. Зависимость вероятности работоспособного состояния кластерной системы $L(t, n)$ от наработки при различном числе развернутых контейнеров приведена на рисунке 8.

Рост числа контейнеров n в кластере приводит к существенному увеличению вероятности работоспособного состояния системы, так как для возникновения полного отказа требуется одновременный отказ всех параллельно соединенных контейнеров. Возведение вероятности отказа отдельного контейнера $(1 - R_2(t))$ в степень n приводит к быстрому стремлению этого значения к нулю, тем самым

максимизируя итоговый показатель $L(t, n)$. Это смещает область высокой готовности вправо по оси времени, обеспечивая системе дополнительный запас надежности даже при высокой интенсивности программного старения отдельных компонентов.

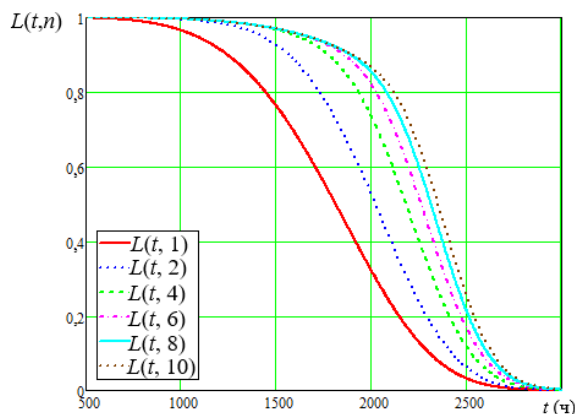


Рис. 8. Зависимость вероятности работоспособного состояния кластерной системы от наработки при различном числе развернутых контейнеров

На основе упрощенной модели построена диаграмма состояний и переходов кластерной системы, представленная на рисунке 9. Приняты следующие обозначения состояний системы: S_0 – работоспособность системы; S_1 – проактивный перезапуск и миграция контейнеров; S_2 – отказ системы.

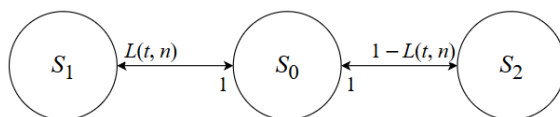


Рис. 9. Диаграмма состояний и переходов упрощенной полумарковской модели системы

Основой для расчета всех показателей служит матрица вероятностей переходов $M_4(t, n)$, которая описывает логику взаимодействия между работоспособным состоянием, состоянием проактивного восстановления и состоянием полного отказа. Матрица переходов имеет вид:

$$M_4(t, n) = \begin{pmatrix} 0 & L(t, n) & 1 - L(t, n) \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Важно подчеркнуть, что процесс программного старения рассматривается как непрерывная деградация ресурсов, происходящая непосредственно в состоянии работоспособности S_0 . Переход системы в состояние проактивного восстановления S_1 осуществляется исключительно в детерминированный момент времени $t = t_0$. Таким образом, состояние S_1 интерпретируется не как процесс старения, а как активная управляющая процедура, направленная на устранение последствий деградации. Вероятность $L(t_0, n)$ в данном контексте определяет долю успешно доживших до момента проактивного восстановления. Такое разделение позволяет четко разграничить естественный износ системы в S_0 и искусственно инициируемый цикл её омоложения в S_1 , что исключает неопределенность при расчете оптимального интервала проактивного перезапуска контейнеров.

На основе данной матрицы вычисление компонентов вектора $v(t, n)$ проводится аналогично описанному выше подходу. В результате получены следующие выражения для компонентов вектора $v(t, n)$:

$$v_0(t, n) = \frac{1}{2}; v_1(t, n) = \frac{L(t, n)}{2}; v_2(t, n) = \frac{1 - L(t, n)}{2}.$$

Для кластерной системы, включающей общее ПО и n изолированных контейнеров, коэффициент готовности определяется следующим выражением:

$$K_4(t_0, n) = \frac{\int_0^{t_0} R_1(t) \left[1 - (1 - R_2(t))^n \right] dt}{\int_0^{t_0} R_1(t) \left[1 - (1 - R_2(t))^n \right] dt + v_1(t_0, n)(t_1 + nt_2) + v_2(t_0, n)(t_3 + nt_2)},$$

где t_0 – интервал проактивного перезапуска контейнеров, t_1 – время перезагрузки контейнеров, t_2 – время миграции контейнеров, t_3 – время восстановления общего ПО.

В связи с этим возникает задача определения оптимального числа контейнеров и интервала проактивного перезапуска контейнеров t_0 для достижения максимального коэффициента готовности системы. Для решения данной задачи на листинге 1 представлена программа, написанная в среде системы компьютерной математики Mathcad.

```

Max(n, m){
    t0 := 1;
    n0 := 1;
    Amax := K4(t0, n0);
    for(x ∈ 1,2,...,m){
        for(y ∈ 1,2,...,n){
            if(G2(x,y) > Amax){
                Amax := K4(x, y);
                t0 := x;
                n0 := y;
            }
        }
    }
    Return (t0, n0, Amax);
}

```

Листинг 1. Оптимизация коэффициента готовности системы по параметрам n и t_0

Реализация численного алгоритма в среде Mathcad позволила получить аналитическое решение задачи оптимизации и визуализировать поведение целевой функции. На рисунке 10 представлена трехмерная поверхность зависимости коэффициента готовности кластерной системы от варьируемых параметров: интервала проактивного перезапуска контейнеров и количества контейнеров.

Визуальный анализ данной поверхности позволяет наглядно определить область глобального максимума и оценить чувствительность системы к изменениям каждого из параметров. Характер представленной зависимости подтверждает существование единственной точки оптимума, обеспечивающей наилучший баланс между отказоустойчивостью инфраструктуры и минимизацией временных затрат на техническое обслуживание. В результате проведенных расчетов установлено, что оптимальное количество контейнеров составляет $n = 4$, а оптимальный интервал $t_0 = 1659$ ч.

Увеличение числа контейнеров в кластере способствует повышению общей отказоустойчивости системы. Однако чрезмерное увеличение числа контейнеров имеет и негативную сторону: возрастает время задержки при выполнении процедур миграции

и перезапуска, а также увеличиваются накладные расходы на управление инфраструктурой. Найденный в ходе исследования экстремум функции демонстрирует баланс между этими двумя факторами. Результаты позволяют определить оптимальное количество контейнеров и наилучший момент времени для проведения профилактики. Таким образом, разработанная модель, учитывающая как старение прикладных контейнеров, так и надежность базовой ОС, представляет собой комплексный инструмент для проектирования высоконадежных облачных сред. Полученные результаты создают теоретическую и практическую базу для построения отказоустойчивых контейнеризированных систем, обеспечивая их стабильное функционирование и эффективное управление ресурсами в условиях отказов ПО. В качестве дальнейшего развития предложенных моделей предполагается модернизация для распределенных кластерных систем при перераспределении нагрузки и миграции виртуальных контейнеров через сеть [50, 51]. Такая модернизация должна учитывать реконфигурацию и накопление отказов узлов многопутевой сети [50, 51], в том числе при организации граничных, туманных и облачных вычислений [13, 14]. Также предполагается развитие работ в части постановки и решения задачи векторной оптимизации [52, 53] с учетом надежности, производительности, стоимости реализации, эксплуатационных расходов и энергопотребления.

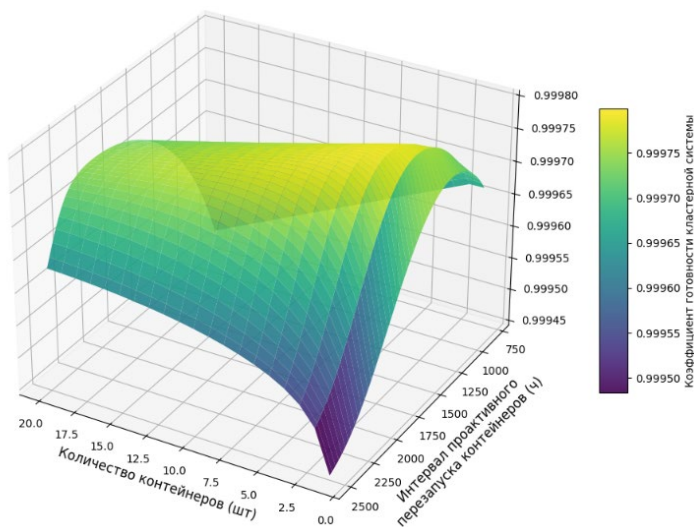


Рис. 10. Зависимость коэффициента готовности кластера от параметров n и t_0

6. Заключение. На основе полумарковских процессов разработана, верифицирована и оптимизирована комплексная математическая модель надежности программного обеспечения в компьютерной системе с контейнерной виртуализацией, позволяющая с высокой точностью прогнозировать периоды деградации программных ресурсов и обосновать необходимость внедрения стратегий проактивного восстановления.

Показано, что стратегии проактивного восстановления являются эффективным инструментом борьбы со старением ПО. Реализация стратегии проактивного восстановления в диапазоне от 1200 до 2400 часов обеспечивает существенный прирост коэффициента готовности (рост на 0,16%) и достигает максимума при полной вероятности проактивного восстановления ($g = 1$).

Разработана упрощенная полумарковская модель, характеризующаяся пренебрежимо малой погрешностью при снижении вычислительной сложности, что позволяет рассматривать её как эффективный инструмент для оперативных инженерных расчетов показателей готовности в реальных эксплуатационных условиях.

Представлена расширенная модель кластера, учитывающая надежность общего системного ПО и ПО, поддерживающего функционирование контейнеров.

Показано существование экстремума функции коэффициента готовности в зависимости от числа развернутых в системе контейнеров и периодичности их проактивного перезапуска. В результате проведенных расчетов установлено, что оптимальное количество контейнеров составляет $n = 4$, а оптимальный интервал $t_0 = 1659$ ч.

В отличие от модели Джелинского-Моранды [34], предложенный подход описывает возрастающую интенсивность отказов через распределение Вейбулла, что адекватно отражает накопление ресурсных ошибок в долгоживущих контейнерах. По сравнению с монолитными моделями Мусы-Шумана [36] и Миллса [38], модель явно разделяет деградацию общего ПО и контейнеров, выявляя отказ общего ПО как доминирующий фактор потери готовности. В отличие от подхода [48], рассматривающего совместный анализ производительности и готовности в рамках обобщённых марковских моделей без учёта специфики программного старения, предложенная модель явно включает механизм деградации через распределение Вейбулла и аналитически решает задачу совместной оптимизации числа контейнеров n и интервала проактивного перезапуска t_0 . По сравнению с работой [47], являющейся методологическим прототипом применения аппарата SMP к оценке надёжности ПО, предложенная модель заменяет

экспоненциальное распределение времени до отказа (постоянная интенсивность) на распределение Вейбулла ($\beta > 1$), разделяет иерархические уровни общего ПО хоста и изолированных контейнеров. При этом дополняет реактивное восстановление механизмом проактивного омоложения с управляемым параметром g , а также решает задачу совместной оптимизации конфигурационных параметров кластера – тогда как работа [47] ориентирована на монолитное ПО с реактивным восстановлением и статичной архитектурой.

Направления дальнейших исследований включают расширение модели для учета неоднородности нагрузки на различные узлы кластера, а также разработку программных агентов для мониторинга ресурсов в реальном времени с целью динамической корректировки интервалов проактивного перезапуска контейнера t_0 . Предложенные решения могут быть внедрены в системы управления облачными дата-центрами для повышения их отказоустойчивости и автоматизации процессов технического обслуживания.

Литература

1. Dong H., Zhang C., Li G., Zhang H. Cloud-Native Databases: A Survey // IEEE Transactions on Knowledge and Data Engineering. vol. 36. no. 12. pp. 7772–7791. 2024. DOI: 10.1109/TKDE.2024.3397508.
2. Goyal P., Deora S.S. Reliability of Trust Management Systems in Cloud Computing // Indian Journal of Cryptography and Network Security. 2022. vol. 2. no. 1. pp. 1–5. DOI: 10.54105/ijcns.C1417.051322.
3. Assiri B., Sheneamer A. Fault tolerance in distributed systems using deep learning approaches // PLOS One. 2025. vol. 20. no. 1. DOI: 10.1371/journal.pone.0310657.
4. Херсонский Н.С., Большедворская Л.Г. Анализ развития статистических методов оценки надежности как свойства качества объектов и продукции // Crede Experto: транспорт, общество, образование, язык. 2024. №1. С. 6–23. DOI: 10.51955/2312-1327_2024_1_6.
5. Gurjanov A.V., Korobeynikov A.G., Zharinov I.O., Zharinov O.O. Edge, fog and cloud computing in the cyber-physical systems networks // Proceedings of the III International Workshop on Modeling, Information Processing and Computing. 2021. pp. 103–108. DOI: 10.47813/dnit-mip3/2021-2899-103-108
6. Kushchazli A., Safargalieva A., Kochetkova I., Gorshenin A. Queuing Model with Customer Class Movement across Server Groups for Analyzing Virtual Machine Migration in Cloud Computing // Mathematics. 2024. vol. 12. no. 3. 468 p. DOI: 10.3390/math12030468.
7. Damarapati A. Containers vs. Virtual Machines: Understanding the Shift to Kubernetes // World Journal of Advanced Engineering Technology and Sciences. 2025. vol. 15. no. 1. pp. 852–861. DOI: 10.30574/wjaets.2025.15.1.0305.
8. Kozhimbayev Z., Sinnott R.O. A performance comparison of container-based technologies for the Cloud // Future Generation Computer Systems. 2017. vol 68. pp. 175–182. DOI: 10.1016/j.future.2016.08.025.
9. Mirna A., Nadja K., Aris L. Utilization prediction-based VM consolidation approach // Journal of Parallel and Distributed Computing. 2022. vol. 170. pp. 24–38. DOI: 10.1016/j.jpdc.2022.08.001.

10. Половко А.М., Гуров С.В. Основы теории надежности // Санкт-Петербург: БХВ-Петербург, 2008. 702 с.
11. Tatamikova T.M., Arkhiptsev E.M. Designing Fault-tolerant Systems with Micro-service Architecture // Proceedings of the 27-th International Conference on Soft Computing and Measurements (SCM). 2024. pp. 348–351. DOI: 10.1109/SCM62608.2024.10554143.
12. Koren I., Krishna S.M. Fault-Tolerant Systems // Boston: Morgan Kaufmann. 2007. 400 p.
13. Yermakov S.G., Khalil M.M., Khomonenko A.D., Bukharova K.A. Evaluating the efficiency of fog computing on the Internet of Things using a Non-Markov model // T-Comm. 2022. vol. 16. no. 12. pp. 46–52.
14. Хомоненко А.Д., Кириенко А.Б., Пугачев С.В. Оценивание оперативности туманных вычислений в геоинформационных системах // Инженерный вестник Дона. 2024. №6. С. 364–382.
15. Хомоненко А.Д., Благовещенская Е.А., Проурзин О.В., Андрук А.А. Прогноз надежности кластерной вычислительной системы с помощью полумарковской модели альтернирующих процессов и мониторинга // Наукоемкие технологии в космических исследованиях Земли. 2018. Т. 10. №4. С. 72–82. DOI: 10.24411/2409-5419-2018-10099.
16. Богатырев В.А., Богатырев С.В., Богатырев А.В. Оценка готовности компьютерной системы к своевременному обслуживанию запросов при его совмещении с информационным восстановлением памяти после отказов // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23. №3. С. 608–617. DOI: 10.17586/2226-1494-2023-23-3-608-617.
17. Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Recovery of Real-Time Clusters with the Division of Computing Resources into the Execution of Functional Queries and the Restoration of Data Generated Since the Last Backup // Lecture Notes in Computer Science. Springer. 2024. vol. 14123. pp. 236–250. DOI: 10.1007/978-3-031-50482-2_19.
18. Bogatyrev V.A., Vinokurova M.S. Control and Safety of Operation of Duplicated Computer Systems // Communications in Computer and Information Science. Springer. 2017. vol. 700. pp. 331–342. DOI: 10.1007/978-3-319-66836-9_28.
19. Алексанков С.М. Модели динамической миграции с итеративным подходом и сетевой миграции виртуальных машин // Научно-технический вестник информационных технологий, механики и оптики. 2015. Т. 15. №6. С. 1098–1104. DOI: 10.17586/2226-1494-2015-15-6-1098-1104.
20. Bogatyrev V.A., Derkach A.N. Evaluation of a Cyber-Physical Computing System with Migration of Virtual Machines During Continuous Computing // Computers. 2020. vol. 9. no 2. pp. 42. DOI: 10.3390/computers9020042.
21. Богатырев В.А., Фунг В.К. Оценка надежности восстанавливаемого кластера контейнерной виртуализацией // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25. №5. С. 988–995. DOI: 10.17586/2226-1494-2025-25-5-988-995.
22. Богатырев В.А., Богатырев С.В., Богатырев А.В. Граничная оценка надежности кластерных систем на основе декомпозиции марковской модели при ограниченном восстановлении узлов с накоплением отказов // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25. №3. С. 574–583. DOI: 10.17586/2226-1494-2025-25-3-574-583.
23. Rodriguez M.A., Buuya R. Container-based cluster orchestration systems: A taxonomy and future directions // Software: Practice and Experience. 2019. vol. 49. pp. 698–719. DOI: 10.1002/spe.2660.
24. Application Container Market Size And Forecast // Verified Market Research. URL: www.verifiedmarketresearch.com/product/application-container-market. (дата обращения: 16.06.2026).

25. Application Container Market (2025-2023) // Grand View Research. URL: www.grandviewresearch.com/industry-analysis/application-container-market. (дата обращения: 16.06.2026).
26. Marella V. Comparative Analysis of Container Orchestration Platforms: Kubernetes vs. Docker Swarm // *International Journal of Scientific Research in Science and Technology*. 2024. vol. 11. no. 5. pp. 526–543. DOI: 10.32628/IJSRST24105254.
27. Grottko M., Trivedi K.S. Fighting bugs: remove, retry, replicate, and rejuvenate // *Computer*. 2007. vol. 40. no. 2. pp. 107–109. DOI: 10.1109/MC.2007.55.
28. Cotroneo D., Natella R., Pietrantuono R., Russo S. Software Aging Analysis of the Linux Operating System // *Proceedings of the IEEE 21st International Symposium on Software Reliability Engineering (ISSRE)*. 2010. pp. 71–80. DOI: 10.1109/ISSRE.2010.24.
29. Nascimento M.G., Moura R.J., Machida F., Andrade E. Comparison of Machine Learning Algorithms for Detecting Software Aging in SQL Server // *Proceedings of the 13th Latin-American Symposium on Dependable and Secure Computing*. 2024. pp. 159–167. DOI: 10.1145/3697090.3699798.
30. Oliveira F., Araujo J., Matos R., Maciel P. Software Aging in Container-Based Virtualization: An Experimental Analysis // *Proceedings of the 16th Iberian Conference on Information Systems and Technologies (CISTI)*. 2021. pp. 1–7. DOI: 10.23919/CISTI52073.2021.9476625.
31. Nguyen T.A., Min D., Choi E., Tran T.D. Reliability and Availability Evaluation for Cloud Data Center Networks Using Hierarchical Models // *IEEE Access*. 2019. vol. 7. pp. 9273–9313. DOI: 10.1109/ACCESS.2019.2891282.
32. Shruthi P., Nagaraj G.C. Software aging prediction – a new approach // *International Journal of Electrical and Computer Engineering*. 2023. vol. 13. no. 2. pp. 1773–1781. DOI: 10.11591/ijece.v13i2.pp1773-1781.
33. Pietrantuono R., Russo S. A survey on software aging and rejuvenation in the cloud // *Software Quality Journal*. 2020. vol. 28. pp. 7–38. DOI: 10.1007/s11219-019-09448-3.
34. Turk L.I.A., Alsolami E.G. Jelinski-Moranda Software Reliability Growth Model: A Brief Literature and Modification // *International Journal of Software Engineering & Applications (IJSEA)*. 2016. vol. 7. no. 2. pp. 33–44. DOI: 10.5121/ijsea.2016.7204.
35. Li N., Tan Y., Wang X., Li B., Luo J. Dynamic Resource Allocation for Containerized Applications in Edge Computing // *Lecture Notes in Electrical Engineering*. Springer. 2023. vol. 1015. pp. 121–130. DOI: 10.1007/978-981-99-0416-7_12.
36. Musa J.D., Iannino A., Okumoto K. Software Reliability: Measurement, Prediction, Application // New York: McGraw-Hill. 1987. 621 p.
37. Dohi T., Goseva-Popstojanova K., Trivedi K.S. Statistical non-parametric algorithms to estimate the optimal software rejuvenation schedule // *Proceedings of the IEEE Pacific Rim International Symposium on Dependable Computing*. 2000. pp. 77–84. DOI: 10.1109/PRDC.2000.897287.
38. Machida F., Kim D.S., Trivedi K.S. Modeling and analysis of software rejuvenation in a server virtualized system with live VM migration // *Performance Evaluation*. 2013. vol. 70. no. 3. pp. 212–230. DOI: 10.1016/j.peva.2012.09.003.
39. Zhang Y., Zheng Z., Lyu M.R. BFTCloud: A Byzantine Fault Tolerance Framework for Voluntary-Resource Cloud Computing // *Proceedings of the IEEE 4th International Conference on Cloud Computing*. 2011. pp. 444–451. DOI: 10.1109/CLOUD.2011.16.
40. Manker H., Hemant K.S., Iti S. Reliability analysis of a system using semi Markov process // *International Journal of Statistics and Applied Mathematics*. 2024. vol. 9. no. 4. pp. 145–151. DOI: 10.22271/math.2024.v9.i4b.1785.

41. Martyna J. Semi-Markov Model of Processing Requests Reliability and Availability in Mobile Cloud Computing Systems // *Advances in Intelligent Systems and Computing*. Springer. 2021. vol. 1389. pp. 264–272. DOI: 10.1007/978-3-030-76773-0_26.
42. Howard R.A. *Dynamic Probabilistic Systems: Semi-Markov and Decision Processes* // New York: Wiley. 1971. 564 p.
43. Machida F., Miyoshi N. Analysis of an optimal stopping problem for software rejuvenation in a deteriorating job processing system // *Reliability Engineering & System Safety*. 2007. vol. 168. pp. 128–135. DOI: 10.1016/j.res.2017.05.019.
44. Weibull W. A Statistical Distribution Function of Wide Applicability // *Journal of Applied Mechanics*. 1951. vol. 18. no. 3. pp. 293–297. DOI: 10.1115/1.4010337.
45. Rinne H. *The Weibull Distribution: A Handbook* // New York: CRC Press. 2008. 808 p. DOI: 10.1201/9781420087444.
46. Murthy D.N.P., Xie M., Jiang R. *Weibull Models* // New York: John Wiley & Sons. 2003. 416 p. DOI: 10.1002/047147326X.
47. Новиков И.С. Методы расчёта количественных показателей надёжности сложных программных комплексов на стадии проектирования и разработки // *Труды СПИРАН*. 2008. №6. С. 86–111. DOI: 10.15622/sp.6.8.
48. Okamura H., Dohi T. Faster Maximum Likelihood Estimation Algorithms for Markovian Arrival Processes // *Proceedings of the IEEE International Conference on Quantitative Evaluation of Systems*. 2009. pp. 73–82. DOI: 10.1109/QEST.2009.28.
49. Trivedi K., Andrade E., Machida F. Combining Performance and Availability Analysis in Practice // *Advances in Computers*. 2012. vol. 84. pp. 1–38. DOI: 10.1016/B978-0-12-396525-7.00001-0.
50. Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Control of Multipath Transmissions in the Nodes of Switching Segments of Reserved Paths // *Proceedings of the International Conference on Information, Control, and Communication Technologies (ICCT)*. 2022. pp. 1–5. DOI: 10.1109/ICCT56057.2022.9976839.
51. Bogatyrev V.A., Bogatyrev A.V., Bogatyrev S.V. Multipath Transmission of Heterogeneous Traffic in Acceptable Delays with Packet Replication and Destruction of Expired Replicas in the Nodes that Make Up the Path // *Communications in Computer and Information Science*. Springer. 2023. vol. 1748. pp. 104–121. DOI: 10.1007/978-3-031-30648-8_9.
52. Уткин Л.В., Мулоха В.А., Лукашин А.А., Константинов А.В. Принятие решений в условиях неопределённости // *Санкт-Петербург: Политех-Пресс*, 2025. 49 с.
53. Уткин Л.В. Анализ риска и принятие решений при неполной информации // *Санкт-Петербург: Наука*, 2007. 403 с.

До Мань Киём — аспирант, факультет программной инженерии и компьютерной техники, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО» (ИТМО). Область научных интересов: теория надёжности, теория массового обслуживания, отказоустойчивости и эффективности компьютерных систем и сетей. Число научных публикаций — 2. domanhkiemnd@gmail.com; Кронверкский пр., 49А, 197101, Санкт-Петербург, Россия; р.т.: +7(911)199-8063.

Богатырев Владимир Анатольевич — д-р техн. наук, профессор, факультет программной инженерии и компьютерной техники, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО» (ИТМО). Область научных интересов: теория и методы обеспечения надёжности, отказоустойчивости и эффективности компьютерных систем и сетей. Число научных публикаций — 497. vladimir.bogatyrev@gmail.com; Кронверкский пр., 49А, 197101, Санкт-Петербург, Россия; р.т.: +7(812)232-5278.

M.K. DO, V. BOGATYREV
**SEMI-MARKOV RELIABILITY MODEL OF A COMPUTER
SYSTEM WITH CONTAINER VIRTUALIZATION**

Do M.K., Bogatyrev V. Semi-Markov Reliability Model of a Computer System with Container Virtualization.

Abstract. In modern cloud infrastructures and microservice architectures, ensuring high reliability, fault tolerance, and continuity of application operation is a critically important task. The primary factor reducing the efficiency of such systems is the phenomenon of software aging – the gradual degradation of system resources (memory leaks, data fragmentation) during prolonged operation, leading to a nonlinear increase in failure intensity. The complexity of reliability prediction under such conditions is обусловлена by the hierarchical dependency between the host's shared software and isolated containers, as well as the dynamic nature of recovery processes. This paper addresses the pressing problem of ensuring software reliability in a computer system with container virtualization subject to the phenomenon of software aging. The scientific novelty lies in the development of a comprehensive mathematical model based on semi-Markov processes, which accounts for the hierarchical structure of the system (host shared software and isolated containers) and enables the optimization of proactive recovery strategies. The Weibull distribution is used to describe the aging process, reflecting an increasing failure rate. A semi-Markov model of states and transitions is proposed, incorporating mechanisms of proactive container restart and migration. To reduce computational complexity, a simplified model is developed based on the assumption of negligibly low probability of secondary failures during short recovery intervals. Calculations of steady-state probabilities and availability are implemented in the Mathcad environment. The conducted numerical experiments showed that implementing a proactive recovery strategy within the interval range of 1200-2400 hours substantially improves the system availability. It is established that the error of the simplified model is negligible compared to the rigorous model. The proposed model allows determining the maximum availability value depending on the number of deployed containers and proactive container restart intervals. A comparison of the developed approach with traditional reactive recovery models confirms the advantage of proactive methods in minimizing downtime caused by error accumulation (memory leaks, resource fragmentation). The proposed model can be applied in the design of highly reliable cloud infrastructures and DevOps pipelines.

Keywords: reliability, availability factor, software, container virtualization, semi-Markov model, proactive recovery, container restart.

References

1. Dong H., Zhang C., Li G., Zhang H. Cloud-Native Databases: A Survey. IEEE Transactions on Knowledge and Data Engineering. vol. 36. no. 12. pp. 7772–7791. 2024. DOI: 10.1109/TKDE.2024.3397508.
2. Goyal P., Deora S.S. Reliability of Trust Management Systems in Cloud Computing. Indian Journal of Cryptography and Network Security. 2022. vol. 2. no. 1. pp. 1–5. DOI: 10.54105/ijcns.C1417.051322.
3. Assiri B., Sheneamer A. Fault tolerance in distributed systems using deep learning approaches. PLOS One. 2025. vol. 20. no. 1. DOI: 10.1371/journal.pone.0310657.
4. Khersonsky N.S., Bolshedvorskaya L.G. [Analysis of the development of statistical methods for assessing reliability as a quality property of objects and products]. Crede Experto: transport, obshchestvo, obrazovanie, yazyk – Crede Experto: transport, society,

- education, language. 2024. no. 1. pp. 6–23. DOI: 10.51955/2312-1327_2024_1_6. (In Russ.).
5. Gurjanov A.V., Korobeynikov A.G., Zharinov I.O., Zharinov O.O. Edge, fog and cloud computing in the cyber-physical systems networks. Proceedings of the III International Workshop on Modeling, Information Processing and Computing. 2021. pp. 103–108. DOI: 10.47813/dnit-mip3/2021-2899-103-108
6. Kushchazli A., Safargalieva A., Kochetkova I., Gorshenin A. Queuing Model with Customer Class Movement across Server Groups for Analyzing Virtual Machine Migration in Cloud Computing. Mathematics. 2024. vol. 12. no. 3. 468 p. DOI: 10.3390/math12030468.
7. Damarapati A. Containers vs. Virtual Machines: Understanding the Shift to Kubernetes. World Journal of Advanced Engineering Technology and Sciences. 2025. vol. 15. no. 1. pp. 852–861. DOI: 10.30574/wjaets.2025.15.1.0305.
8. Kozhirkbayev Z., Sinnott R.O. A performance comparison of container-based technologies for the Cloud. Future Generation Computer Systems. 2017. vol 68. pp. 175–182. DOI: 10.1016/j.future.2016.08.025.
9. Mirna A., Nadja K., Aris L. Utilization prediction-based VM consolidation approach. Journal of Parallel and Distributed Computing. 2022. vol. 170. pp. 24–38. DOI: 10.1016/j.jpdc.2022.08.001.
10. Polovko A.M., Gurov S.V. Osnovy' teorii nadezhnosti [Fundamentals of reliability theory]. Sankt-Peterburg: BXV-Peterburg. 2008. 702 p. (In Russ.).
11. Tatarnikova T.M., Arkhipov E.M. Designing Fault-tolerant Systems with Micro-service Architecture. Proceedings of the 27-th International Conference on Soft Computing and Measurements (SCM). 2024. pp. 348–351. DOI: 10.1109/SCM62608.2024.10554143.
12. Koren I., Krishna C.M. Fault-Tolerant Systems. Boston: Morgan Kaufmann. 2007. 400 p.
13. Yermakov S.G., Khalil M.M., Khomonenko A.D., Bukharova K.A. Evaluating the efficiency of fog computing on the Internet of Things using a Non-Markov model. T-Comm. 2022. vol. 16. no. 12. pp. 46–52.
14. Khomonenko A.D., Kiriynko A.B., Pugachev S.V. [Evaluation of the efficiency of fog computing in geoinformation systems]. Inzhenerny'j vestnik Dona – Engineering Bulletin of the Don. 2024. no. 6. pp. 364–382. (In Russ.).
15. Khomonenko A.D., Blagoveshchenskaya E.A., Prouzin O.V., Andruk A.A. [Forecasting the reliability of a cluster computing system using a semi-Markov model of alternative processes and monitoring]. Naukoemkie tekhnologii v kosmicheskix issledovaniyax Zemli – H&ES Research. 2018. vol. 10. no. 4. pp. 72–82. DOI: 10.24411/2409-5419-2018-10099. (In Russ.).
16. Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. [Assessment of the computer system's readiness for timely query service when combined with information recovery of memory after failures]. Nauchno-tekhnicheskij vestnik informacionny'x tekhnologij, mexaniki i optiki – Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2023. vol. 23. no. 3. pp. 608–617. DOI: 10.17586/2226-1494-2023-23-3-608-617. (In Russ.).
17. Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Recovery of Real-Time Clusters with the Division of Computing Resources into the Execution of Functional Queries and the Restoration of Data Generated Since the Last Backup. Lecture Notes in Computer Science. Springer. 2024. vol. 14123. pp. 236–250. DOI: 10.1007/978-3-031-50482-2_19
18. Bogatyrev V.A., Vinokurova M.S. Control and Safety of Operation of Duplicated Computer Systems. Communications in Computer and Information Science. Springer. 2017. vol. 700. pp. 331–342. DOI: 10.1007/978-3-319-66836-9_28.
19. Aleksankov S.M. [Models of dynamic migration with an iterative approach and network migration of virtual machines]. Nauchno-tekhnicheskij vestnik informacionny'x tekhnologij, mexaniki i optiki – Scientific and Technical Journal

- of Information Technologies, Mechanics and Optics. 2015. vol. 15. no. 6. pp. 1098–1104. DOI: 10.17586/2226-1494-2015-15-6-1098-1104. (In Russ.).
20. Bogatyrev V.A., Derkach A.N. Evaluation of a Cyber-Physical Computing System with Migration of Virtual Machines During Continuous Computing. *Computers*. 2020. vol. 9. no 2. pp. 42. DOI: 10.3390/computers9020042.
21. Bogatyrev V.A., Fung V.K. [Assessment of the reliability of the cluster being restored by container virtualization]. *Nauchno-texnicheskij vestnik informacionny'x texnologij, mexaniki i optiki* – Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2025. vol. 25. no. 5. pp. 988–995. DOI: 10.17586/2226-1494-2025-25-5-988-995. (In Russ.).
22. Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. [Boundary estimation of cluster system reliability based on Markov model decomposition with limited node recovery with accumulated failures]. *Nauchno-texnicheskij vestnik informacionny'x texnologij, mexaniki i optiki* – Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2025. vol. 25. no. 3. pp. 574–583. DOI: 10.17586/2226-1494-2025-25-3-574-583. (In Russ.).
23. Rodriguez M.A., Buyya R. Container-based cluster orchestration systems: A taxonomy and future directions. *Software: Practice and Experience*. 2019. vol. 49. pp. 698–719. DOI: 10.1002/spe.2660.
24. Application Container Market Size And Forecast. Verified Market Research. Available at: www.verifiedmarketresearch.com/product/application-container-market. (accessed 16.06.2026). (In Russ.).
25. Application Container Market (2025-2023). Grand View Research. Available at: www.grandviewresearch.com/industry-analysis/application-container-market. (accessed 16.06.2026). (In Russ.).
26. Marella V. Comparative Analysis of Container Orchestration Platforms: Kubernetes vs. Docker Swarm. *International Journal of Scientific Research in Science and Technology*. 2024. vol. 11. no. 5. pp. 526–543. DOI: 10.32628/IJSRST24105254.
27. Grotke M., Trivedi K.S. Fighting bugs: remove, retry, replicate, and rejuvenate. *Computer*. 2007. vol. 40. no. 2. pp. 107–109. DOI: 10.1109/MC.2007.55.
28. Cotroneo D., Natella R., Pietrantuono R., Russo S. Software Aging Analysis of the Linux Operating System. *Proceedings of the IEEE 21st International Symposium on Software Reliability Engineering (ISSRE)*. 2010. pp. 71–80. DOI: 10.1109/ISSRE.2010.24.
29. Nascimento M.G., Moura R.J., Machida F., Andrade E. Comparison of Machine Learning Algorithms for Detecting Software Aging in SQL Server. *Proceedings of the 13th Latin-American Symposium on Dependable and Secure Computing*. 2024. pp. 159–167. DOI: 10.1145/3697090.3699798.
30. Oliveira F., Araujo J., Matos R., Maciel P. Software Aging in Container-Based Virtualization: An Experimental Analysis. *Proceedings of the 16th Iberian Conference on Information Systems and Technologies (CISTI)*. 2021. pp 1–7. DOI: 10.23919/CISTI52073.2021.9476625.
31. Nguyen T.A., Min D., Choi E., Tran T.D. Reliability and Availability Evaluation for Cloud Data Center Networks Using Hierarchical Models. *IEEE Access*. 2019. vol. 7. pp. 9273–9313. DOI: 10.1109/ACCESS.2019.2891282.
32. Shruthi P., Nagaraj G.C. Software aging prediction – a new approach. *International Journal of Electrical and Computer Engineering*. 2023. vol. 13. no 2. pp. 1773–1781. DOI: 10.11591/ijece.v13i2.pp1773-1781.
33. Pietrantuono R., Russo S. A survey on software aging and rejuvenation in the cloud. *Software Quality Journal*. 2020. vol 28. pp. 7–38. DOI: 10.1007/s11219-019-09448-3.

34. Turk L.I.A., Alsolami E.G. Jelinski-Moranda Software Reliability Growth Model: A Brief Literature and Modification. *International Journal of Software Engineering & Applications (IJSEA)*. 2016. vol. 7. no. 2. pp. 33–44. DOI: 10.5121/ijsea.2016.7204.
35. Li N., Tan Y., Wang X., Li B., Luo J. Dynamic Resource Allocation for Containerized Applications in Edge Computing. *Lecture Notes in Electrical Engineering*. Springer. 2023. vol. 1015. pp. 121–130. DOI: 10.1007/978-981-99-0416-7_12.
36. Musa J.D., Iannino A., Okumoto K. *Software Reliability: Measurement, Prediction, Application*. New York: McGraw-Hill. 1987. 621 p.
37. Dohi T., Goseva-Popstojanova K., Trivedi K.S. Statistical non-parametric algorithms to estimate the optimal software rejuvenation schedule. *Proceedings of the IEEE Pacific Rim International Symposium on Dependable Computing*. 2000. pp. 77–84. DOI: 10.1109/PRDC.2000.897287.
38. Machida F., Kim D.S., Trivedi K.S. Modeling and analysis of software rejuvenation in a server virtualized system with live VM migration. *Performance Evaluation*. 2013. vol. 70. no. 3. pp. 212–230. DOI: 10.1016/j.peva.2012.09.003.
39. Zhang Y., Zheng Z., Lyu M.R. BFTCloud: A Byzantine Fault Tolerance Framework for Voluntary-Resource Cloud Computing. *Proceedings of the IEEE 4th International Conference on Cloud Computing*. 2011. pp. 444–451. DOI: 10.1109/CLOUD.2011.16.
40. Manker H., Hemant K.S., Iti S. Reliability analysis of a system using semi Markov process. *International Journal of Statistics and Applied Mathematics*. 2024. vol. 9. no. 4. pp. 145–151. DOI: 10.22271/math.2024.v9.i4b.1785.
41. Martyna J. Semi-Markov Model of Processing Requests Reliability and Availability in Mobile Cloud Computing Systems. *Advances in Intelligent Systems and Computing*. Springer. 2021. vol. 1389. pp. 264–272. DOI: 10.1007/978-3-030-76773-0_26.
42. Howard R.A. *Dynamic Probabilistic Systems: Semi-Markov and Decision Processes*. New York: Wiley. 1971. 564 p.
43. Machida F., Miyoshi N. Analysis of an optimal stopping problem for software rejuvenation in a deteriorating job processing system. *Reliability Engineering & System Safety*. 2007. vol. 168. pp. 128–135. DOI: 10.1016/j.res.2017.05.019.
44. Weibull W. A Statistical Distribution Function of Wide Applicability. *Journal of Applied Mechanics*. 1951. vol. 18. no. 3. pp. 293–297. DOI: 10.1115/1.4010337.
45. Rinne H. *The Weibull Distribution: A Handbook*. New York: CRC Press. 2008. 808 p. DOI: 10.1201/9781420087444.
46. Murthy D.N.P., Xie M., Jiang R. *Weibull Models*. New York: John Wiley & Sons. 2003. 416 p. DOI: 10.1002/047147326X.
47. Novikov I.C. [Methods of calculation of quantity indicators of reliability of difficult program complexes on a design stage and development]. *Trudy SPIRAN – SPIRAS Proceedings*. 2008. no. 6. pp. 86–111. DOI: 10.15622/sp.6.8. (In Russ.).
48. Okamura H., Dohi T. Faster Maximum Likelihood Estimation Algorithms for Markovian Arrival Processes. *Proceedings of the IEEE International Conference on Quantitative Evaluation of Systems*. 2009. pp. 73–82. DOI: 10.1109/QEST.2009.28.
49. Trivedi K., Andrade E., Machida F. Combining Performance and Availability Analysis in Practice. *Advances in Computers*. 2012. vol. 84. pp. 1–38. DOI: 10.1016/B978-0-12-396525-7.00001-0.
50. Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Control of Multipath Transmissions in the Nodes of Switching Segments of Reserved Paths. *Proceedings of the International Conference on Information, Control, and Communication Technologies (ICCT)*. 2022. pp. 1–5. DOI: 10.1109/ICCT56057.2022.9976839.
51. Bogatyrev V.A., Bogatyrev A.V., Bogatyrev S.V. Multipath Transmission of Heterogeneous Traffic in Acceptable Delays with Packet Replication and Destruction of Expired Replicas in the Nodes that Make Up the Path.

- Communications in Computer and Information Science. Springer. 2023. vol. 1748. pp. 104–121. DOI: 0.1007/978-3-031-30648-8_9.
52. Utkin L.V., Mulukha V.A., Lukashin A.A., Konstantinov A.V. Prinyatie reshenij v usloviyax neopredelennosti [Decision-making in conditions of uncertainty]. Sankt-Peterburg: Politekh-Press. 2025. 49 p. (In Russ.).
 53. Utkin L.V. Analiz riska i prinyatie reshenij pri nepolnoj informacii [Risk analysis and decision-making with incomplete information]. Sankt-Peterburg: Nauka. 2007. 403 p. (In Russ.).

Do Manh Kiem — Graduate Student, Department of Software Engineering and Computer Science, ITMO University. Research interests: reliability theory, queueing theory, fault tolerance, and the efficiency of computer systems and networks. The number of publications — 2. domanhkiemnd@gmail.com; 49A, Kronverkskiy Ave., 197101, St. Petersburg, Russia; office phone: +7(911)199-8063.

Bogatyrev Vladimir — Ph.D., Dr.Sci., Professor, Department of Software Engineering and Computer Science, ITMO University. Research interests: theory and methods for ensuring the reliability, fault tolerance, and efficiency of computer systems and networks. The number of publications — 497. vladimir.bogatyrev@gmail.com; 49A, Kronverkskiy Ave., 197101, St. Petersburg, Russia; office phone: +7(812)232-5278.

С.Ф. БОЕВ, А.С. ЗАХАРОВ, А.Ю. ПЕРЛОВ, А.В. ТИМОШЕНКО
**МОДЕЛЬ КЛАССИФИКАЦИИ ПОДВИЖНЫХ ОБЪЕКТОВ
ПО ДАННЫМ РАДИОЛОКАЦИОННЫХ ПОРТРЕТОВ
КРУПНОАПЕРТУРНЫХ РЛС L-ДИАПАЗОНА С УЧЕТОМ
СТОХАСТИЧЕСКИХ НЕГАТИВНЫХ ФАКТОРОВ
РАЗЛИЧНОЙ ПРИРОДЫ**

Боев С.Ф., Захаров А.С., Перлов А.Ю., Тимошенко А.В. Модель классификации подвижных объектов по данным радиолокационных портретов крупноапертурных РЛС L-диапазона с учетом стохастических негативных факторов различной природы.

Аннотация. Приведены результаты исследований по совершенствованию моделей и методов оперативной классификации космических аппаратов на основе построения радиолокационных портретов. Предложена гипотеза о том, что минимально необходимое для правильной классификации число радиолокационных портретов является функционалом оригинального показателя интегрального риска, учитывающего стохастический характер негативных факторов, снижающих точностные характеристики радиолокационной станции. Предложена модель классификации подвижных объектов по данным радиолокационных портретов крупноапертурных РЛС L-диапазона с учетом стохастических негативных факторов различной природы. Приведена оригинальная формализация процесса информационного обеспечения классификации космических аппаратов по радиолокационным портретам с использованием двумерного неоднородного дифференциального уравнения для оценки функции плотности вероятности воздействия негативных факторов различной природы на точностные характеристики РЛС. Новизна модели заключается в комплексном учете разнородных факторов на процесс классификации КА, интеграции с риск-анализом и иерархической многоуровневой структуре модели в интересах оперативной классификации. Показано, что представление модели классификации космических аппаратов на основе усовершенствованного риск-анализа обеспечивает получение аналитической зависимости числа радиолокационных портретов с учетом воздействия факторов различной природы, которые формализованы в виде усовершенствованных уравнений Фоккера-Планка (для формализации внутренних факторов) и уравнения Фейнмана-Каца (для внешних факторов, включая динамику реального космического фона). Показано, что совершенствование теории риск-анализа научной школы академика Ю.И. Шокина с учетом результатов решения системы уравнений позволяет существенно улучшить информационное обеспечение задач классификации. Приведены результаты вычислительного эксперимента по обработке радиолокационных портретов, показывающие потенциальную возможность повышения оперативности и точности классификации.

Ключевые слова: радиолокационная станция, подвижный объект, космический аппарат, радиолокационное портретирование, классификация, стохастические факторы.

1. Введение. Неуклонный рост числа выводимых в околоземное пространство подвижных объектов ведет к повышению сложности решения актуальной задачи классификации космических аппаратов (КА), а оснащение нового поколения низкоорбитальных аппаратов

двигателями малой тяги, позволяющими оперативно изменять параметры орбиты, приводит к еще большему возрастанию требований к системе информационного обеспечения решения задач классификации в интересах поддержания каталога КА [1].

В настоящее время классификация КА в основном проводится с использованием оптоэлектронных средств (ОЭС), возможности которых в ряде случаев существенно ограничены из-за снижения оптической видимости при ухудшении погодных условий [2 – 5].

В этой связи актуальной задачей является совершенствование методов классификации по данным радиолокационных станций мониторинга космического пространства (РЛС МКП), работающих с широкополосными сигналами для построения радиолокационных портретов (РЛП) КА, являющихся источником информации о типе и форме объекта для классификации КА, а также о характере движения КА, в том числе о маневрировании.

Современные РЛС L-диапазона, использующие широкополосные линейно-частотно модулированные (ЛЧМ) сигналы большой длительности, обладают потенциальными возможностями для решения задачи построения радиолокационных портретов (РЛП) в интересах классификации КА [6 – 9].

В настоящей работе под классификацией понимается распознавание типа объекта с заданной достоверностью. В качестве основного информативного признака для классификации используется дальностный радиолокационный портрет, полученный в ходе сжатия ЛЧМ сигналов. Обработка пачек широкополосных ЛЧМ сигналов при сопровождении КА позволяет определить геометрическую форму и размер объекта для его распознавания путем сравнения с эталонным портретом из каталога КА.

Теоретические исследования процессов учета и компенсации искажений и нестабильности параметров устройств формирования и усиления широкополосных сигналов достаточно подробно представлены в работах [7, 10], где показано, что воздействие стохастических негативных факторов различной природы приводит к снижению точности измерения координат КА и соответственно к ухудшению РЛП и оперативности классификации КА.

Известные методы калибровки и юстировки РЛС различного типа в определенной степени обеспечивают решение задачи компенсации по каждому фактору в отдельности для построения РЛП целей с одной стороны [7, 8]. Однако, отсутствие методов комплексной оценки и оперативной компенсации негативного влияния на точностные характеристики всех факторов не позволяет обеспечить

требуемую оперативность классификации КА. Данное обстоятельство обуславливает необходимость разработки модели информационного обеспечения классификации КА, позволяющей представить процесс классификации с учетом разнородного влияния стохастических факторов на требуемое количество РЛП для совершенствования методов оперативной классификации КА.

Существующие методы калибровки и юстировки в РЛС направлены на компенсацию систематических составляющих ошибок и слабо применимы для оперативной компенсации быстрых случайных воздействий, имеющих стохастический характер. В этой связи актуальной является разработка динамической модели, которая позволила бы определить аналитическую зависимость влияния негативных факторов на информационную достаточность последовательности РЛП для решения задачи классификации.

В наибольшей степени [11 – 36] на точностные характеристики влияют факторы, связанные с аппаратурой, такие как неравномерность АЧХ, стохастические изменения спектральной плотности мощности фазовых шумов (СПМФШ), потери коэффициента усиления, вызванные тепловыми процессами и случайными отказами. Кроме этого, текущие изменения условий функционирования и среды распространения в ряде случаев также приводят к критическому снижению точности измерения координат КА.

Анализ функционирования РЛС [11 – 18] при построении РЛП КА показал, что необходимо также учитывать скорость и степень влияния факторов на число РЛП. Зависимость числа РЛП от доверительного уровня представлена на рисунке 1.

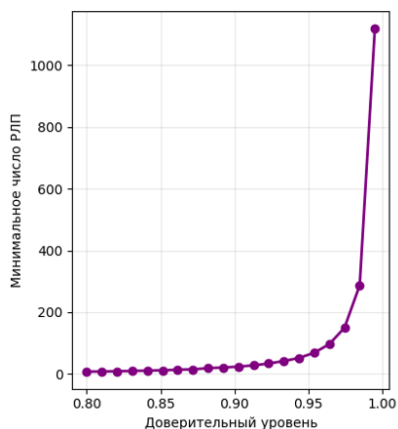


Рис. 1. Зависимость числа РЛП от доверительного уровня

Нелинейная зависимость минимального необходимого числа РЛП от требуемого доверительного уровня демонстрирует экспоненциальный рост объемов данных при стремлении вероятности охвата к единице. Точка перегиба при уровне доверия 0.92 соответствует переходу от экономичного к надежному режиму работы системы. Асимптотическое поведение при $P \rightarrow 1$ подтверждает фундаментальное ограничение на гарантированное исключение сколь угодно редких событий при конечных объемах выборки.

Оперативное решение задачи классификации КА в значительной степени определяется влиянием на РЛС МКП следующих факторов [7, 13, 14]:

- ухудшением параметров технического и функционального состояния РЛС, в том числе текущей неидентичности амплитудно-фазового распределения (АФР), нестабильности температурных параметров приемо-передающих модулей, неравномерности амплитудно-частотной характеристики (АЧХ) трактов, что в совокупности с параметрами зондирующих сигналов (ширина спектра, тип модуляции, длительность, скважность) формирует стохастические изменения параметров внутри РЛС;

- динамическими и стохастическими изменениями параметров среды (ионосферные и тропосферные изменения);

- неопределенностью параметров движения КА (локальные неоднородности, флуктуации фазового центра, вращение, маневрирование).

Для проверки гипотезы о возможности применения метода риск-анализа при оценке влияния факторов на время классификации КА сгруппируем представленные выше негативные факторы на две группы (внешние факторы и внутренние факторы), рисунок 2.

Из приведенной схемы следует, что модель классификации в данном случае должна включать в свой состав частные модели всех указанных составляющих процесса классификации (внешние факторы, внутренние факторы, характеристики РЛП), что должно обеспечить выполнение целевой задачи в режиме построения РЛП – задачи оперативной классификации КА при воздействии негативных факторов на точностные характеристики РЛС.

Объектом исследования выступает крупноапертурная РЛС L-диапазона нового поколения, что требует теоретического описания процессов в условиях фактического отсутствия информации о функционировании аналогов. Данное обстоятельство также ограничивает возможность разработки цифрового двойника РЛС и его компонентов, что определяет актуальность разработки

математической модели на основе формализованного процесса классификации от влияния ключевых факторов на качество РЛП.

Целью работы является разработка теоретической модели классификации подвижных объектов по данным радиолокационных портретов крупноапертурных РЛС L-диапазона с учетом стохастических негативных факторов различной природы.

В качестве теоретической основы для описания внутренних факторов в работе используется уравнение Фоккера-Планка. Применение данного уравнения обосновано тем, что такие параметры, как температурное распределение по полотну АФАР или деградация коэффициента усиления приемо-передающих модулей (ППМ), изменяют свои значения во времени под действием детерминированного тренда (сноса) и стохастических флуктуаций (диффузии) вследствие высокого уровня мощности сигналов и неидентичностей в приемо-передающих трактах.

Для формализации внешних факторов, связанных с возмущениями среды распространения и изменением параметров движения цели, предлагается совершенствование уравнения Фейнмана-Каца.



Рис. 2. Представление деструктивных факторов в виде двух групп – внешних и внутренних

2. Формализованная постановка задачи. Будем полагать, что процесс классификации по данным РЛП с требуемым уровнем точности измерения координат КА при заданных параметрах зондирующих сигналов и характеристик приемо-передающих каналов РЛС можно представить как процесс влияния факторов на достоверность полученной информации, определяющейся числом

РЛП. Формализованная постановка задачи определения необходимого объема выборки представлена в следующем виде:

$$T = \operatorname{argmin} N [P > 0,95; S; A; T; K; R], \quad (1)$$

где P – вероятность правильной классификации; N – необходимое число портретов РЛП для правильной классификации КА; T – вектор характеристик (сигнал/шум, разрешающая способность по дальности и углам) в режиме радиопортретирования; S – вектор параметров зондирующих сигналов РЛС; K – вектор параметров приемо-передающего тракта РЛС; A – вектор параметров условий функционирования (ошибки юстировки и возмущения среды); R – риск изменения точности измерения координат.

В такой постановке целевой функцией является число РЛП, которое изменяется в зависимости от условий функционирования РЛС. Время классификации определяется как:

$$T_{class} = \min t \quad \text{при} \quad P(\hat{y}(t)) \geq \theta, \quad (2)$$

где $\hat{y}(t)$ – спрогнозированный класс КА; P – вероятность правильной классификации; θ – заданный порог доверия; t – время выполнения операций классификации, измерения и накопления полезного сигнала. Вероятность классификации зависит от последовательности качественных РЛП с минимальным уровнем ошибок измерения координат.

С целью определения минимально необходимого числа РЛП предложено применение математического аппарата риск-анализа для комплексной оценки влияния разнородных и стохастических негативных факторов на снижение точностных характеристик при сопровождении КА.

Проведенный анализ влияния деструктивных факторов на процесс информационного обеспечения решения задач классификации КА показал, что для разработки модели классификации КА предлагается провести:

- формализованное описание влияния внутренних факторов на основе совершенствования уравнения Фоккера-Планка;
- формализованное описание влияния внешних факторов с использованием модифицированного уравнения Фейнмана-Каца;
- на основе совместного решения уравнений определить функцию распределения суммарной ошибки измерения координат;

– установить аналитическую зависимость минимально необходимого числа РЛП от интегрального риска и требуемой вероятности правильной классификации КА.

Общая схема работы модели при решении задачи оперативной классификации представлена на рисунке 3.

Основой предлагаемой модели является применение двух дифференциальных уравнений (Фоккера-Планка и Фейнмана-Каца) для описания стохастических процессов с последующим использованием аппарата риск-анализа для получения аналитической зависимости вероятность правильной классификации от числа РЛП.

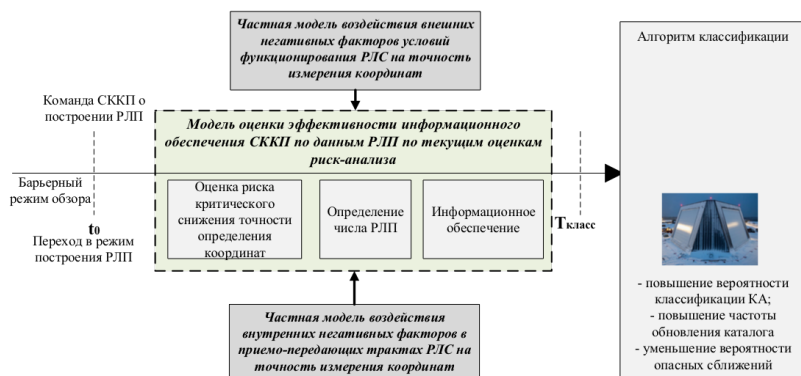


Рис.3. Общая схема работы модели при решении задачи оперативной классификации

3. Математическая модель риск-анализа для оценки числа РЛП для эффективной классификации. Риск-анализ представляет собой систематический процесс использования информации для идентификации опасностей, оценки потенциального вреда и разработки мер по управлению [37]. Согласно концепции Ю.И. Шокина [38] риск является не статической, а динамической величиной, которая существенно зависит от объема и качества доступных знаний I . В указанных работах обоснована необходимость рассмотрения риска как функции $R=\{A, C, Q; I\}$, где A – сценарии, C – последствия, Q – мера неопределенности, а I – совокупность информации.

Любая задача оптимизации, связанная с оценкой потерь при вычислении точности в зависимости от принятого решения, должна учитывать не только ожидаемые потери, но и риски, обусловленные самим решением [39, 40]. При решении задачи оперативной классификации КА под риском будем понимать число,

характеризующее уровень воздействия деструктивных факторов на аппаратуру РЛС.

Исследования показали [40, 41], что риск-анализ позволяет решать задачи оценивания потенциальных изменений точностных характеристик РЛС. Поэтому исходя из теории риск-анализа [39], процесс построения РЛП может быть представлен как риск критического снижения точностных характеристик, который определяется, как функция изменения точности измерения координат $\Delta(r, \theta, \phi)$ и плотности вероятности изменения w для быстрой и медленной составляющей воздействий:

$$R_{\Delta(r, \theta, \phi)} = \sum \Delta_i(r, \theta, \phi) \prod_{f,s} w_j, \quad (3)$$

где Δ_i – изменения точностных характеристик по координатам.

Риск в данном случае является не вероятностной характеристикой, а его размерность определяется той характеристикой, изменение которой надо найти. В данной работе будем использовать подход, в котором размерный риск является функцией потерь, среднее значение которой определяется для оценки воздействия деструктивных факторов.

Детализируя вектор функциональных и технических характеристик на примере крупноапертурной РЛС L-диапазона, состоящий из разрешающей способности по угловым координатам и отношения сигнал/шум (по мощности или по амплитуде), а также внутренних и внешних факторов, процесс классификации КА по данным РЛП может быть также представлен следующим образом:

$$P_{\text{класс}} = \operatorname{argmin}_\tau (N_{\text{РЛП}}(R_{\Delta(r, \theta, \phi)})), \quad (4)$$

$$R_{\Delta(r, \theta, \phi)} = R_{\Delta(r, \theta, \phi)}(\delta r, \delta \theta, \delta \phi), \quad (5)$$

$$\begin{aligned} \delta r &= \delta r(B(K(\partial H))); \delta \theta = \delta \theta(I_{\text{АФАР}}, H_{\text{тракта}}); \\ \delta \phi &= \delta \phi(I_{\text{АФАР}}, H_{\text{тракта}}), \end{aligned} \quad (6)$$

где $P_{\text{класс}}$ – вероятность правильной классификации КА, $N_{\text{РЛП}}$ – число РЛП КА необходимых для точной классификации; τ – время классификации КА; δ_r – разрешающая способность по дальности; B – ширина спектра сигнала; ∂H – неравномерность АЧХ приемопередающего тракта; δ_θ , δ_ϕ – разрешающая способность

по азимуту и углу места; $I_{\text{АФАР}}$ – изменения амплитудно-фазового распределения на полотне АФАР; $H_{\text{тракта}}$ – искаженное АЧХ тракта.

Функция аргумента минимизации в данном случае выбрана по причине необходимости минимизации времени классификации при сопровождении КА в зоне обзора РЛС. Время классификации зависит от числа портретов неинвариантно, поскольку каждый портрет имеет разное качество, а значит при наличии серьезных искажений необходимо провести дополнительный цикл зондирования и выбора оптимальных сигналов РЛС. Подобная оптимизация может быть достигнута на основе градиентного спуска.

Для учета влияния негативных факторов на процесс классификации КА введем обобщенный функционал риска:

$$R_{\text{общ}} = \alpha R_{\text{внут}} + \beta R_{\text{внеш}} + \gamma R_{\text{взаим}}, \quad (7)$$

где α , β , γ – весовые коэффициенты, учитывающие значимость факторов, определяемые эмпирически и изменяющиеся от 0 до 1 с условием $\alpha + \beta + \gamma = 1$. Данное представление дает основу для представления всех факторов в иерархию, где коэффициенты определяют уровень значимости.

Представленный процесс определения необходимого числа РЛП космических аппаратов обуславливает необходимость описания стохастических факторов различной природы для применения метода риск-анализа с целью оценки изменения ошибок измерения координат.

4. Иерархический процесс свертки параметров модели классификации КА. Описание процесса классификации необходимо провести с учетом временных ограничений на проводку КА и юстировку, то есть разделить их на внутренние и внешние, а также по скорости взаимодействия.

В соответствии с анализом влияния деструктивных факторов на аппаратуру РЛС процесс информационного обеспечения представим как многоуровневый (Рис. 4).

В рамках предложенного представления построена структурированная система оценки, в которой риск агрегируется по принципу «снизу-вверх»: от оценок частных рисков на уровне отдельных факторов (на основе решений базовых уравнений) к комплексному интегральному риску через линейную комбинацию:

$$R_{\text{интегр}} = \sum_i w_i R_i, \quad (8)$$

где w_i – коэффициенты, аналогичные α , β , γ ; R_i – риск изменения точностных характеристик на каждом уровне.

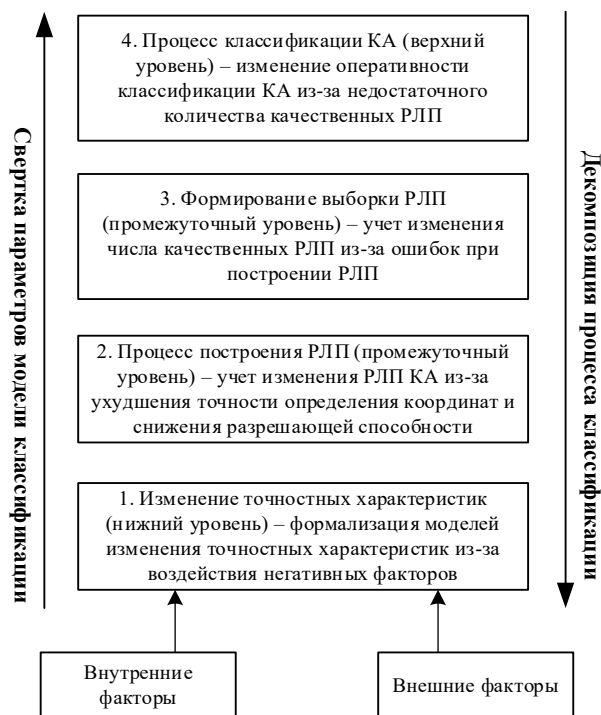


Рис. 4. Многоуровневая модель классификации КА с учетом негативных факторов, влияющих на качество РЛП

Оценка определения необходимого для правильной классификации КА числа портретов является итеративной процедурой с оценкой риска неполучения РЛП на каждой итерации:

$$N_{РЛП} \rightarrow N_{РЛП} + 1: \frac{R\delta r}{\Delta r_i} + \frac{R\delta\theta}{\Delta\theta_i} + \frac{R\delta\phi}{\Delta\phi_i} > 3, \quad (9)$$

где Δr_i , $\Delta\theta_i$, $\Delta\phi_i$ – разница между блестящими точками. Такое представление справедливо из-за того, что в случае повышения риска увеличения разрешающей способности возможно размытие блестящих точек, что приведет к неверной классификации. При этом вероятна такая ситуация, при которой по одной из координат возможно

значительное превышение, что также приводит к необходимости коррекции. Значение критерия, при котором сумма трех составляющих в общем больше 3 выбрана из-за того, что по каждой из координат строится РЛП с последующей классификацией, поэтому даже при наличии одного портрета с размытием вклад в вероятность классификации все равно будет положителен. Если же по каждой координате риск изменения будет больше разрешающей способности, то блестящие точки будут сливаться на каждом портрете, поэтому данный критерий достаточен для увеличения числа РЛП.

Разработанная иерархическая структура модели обеспечивает декомпозицию сложной задачи классификации КА по данным РЛП на частные модели негативных факторов, позволяя последовательно анализировать влияние стохастических воздействий различной природы (от аппаратных нестабильностей до ионосферных возмущений) на итоговое качество распознавания.

При этом в разработке модели необходимо учесть временную динамику негативных факторов на качество построения РЛП при работе с широкополосными сигналами в сложных условиях функционирования (например, длительное отсутствие юстировочного аппарата в зоне обзора РЛС, а также при возмущении ионосферы).

С целью описания процесса построения РЛП во времени декомпозируем ошибку формирования РЛП на медленную (S) и быструю (F) составляющие. При работе РЛС в режиме РЛП негативные факторы можно охарактеризовать как быстрые процессы, соответствующие внешним факторам, а также медленные процессы, которые соответствуют внутренним факторам, влияющим на точность измерения координат. Данное представление основано на оценке скорости процессов при зондировании КА, что характеризуется изменениями ракурса цели за 1 цикл излучения и приема пачки широкополосных сигналов. Быстрая составляющая $F(r,t)$ описывает оперативные ошибки измерения координат, меняющиеся в течение одного цикла зондирования КА: пассивные помехи в области возмущения ионосферы, ошибки в измерениях координат из-за отсутствия юстировочных КА.

Медленные процессы соответствуют изменениям при работе РЛС с широкополосными сигналами за несколько циклов зондирования и характеризуется незначительными изменениями характеристик за цикл зондирования. При этом медленные процессы имеют свойства накопления. Медленная составляющая $S(r,t)$ описывает инерционные изменения в аппаратуре РЛС: изменение амплитудно-фазового

распределения (АФР) в АФАР из-за температурных изменений, а также изменения характеристик приемо-передающих модулей (ППМ).

Такое представление модели классификации с учетом ошибок на одной временной оси с декомпозицией по скорости изменения позволяет применить метод риск-анализа, доказавший свою эффективность при описании функционирования сложных информационных систем [34] в качестве показателя для оценки влияния внешних и внутренних негативных факторов на качество классификации по данным РЛП.

5. Частная модель оценки воздействия медленных составляющих деструктивных факторов внутренней природы на основе уравнений Фоккера-Планка. В рамках задачи оперативной классификации КА по данным РЛП, формируемых крупноапертурными РЛС L-диапазона, воздействие внутренних негативных факторов (таких как тепловые процессы в каналах усиления, старение компонентов, неидентичность характеристик приемо-передающих модулей – ППМ) носит активный, негативный и диффузионный характер. Эти процессы, в отличие от быстрых внешних возмущений, обладают значительной временной постоянной (от нескольких минут до часов и суток) и накапливаются в течение многих циклов зондирования, систематически искажая амплитудно-фазовое распределение (АФР) и амплитудно-частотную характеристику (АЧХ) трактов.

Для формализации подобных стохастических, но инерционных воздействий возможным математическим аппаратом является уравнение Фоккера-Планка. Его фундаментальное преимущество в данном контексте заключается в способности описывать эволюцию во времени плотности вероятности $w_S(\vec{x}, t)$ ключевых параметров состояния системы $\vec{x} = \{x_1, \dots, x_N\}$, где под x понимается многомерный вектор, включающий:

- фазовые сдвиги в каналах АФАР, вызванные градиентами температуры;
- коэффициенты усиления ППМ;
- параметры нелинейности и неравномерности АЧХ;
- уровни собственных шумов аппаратуры.

Уравнение Фоккера-Планка является кинетическим уравнением для марковских диффузионных процессов, что соответствует физике медленных деградационных изменений в РЛС: будущее состояние системы зависит только от текущего, а ее эволюция складывается из детерминированного тренда («сноса») и случайных блужданий («диффузии») [41 – 43]. Марковскими процессами в аппаратуре РЛС

являются процессы восстановления, старения и каскадных повреждений от единичного деструктивного воздействия. Их использование позволяет перейти от качественного описания к количественному прогнозированию надежности и готовности сложнейших радиолокационных комплексов. Марковское свойство предполагается выполненным в силу того, что временные константы рассматриваемых процессов значительно превышают характерные времена корреляции внешних возмущений.

Формализуем влияние медленных составляющих на точностные характеристики на основе уравнения Фоккера-Планка, поскольку для нашего случая уравнение описывает изменения во времени плотности вероятности характеристик для таких диффузионных процессов, как изменение точности измерения координат, влияющее на число РЛП для правильной классификации КА.

Под коэффициентом сноса в статье понимаются значения смещения вероятности отказа аппаратуры из-за неидентичностей характеристик в приеме-передающих трактах, а коэффициент диффузии показывает, как деструктивные факторы распространяются по выделенному объему аппаратуры. Коэффициенты сноса и диффузии могут быть оценены на основе экспериментальных данных с использованием методов статистического анализа временных рядов, например, метода максимального правдоподобия для стохастических дифференциальных уравнений.

Применение уравнения Фоккера-Планка позволит анализировать и прогнозировать, как будет изменяться распределение ключевых параметров системы (температуры, коэффициентов усиления, фазовых сдвигов) в течение многих циклов зондирования. Данная формула применима, поскольку в РЛС МКП имеется множество однотипных блоков, то есть имеется возможность описать ансамбль объектов.

Эти процессы по своей природе являются марковскими диффузионными, где состояние системы (например, параметры ППМ) плавно изменяется во времени под влиянием детерминированного тренда (например, нагрева) и стохастических флуктуаций (например, шумов). Уравнение Фоккера-Планка для описания внутренних факторов представим в следующем виде [42 – 44]:

$$\frac{\partial w_S}{\partial t} = - \sum_{i=1}^2 \frac{\partial}{\partial x_i} (\sigma_i w_S) + \sum_{i=1}^2 \sum_{j=1}^2 \frac{\partial^2}{\partial x_i \partial x_j} (D_{i,j} w_S), \quad (10)$$

где σ – коэффициент сноса, который характеризует изменение среднего значения плотности вероятности воздействия деструктивных факторов; D_{ij} – коэффициент диффузии, который характеризует изменение плотности вероятности воздействия деструктивных факторов в пространстве аппаратуры РЛС. Усовершенствование уравнений Фоккера-Планка состоит в учете импульсного режима работы РЛС, характеризуемого сменой режима работы с учетом планировщика тактового расписания. С точки зрения уравнений Фоккера-Планка смену режима работы можно описать как переход из состояния x' в x . Если имеется определенная периодичность смены режимов, то можно построить интенсивность перехода λ . Тогда:

$$\begin{aligned} \frac{\partial w_S}{\partial t} = & - \sum_{i=1}^2 \frac{\partial}{\partial x_i} (\sigma_i w_S) + \sum_{i=1}^2 \sum_{j=1}^2 \frac{\partial^2}{\partial x_i \partial x_j} (D_{i,j} w_S) + \\ & + \int \left[(\lambda(x', t) w_S(x', t)) - (\lambda(x, t) w_S(x, t)) \right] dx' \end{aligned} \quad (11)$$

Будучи фундаментальным уравнением необратимой термодинамики, уравнение Фоккера-Планка подходит для моделирования температурных дрейфов в каналах усиления РЛС. Динамику температуры приемо-передающих модулей (ППМ) можно описать как:

$$\frac{\partial P(T, t)}{\partial t} = \frac{-\partial}{\partial T} \left[\frac{k(T_{env} - T) P \omega}{W_{пол}} \right] + \frac{\partial^2}{\partial T^2} [D_T P], \quad (12)$$

где κ – коэффициент теплопередачи ППМ, T_{env} – температура окружающей среды, D_T (10^{-3} - 10^{-2} K²/с) – коэффициент тепловой диффузии, $W_{пол}$ – полная мощность излучения, ω – круговая частота. Физически это соответствует закону Фурье с учётом стохастических флуктуаций, что точно отражает реальный тепловой режим аппаратуры.

Процессы деградации компонентов РЛС, такие как старение транзисторов или деградация амплитудно-частотной характеристики (АЧХ), обладают марковским свойством. Уравнение Фоккера-Планка для плотности вероятности $w(x, t)$ формализует это:

$$\frac{\partial w}{\partial t} = \frac{-\partial(Aw)}{\partial x} + \frac{1}{2} \frac{\partial^2(Bw)}{\partial x^2}, \quad (13)$$

где A – вектор сноса, описывающий систематическую деградацию, а B – тензор диффузии, характеризующий случайные изменения. Деградация активных фазированных антенных решеток (АФАР)

следует такому марковскому процессу с коэффициентом сноса порядка 0.001-0.01 дБ/час.

Для моделирования сложных систем, таких как крупноапертурные активные фазированные антенные решетки (АФАР), уравнение обобщается с учётом пространственных координат:

$$\frac{\partial w(\phi, r, t)}{\partial t} = \frac{-\partial}{\partial \phi} (v_{\phi} w) - \frac{\partial}{\partial r} (v_r w) + D_{\phi} \frac{\partial^2 w}{\partial \phi^2} + D_r \frac{\partial^2 w}{\partial r^2}, \quad (14)$$

где ϕ – фазовый сдвиг, а r – амплитуда в точке решетки. Это позволяет учесть пространственную корреляцию между элементами.

В стационарном режиме ($\partial w / \partial t = 0$) для линейных задач уравнение часто допускает аналитическое решение. Например, для системы с линейным дрейфом стационарное распределение является гауссовским:

$$w_{st}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad (15)$$

где параметры μ и σ^2 определяются коэффициентами уравнения. Это подтверждается практикой: распределение температуры ППМ в установившемся режиме близко к гауссовскому со стандартным отклонением $\sigma = 2-5$ К.

Таким образом, уравнение Фоккера-Планка предоставляет универсальный, физически обоснованный и математически строгий инструмент для комплексного моделирования, прогнозирования и оптимизации работы радиолокационных систем на всех этапах их жизненного цикла.

6. Частная параболическая модель быстрых составляющих деструктивных факторов. С целью формализации быстрых составляющих деструктивных факторов запишем выражение для изменения зондирующего сигнала:

$$\delta I = \operatorname{argmin}(u), u \rightarrow \text{const}, \quad (16)$$

где δI – изменение интенсивности сигнала, u – плотность вероятности изменения ионосферы и реального космического фона.

Допустим, в зоне обзора РЛС имеется ионосферная помеха, являющаяся неоднородностью в пространстве. В момент времени t_0 вся плотность вероятности сосредоточена в области помехи, тогда $w = F(r, t)$, где r объединяет значения x, y, z . Изменение коэффициента

преломления волн соответствует эффективному потенциалу для ионосферных неоднородностей, что позволяет в дальнейших уравнениях включить произведение V^*w . При критических значениях потенциала V происходит фазовый переход, соответствующий образованию каналов распространения в ионосфере. Данные о распространении КВ радиоволн подтверждают существование критических условий для образования ионосферных каналов. Со временем ионосферные помехи изменяются с помощью диффузионного течения, поскольку корреляционная длина ионосферных неоднородностей составляет 10-100 км, тогда введя коэффициенты сноса и диффузии, отвечающие за скорость распространения, финальное уравнение определяется выражением:

$$\begin{aligned} \frac{\partial w_F}{\partial t} + \sum_{i=1}^2 \frac{\partial}{\partial x_i} (\sigma_i w_F) + \sum_{i=1}^3 \sum_{j=1}^3 \frac{\partial^2}{\partial x_i \partial x_j} (D_{i,j} w_F) - \\ - V(x_1 \dots x_n, t) w_F + F(x_1 \dots x_n, t) = 0 \end{aligned} \quad (17)$$

Полученное уравнение является адаптацией уравнения Фейнмана-Каца, которое отражает связь между дифференциальными уравнениями в частных производных и случайными процессами. Усовершенствование уравнений Фейнмана-Каца состоит в оценке диффузионных течений на основе оценки корреляционной длины ионосферных неоднородностей, при этом учитывая средний размер неоднородностей [44]. В части адекватности модели примем гипотезу, что значения потенциала V не достигают больших значений.

Уравнение Фейнмана-Каца устанавливает связь между стохастическими дифференциальными уравнениями (описывающими динамику состояния, например, динамику реального космического фона) и детерминированными уравнениями в частных производных параболического типа. Достоинством уравнения Фейнмана-Каца является то, что оно позволяет вычислять математические ожидания ключевых функционалов от траектории системы (например, средний квадрат ошибки положения КА) как решения соответствующего дифференциального уравнения. Данное уравнение позволяет проводить оценки мгновенного и краткосрочного вероятностного воздействия быстрых помех, создаваемых реальным космическим фоном, на точность измерений РЛС в реальном времени.

Сочетание усовершенствованных уравнений Фоккера-Планка для описания внутренних факторов в приемо-передающих трактах РЛС и модифицированного уравнения Фейнмана-Каца для описания внешних факторов реального-космического фона в рамках

предложенного иерархического описания процесса классификации с использованием метода риск-анализа позволит создать эффективную систему компенсации стохастического влияния негативных факторов на качество РЛП.

В соответствии с предложенным описанием выше, необходимо оценить эффективность представленной модели классификации на демонстрационном примере.

7. Демонстрационный эксперимент по построению дальностных портретов типовых объектов наблюдения. С целью подтверждения эффективности предложенной модели классификации использовался следующий сценарий моделирования для построения РЛП заданного объекта наблюдения: в зоне обзора РЛС находится аппарат Starlink размером 4 метра (по продольной дальности). В момент начала построения РЛП на аппаратуру РЛС МКП воздействуют деструктивные факторы и изменился реальный космический фон, которые изменяют точностные характеристики РЛС.

Решение уравнений (5) и (6) позволяет определить распределение плотности вероятности воздействия w_F , w_S в пространстве и времени. После перемножения плотности распределения на изменение разрешающей способности по дальности рассчитывается риск снижения точности определения координат:

$$\Delta(\delta r) = (\delta r_1 + \delta r_2)w_F w_S, \quad (18)$$

где $\delta r_1, \delta r_2$ – разрешающая способность по дальности для быстрых и медленных воздействий. При расчетах негативные факторы разделялись на внутренние и внешние, из-за чего $\gamma=0$. В таблице 1 представлены рассчитанные средние значения изменений разрешающей способности по дальности.

Таблица 1. Значение изменений разрешающей способности по дальности

Медленные составляющие, отн.ед	0,85
Быстрые составляющие, отн.ед	0,8
Весовой коэффициент влияния внутренних факторов α , отн.ед	0,51
Весовой коэффициент влияния внешних факторов β , отн.ед	0,49
Значение риска изменения разрешающей способности по дальности, м	0,75
Изменение разрешающей способности по дальности, м	до 1.5

Полученные значения применяются для расчета РЛП, где шаг расчета по дальности будет изменяться в зависимости от риск-анализа. Полученный идеальный РЛП моделируемого КА в момент отсутствия воздействия внутренних и внешних негативных факторов представлен на рисунке 5.

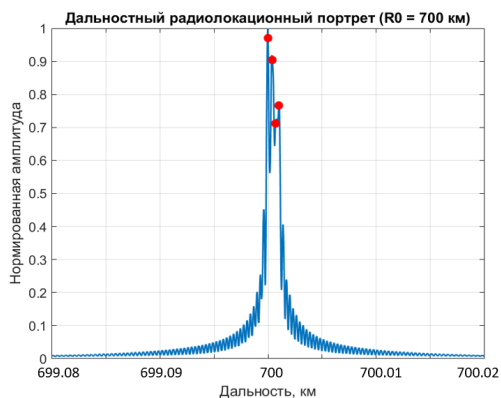


Рис. 5. Идеальный дальностный РЛП Starlink при наклонной дальности 700 км

Приведенный дальностный портрет соответствует минимальному уровню влияния внутренних и внешних негативных факторов на качество РЛП.

По этому распределению по формуле (3) по результатам применения метода риск-анализа проводилась оценка изменения точностных характеристик по дальности для дальнейшего построения РЛП с учетом воздействия негативных факторов, представленная на рисунке 6.

Изменение точности классификации на основе корреляционного анализа, представлено на рисунке 7. Точность классификации оценивалась через оценку коэффициента корреляции между полученным РЛП и эталонным.

Как видно из рисунка, точность классификации представляет собой стохастическое изменение коэффициента корреляции между искаженным и идеальным РЛП, что является следствием воздействия деструктивных факторов. Полученная зависимость характеризует временные интервалы, при которых значения коэффициента корреляции ниже 0,7, что требует ввода корректирующих поправок к аппаратуре или зондирующим сигналам. Классификация при этом происходит в непрерывном режиме с дискретом формирования РЛП.

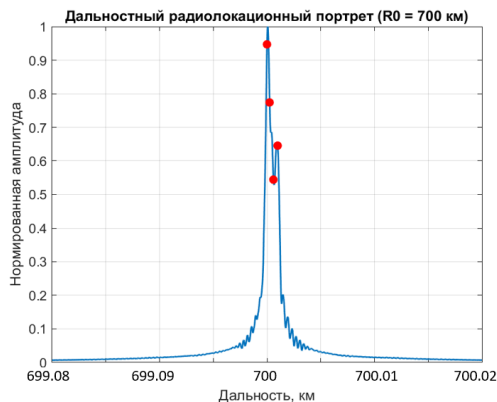


Рис. 6. Дальностный РЛП с искажениями



Рисунок 7. Изменение точности классификации

Ожидается, что для низкоорбитальных КА подобной формы результаты классификации будут иметь сходимость, поскольку РЛС L-диапазона частот позволяет провести косвенную оценку типа цели по дополнительным сигнальным признакам (дальностный портрет) путем оценки расстояний между блестящими точками после сжатия ЛЧМ-сигнала, отраженного от цели.

Проведем сравнение с известной моделью классификации, основанной на корреляционном анализе без учета деструктивных факторов. Сравнительный анализ представлен на рисунке 8.

Представленная модель позволяет повысить оперативность и точность классификации в диапазоне до 40% при построении дальностных портретов за счет учета влияния негативных факторов. Представленные результаты показывают важность учета деструктивных факторов, влияющих на РЛС МКП при построении РЛП.

Усовершенствованное описание уравнений Фоккера-Планка в сочетании с модифицированным уравнением Фейнмана-Каца для описания негативных факторов различной природы позволяет адаптировать параметры РЛС к различным стохастическим воздействиям на точностные характеристики при построении РЛП.

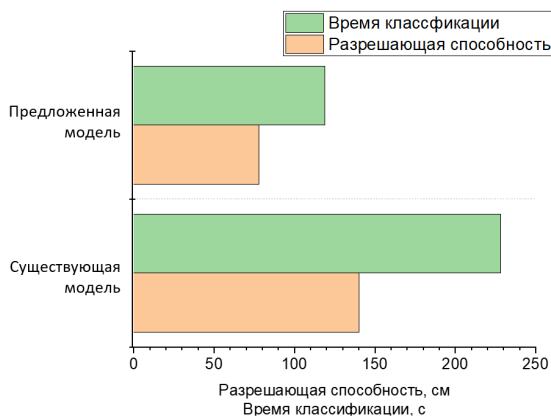


Рис. 8. Сравнительный анализ результатов учета воздействия деструктивных факторов

Связь между подходами выстроена иерархически, на уровне общей структуры. Например, диффузионные процессы хорошо отражают медленные, низкочастотные дрейфы характеристик аппаратуры. Уравнения с потенциалом, наоборот, подходят для моделирования быстрых, высокочастотных помех извне. Риск-анализ интегрирует их влияние в единую количественную меру, определяющую конечное требование к объему данных (числу РЛП).

8. Выводы. Предложенное в статье описание процесса классификации КА по результатам построения РЛП отражает наиболее существенные особенности функционирования крупноапертурных РЛС, а соответствующее математическое описание внешних и внутренних негативных факторов характеризуется стохастическими дифференциальными уравнениями с учетом скорости воздействия факторов на точностные характеристики.

Разработана и экспериментально проверена модель классификации подвижных объектов по данным радиолокационных портретов крупноапертурных РЛС L-диапазона, показывающая возможность повышения оперативности и точности классификации

в диапазоне до 40% по сравнению с моделью классификации без учета деструктивных факторов.

Разработанная модель позволяет определить достаточное число РЛП при классификации подвижных объектов с использованием разнородных математических аппаратов: стохастического дифференциального исчисления (через уравнение Фоккера-Планка), методов математической физики (через уравнение Фейнмана-Каца) и теории принятия решений в условиях неопределенности (через риск-анализ).

Приведена оригинальная формализация процесса информационного обеспечения классификации космических аппаратов по радиолокационным портретам РЛС L-диапазона с использованием двумерного неоднородного дифференциального уравнения для оценки функции плотности вероятности воздействия негативных факторов различной природы на точностные характеристики РЛС.

Формализована модель классификации космических аппаратов с учетом стохастических негативных факторов различной природы на основе уравнений Фоккера-Планка и Фейнмана-Каца, отличающаяся от известных тем, что с целью оценки влияния деструктивных факторов приведено разделение их по типам воздействия (быстрые и медленные) с дальнейшим описанием уравнений оценки. Полученные распределения являются основой для расчета риска обработки дополнительного РЛП, который потом пересчитывается в число необходимых РЛП и времени классификации.

Проведенные исследования по разработке стохастической модели влияния негативных факторов на эффективность классификации КА создают основы для совершенствования метода риск-анализа научной школы Ю.И. Шокина для решения задачи оптимизации процесса классификации подвижных объектов по РЛП в условиях неопределённости деструктивных воздействий.

Дальнейшие исследования связаны с численным решением предложенной модели для оценки вычислительной сложности предложенных уравнений при временных ограничениях на сопровождении КА.

Литература

1. Порсев В.И., Гелесев А.И., Красько А.Г. Угловое сверхразрешение сигналов с использованием «виртуальных» антенных решеток // Вестник Концерна ВКО «Алмаз – Антей». 2019. №4(31). С. 24–34. DOI: 10.38013/2542-0542-2019-4-24-34.
2. Власов С.А., Кочиев С.Г. Расчет орбитальной структуры систем космических аппаратов периодического обзора поверхности Земли в разных постановках // Труды Военно-космической академии имени А.Ф. Можайского. 2025. №697. С. 227–242.

3. Смотохин А.С., Хуторовский З.Н. Определение прогнозного положения ИСЗ при ограниченных ошибках измерений // Космические исследования. 2011. Т. 49. №6. С. 526–537.
4. Перлов А.Ю., Захаров А.С., Тимошенко А.В., Булатов М.Ф. Модель тепловых искажений амплитудно-частотной характеристики приемной системы радиолокационной станции в режиме сверхразрешения // Научные технологии в космических исследованиях Земли. 2024. Т. 16. №5. С. 26–34. DOI: 10.36724/2409-5419-2024-16-5-26-34.
5. Потапов А.А., Кузнецов В.А., Гончаров С.А. Имитационная модель формирования портретов сложных радиолокационных объектов со сниженной радиолокационной заметностью // Радиотехника и электроника. 2025. Т. 70. №6. С. 564–582. DOI: 10.7868/s3034590125060047.
6. Потапов А.А. Зарождение и становление фрактальной радиофизики и фрактальной радиоэлектроники в ИРЭ РАН // Известия высших учебных заведений. Прикладная нелинейная динамика. 2025. Т. 33. №5. С. 748–776. DOI: 10.18500/0869-6632-003183.
7. Григорян Д.С. Определение области сверхразрешения источников излучения по пространству при собственных шумах и неидентичностях приемных трактов // Радиотехника. 2007. №8. С. 43–48.
8. Филонович А.В., Ворначева И.В., Степанова В.В., Войнаш С.А., Соколова В.А., Загидуллин Р.Р., Сабитов Л.С. Методы получения двумерных и трехмерных радиолокационных портретов объектов // Известия Тульского государственного университета. Технические науки. 2023. №4. С. 228–233. DOI: 10.24412/2071-6168-2023-4-228-234.
9. Sharma A.K., Suraj N., Rokade M.V., et al. Study of variations in the absorption of cosmic radio noise using Riometer at Casey station // Library & Information Science Research. 2012. vol. 3. no. 2. pp. 149–155.
10. Загуменов А.В., Кузьмин В.А. Оценка ошибок определения ракурса подвижных протяженных надводных объектов на основе цифровой обработки радиолокационного дальностно-азимутального портрета // Современные проблемы радиоэлектроники и телекоммуникаций. 2023. №6. С. 139.
11. Захарченко В.Д., Коваленко И.Г., Акимов П.А., Денисов З.А. О возможности оценки геометрических характеристик астероида по радиолокационному портрету // Математическая физика и компьютерное моделирование. 2023. Т. 26. №4. С. 69–82. DOI: 10.15688/mpcm.jvolsu.2023.4.6.
12. Дурнаков А.А., Боков А.С., Кутькин А.Е. Оценка требований к разрядности преобразователей сигналов в задаче имитации радиолокационных портретов // СВЧ-техника и телекоммуникационные технологии. 2024. №6. С. 230–231.
13. Манукян А.А. Построение двумерных радиоизображений объектов по неэквидистантным кодированным частотно-временным последовательностям импульсов // Радиотехника и электроника. 2015. Т. 60. №3. С. 282. DOI: 10.7868/S0033849414120134.
14. Филонович А.В., Ворначева И.В., Степанова В.В., Артюхова В.И. К вопросу о получении двумерных портретов целей в многопозиционных радиолокационных системах // Известия Тульского государственного университета. Технические науки. 2023. №7. С. 553–558. DOI: 10.24412/2071-6168-2023-7-533-534.
15. Пафиков Е.А., Тимошенко А.В., Тычков А.Ю., Савчук А.М., Захаров А.С. Пространственно-временные характеристики сигналов радиолокационных станций с синтезированной апертурой с учетом отражения от целей и местных предметов // Известия высших учебных заведений. Авиационная техника. 2025. №1. С. 175–184.

16. Alqahtani A.M. Numerical simulation to study the pattern formation of reaction – diffusion Brusselator model arising in triple collision and enzymatic // *Journal of Mathematical Chemistry*. 2018. vol. 56. pp. 1543–1566. DOI: 10.1007/s10910-018-0859-8.
17. Bashkirtseva I., Pavletsov M., Perevalova T., Ryashko L. Analysis of Noise-Induced Transitions in a Thermo-Kinetic Model of the Autocatalytic Trigger // *Mathematics*. 2023. vol. 11. no. 20. 4302 p. DOI: 10.3390/math11204302.
18. Timoshenko A.V., Perlov A.Y., Zakharov A.S., Sazonov V.V. Model for Calculating Changes in the Radiation Pattern and Amplitude-Phase Distribution in the Subarray of a Large-Aperture APPA Based on a Modified Thermal Conductivity Equation // *Proceedings of the Systems of Signals Generating and Processing in the Field of on Board Communications*. 2024. pp. 1–4. DOI: 10.1109/ieeeeconf60226.2024.10496770.
19. Wang Y., Wang C., Lian P., et al. Effect of Temperature on Electromagnetic Performance of Active Phased Array Antenna // *Electronics*. 2020. no. 9(8). 1211 p. DOI: 10.3390/electronics9081211.
20. Xueming J., Manqing W. Jianmei T. Experimental study on a digital T/R module for phased array radar // *Proceedings of the CIE International Conference on Radar Proceedings*. 2001. pp. 898–902. DOI: 10.1109/ICR.2001.984856.
21. Mou C., Chen J., Peng H. Electromagnetic–Thermal Co-Simulation of Planar Monopole Antenna Based on HIE-FDTD Method // *Electronics*. 2022. no. 11(24). 4167 p. DOI: 10.3390/electronics11244167.
22. Liu H., Wang W., Tang D., et al. Thermal Deformation Modeling for Phased Array Antenna Compensation Control // *Sensors*. 2022. no. 22(6). 2325 p. DOI: 10.3390/s22062325.
23. Bozhko A.N., Livantsov V.E. Optimization of geometric analysis strategy in CAD-systems // *Computer Research and Modeling*. 2024. vol. 16. no. 4. pp. 825–840. DOI: 10.20537/2076-7633-2024-16-4-825-840.
24. Kim C., Choi H.-H., Lee K. Interference Coordination for Multi-UAV-Enabled Communications Under Probabilistic LoS Channels // *IEEE Internet of Things Journal*. 2025. vol. 12. no. 19. pp. 40484–40498. DOI: 10.1109/IJOT.2025.3589565.
25. Lee K., Park G., Chae, S.H. Joint Trajectory and Resource Optimization for Interference Management in Space-Air-Ground Integrated Networks Under Probabilistic LoS Channels // *IEEE Transactions on Communications*. 2026. vol. 74. pp. 5504–5518. DOI: 10.1109/TCOMM.2026.3664437.
26. Heo K., Lee W., Lee K. UAV-Assisted Wireless-Powered Secure Communications: Integration of Optimization and Deep Learning // *IEEE Transactions on Wireless Communications*. 2024. vol. 23. no. 9. pp. 10530–10545. DOI: 10.1109/TWC.2024.3372997.
27. Kim C., Choi H.-H., Lee K. Joint Optimization of Trajectory and Resource Allocation for Multi-UAV-Enabled Wireless-Powered Communication Networks // *IEEE Transactions on Communications*. 2024. vol. 72. no. 9. pp. 5752–5764. DOI: 10.1109/TCOMM.2024.3383113.
28. Zhu Q., Bai F., Pang M., et al. Geometry-Based Stochastic Line-of-Sight Probability Model for A2G Channels Under Urban Scenarios // *IEEE Transactions on Antennas and Propagation*. 2022. vol. 70. no. 7. pp. 5784–5794. DOI: 10.1109/TAP.2022.3161277.
29. Pan Y., Da X., Hu H., et al. Joint Optimization of Trajectory and Resource Allocation for Time-Constrained UAV-Enabled Cognitive Radio Networks // *IEEE Transactions on Vehicular Technology*. 2022. vol. 71. no. 5. pp. 5576–5580. DOI: 10.1109/TVT.2022.3151671.
30. Луковский М.А., Матвеев А.М. Двухэтапный метод калибровки АФАР // *Журнал радиоэлектроники*. 2022. №7. С. 1–23. DOI: 10.30898/1684-1719.2022.7.4
31. Луковский М.А., Матвеев А.М. Сравнительный анализ методов калибровки каналов сигналов антенн моноимпульсной радиолокации. *Журнал радиоэлектроники*. 2022. №6. С. 1–31. DOI: 10.30898/1684-1719.2020.6.5.

32. Корыстин А.А., Кулешов Ю.В., Новиков А.В. Имитационное моделирование прямо-передающего канала коротковолновой радиолокационной станции на основе модели ионосферного канала Уоттерсона // Труды Военно-космической академии имени А.Ф. Можайского. 2025. №697. С. 114–128.
33. Разиньков С.Н., Перлов А.Ю., Захаров А.С., Темник Я.А. Оперативная калибровка активных фазированных антенных решеток наземных радиолокационных станций // Воздушно-космические силы. Теория и практика. 2024. №29. С. 94–102.
34. Timoshenko A.V., Perlov A.Y., Chirov D.S., Zakharov A.S. Algorithm for assessing the readiness of a wideband radar for constructing radar portraits of space objects // T-Comm. 2025. vol. 19. no. 8. pp. 64–71.
35. Babkin Y.V., Zverev G.P., Timoshenko A.V., et al. Management of space surveillance radar temporal resource on fuzzy theory // Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2024. vol. 24. no. 3. pp. 513–519. DOI: 10.17586/2226-1494-2024-24-3-513-519.
36. Korpusov M.O., Perlov A.Y., Timoshenko A.V., Shafir R.S. On the Blow-Up of the Solution of a Nonlinear System of Equations of a Thermal-Electrical Model // Mathematical Notes. 2024. vol. 114. pp. 850–861. DOI: 10.1134/S0001434623110202.
37. ГОСТ Р 51901-2002. Менеджмент Риска. Анализ риска технологических систем // М.: Госстандарт России. 2002.
38. Лепихин А.М., Махутов Н.А., Шокин Ю.И., Юрченко А.В. Концепция риск-анализа технических систем с использованием цифровых двойников // Вычислительные технологии. 2020. Т. 25. №4. С. 99–113. DOI: 10.25743/ICT.2020.25.4.009.
39. Rockafellar R.T., Uryasev S. Conditional value-at-risk for general loss distributions // Journal of banking & finance. 2002. vol. 26. no. 7. pp. 1443–1471. DOI: 10.1016/S0378-4266(02)00271-6.
40. Иванюк В.А., Андропов К.Н., Егорова Н.Е. Классические и современные подходы к оценке и прогнозированию рисков активов // Фундаментальные исследования. 2016. №3-2. С. 380–384.
41. Курушина С.Е., Федорова Е.А., Гуровская Ю.А. Методика анализа шумоиндуцированных явлений в двухкомпонентных стохастических системах реакционно-диффузионного типа со степенной нелинейностью // Компьютерные исследования и моделирование. 2025. Т. 17. №2. С. 277–291. DOI: 10.20537/2076-7633-2025-17-2-277-291.
42. Орлов Ю.Н., Федоров С.Л. Генерация нестационарных траекторий временного ряда на основе уравнения Фоккера-Плanka // Труды Московского физико-технического института. 2016. Т. 8. №2(30). С. 126–133.
43. Орлов Ю.Н., Босов А.Д. Кинетико-гидродинамический подход к прогнозированию нестационарных временных рядов на основе уравнения Фоккера-Плanka // Труды Московского физико-технического института. 2012. Т. 4. №4(16). С. 134–140.
44. Захаров С.М., Митрохин М.А., Климов А.Е. Анализ возможности использования методов прогнозирования временных рядов для адаптации решающих правил средств обнаружения систем безопасности объектов // Известия высших учебных заведений. Поволжский регион. Технические науки. 2015. №2(34). С. 36–44.

Боев Сергей Федотович — д-р техн. наук, д-р эконом. наук, профессор, советник председателя, Публичное акционерное общество «Акционерная финансовая корпорация “Система”». Область научных интересов: создание и оценка эффективности сложных технических систем, риск-анализ, радиолокация. Число научных публикаций — 128. ris@tstu.tver.ru; ул. Моховая, 13, 125009, Москва, Россия; р.т.: +7(482)278-5261; факс: +7(482)278-5261.

Захаров Александр Сергеевич — канд. техн. наук, старший преподаватель, кафедра математики и анализа данных; младший научный сотрудник, институт цифровых технологий факультета информационных технологий и анализа больших данных, Федеральное государственное образовательное бюджетное учреждение высшего образования «Финансовый университет при Правительстве Российской Федерации» (Финуниверситет). Область научных интересов: математическое моделирование, радиотехника, термодинамика, теория вероятности. Число научных публикаций — 29. zakharov.as17@physics.msu.ru; 4-й Вешняковский проезд, 4, 109456, Москва, Россия; р.т.: +7(910)467-0087.

Перлов Анатолий Юрьевич — канд. техн. наук, доцент, ведущий научный сотрудник, институт цифровых технологий факультета информационных технологий и анализа больших данных, Федеральное государственное образовательное бюджетное учреждение высшего образования «Финансовый университет при Правительстве Российской Федерации» (Финуниверситет). Область научных интересов: радиолокация, прогнозирование временных рядов, цифровые двойники, применение методов машинного обучения, имитационное моделирование. Число научных публикаций — 99. laperlov@yandex.ru; ул. 8 марта, 10, 199178, Москва, Россия; р.т.: +7(915)335-1914.

Тимошенко Александр Васильевич — д-р техн. наук, профессор, кафедра информационной безопасности факультета информационных технологий и анализа больших данных, Федеральное государственное образовательное бюджетное учреждение высшего образования «Финансовый университет при Правительстве Российской Федерации» (Финуниверситет). Область научных интересов: онтологическое моделирование, риск-анализ, радиотехника, имитационное моделирование, проектирование крупномасштабных систем. Число научных публикаций — 179. u567ku78@gmail.com; 4-й Вешняковский проезд, 4, 109456, Москва, Россия; р.т.: +7(903)209-7116.

Поддержка исследований. Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финуниверситета.

S. BOEV, A. ZAKHAROV, A. PERLOV, A. TYMOSHENKO
**A CLASSIFICATION MODEL FOR MOVING OBJECTS BASED
ON RADAR PROFILE DATA FROM LARGE-APERTURE L-BAND
RADARS, ACCOUNTING FOR STOCHASTIC NEGATIVE
FACTORS OF VARIOUS NATURES**

Boev S., Zakharov A., Perlov A., Tymoshenko A. **A Classification Model for Moving Objects Based on Radar Profile Data from Large-Aperture L-Band Radars, Accounting for Stochastic Negative Factors of Various Natures.**

Abstract. The results of research on improving models and methods for rapid classification of spacecraft based on radar cross-section profile reconstruction are presented. A hypothesis is proposed that the minimum number of radar profiles required for correct classification is a functional of the original integral risk indicator, which accounts for the stochastic nature of adverse factors degrading the accuracy characteristics of the radar station. A classification model for moving objects based on radar profile data from large-aperture L-band radars is proposed, taking into account stochastic adverse factors of various natures. An original formalization of the information support process for spacecraft classification using radar profiles is presented, employing a two-dimensional inhomogeneous differential equation to estimate the probability density function of the impact of adverse factors of various natures on radar accuracy characteristics. The novelty of the model lies in the comprehensive consideration of heterogeneous factors affecting the spacecraft classification process, integration with risk analysis, and the hierarchical multi-level structure of the model for rapid classification purposes. It is shown that representing the spacecraft classification model based on refined risk analysis provides an analytical relationship for the number of radar profiles required, accounting for the impact of factors of various natures, which are formalized using refined Fokker–Planck equations (for internal factors) and the Feynman–Kac equation (for external factors, including real-space background dynamics). It is demonstrated that refining the risk analysis theory of the scientific school of Academician Yu.I. Shokin, informed by the solution of the system of equations, can significantly improve the information support for classification tasks. The results of a computational experiment on radar profile processing are presented, demonstrating the potential for enhancing the speed and accuracy of classification.

Keywords: radar station, mobile object, spacecraft, radar portraiture, classification, stochastic factors.

References

1. Porsev V.I., Gelesev A.I., Krasko A.G. [Angular Super-Resolution of Signals using "virtual" Antenna Arrays]. *Vestnik Koncerna VKO «Almaz – Antej» – Bulletin of the Almaz – Antey EKR Concern*. 2019. no. 4(31). pp. 24–34. DOI: 10.38013/2542-0542-2019-4-24-34. (In Russ.).
2. Vlasov S.A., Kochiev S.G. [Calculation of the Orbital Structure of Spacecraft Systems for Periodic Survey of the Earth's Surface in Different Settings]. *Trudy Voenno-kosmicheskoy akademii imeni A.F. Mozhajskogo – Proceedings of the Military Space Academy named after A.F. Mozhaisky*. 2025. no. 697. pp. 227–242. (In Russ.).
3. Samotokhin A.S., Khutorovsky Z.N. [Determination of the Projected Position of the Satellite with Limited Measurement Errors]. *Kosmicheskie issledovaniya – Cosmic Research*. 2011. vol. 49. no. 6. pp. 526–537. (In Russ.).

4. Perlov A.Y., Zakharov A.S., Timoshenko A.V., Bulatov M.F. [A Model of Thermal Distortion of the Amplitude-Frequency Response of the Radar Receiving System in the Super-Resolution Mode]. *Naukoemkie tekhnologii v kosmicheskix issledovaniyax Zemli – High Technologies in Earth Space Research*. 2024. vol. 16. no. 5. pp. 26–34. DOI: 10.36724/2409-5419-2024-16-5-26-34. (In Russ.).
5. Potapov A.A., Kuznetsov V.A., Goncharov S.A. [Simulation Model of Forming Portraits of Complex Radar Objects with Reduced Radar Visibility]. *Radiotekhnika I elektronika – Journal of Communications Technology and Electronics*. 2025. vol. 70. no. 6. pp. 564–582. DOI: 10.7868/s3034590125060047. (In Russ.).
6. Potapov A.A. [The Origin and Formation of Fractal Radiophysics and Fractal Radio Electronics at the IRE RAS]. *Izvestiya vysshix uchebnykh zavedenij. Prikladnaya nelinejnaya dinamika – News of Higher Educational Institutions. Applied Nonlinear Dynamics*. 2025. vol. 33. no. 5. pp. 748–776. DOI: 10.18500/0869-6632-003183. (In Russ.).
7. Grigoryan D.S. [Determination of the Area of Super-Resolution of Radiation Sources in Space with Intrinsic Noise and Non-Identity of the Receiving Paths]. *Radiotekhnika – Radio Engineering*. 2007. no. 8. pp. 43–48. (In Russ.).
8. Filonovich A.V., Vornacheva I.V., Stepanova V.V., Voynash S.A., Sokolova V.A., Zagidullin R.R., Sabitov L.S. [Methods for Obtaining Two-Dimensional and Three-Dimensional Radar Portraits of Objects]. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Texnicheskie nauki – Proceedings of Tula State University. Technical sciences*. 2023. no. 4. pp. 228–233. DOI: 10.24412/2071-6168-2023-4-228-234. (In Russ.).
9. Sharma A.K., Suraj N., Rokade M.V., et al. Study of variations in the absorption of cosmic radio noise using Riometer at Casey station. *Library & Information Science Research*. 2012. vol. 3. no. 2. pp. 149–155.
10. Zagumenov A.V., Kuzmin V.A. [Estimation of Errors in Determining the Angle of Mobile Extended Surface Objects Based on Digital Processing of a Radar Range-Azimuth Portrait]. *Sovremennyye problemy radioelektroniki i telekommunikacij – Modern Problems of Radio Electronics and Telecommunications*. 2023. no. 6. 139 p. (In Russ.).
11. Zakharchenko V.D., Kovalenko I.G., Akimov P.A., Denisov Z.A. [On the Possibility of Estimating the Geometric Characteristics of an Asteroid from a Radar Portrait]. *Matematicheskaya fizika i komp'yuternoe modelirovanie – Mathematical Physics and Computer Simulation*. 2023. vol. 26. no. 4. pp. 69–82. DOI: 10.15688/mpcm.jvolsu.2023.4.6. (In Russ.).
12. Durnakov A.A., Bokov A.S., Kutkin A.E. [Assessment of the Requirements for the Bit Depth of Signal Converters in the Task of Simulating Radar Portraits]. *SVCh-tehnika i telekommunikacionnyye tekhnologii – Microwave and Telecommunication Technology*. 2024. no. 6. pp. 230–231. (In Russ.).
13. Manukyan A.A. [Construction of Two-Dimensional Radio Images of Objects Based on Nonequidistant Encoded Time-Frequency Pulse Sequences]. *Radiotekhnika I elektronika – Journal of Communications Technology and Electronics*. 2015. vol. 60. no. 3. 282 p. DOI: 10.7868/S0033849414120134. (In Russ.).
14. Filonovich A.V., Vornacheva I.V., Stepanova V.V., Artyukhova V.I. [On the Issue of Obtaining Two-Dimensional Portraits of Targets in Multi-Position Radar Systems]. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Texnicheskie nauki – Proceedings of Tula State University. Technical sciences*. 2023. no. 7. pp. 553–558. DOI: 10.24412/2071-6168-2023-7-533-534. (In Russ.).
15. Papikov E.A., Timoshenko A.V., Tychkov A.Y., Savchuk A.M., Zakharov A.S. [Spatial and Temporal Characteristics of Signals from Synthesized Aperture Radar Stations Taking into Account Reflection from Targets and Local Objects]. *Izvestiya*

- vy'sshix uchebny'x zavedenij. Aviacionnaya texnika – News of Higher Educational Institutions. Aviation Equipment. 2025. no. 1. pp. 175–184. (In Russ.).
16. Alqahtani A.M. Numerical simulation to study the pattern formation of reaction – diffusion Brusselator model arising in triple collision and enzymatic. *Journal of Mathematical Chemistry*. 2018. vol. 56. pp. 1543–1566. DOI: 10.1007/s10910-018-0859-8.
17. Bashkirtseva I., Pavletsov M., Perevalova T., Ryashko L. Analysis of Noise-Induced Transitions in a Thermo-Kinetic Model of the Autocatalytic Trigger. *Mathematics*. 2023. vol. 11. no. 20. 4302 p. DOI: 10.3390/math11204302.
18. Timoshenko A.V., Perlov A.Y., Zakharov A.S., Sazonov V.V. Model for Calculating Changes in the Radiation Pattern and Amplitude-Phase Distribution in the Subarray of a Large-Aperture APPA Based on a Modified Thermal Conductivity Equation. *Proceedings of the Systems of Signals Generating and Processing in the Field of on Board Communications*. 2024. pp. 1–4. DOI: 10.1109/ieeeconf60226.2024.10496770.
19. Wang Y., Wang C., Lian P., et al. Effect of Temperature on Electromagnetic Performance of Active Phased Array Antenna. *Electronics*. 2020. no. 9(8). 1211 p. DOI: 10.3390/electronics9081211.
20. Xueming J., Manqing W. Jianmei T. Experimental study on a digital T/R module for phased array radar. *Proceedings of the CIE International Conference on Radar Proceedings*. 2001. pp. 898–902. DOI: 10.1109/ICR.2001.984856.
21. Mou C., Chen J., Peng H. Electromagnetic-Thermal Co-Simulation of Planar Monopole Antenna Based on HIE-FDTD Method. *Electronics*. 2022. no. 11(24). 4167 p. DOI: 10.3390/electronics11244167.
22. Liu H., Wang W., Tang D., et al. Thermal Deformation Modeling for Phased Array Antenna Compensation Control. *Sensors*. 2022. no. 22(6). 2325 p. DOI: 10.3390/s22062325.
23. Bozhko A.N., Livantsov V.E. Optimization of geometric analysis strategy in CAD-systems. *Computer Research and Modeling*. 2024. vol. 16. no. 4. pp. 825–840. DOI: 10.20537/2076-7633-2024-16-4-825-840.
24. Kim C., Choi H.-H., Lee K. Interference Coordination for Multi-UAV-Enabled Communications Under Probabilistic LoS Channels. *IEEE Internet of Things Journal*. 2025. vol. 12. no. 19. pp. 40484–40498. DOI: 10.1109/IJOT.2025.3589565.
25. Lee K., Park G., Chae, S.H. Joint Trajectory and Resource Optimization for Interference Management in Space-Air-Ground Integrated Networks Under Probabilistic LoS Channels. *IEEE Transactions on Communications*. 2026. vol. 74. pp. 5504–5518. DOI: 10.1109/TCOMM.2026.3664437.
26. Heo K., Lee W., Lee K. UAV-Assisted Wireless-Powered Secure Communications: Integration of Optimization and Deep Learning. *IEEE Transactions on Wireless Communications*. 2024. vol. 23. no. 9. pp. 10530–10545. DOI: 10.1109/TWC.2024.3372997.
27. Kim C., Choi H.-H., Lee K. Joint Optimization of Trajectory and Resource Allocation for Multi-UAV-Enabled Wireless-Powered Communication Networks. *IEEE Transactions on Communications*. 2024. vol. 72. no. 9. pp. 5752–5764. DOI: 10.1109/TCOMM.2024.3383113.
28. Zhu Q., Bai F., Pang M., et al. Geometry-Based Stochastic Line-of-Sight Probability Model for A2G Channels Under Urban Scenarios. *IEEE Transactions on Antennas and Propagation*. 2022. vol. 70. no. 7. pp. 5784–5794. DOI: 10.1109/TAP.2022.3161277.
29. Pan Y., Da X., Hu H., et al. Joint Optimization of Trajectory and Resource Allocation for Time-Constrained UAV-Enabled Cognitive Radio Networks. *IEEE Transactions on Vehicular Technology*. 2022. vol. 71. no. 5. pp. 5576–5580. DOI: 10.1109/TVT.2022.3151671.
30. Lukovsky M.A., Matveev A.M. [Two-Stage AFAR Calibration Method]. *Zhurnal radioelektroniki – Journal of Radio Electronics*. 2022. no. 7. pp. 1–23. DOI: 10.30898/1684-1719.2022.7.4. (In Russ.).

31. Lukovsky M.A., Matveev A.M. [Comparative Analysis of Methods for Calibrating Signal Channels of Monopulse Radar Antennas]. *Zhurnal radioelektroniki – Journal of Radio Electronics*. 2022. no. 6. pp. 1–31. DOI: 10.30898/1684-1719.2020.6.5. (In Russ.).
32. Korystin A.A., Kuleshov Y.V., Novikov A.V. [Simulation of the Receiving and Transmitting Channel of a Shortwave Radar Station Based on the Watterson Ionospheric Channel Model]. *Trudy Voenno-kosmicheskoy akademii imeni A.F. Mozhajskogo – Proceedings of the Military Space Academy named after A.F. Mozhaisky*. 2025. no. 697. pp. 114–128. (In Russ.).
33. Razinkov S.N., Perlov A.Y., Zakharov A.S., Temnik Y.A. [Operational Calibration of Active Phased Antenna Arrays of Ground-Based Radar Stations]. *Vozdushno-kosmicheskie sily. Teoriya i praktika – Aerospace Forces. Theory and Practice*. 2024. no. 29. pp. 94–102. (In Russ.).
34. Timoshenko A.V., Perlov A.Y., Chirov D.S., Zakharov A.S. Algorithm for assessing the readiness of a wideband radar for constructing radar portraits of space objects. *T-Comm*. 2025. vol. 19. no. 8. pp. 64–71.
35. Babkin Y.V., Zverev G.P., Timoshenko A.V., et al. Management of space surveillance radar temporal resource on fuzzy theory. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2024. vol. 24. no. 3. pp. 513–519. DOI: 10.17586/2226-1494-2024-24-3-513-519.
36. Korpusov M.O., Perlov A.Y., Timoshenko A.V., Shafir R.S. On the Blow-Up of the Solution of a Nonlinear System of Equations of a Thermal-Electrical Model. *Mathematical Notes*. 2024. vol. 114. pp. 850–861. DOI: 10.1134/S0001434623110202.
37. GOST R 51901-2002. [Risk Management. Risk Analysis of Technological Systems]. M.: Gosstandart Rossii. 2002. (In Russ.).
38. Lepikhin A.M., Makhutov N.A., Shokin Y.I., Yurchenko A.V. [The Concept of Risk Analysis of Technical Systems using Digital Twins]. *Vychislitel'nye tekhnologii – Computational Technologies*. 2020. vol. 25. no. 4. pp. 99–113. DOI: 10.25743/ICT.2020.25.4.009. (In Russ.).
39. Rockafellar R.T., Uryasev S. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*. 2002. vol. 26. no. 7. pp. 1443–1471. DOI: 10.1016/S0378-4266(02)00271-6.
40. Ivanyuk V.A., Andropov K.N., Egorova N.E. [Classical and Modern Approaches to Asset Risk Assessment and Forecasting]. *Fundamental'ny'e issledovaniya – Fundamental Research*. 2016. no. 3-2. pp. 380–384. (In Russ.).
41. Kurushina S.E., Fedorova E.A., Gurovskaya Y.A. [A Method for Analyzing Noise-Induced Phenomena in Two-Component Stochastic Systems of the Reaction-Diffusion Type with Power-Law Nonlinearity]. *Komp'yuternye issledovaniya i modelirovanie – Computer Research and Modeling*. 2025. vol. 17. no. 2. pp. 277–291. DOI: 10.20537/2076-7633-2025-17-2-277-291. (In Russ.).
42. Orlov Y.N., Fedorov S.L. [Generation of Non-Stationary Time Series Trajectories Based on the Fokker-Planck Equation]. *Trudy Moskovskogo fiziko-texnicheskogo instituta – Proceedings of Moscow Institute of Physics and Technology*. 2016. vol. 8. no. 2(30). pp. 126–133. (In Russ.).
43. Orlov Y.N., Bosov A.D. [Kinetic-Hydrodynamic Approach to Forecasting Nonstationary Time Series Based on the Fokker-Planck Equation]. *Trudy Moskovskogo fiziko-texnicheskogo instituta – Proceedings of Moscow Institute of Physics and Technology*. 2012. vol. 4. no. 4(16). pp. 134–140. (In Russ.).
44. Zakharov S.M., Mitrokhin M.A., Klimov A.E. [Analysis of the Possibility of using Time Series Forecasting Methods to Adapt Crucial Rules for Detecting Facility Security Systems]. *Izvestiya vysshix uchebnykh zavedenij. Povolzhskij region. Texnicheskie nauki – University Proceedings. Volga Region. Technical Sciences*. 2015. no. 2(34). pp. 36–44. (In Russ.).

Boev Sergei — Ph.D., Professor, Advisor to the Chairman, Public Joint Stock Financial Corporation “Sistema“. Research interests: creation and evaluation of the effectiveness of complex technical systems, risk analysis, radar. The number of publications — 128. ris@tsu.tver.ru; 13, Mokhovaya St., 125009, Moscow, Russia; office phone: +7(482)278-5261; fax: +7(482)278-5261.

Zakharov Alexander — Ph.D., Senior Lecturer, Department of Mathematics and Data Analysis; Junior Research Assistant, Institute of Digital Technologies, Faculty of Information Technologies and Big Data Analysis, Federal State Educational Budgetary Institution of Higher Education "Financial University under the Government of the Russian Federation" (Financial University). Research interests: mathematical modeling, radio engineering, thermodynamics, probability theory. The number of publications — 29. zakharov.as17@physics.msu.ru; 4, 4th Veshnyakovsky passage, 109456, Moscow, Russia; office phone: +7(910)467-0087.

Perlov Anatoliy — Ph.D., Associate Professor, Leading Researcher, Institute of Digital Technologies, Faculty of Information Technologies and Big Data Analysis, Federal State Educational Budgetary Institution of Higher Education "Financial University under the Government of the Russian Federation" (Financial University). Research interests: radar, time series forecasting, digital twins, application of machine learning methods, simulation modeling. The number of publications — 99. laperlov@yandex.ru; 10, 8 Marta st., 199178, Moscow, Russia; office phone: +7(915)335-1914.

Tymoshenko Alexander — Ph.D., Professor, Department of Information Security of the Faculty of Information Technology and Big Data Analysis, Federal State Educational Budgetary Institution of Higher Education "Financial University under the Government of the Russian Federation" (Financial University). Research interests: ontological modeling, risk analysis, radio engineering, simulation modeling, design of large-scale systems. The number of publications — 179. u567ku78@gmail.com; 4, 4th Veshnyakovsky passage, 109456, Moscow, Russia; office phone: +7(903)209-7116.

Acknowledgements. The article was prepared based on the results of research carried out using budgetary funds under the state assignment of the Financial University.

С.В. ДВОРНИКОВ, Д. В. ВАСИЛЬЕВА
**РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ РАЗЛИВОВ
НЕФТЕПРОДУКТОВ НА ОСНОВЕ КРАТНОМАСШТАБНОЙ
ОБРАБОТКИ ГИСТОГРАММ РАСПРЕДЕЛЕНИЯ ИХ ЯРКОСТИ**

Дворников С.В., Васильева Д. В. Распознавание изображений разливов нефтепродуктов на основе кратномасштабной обработки гистограмм распределения их яркости.

Аннотация. Особую актуальность проблема обнаружения аномалий на обрабатываемых изображениях, получаемых с аппаратуры визуального контроля, приобретает в ходе экологического мониторинга акваторий мирового океана при поиске последствий техногенных катастроф, вызванных разливами нефтепродуктов. Под аномалиями в представленном исследовании рассматриваются фрагменты изображения, характеризующие разливы нефтепродуктов. В качестве предмета исследования определены элементы теории распознавания образов, в соответствии с которыми обрабатываемые изображения представляются в виде векторов признаков. Обоснован выбор гистограмм, характеризующих распределения битов яркости обрабатываемых изображений, которые используются в качестве первичных векторов признаков. Установлено, что высокая вариативность фона изображений водных поверхностей приводит к значительной дисперсии векторов признаков, формируемых на основе гистограмм распределения яркости даже для одного общего класса объектов распознавания. Поэтому предложено подвергать первичные векторы признаков декоррелирующей обработке на основе вейвлет-преобразований. Рассмотрены виды вейвлет-преобразования и различные типы формирующих функций, среди которых обоснован выбор вейвлета Гаусса 1-го порядка, позволившего обеспечить наименьшие различия между векторами признаков обрабатываемых изображений, характеризующих один и тот же класс. Предложен подход к формированию вторичных признаков распознавания путем усреднения компонентов матриц кратномасштабного преобразования гистограмм распределения яркости изображений вдоль оси их параметра сдвига. Обоснован переход к выбору избыточного числа масштабов декомпозиции в вейвлет-преобразованиях при формировании вторичных векторов признаков. Демонстрируются результаты проведенного эксперимента, подтверждающие правомерность такого перехода. Представлены тестовые и эталонные изображения, используемые в ходе эксперимента, а также формируемые на их основе первичные и вторичные векторы признаков. Приведены численные результаты, характеризующие полученную оценку различий векторов признаков как одного класса, так и альтернативных классов. Представлена вероятностная оценка результатов распознавания аномалий на обрабатываемых изображениях. Обосновано использование коэффициента оценки различий векторов признаков в разработанной формуле вероятности правильного обнаружения аномалий в обрабатываемых изображениях. Приведены графики, характеризующие значение вероятностной оценки в зависимости от соотношения размеров аномалии к общей площади кадра для различных значений коэффициента, повышающего различия векторов признаков.

Ключевые слова: аномалии изображений, разливы нефтепродуктов, векторы признаков, распознаваемые классы изображений, гистограммы распределения яркости изображений.

1. Введение. Повышение эффективности экологического мониторинга является одной из глобальных задач, решаемых

международной научной общественностью. Для сбора и констатации фактов, характеризующих последствия загрязнения окружающей среды, службы экологического контроля активно используют системы видеомониторинга. Вместе с тем возрастающий поток видеoinформации существенно усложняет процедуру ее обработки, о чем свидетельствуют результаты исследований, представленных в [1–3], что стимулирует к поиску новых подходов к ее структуризации.

При этом особую актуальность данная проблема приобретает в ходе поиске абстрактных объектов в базах данных обрабатываемых изображений [4, 5]. Примером таких объектов являются результаты разливов нефтепродуктов в акваториях морей и океанов, поскольку они не имеют четкой структуры, позволяющей осуществить их единую формализацию, для последующей обработки на основе методов машинного обучения [6, 7]. Кроме того, сложность своевременного обнаружения результатов такого рода техногенных катастроф усугубляется необходимостью мониторинга огромных площадей контролируемых акваторий [8, 9].

Именно поэтому поиск научных подходов, позволяющих решать задачи обработки изображений в интересах своевременного обнаружения последствий техногенных катастроф, в том числе разливов нефтепродуктов, непосредственно по мере их поступления с аппаратуры фото и видео фиксации, устанавливаемой на летно-подъемных платформах, является актуальной задачей.

Вместе с тем удаленность летно-подъемных платформ от пунктов экологического контроля с телекоммуникационным оборудованием предполагает необходимость организации высокопроизводительных каналов передачи информации [10, 11]. Но это не всегда возможно в силу различных обстоятельств. Поэтому рациональным решением в такой ситуации является предварительная селекция информационных данных непосредственно на борту летно-подъемных платформ, что позволило бы существенно снизить объем передаваемого канального трафика, предоставляемого операторам для принятия решения о результатах обнаружения разлива нефтепродуктов в контролируемых акваториях [12]. Реализация такого подхода видится возможной только при условии наличия высокоэффективных алгоритмов обработки изображений, обеспечивающих автоматическое обнаружение последствий разливов нефтепродуктов. Однако проведенный анализ подходов к организации экологического мониторинга, рассмотренных в [6, 7, 9–11] показал, что они, преимущественно, используют обработку изображений в отложенном режиме на основе методов машинного обучения.

Указанные обстоятельства определяют необходимость поиска и разработки новых методов и способов, позволяющих успешно разрешить рассмотренные проблемы.

Так, в представленной статье демонстрируются возможности предлагаемого метода вероятностной оценки, характеризующего результаты поиска последствий техногенных катастроф, вызванных разливами нефтепродуктов в акваториях морей и океанов, по данным обработки фото и видеоматериалов.

Основной целью исследования является разработка аналитического аппарата, обеспечивающего вероятность правильного обнаружения разливов нефтепродуктов, представляющих собой аномалии обрабатываемых кадров изображений, на уровне 0,93-0,98.

2. Формализация обрабатываемых изображений и формирование первичных векторов признаков. Основополагающим фактором, определяющим достоверность обнаружения объектов на обрабатываемых изображениях, является форма представления исходных данных. Поэтому важным этапом является выбор подхода к их формализации, позволяющей, с одной стороны, получить компактное описание, а с другой стороны, обеспечить существенные различия между изображениями, относящимися к альтернативным классам. В данном разделе представлен подход к формализации обрабатываемых изображений для формирования их первичных векторов признаков.

В [8, 13] обоснована целесообразность представления данных кадров обрабатываемых изображений в виде гистограмм распределения их яркости. Такой выбор, прежде всего, обусловлен не только компактностью представления исходных данных, но и достаточно высокой их чувствительностью к возникновению аномалий [14, 15]. Под аномалиями в рамках проведенного исследования будем понимать фрагмент изображения, вызванный последствием разливов нефтепродуктов в контролируемой акватории [8].

В качестве примера на рисунке 1 демонстрируются изображения формата *.jpeg с разрешением 640×480 , на которых представлены фрагменты акватории Черного моря, без разливов нефтепродуктов (Рис. 1(а)) и с последствиями техногенной катастрофы (Рис. 1(б)).

На рисунке 2 изображены, соответственно, матрицы амплитудных значений изображений, представленных на рисунке 1.

Очевидно, что такой формат представления данных достаточно громоздкий для его обработки, требующий существенных вычислительных ресурсов при формировании векторов признаков, имеющих $307 \cdot 200$ элементов. Переход к векторному представлению

за счет усреднения амплитудных значений матриц вдоль одной из осей ведет к снижению информативности такого представления, что является нежелательным моментом [16, 17].

Вместе с тем в [18] обосновано представление изображений для их последующей обработки в виде векторов распределения яркости.

Учитывая, что при 8-битовом кодировании яркость представляется всего лишь 256 значениями, то вектор будет иметь такую же размерность. На рисунке 3 представлены гистограммы распределения яркости для ТИ-1 и ТИ-2 (далее по тексту – гистограммы) в одинаковом масштабе.

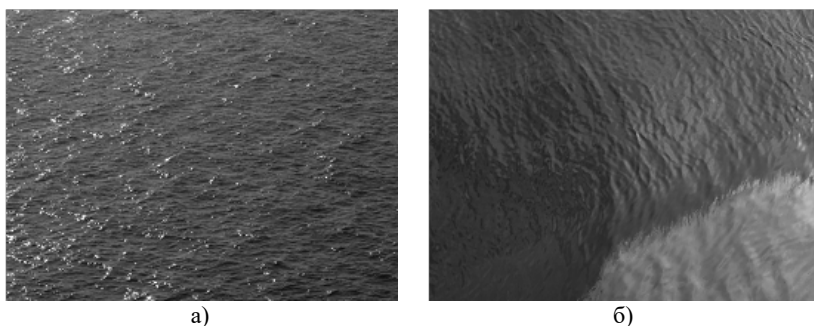


Рис. 1. Тестовое изображение: а) ТИ-1; б) ТИ-2

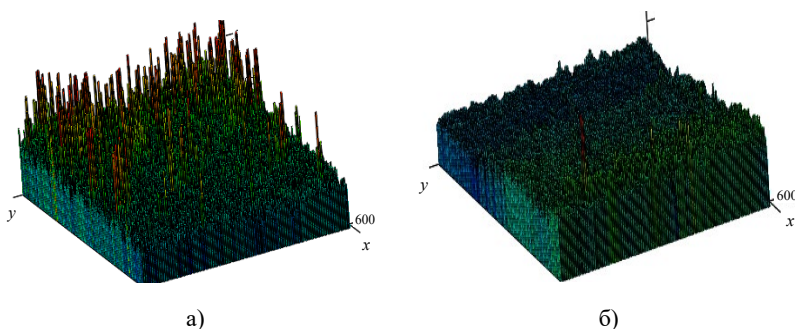


Рис. 2. Матрица амплитудных значений тестового изображения:
а) ТИ-1; б) ТИ-2

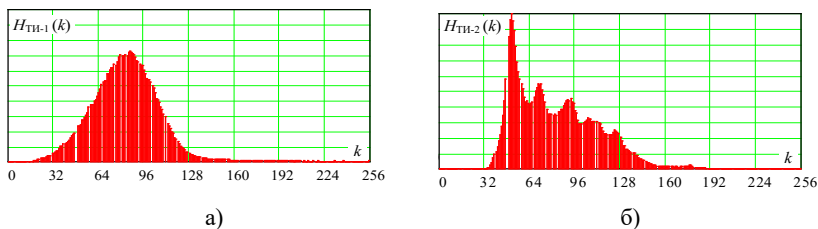


Рис. 3. Гистограмма тестового изображения: а) ТИ-1; б) ТИ-2

Согласно результатам исследования, представленным в [6, 8, 17], векторы признаков, сформированные на основе гистограмм, обеспечивают наилучшие различия по отношению к другим форматам представления амплитудных значений матрицы обрабатываемых изображений. Однако высокая вариативность водной поверхности даже в условиях отсутствия разливов нефтепродуктов приводит к тому, что характеризующие их изображения векторов признаков на основе гистограмм распределения яркости имеют весьма высокую дисперсию, существенно усложняющую получения высоких результатов достоверности выявления аномалий [6].

В частности, для эталонного изображения ТИ-1 (Рис. 1(а)) с водной поверхностью, не загрязненной разливами нефтепродуктов, были подготовлены кадры с различной освещенностью при незначительном изменении характера волнения. Всего для эксперимента было подготовлено 200 кадров, из которых для исследования были отобраны тестовые изображения ТИ-11, ТИ-12, ТИ-13 и ТИ-14 (Рис. 4). И соответствующие им векторы признаков $H_{\text{ТИ-11}}$, $H_{\text{ТИ-12}}$, $H_{\text{ТИ-13}}$, $H_{\text{ТИ-14}}$ (Рис. 5).

Тестовые изображения отбирались следующим образом. Подготовленные 200 кадров первоначально были разделены на четыре группы, в зависимости от величины разброса компонентов их гистограмм яркости. Затем из каждой группы были отобраны изображения, имеющие наибольшие разбросы компонент яркости в пределах каждой из групп. Таковыми явились тестовые изображения ТИ-11, ТИ-12, ТИ-13 и ТИ-14. Последующий выбор изображения ТИ-1 в качестве эталонного обусловлен тем, что оно характеризует фрагмент водной акватории вблизи разливов нефтепродуктов.

Сопоставительный анализ тестовых изображений и соответствующих им гистограмм позволяет сделать следующее заключение: формирование устойчивых векторов признаков, характеризующих один и тот же класс распознаваемых объектов (водную акваторию без разливов нефтепродуктов), связано

с определенными сложностями, обусловленными значительным разбросом амплитудных значений их гистограмм.

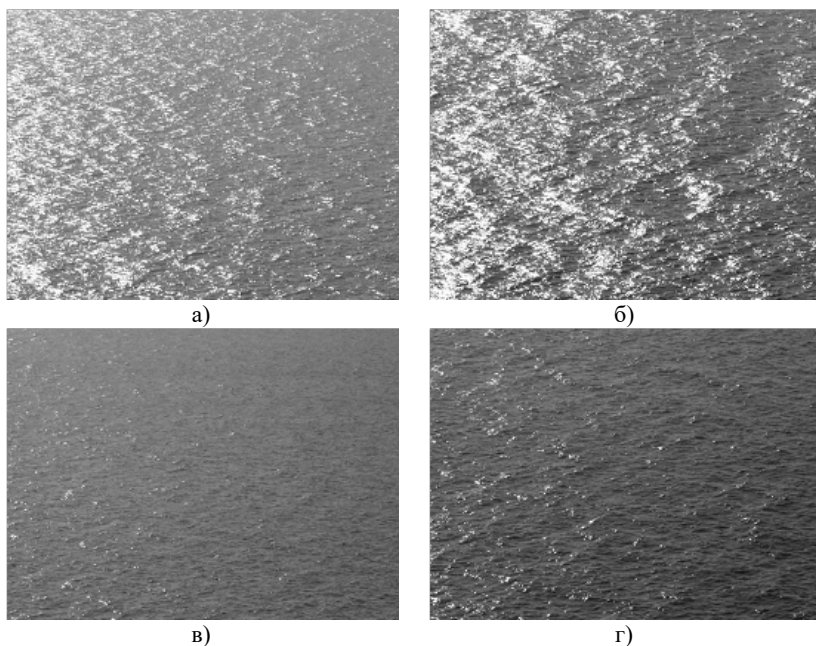


Рис. 4. Тестовое изображение: а) ТИ-11; б) ТИ-12; в) ТИ-13; г) ТИ-14

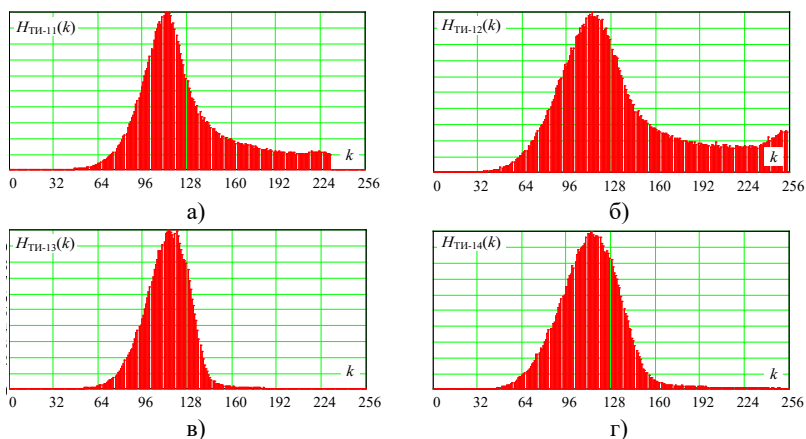


Рис. 5. Гистограмма тестового изображения:
а) ТИ-11; б) ТИ-12; в) ТИ-13; г) ТИ-14

Заметим, что изменение освещенности ведет к увеличению амплитудных значений слева и справа от основного тренда (смещение в сторону нуля происходит при затемнении, а в сторону 255 – при повышении освещенности). Следовательно, любое изменение освещенности объекта по отношению к исходному изображению ведет к увеличению разброса компонентов яркости.

В качестве примера оценим вариативность гистограмм $H_{\text{ТИ-11}}$, $H_{\text{ТИ-12}}$, $H_{\text{ТИ-13}}$, $H_{\text{ТИ-14}}$, характеризующих тестовое изображение ТИ-1 в других условиях освещенности. Для этого построим функцию различий в соответствии с выражением:

$$\Delta r_{0,N}(k) = |H_{\text{ТИ-1}}(k) - H_{\text{ТИ-N}}(k)|, \quad (1)$$

где N – текущий номер гистограммы обрабатываемого изображения, сравниваемый с гистограммой изображения ТИ-1.

На рисунке 6 представлены графики функций различий $\Delta r_{0,N}(k)$, для гистограмм тестовых изображений ТИ-11, ТИ-12, ТИ-13 и ТИ-14.

На рисунке 6 введены следующие обозначения:

$\Delta r_{0,1}(k)$ – функция различий между изображением ТИ-1 и ТИ-11;

$\Delta r_{0,2}(k)$ – функция различий между изображением ТИ-1 и ТИ-12;

$\Delta r_{0,3}(k)$ – функция различий между изображением ТИ-1 и ТИ-13;

$\Delta r_{0,4}(k)$ – функция различий между изображением ТИ-1 и ТИ-14;

$\Delta r_{0,0}(k)$ – усредненная функция различий.

Следует отметить, что все рассчитанные функции различий равны нулю в тех позициях, где элементы гистограмм как эталонного, так и текущего описания равны нулю, а также в позиции совпадения их максимальных значений. На рисунке 6 представлена и усредненная функция различий $\Delta r_{0,0}(k)$. Ее положение как раз и позволяет оценить параметры вариативности. Для оценки усредненной функции различий рассмотренных изображений дополнительно были рассчитаны функции, характеризующие ее границы сверху и снизу (Рис. 7).

В таблице 1 приведены результаты оценки различий гистограмм тестовых изображения, полученных на основе выражений:

$$\Delta r = \overline{|H_{\text{ТИ-1}} - H_{\text{ТИ-N}}|}, \quad (2)$$

$$\Delta r [\%] = \overline{|H_{\text{ТИ-1}} - H_{\text{ТИ-N}}|} \times \left(\sum_{k=0}^{255} H_{\text{ТИ-1}}(k) \right)^{-1} \times 100\%. \quad (3)$$

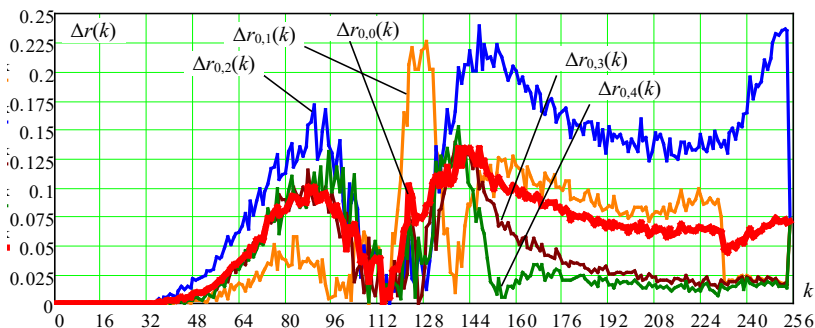


Рис. 6. Графики функций различий гистограмм тестовых изображений по отношению к эталонному описанию ТИ-1



Рис. 7. График усредненной функции различий

В формулах (2) и (3) представлена операция поэлементного вычитания векторов $H_{\text{ТИ-1}} - H_{\text{ТИ-N}}$.

Величина $H_{\text{ТИ-1}} - H_{\text{ТИ-2}}$, характеризующая различия между эталонным изображением ТИ-1 и изображением с разливами нефтепродуктов ТИ-2, рассчитанная в соответствии с выражением (3), составляет 7,6%.

Таблица 1. Оценка различий гистограмм тестовых изображений

Различия гистограмм	$H_{\text{ТИ-1}} - H'_{\text{ТИ-11}}$	$H_{\text{ТИ-1}} - H_{\text{ТИ-12}}$	$H_{\text{ТИ-1}} - H_{\text{ТИ-13}}$	$H_{\text{ТИ-1}} - H_{\text{ТИ-14}}$
Δr	1,187	2,031	0,815	0,782
$\Delta r, \%$	2,5	4,3	1,7	1,6

Отметим, что для рассматриваемых тестовых изображений величина разброса функции различий, согласно таблице 1, достигает 4%. Очевидно, что при более сильном освещении (затемнении) изображения на кадре по отношению к исходному, рассмотренный эффект усилится.

Вместе с тем, согласно [6], даже при величине разброса функции различий порядка 4% потенциальная вероятность правильного обнаружения не превысит значения 0,96. Для усредненной функции различий гистограмм среднее значение Δr будет равно 1,204. А общая величина разброса достигает значения 0,338.

Тогда, если для обнаружения разливов нефтепродуктов использовать ранговый классификатор, то аналитический алгоритм принятия решения можно представить в следующем виде:

$$A_0 \underset{A_1}{\overset{A_2}{>}} 0, \quad (4)$$

где,

$$A_0 = (d_{01} - d_{02}).$$

Здесь: d_{01} – нормированная абсолютная величина различий между вектором эталонного описания $H_{\text{ТИ-1}}$, принадлежащего классу A_1 , соответствующего изображению с отсутствием разливов нефтепродуктов, и одного из векторов обрабатываемых изображений $H_{\text{ТИ-N}}$ этого же класса A_1 ; M – размерность векторов; d_{02} – нормированная абсолютная величина различий между вектором эталонного описания $H_{\text{ТИ-1}}$, принадлежащего классу A_1 , и одного из векторов обрабатываемых изображений $H_{\text{ТИ-N}}$, принадлежащих классу A_2 , т.е. к классу, который характеризует изображения с разливами нефтепродуктов.

Расчет значений d_{01} и d_{02} осуществляют в соответствии с одним и тем же выражением:

$$d_{01(02)} = \frac{1}{M} \sum_{k=1}^M |H_{\text{ТИ-1}}(k) - H_{\text{ТИ-N}}(k)|. \quad (5)$$

Рассмотренный подход к формализации изображений в виде гистограмм распределения яркости позволяет формировать векторы

признаков, которые удобно использовать в ранговых классификаторах. Однако такие вектора признаков весьма чувствительны к изменению освещенности объектов, что затрудняет получение устойчивого эталонного описания, и как результат – высокой достоверности распознавания.

Следовательно, для повышения достоверности необходим поиск подходов к снижению вариативности векторов признаков, характеризующих один и тот же класс распознаваемых изображений.

3. Предложение по снижению вариативности векторов признаков на основе кратномасштабной обработки гистограмм.

Одной из причин высокой вариативности гистограмм является изначальная избыточность тестовых изображений, обусловленная достаточно сильной пространственной корреляцией их элементов [15, 19]. Поэтому предлагается к формируемым на их основе гистограммам распределения яркости применить декоррелирующие преобразования.

Согласно [20], основной задачей декоррелирующих преобразований является снижение объема представляемых для обработки данных при сохранении их информационного наполнения.

Среди декоррелирующих преобразований широкое распространение получили следующие: сингулярное преобразование [21]; дискретное преобразование Карунена-Лоэва [22]; дискретное косинусное преобразование [23]; и дискретное вейвлет-преобразование [24]. Вместе с тем анализ результатов исследования [25–29] показал, что для изображений с высокой динамикой изменения фона указанные декоррелирующие преобразования проигрывают ДВП по компактности представления, ошибки аппроксимации и помехоустойчивости. При этом, согласно результатам исследований, представленным в [30], ДВП является наиболее эффективным декоррелирующим преобразованием, обеспечивающим наилучшие различия между векторами признаков, характеризующим альтернативные классы при минимизации их размерности.

В общем случае дискретное вейвлет-преобразование представляет собой результат декомпозиции обрабатываемого процесса в базисах вейвлет-функций (англ. *wavelet*) [31].

Свойство масштабируемости вейвлет-функций открывает возможность локализовать различные по характеру изменения исследуемого процесса в соответствующих проекциях масштаба, что позволяет рассматривать их при одинаковом разрешении, независимо от реальных размеров. В общем случае реализация ДВП $W(k, n)$ представляет собой свертку входной реализации

обрабатываемого процесса $x(t)$ и формирующего материнского вейвлета $\psi(t)$ [32]:

$$W(k, n) = \frac{1}{\sqrt{2^k}} \int_0^{N-1} x(t) \psi\left(\frac{t-2^k n}{2^k}\right) dt, \quad (6)$$

где n – переменная сдвига; k – переменная масштаба; N – величина временного интервала обработки.

Таким образом, в соответствии с (6), ДВП представляет собой результат декомпозиции входного процесса путем его последовательной ВЧ и НЧ-фильтрации с временным сдвигом n и кратномасштабным изменением амплитудно-частотной характеристики НЧ-фильтра $g(2n - t)$ и ВЧ-фильтра $h(2n - t)$.

При этом результат ВЧ-фильтрации $h(2n - k)$ позволяет получить детализирующие значения, определяемые как вейвлет-коэффициенты $y_{ВЧ}(n)$, а НЧ-фильтрация $g(2n - t)$ ведет к формированию аппроксимирующих коэффициентов $y_{НЧ}(n)$:

$$y_{НЧ}(n) = \sum_{t=0}^{N-1} x(t) g(2n - t), y_{ВЧ}(n) = \sum_{t=0}^{N-1} x(t) h(2n - t). \quad (7)$$

Фильтры, описываемые посредством выражений (10), получили название квадратурных зеркальных фильтров [33]. Последовательная реализация процедур, в соответствии с выражениями (10), ведет к формированию двумерной матрицы вейвлет-коэффициентов, которая представляет собой дискретное вейвлет-преобразование [34].

С представленных позиций ДВП можно рассматривать как распределение энергии (в рассматриваемом случае амплитудных значений гистограмм) в пространстве их масштабных преобразований и временных сдвигов, реализуемых на основе формирующих (порождающих) материнских вейвлетов.

Вместе с тем, анализ выражения (6) показывает, что ДВП обеспечивает кратномасштабный сдвиг, пропорциональный множителю 2^k , в пределах которого как раз и происходит усреднение обрабатываемых значений вдоль оси n . Для исключения потери информативности первичных векторов признаков в результате их усреднения на каждом последующем сдвиге 2^k , без потери общности воспользуемся непрерывной формой кратномасштабного вейвлет-преобразования, определяемого следующим выражением [31]:

$$W(k, n) = \frac{1}{\sqrt{k}} \int_0^{N-1} x(t) \psi\left(\frac{t-n}{k}\right) dt. \quad (8)$$

Отметим, что кратномасштабное преобразование, описываемое выражением (8), также позволяет формировать двумерные матрицы вейвлет-коэффициентов, структура которых будет определяться типом формирующей вейвлет-функции $\psi(*)$.

Свойства вейвлет-преобразований во многом определяются выбором исходной порождающей функции. Это обусловлено тем, что по своей сущности матрица кратномасштабного преобразования представляет собой результат корреляции обрабатываемых данных с различными формами материнского вейвлета. При этом формы порождающей функции формируют в результате изменения коэффициента частотного масштаба и временного сдвига. Поэтому необходимо среди множества видов материнских вейвлетов выбрать тот, который позволит формировать кратномасштабные преобразования векторов признаков обрабатываемых изображений, наиболее соответствующие решаемым задачам. Простейшей формирующей функцией (далее по тексту – материнский вейвлет), используемой для кратномасштабного преобразования, является вейвлет Хаара, описываемый следующим выражением [35]:

$$\psi_X(t) = \begin{cases} 1, & 0 \leq t < 0,5; \\ -1, & 0,5 \leq t < 1; \\ 0, & t \notin [0, 1). \end{cases} \quad (9)$$

А масштабирующая функция для вейвлета Хаара (далее по тексту – отцовский вейвлет) определяется формулой:

$$\varphi_X(t) = \begin{cases} 1, & 0 \leq t < 1; \\ 0, & t \notin [0, 1). \end{cases} \quad (10)$$

Здесь и далее нижний индекс $\psi(*)$ (t) определяет первую букву названия функции вейвлета.

В соответствии с аналитической формой представления формирующей функции вейвлета Хаара $\psi_X(t)$ (9), элементы входной реализации группируются по два. А затем для каждой группы вычисляется их сумма и разность. Указанная последовательность процедур реализуется рекурсивно, с образованием нового уровня масштабного преобразования в пространстве временных сдвигов [32].

В результате выполнения указанных процедур получают 2^{k-1} уровней декомпозиций и одну общую сумму в виде ошибки (остатка) аппроксимации.

В ДВП активное применение находят вейвлеты Добеши. Это семейство ортогональных функций, не имеющие аналитического описания, коэффициенты которых рассчитывают на основе итерационных алгоритмов. Представляют их следующим образом [36]:

$$\psi_d(n) = \sqrt{2} \sum_{t=0}^{N-1} g_t \varphi(2n - t), \quad (11)$$

$$\varphi_d(n) = \sqrt{2} \sum_{t=0}^{N-1} h_t \varphi(2n - t). \quad (12)$$

Расчет отцовского вейвлета $\varphi(n)$ происходит при использовании вычисленных значений ортогональных нормированных коэффициентов, получивших название коэффициенты Добеши низких порядков [36]. В частности, на рисунке 9 представлены формирующий и масштабирующий вейвлеты Добеши $\psi_d(n)$ и $\varphi_d(n)$ 4-го порядка.

Поскольку масштабирующая функция (отцовский вейвлет) служит для построения кратномасштабного базиса декомпозиции с возможностью последующего восстановления анализируемого процесса, что не является предметом данного исследования, то в дальнейшем будем рассматривать лишь формирующие материнские функции основных видов вейвлетов.

Для дискретных данных достаточно часто используют, так называемые, дуальные (двойные) вейвлеты, синтезируемые на основе передаточных характеристик низкочастотных и высокочастотных фильтров, или R -функций, связанных следующим условием [37]:

$$\psi_{Дв}(n) = \varphi_{Дв}(n) + z\varphi_{Дв}(2n), \quad (13)$$

где z – некоторое комплексное число.

В обработке изображений широкое применение находят биортогональные вейвлеты (B -сплайновые (англ. *B-spline*) вейвлеты) [38]. Главная особенность биортогональных вейвлетов заключается в том, что формируемый на их основе кратномасштабный базис не является ортогональным [37]. Однако применение биортогональных вейвлетов допускает больше возможных степеней свободы для построения симметричных вейвлет-функций.

В преобразованиях на основе биортогональных вейвлетов применяют две формирующие ψ_B , ψ'_B и две масштабирующие φ_B , φ'_B функции.

Аналитически это можно представить следующим образом:

$$\varphi_B(n) = \sum_{t=0}^{N-1} p_t \varphi(2n - t), \quad \varphi'_B(n) = \sum_{t=0}^{N-1} p'_t \varphi'(2n - t), \quad (14)$$

$$\begin{aligned} \psi_B(n) &= \sum_{t=0}^{N-1} (-1)^{t+1} p_{-t+1} \varphi(2n - t), \\ \psi'_B(n) &= \sum_{t=0}^{N-1} (-1)^{t+1} p_{-t+1} \varphi'(2n - t). \end{aligned} \quad (15)$$

В (17) и (18) коэффициенты p_t и p'_t являются расчетными, значения которых представлены в [39]. Наиболее широкое применение в кратномасштабном анализе получили вейвлеты Гаусса различных порядков [26, 28, 30, 31, 40]. В аналитическом виде вейвлет Гаусса n -го порядка можно представить на основе следующего выражения:

$$\psi_{\Gamma n}(t) = (-1)^n \frac{d^n}{dt^n} \exp(-t^2/2), \quad (16)$$

где n – порядок дифференцирования.

Согласно выражению (16) повышение порядка дифференцирования ведет к увеличению количества осцилляций функции на длительности ее элемента. Поэтому учитывая, что гистограммы представляют собой векторы дискретных значений, применение вейвлет-функций выше 2-го порядка не целесообразно.

С учетом такого ограничения, аналитические выражения, описывающие вейвлеты Гаусса 1-го и 2-го порядка, будут иметь вид:

$$\psi_{\Gamma 1}(t) = -t \exp(-t^2/2), \quad (17)$$

$$\psi_{\Gamma 2}(t) = (1 - t^2) \exp(-t^2/2). \quad (18)$$

4. Результаты эксперимента. В интересах выбора вида формирующей вейвлет-функции для дальнейшего исследования были отобраны два изображения ТИ-1 и ТИ-12, которые характеризуют один и тот же альтернативный класс, но при этом ТИ-12, согласно таблице 1, имеет наибольшие различия с ТИ-1, по отношению к другим рассматриваемым изображениям.

На рисунках 8 и 9 представлены матрицы кратномасштабного преобразования (КМП) гистограмм тестовых изображений ТИ-1 и ТИ-12, рассчитанные на основе алгоритма, реализующего выражение (6) для рассмотренных выше формирующих вейвлет-функций.

На рисунках 8 и 9 (и далее) верхний левый индекс в обозначении матрицы указывает на тип формирующего вейвлета: Х – вейвлет Хаара; Г1 – вейвлет Гаусса 1-го порядка; Г2 – вейвлет Гаусса 2-го порядка; Д – вейвлет Добеши 4-го порядка; Б – биортогональный вейвлет; Дв – дуальный (двойной) вейвлет.

Матрицы кратномасштабного преобразования, представленные на рисунках 8 и 9, построены с 4-х кратным прореживанием по оси времени, для лучшей наглядности их структуры. Дальнейшие расчеты будут представлены для полноразмерных матриц.

Проведенный визуальный анализ сформированных матриц указывает на то, что кратномасштабное преобразование тестового изображения ТИ-12 имеет ярко выраженные коэффициенты в интервале значений параметра $n \in [245; 255]$ в отличие от матриц эталонного изображения ТИ-1. Именно эти коэффициенты как раз и приводят к высокой дисперсии различий гистограмм ТИ-1 и ТИ-12.

Формируемые матрицы имеют две координаты по оси k и оси n , что открывается возможность построения вторичных векторов признаков путем их усреднения или вдоль оси пространственного сдвига n или вдоль оси масштабирования k . На рисунке 10 представлены вторичные векторы признаков, формируемые путем усреднения вдоль оси пространственного сдвига n для матрицы изображения ТИ-1.

На рисунке 11 демонстрируются вторичные векторы признаков, формируемые путем усреднения вдоль оси масштабирования k для матрицы изображения ТИ-1.

Поскольку для исследования значение имеет не сам вид матрицы преобразования в соответствии с выбранным видом вейвлет-функции, а возможность снизить разброс формируемых на ее основе векторов признаков, описывающих изображения, характеризующих один и тот же распознаваемый класс, то для сравнительной оценки аналогичным образом были построены вторичные векторы признаков по тестовому изображению ТИ-12.

Важно отметить, что причиной выбора для исследования гистограммы тестового изображения ТИ-12 послужило то, что она имеет наибольшие структурные отличия от гистограммы эталонного изображения ТИ-1.

В таблице 2 приведены значения оценки различий гистограмм, полученных на основе выражения (2), для эталонного изображения ТИ-1 и тестового изображения ТИ-12, сформированных в различных базах вейвлет-функций.

В последнем столбце приведены значения разности вторичных векторов признаков между эталонным изображением и изображением ТИ-2, характеризующим результаты разливов нефтепродуктов.

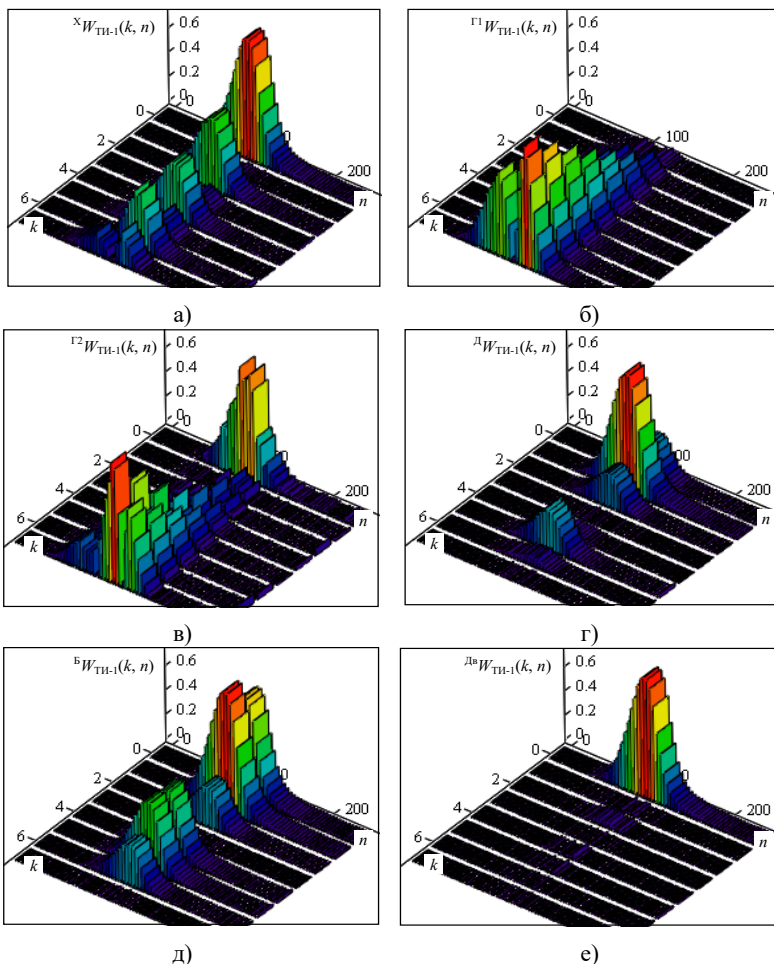


Рис. 8. Трехмерная матрица КМП гистограммы изображения ТИ-1 на основе: а) вейвлета Хаара; б) вейвлета Гаусса 1-го порядка; в) вейвлета Гаусса 2-го порядка; г) вейвлета Добеши 4-го порядка; д) биортогонального вейвлета; е) дуального вейвлета

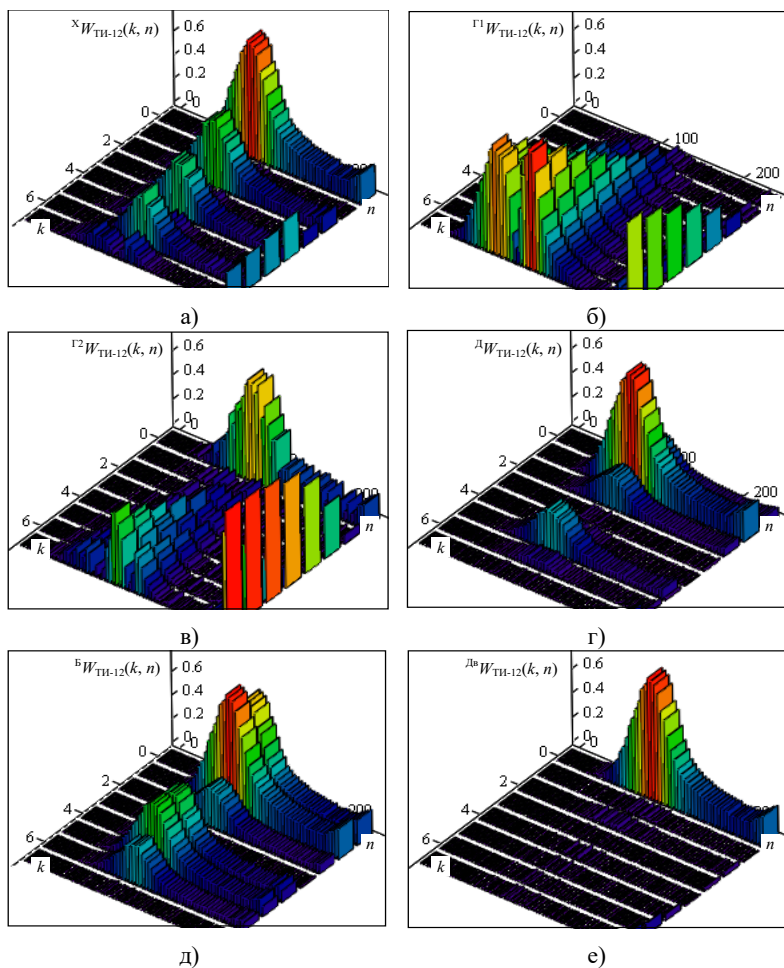


Рис. 9. Трехмерная матрица КМП гистограммы изображения ТИ-12 на основе: а) вейвлета Хаара; б) вейвлета Гаусса 1-го порядка; в) вейвлета Гаусса 2-го порядка; г) вейвлета Добеши 4-го порядка; д) биортогонального вейвлета; е) дуального вейвлета

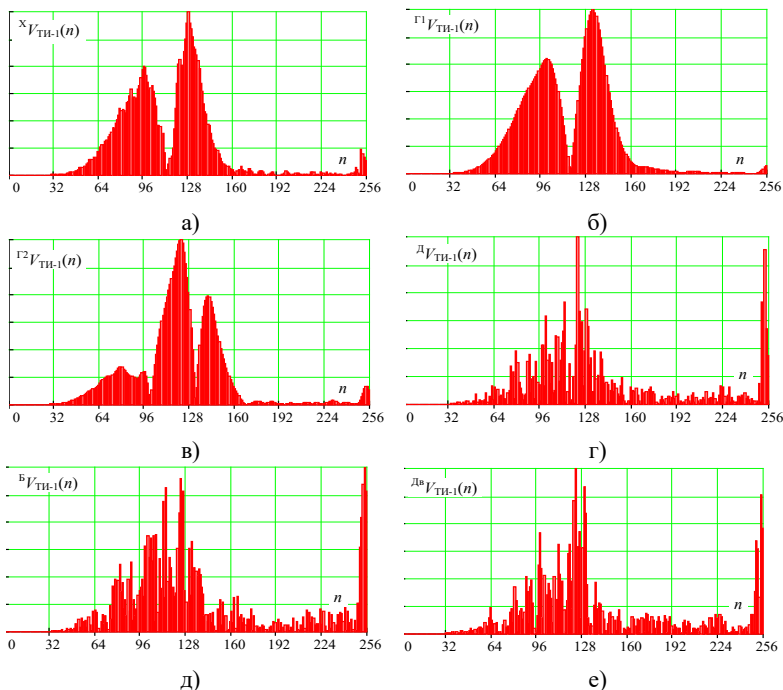


Рис. 10. Вектор средних значений матрицы КМП на основе:
 а) вейвлета Хаара изображения ТИ-1 (по оси n);
 б) вейвлета Гаусса 1-го порядка изображения ТИ-1 (по оси n);
 в) вейвлета Гаусса 2-го изображения ТИ-1 (по оси n);
 г) вейвлета Добеши 4-го порядка изображения ТИ-1 (по оси n);
 д) биортогонального вейвлета ТИ-1 (по оси n);
 е) дуального вейвлета изображения ТИ-1 (по оси n)

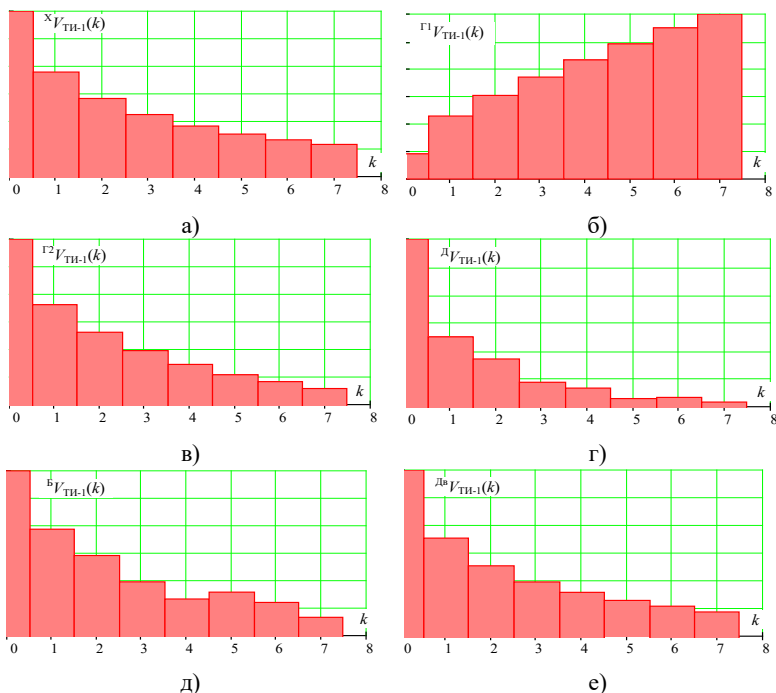


Рис. 11. Вектор средних значений матрицы КМП на основе:

- а) вейвлета Хаара изображения ТИ-1 (по оси k);
- б) вейвлета Гаусса 1-го порядка изображения ТИ-1 (по оси k);
- в) вейвлета Гаусса 2-го изображения ТИ-1 (по оси k);
- г) вейвлета Добеши 4-го порядка изображения ТИ-1 (по оси k);
- д) биортогонального вейвлета ТИ-1 (по оси k);
- е) дуального вейвлета изображения ТИ-1 (по оси k)

Таблица 2. Разность векторов признаков изображений ТИ-1 и ТИ-12, полученных усреднением матрицы КМП гистограммы по оси n

Матрицы КМП	$X_{V_{TI-1}} - X_{V_{TI-12}}$	$\Gamma^1 V_{TI-1} - \Gamma^1 V_{TI-12}$	$\Gamma^2 V_{TI-1} - \Gamma^2 V_{TI-12}$	$D^1 V_{TI-1} - D^1 V_{TI-12}$	$B^1 V_{TI-1} - B^1 V_{TI-12}$	$D^2 V_{TI-1} - D^2 V_{TI-12}$	$\Gamma^1 V_{TI-1} - \Gamma^1 V_{TI-2}$
Δr	2,586	2,972	3,179	2,365	2,963	2,704	4,533
$\Delta r, \%$	6,3	6,0	8,0	9,9	8,5	9,8	9,5

Представленные расчеты проведены в соответствии с выражениями (2) и (3). В таблице 3 представлены значения разности между векторами эталонного ТИ-1 и тестового ТИ-12 изображений, сформированных в различных базисах вейвлет-функций, но усредненных по оси масштабирования k .

Таблица 3. Разность векторов признаков изображений ТИ-1 и ТИ-12, полученных усреднением матрицы КМП гистограммы по оси k

Матрицы КМП	${}^X V_{\text{ТИ-1}} - {}^X V_{\text{ТИ-12}}$	${}^{\Gamma^1} V_{\text{ТИ-1}} - {}^{\Gamma^1} V_{\text{ТИ-12}}$	${}^{\Gamma^2} V_{\text{ТИ-1}} - {}^{\Gamma^2} V_{\text{ТИ-12}}$	${}^D V_{\text{ТИ-1}} - {}^D V_{\text{ТИ-12}}$	${}^B V_{\text{ТИ-1}} - {}^B V_{\text{ТИ-12}}$	${}^{D^B} V_{\text{ТИ-1}} - {}^{D^B} V_{\text{ТИ-12}}$	${}^{\Gamma^1} V_{\text{ТИ-1}} - {}^{\Gamma^1} V_{\text{ТИ-2}}$
Δr	1,649	0,204	0,33	1,423	1,575	1,506	0,536
$\Delta r, \%$	37,7	4,0	8,2	28,2	37,4	36,6	10,5

Анализ полученных результатов показывает, что наилучший результат обеспечивает кратномасштабная обработка гистограмм в базисе вейвлета Гаусса 1-го порядка, так как он обеспечил наименьшие различия между эталонным изображением ТИ-1 и тестовым изображением ТИ-12. Следует отметить, что ТИ-12 было отобрано как изображение, имеющее наибольшие различия в структуре гистограммы по отношению к гистограмме эталона ТИ-1 среди 200 отобранных тестовых изображений.

Переход к оценке, отражающей разницу векторов признаков, выраженную в процентах по отношению к эталонному значению, обусловлен различием амплитудных значений коэффициентов, сформированных в различных функциональных базисах.

5. Обоснование размерности матрицы кратномасштабного преобразования для построения векторов признаков. Проведенный дополнительный анализ показал, что декоррелирующая обработка гистограмм, наряду с уменьшением абсолютных различий между векторами изображений ТИ-1 и ТИ-12 в пределах одного альтернативного класса, привела к снижению различий между векторами различных классов, что является негативным моментом.

Следовательно, необходим поиск научных решений, позволяющих не только снизить дисперсию векторов признаков изображений, принадлежащих одному классу, но и повысить общие различия признакового пространства, под которым в [41] определены абсолютные наименьшие различия между векторами признаков, описывающих любые альтернативные классы.

В интересах реализации данной концепции предложено изменить масштабирующий множитель 2^k в выражения (8). Это обусловлено тем, что в рассматриваемой ситуации выбор значения масштаба не является принципиальным моментом. Так как размерность гистограммы в общем случае может быть любой, то предлагается в выражении (8) вейвлет $\psi(*)$ масштабировать произвольным образом. В качестве примера на рисунке 12 представлены матрицы КМП эталонного изображения ТИ-1,

Таблица 5. Оценка различий векторов признаков
тестовых изображений для $k = 32$

Матрицы КМП	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-21}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-22}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-23}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-24}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-25}}$
Δr	1,376	1,601	0,84	0,795	5,408
$\Delta r, \%$	1,7	2,0	1,1	0,9	6,8

Таблица 6. Оценка различий векторов признаков
тестовых изображений для $k = 64$

Матрицы КМП	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-21}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-22}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-23}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-24}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-25}}$
Δr	1,609	1,878	0,651	0,486	6,041
$\Delta r, \%$	1,3	1,5	0,5	0,4	4,9

Таблица 7. Оценка различий векторов признаков
тестовых изображений для $k = 128$

Матрицы КМП	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-21}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-22}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-23}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-24}}$	$\Gamma^1 V_{\text{ТИ-20}} - \Gamma^1 V_{\text{ТИ-25}}$
Δr	1,618	2,37	0,591	0,526	4,927
$\Delta r, \%$	0,9	1,3	0,33	0,29	2,7

Значения $\Delta r, \%$ в таблицах 4-7 рассчитаны в соответствии с выражением:

$$\Delta r = \frac{|\Gamma^1 V_{\text{ТИ-1}} - \Gamma^1 V_{\text{ТИ-**}}|}{\sum_{n=0}^{N-1} \Gamma^1 V_{\text{ТИ-1}}(n)} \times 100\%, \quad (19)$$

где $\Gamma^1 V_{\text{ТИ-**}}$ – вторичный вектор признаков текущего тестового изображения, сравниваемого с аналогичным вектором эталонного описания ТИ-1; $|\Gamma^1 V_{\text{ТИ-1}} - \Gamma^1 V_{\text{ТИ-**}}|$ – модуль разницы векторов.

Выражение (19) является модификацией формулы (3) при переходе от гистограмм тестовых изображения к вторичным векторам признаков изображений. Для окончательного выбора размерности матрицы кратномасштабного преобразования, необходимой для формирования векторов признаков, воспользуемся формулой:

$$\delta_k = \frac{\Delta r_{\text{ТИ-1,ТИ-2}}}{\Delta r_{\text{ТИ-1,ТИ-12}}}, \quad (20)$$

где $\Delta r_{\text{ТИ-1,ТИ-2}}$ и $\Delta r_{\text{ТИ-1,ТИ-12}}$ – расчетные значения из таблиц 4-7.

Результаты расчетов значения δ при различной величине k , представлены в таблице 8. Полученный результат характеризует различия векторов признаков изображений, используемых при обнаружении аномалий для рассматриваемых условий.

Таблица 8. Оценка различий векторов признаков для различных k

Размерность вектора	Для $k = 8$	Для $k = 16$	Для $k = 32$	Для $k = 64$	Для $k = 128$
δ_k	1,77	1,93	3,4	3,27	2,08

Согласно результатам, представленным в таблице 8, наилучшие различия между вторичными векторами признаков для рассматриваемых изображений обеспечивают векторы, полученные усреднением по оси временных сдвигов матриц КМП размерностью 256×32 , полученных на основе вейвлетов Гаусса 1-го порядка.

И хотя полученный результат в общем случае применим для используемых в исследовании изображений, он позволяет сделать следующее заключение: применение процедур кратномасштабной обработки позволяет повысить различие векторов признаков.

Для оценки эффективности представленного подхода предложено ввести коэффициент повышения различий (КПР) $K_{\text{ПК}}$, значение которого определяется следующей формулой:

$$K_{\text{ПК}} = \frac{d_{H_3, H_T}}{d_{V_3, V_T}}, \quad (21)$$

где d_{H_3, H_T} – разность значений элементов гистограмм эталонного и текущего изображений до реализации процедур КМП; d_{V_3, V_T} – разность, полученная после реализации указанных процедур.

Результаты расчета КПР для $k = 32$ дают значение $K_{\text{ПК}} = 3,4$.

Тогда формула оценки вероятности правильного обнаружения аномалий с учетом коэффициента оценки различий, полученная в [6], будет иметь вид:

$$P_{\text{по}} = \begin{cases} 0,5 \left[1 + \Phi \left(\frac{\delta_A K_{\text{ПК}} - \sigma_p^2}{\sigma_p^2} \right) \right], & \frac{\delta_A K_{\text{ПК}}}{\sigma_p^2} \geq \sigma_p^2 \\ 0,5 \left[1 - \Phi \left(\frac{\sigma_p^2 - \delta_A K_{\text{ПК}}}{\sigma_p^2} \right) \right], & \frac{\delta_A K_{\text{ПК}}}{\sigma_p^2} < \sigma_p^2 \end{cases}, \quad (22)$$

где δ_A – коэффициент, характеризующий относительные размеры аномалии к общей площади кадра обрабатываемого изображения,

$\delta_A \in [0; 1]$; σ_p^2 – дисперсия различий между векторами признаков эталонного изображения и текущего; σ_z^2 – дисперсия вектора признаков изображения, рассматриваемого в качестве эталонного.

В интересах оценки влияния значения КПР на величину вероятности правильного обнаружения, на рисунке 13 представлены графики $P_{по}(\delta_A)$ для значений $K_{ПК} = 1; 1,5; 3,4$.

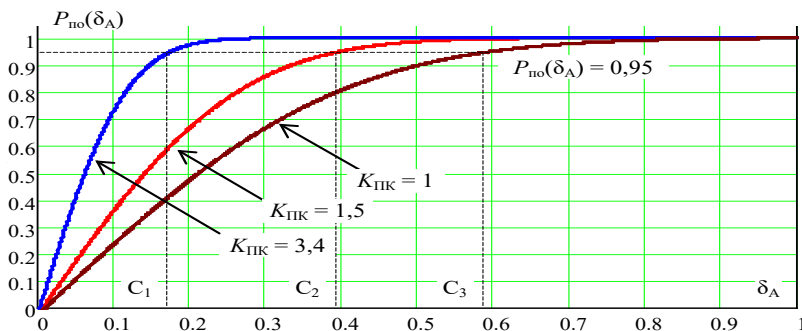


Рис. 13. Графики вероятности правильного обнаружения аномалии изображения при различных значениях $K_{ПК}$

Графики, представленные на рисунке 13 получены для условия $\sigma_p^2 = 0,3$; пунктиром показан уровень, соответствующий вероятности правильного распознавания равно $P_{по} = 0,95$.

Следует отметить, что при $K_{ПК} = 1,5$ заданное значение вероятности $P_{по} = 0,95$ будет обеспечено при снижении размера аномалии в 1,5 по отношению к условиям, при которых $K_{ПК} = 1$ (точки C_3 и C_2).

В тоже время повышение различий до $K_{ПК} = 3,4$, позволит обнаруживать аномалии в 3,3 раза меньшего размера с тем же значением вероятности. Результат подтвержден по результатам обработки 252 фрагментов различных изображений при значении вероятности правильного обнаружения $P_{по} = 0,93-0,98$.

6. Анализ возможностей методов машинного обучения в решении задач обнаружения разливов нефтепродуктов.

Альтернативным подходом к решению задачи обнаружения разливов нефтепродуктов является применение методов машинного обучения.

Анализ работ [42–45], посвященных разработке и исследованию возможностей глубоких сверточных нейронных сетей CNNs, позволил обобщить сведения об их эффективности и производительности применительно к задачам обнаружения разливов нефтепродуктов, которые сведены в таблице 9.

Таблица 9. Параметры эффективности и производительность нейронных сетей

Тип сети	AlexNet	VGG-16	GoogLeNet	ResNet	SENet
Число операций	724×10^6	$15,5 \times 10^9$	$1,43 \times 10^9$	$3,6 \times 10^9$	$5,8 \times 10^9$
Достоверность	0,84	0,92	0,93	0,96	0,965

При этом размеры изображений в среднем составили всего 227×227 пикселей. А распознаванию подвергались детерминированные объекты. Наибольшие успехи в решении задач экологического мониторинга отмечены в [46], где решалась задача поиска разливов нефтепродуктов на основе нейронных сетей глубокого обучения по результатам обработки изображений из баз данных Европейского космического агентства, полученных от спутникового радара с синтезированной апертурой (*Synthetic Aperture Radar* – SAR).

Обработке подвергались фрагменты изображений с разрешением 1252×609 пикселей. При этом в ходе эксперимента были определены два эталонных класса изображений: разливы нефтепродуктов; отсутствие разливов.

Результаты работы классификатора представлены в таблице 10.

Таблица 10. Результаты классификации при выделении трех семантических сегментов

Вид семантического сегмента	Overall	Oil spills	Look-alikes
Достоверность классификации	0,8063	0,8621	0,7588

В таблице 10 использованы следующие обозначения: overall – общая вероятность классификации изображения; oil spills – достоверное наличие разливов нефтепродуктов; Look-alikes – полное совпадение результатов работы сети с данными визуального анализа.

При этом допустимый порог классификации для разделения классов составлял порядка 2% пикселей, соответствующих альтернативному классу. Следует отметить, что применение для решения данной задачи алгоритма tree forest позволило повысить точность классификации до 0,85. К сожалению, доступ к базам данных Европейского космического агентства закрыт, что не позволяет осуществить непосредственную проверку разработанного подхода с известными результатами.

7. Заключение. Полученные результаты позволяют сделать вывод, что в ходе проведенного исследования разработан аналитический аппарат, позволяющий получить вероятностную оценку результатов обработки изображений. Несмотря на то, что объектом исследования являлись результаты разливов нефтепродуктов в акваториях морей и океанов, разработанный аналитический аппарат может быть применен и в других областях науки и техники, связанных с обработкой кадров фото и видеоизображений.

В работе обосновано понятие аномалий, под которыми, в широком аспекте, авторы понимают элементы фона (фрагменты) графического изображения, возникающие в результате внешних факторов, в частности, последствий разливов нефтепродуктов.

Это позволило разработать подход к формированию первичных и вторичных векторов признаков обрабатываемых изображений на основе гистограмм распределения яркости, с последующей их кратномасштабной обработкой. Выбор кратномасштабной обработки в качестве декоррелирующего преобразования обусловлен высокой зависимостью дисперсии векторов признаков, формируемых на основе гистограмм распределения яркости от вариативности фона изображений, относящихся к одному и тому же альтернативному классу распознаваемых объектов.

Это позволило построить классификатор, работающий с компактными данными, характеризующими изображения.

В работе обоснован выбор вейвлета Гаусса 1-го порядка в декоррелирующей обработке векторов признаков. И хотя его эффективность доказана только для частного случая, полученные результаты, по аналогии с рекомендациями [24, 25] могут быть обобщены на более широкий класс изображений, при условии соответствия закона распределения их амплитудных значений яркости, с представленными примерами изображений.

Состоятельность предложений основана на численных результатах, характеризующих вероятностную оценку, которые получили подтверждение в эксперименте по обнаружению аномалий на 252 фрагментах изображений. Средняя вероятность достоверного принятия решения по полученным результатам составляет 0,95. Это обеспечивается за счет использования коэффициента повышения различий, значение которого целесообразно каждый раз уточнять перед проведением расчетов. Практическую направленность полученных результатов исследований авторы связывают

с их применением в решении задач поиска разливов нефтепродуктов по результатам обработки изображений.

Литература

1. Козловский А.В. Распараллеливание обработки информации при формировании составных изображений // Известия ЮФУ. Технические науки. 2025. №1(243). С. 92–103. DOI: 10.18522/2311-3103-2025-1-92-103.
2. Ушенкин В.А. Метод высокоточной обработки топографических интерферометрических изображений с привлечением низкодетальной стереофотограмметрической информации от оптико-электронных систем ДЗЗ // Цифровая обработка сигналов. 2025. №1. С. 59–66.
3. Хомяков О.О., Панищев В.С., Титов В.С., Ватутин Э.И. Математическая модель и параллельный алгоритм обработки изображений, содержащих символьную информацию // Труды МАИ. 2024. №137. С. 1–25.
4. Копейкин Р.Е., Перов А.Ю., Соловьев М.А. Алгоритмы поляризационной селекции для обработки информации в изображениях, полученных с помощью радара с синтезированной апертурой // Научно-технический вестник Поволжья. 2023. №12. С. 301–303.
5. Хомколов И.В., Медиевский А.В., Зотин А.Г., Симонов К.В., Кругляков А.С. Информации алгоритмическое обеспечение обработки и анализа изображений медицинского эксперимента // Информатизация и связь. 2023. №3. С. 35–42. DOI: 10.34219/2078-8320-2023-14-3-35-42.
6. Дворников С.В., Васильева Д.В. Повышение достоверности выявления аномалий на изображениях при формировании их векторов признаков в базах вейвлетов // Информатика и автоматизация. 2024. Т. 23. №6. С. 1698–1729. DOI: 10.15622/ia.23.6.5.
7. Abbasov I.B. Image Recognition in Agriculture and Landscape Protection // International Journal of Science and Research. 2020. vol. 9. no. 12. pp. 757–763. DOI: 10.21275/SR201212144831.
8. Васильева Д.В., Дворников С.В. Структуризация первичных признаков изображений систем мониторинга окружающей среды // Радиотехника. 2024. Т. 88. №8. С. 5–17. DOI: 10.18127/j00338486-202408-01.
9. Пахлян И.А., Омелянюк М.В. Разработка технических и технологических решений для очистки акваторий от разлива нефтепродуктов // Нефтяное хозяйство. 2025. №8. С. 94–97. DOI: 10.24887/0028-2448-2025-8-94-97.
10. Маценко С.В., Блиновская Я.Ю. Показатели эффективности мероприятий по ликвидации разливов нефти и нефтепродуктов на морских акваториях: организационные и технологические аспекты // Защита окружающей среды в нефтегазовом комплексе. 2023. №2(311). С. 5–11.
11. Vasilyeva D.V., Dvornikov S.V., Yakushenko S.A., Dvornikov S.S. Automation of detection procedures based on the results of processing images from video surveillance systems // Collection of abstracts of reports of the III International Forum “Mathematical methods and models in high-tech production”. 2023. 157–158.
12. Петрова Е.Г. Чрезвычайные ситуации с аварийным выбросом опасных химических веществ // Экологические системы и приборы. 2023. №5. С. 47–54. DOI: 10.25791/esip.5.2023.1372.
13. Подрезов Р.В., Райфельд М.А. Непараметрический метод определения числа классов в задаче сегментации изображений // Автотметрия. 2020. Т. 56. №3. С. 81–90. DOI: 10.15372/AUT20200309.
14. Андриянов Н.А., Гаврилина Ю.Н. Исследование алгоритма обнаружения детерминированных аномалий на сложных изображениях с помощью дважды

- стохастической модели // Ural Radio Engineering Journal. 2020. Т. 4. №1. С. 18–32. DOI: 10.15826/urej.2020.4.1.002.
15. Li J., Li X., Li H., et al. Fabrication of the Cu/AgCuTi/Nb composite for superconducting radio-frequency material under extreme service conditions based on electroplating additive manufacturing // Materials & Design. 2024. vol. 244. 113220 p. DOI: 10.1016/j.matdes.2024.113220.
16. Белов А.М., Денисова А.Ю. Классификация изображений ДЗЗ с использованием алгоритма комплексирования данных различных сенсоров // Компьютерная оптика. 2020. Т. 44. №4. С. 627–635. DOI: 10.18287/2412-6179-CO-735.
17. Васильева Д.В., Дворников С.В., Якушенко С.А., Дворников С.С. Автоматизация процедур обнаружения лесных пожаров по результатам обработки видео // Вестник Санкт-Петербургского университета государственной противопожарной службы МЧС России. 2023. №4. С. 47–58. DOI: 10.61260/2218-130X-2024-2023-4-47-58.
18. Сурин В.А. Об обработке зашумленных контрастных изображений // Вестник Южно-Уральского государственного университета. 2021. Т. 13. №1. С. 14–21. DOI: 10.14529/mmph210102.
19. Велигоша А.В., Малышко Н.Н., Струков Р.И. Применение искусственных нейронных сетей для снижения избыточности дискретного вейвлет-преобразования изображений // Теория и техника радиосвязи. 2020. №2. С. 5–12.
20. Чебачев В.О. Обработка сигналов в системах радиочастотной идентификации с помощью декоррелирующих преобразований // Вопросы радиоэлектроники. 2016. №9. С. 35–40.
21. Булатов Ю.Н. Теорема о норме сингулярных J-псевдодифференциальных операторов Киприянова-Катрахова // Вестник Воронежского государственного университета. 2025. №2. С. 38–48.
22. Осанов В.А. Сравнительный анализ методов декорреляции сетевого трафика // Вестник Рязанского государственного радиотехнического университета. 2025. №92. С. 25–38. DOI: 10.21667/1995-4565-2025-92-25-38.
23. Семенов В.И. Сравнение алгоритма кратномасштабного анализа изображений в частотной области с алгоритмом Малла // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. №59. С. 73–82. DOI: 10.17223/19988605/59/8.
24. Sheikh H.R., Bovik A.C. Image information and visual quality // IEEE Transactions on Image Processing. 2006. vol. 15. no. 2. pp. 430–444. DOI: 10.1109/tip.2005.859378.
25. Bisgard J. Analysis and Linear Algebra: The Singular Value Decomposition and Applications / Student Mathematical Library // AMS. 2021. 217 p.
26. Дворников С.В., Дворников С.С., Коноплев М.А. Алгоритм распознавания сигналов радиосвязи на основе симметрических матриц // Информационные технологии. 2010. №9. С. 75–77.
27. Harvey Mudd College. Karhunen-Loeve Transformation. URL: <https://pages.hmc.edu/ruye/MachineLearning/lectures/ch8/node4.html> (дата обращения: 02.10.2025).
28. Ahmed N., Natarajan T., Rao K.R. Discrete Cosine Transform // IEEE Transactions on Computers. 1974. vol. 23. no. 1. pp. 90–93. DOI: 10.1109/T-C.1974.223784.
29. Loeffler C., Ligtenberg A., Moschytz G.S. Practical Fast 1-D DCT Algorithms with 11 Multiplications // Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP). 1989. pp. 988–991.
30. Дворников С.В., Степьяин Д.В., Дворников А.С., Букарева А.П. Формирование векторов признаков сигналов из вейвлет-коэффициентов их фреймовых преобразований // Информационные технологии. 2013. №5. С. 46–49.

31. Flores-Soriano M. Solar radio bursts impact on the International GNSS Service Network during Solar Cycle 24 // *Journal of Space Weather and Space Climate*. 2024. DOI: 10.1051/swsc/2024034.
32. Васильева Д.В. Исследование чувствительности векторов признаков, сформированных на основе кратномасштабных преобразований обрабатываемых изображений // *Известия высших учебных заведений России. Радиоэлектроника*. 2024. Т. 27. №3. С. 20–29. DOI: 10.32603/1993-8985-2024-27-3-20-29.
33. Singh U., Dua A., Kumar N., Guizani M. QoS Aware Uplink Scheduling for M2M Communication in LTE/LTE-A Network: A Game Theoretic Approach // *IEEE Transactions on Vehicular Technology*. 2022. vol. 71. no. 4. pp. 4156–4170. DOI: 10.1109/TVT.2021.3132535.
34. Сай С.В., Зинкевич А.В., Фомина Е.С. Сравнение дискретного косинус и вейвлет-преобразований в системах сжатия RAW-изображений // *Компьютерная оптика*. 2022. Т. 46. №6. С. 929–938. DOI: 10.18287/2412-6179-СО-1094.
35. Велигоша А.В., Малышко Н.Н., Струков Р.И. Применение искусственных нейронных сетей для снижения избыточности дискретного вейвлет-преобразования изображений // *Теория и техника радиосвязи*. 2020. №2. С. 5–12.
36. Гришпенцев А.Ю., Коровкин Н.В., Островский Д.П. Вейвлеты Эрмита-Гаусса: синтез дискретных форм и исследование свойств // *Научно-технический вестник информационных технологий, механики и оптики*. 2025. Т. 25. №4. С. 789–796. DOI: 10.17586/2226-1494-2025-25-4-789-796.
37. Chui C.K. An Introduction to Wavelets / Wavelet Analysis & Its Applications // San Diego: Academic Press. 1992. 266 p.
38. Семенов В.И., Чумаров С.Г. От конструирования вейвлетов на основе производных функции Гаусса к синтезу фильтров с конечной импульсной характеристикой // *Научно-технический вестник информационных технологий, механики и оптики*. 2024. Т. 24. №2. С. 306–313. DOI: 10.17586/2226-1494-2024-24-2-306-313.
39. Беляев А.А., Кононов Д.П., Кудрин А.М. Исследование выбора материнской вейвлет-функции при вибродиагностике тепловозных двигателей // *Известия Петербургского университета путей сообщения*. 2025. Т. 22. №2. С. 290–299. DOI: 10.20295/1815-588X-2025-2-290-299.
40. Васильева Д.В., Дворников С.С., Толстуха Ю.Е., Обрезков П.С., Дворников С.В. Формирование векторов признаков для систем видеонаблюдения // *Вопросы радиоэлектроники*. 2023. №4. С. 62–68.
41. Dvornikov S.S., Zheglov K.D., Dvornikov S.V. SSB signals with controlled pilot level // *T-Comm*. 2023. vol. 17. no. 3. pp. 41–47. DOI: 10.36724/2072-8735-2023-17-3-41-47.
42. Jiao Z., Jia G., Cai Y. A new approach to oil spill detection that combines deep learning with unmanned aerial vehicles // *Computers & Industrial Engineering*. 2019. vol. 135. pp. 1300–1311.
43. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778. DOI: 10.1109/CVPR.2016.90.
44. Howard A.G., Zhu M., Chen B., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications // *arXiv preprint arXiv:1704.04861*. 2017.
45. Venkatesan R., Li B. Convolutional Neural Networks in Visual Computing: A Concise Guide // Boca Raton: CRC Press. 2017. 186 p.
46. Orfanidis G., Ioannidis K., Avgerinakis K., et al. A Deep Neural Network for Oil Spill Semantic Segmentation in Sar Images // *Proceedings of 25th IEEE International Conference on Image Processing (ICIP)*. 2018. pp. 3773–3777. DOI: 10.1109/ICIP.2018.8451113.

Дворников Сергей Викторович — д-р техн. наук, профессор, кафедра конструирования и технологий электронных и лазерных средств института радиотехники и инфокоммуникационных технологий, Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет аэрокосмического приборостроения» (ФГАОУ ВО ГУАП); профессор, кафедра радиосвязи, Военная академия орденов Жукова и Ленина Краснознамённая академия связи имени Маршала Советского Союза С.М. Будённого (ВАС им. С.М. Будённого). Область научных интересов: построение помехозащищенных систем радиосвязи, формирование и обработка сигналов сложных структур. Число научных публикаций — 700. practicdsv@yandex.ru; ул. Большая Морская, 67, 190000, Санкт-Петербург, Россия; р.т.: +7(812)247-9400.

Васильева Дина Владимировна — старший преподаватель, кафедра радиотехнических систем института радиотехники и инфокоммуникационных технологий, Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет аэрокосмического приборостроения» (ФГАОУ ВО ГУАП). Область научных интересов: цифровая обработка сигналов. Число научных публикаций — 61. dolli.dina@mail.ru; ул. Большая Морская, 67, 190000, Санкт-Петербург, Россия; р.т.: +7(931)385-9181.

Поддержка исследований. Исследование выполнено за счет гранта Российского научного фонда (проект № 24-79-10259).

S. DVORNIKOV, D. VASILIEVA

OIL SPILL IMAGE RECOGNITION BASED ON MULTIPLE-SCALE PROCESSING OF BRIGHTNESS DISTRIBUTION HISTOGRAMS

Dvornikov S., Vasilieva D. Oil Spill Image Recognition Based on Multiple-Scale Processing of Brightness Distribution Histograms.

Abstract. The problem of detecting anomalies in processed images acquired from visual monitoring equipment becomes particularly relevant in the context of ecological monitoring of the world ocean's water areas, specifically when searching for the consequences of technogenic catastrophes caused by oil spills. In the present study, anomalies are understood as image fragments characterizing oil product spills. The subject of the research is defined by elements of pattern recognition theory, according to which the processed images are represented as feature vectors. The rationale for selecting histograms characterizing the brightness bit distributions of the processed images, which are used as primary feature vectors, is substantiated. It is established that the high variability of the background of water surface images leads to significant variance in the feature vectors formed on the basis of brightness distribution histograms, even for a single common class of recognition objects. Therefore, it is proposed to subject the primary feature vectors to decorrelating processing based on wavelet transforms. Types of wavelet transforms and various types of generating functions are considered, among which the rationale for selecting the first-order Gaussian wavelet is provided, as it ensures the smallest differences between the feature vectors of the processed images characterizing the same class. An approach to the formation of secondary recognition features is proposed by averaging the components of multiscale transformation matrices of image brightness distribution histograms along the axis of their shift parameter. The rationale for transitioning to the selection of a redundant number of decomposition scales in wavelet transforms when forming secondary feature vectors is substantiated. The results of the conducted experiment are demonstrated, confirming the validity of such a transition. Test and reference images used in the experiment are presented, as well as the primary and secondary feature vectors formed on their basis. Numerical results characterizing the obtained estimate of differences between feature vectors of both the same class and alternative classes are provided. A probabilistic assessment of the anomaly recognition results on the processed images is presented. The rationale for using the coefficient of estimating differences between feature vectors in the developed formula for the probability of correct anomaly detection in the processed images is substantiated. Graphs characterizing the value of the probabilistic estimate as a function of the ratio of anomaly size to the total frame area for various values of the coefficient enhancing the differences between feature vectors are presented.

Keywords: image anomalies, oil product spills, feature vectors, recognizable image classes, histograms of image brightness distribution.

References

1. Kozlovsky A.V. [Parallelization of information processing in the formation of composite images]. *Izvestiya YuFU. Tekhnicheskie nauki – Izvestiya SFedU. Engineering Sciences*. 2025. no. 1(243), pp. 92–103. DOI: 10.18522/2311-3103-2025-1-92-103. (In Russ.).
2. Ushenkin V.A. [Method for high-precision processing of topographic interferometric images using low-resolution stereophotogrammetric information from

- optical-electronic remote sensing systems.]. *Cifrovaya obrabotka signalov – Digital Signal Processing*. 2025. no. 1. pp. 59–66. (In Russ.).
3. Khomyakov O.O., Panishchev V.S., Titov V.S., Vatutin E.I. [Mathematical model and parallel algorithm for processing images containing character information.]. *Trudy MAI – Proceedings of MAI*. 2024. no. 137. pp. 1–25. (In Russ.).
4. Kopeykin R.E., Perov A.Y., Solovyov M.A. [Polarization selection algorithms for processing information in images obtained using synthetic aperture radar.]. *Nauchno-texnicheskij vestnik Povolzh'ya – Scientific and Technical Volga region Bulletin*. 2023. no. 12. pp. 301–303. (In Russ.).
5. Khomkolov I.V., Medievsky A.V., Zotin A.G., Simonov K.V., Kruglyakov A.S. [Information and algorithmic support for processing and analyzing images in a medical experiment.]. *Informatizatsiya i svyaz – Informatization and Communication*. 2023. no. 3. pp. 35–42. DOI: 10.34219/2078-8320-2023-14-3-35-42. (In Russ.).
6. Dvornikov S., Vasilyeva D. [Increase of Reliability of Anomalies Detection on Images at Formation of Their Feature Vectors in Wavelet Bases]. *Informatika i avtomatizatsiya – Informatics and Automation*. 2024. vol. 23. no. 6. pp. 1698–1729. DOI: 10.15622/ia.23.6.5. (In Russ.).
7. Abbasov I.B. Image Recognition in Agriculture and Landscape Protection. *International Journal of Science and Research*. 2020. vol. 9. no. 12. pp. 757–763. DOI: 10.21275/SR201212144831.
8. Vasilyeva D.V., Dvornikov S.V. [Structuring Primary Image Features of Environmental Monitoring Systems]. *Radiotekhnika – Radio engineering*. 2024. vol. 88. no. 8. pp. 5–17. DOI: 10.18127/j00338486-202408-01. (In Russ.).
9. Pakhlyan I.A., Omelyanyuk M.V. [Development of technical and technological solutions for cleaning water areas from oil spills]. *Neftyanoe xozyajstvo – The oil industry*. 2025. no. 8. pp. 94–97. DOI: 10.24887/0028-2448-2025-8-94-97. (In Russ.).
10. Matsenko S.V., Blinovskaya Y.Y. [Indicators of the effectiveness of measures to eliminate oil and petroleum product spills in marine areas: organizational and technological aspects.]. *Zashhita okruzhayushhej sredy v neftegazovom komplekse – Environmental Protection in Oil and Gas Complex*. 2023. no. 2(311). pp. 5–11. (In Russ.).
11. Vasilyeva D.V., Dvornikov S.V., Yakushenko S.A., Dvornikov S.S. Automation of detection procedures based on the results of processing images from video surveillance systems. Collection of abstracts of reports of the III International Forum “Mathematical methods and models in high-tech production”. 2023. 157–158.
12. Petrova E.G. [Emergency situations involving the accidental release of hazardous chemicals]. *E'kologicheskie sistemy i pribory – Ecological systems and devices*. 2023. no. 5. pp. 47–54. DOI: 10.25791/esip.5.2023.1372. (In Russ.).
13. Podrezov R.V., Raifeld M.A. [Nonparametric method for determining the number of classes in the image segmentation problem.]. *Avtometriya – Autometry*. 2020. vol. 56. no. 3. pp. 81–90. DOI: 10.15372/AUT20200309. (In Russ.).
14. Andriyanov N.A., Gavrilina Yu.N. [Research of an algorithm for detecting deterministic anomalies in complex images using a doubly stochastic model.]. *Ural'skij radiotekhnicheskij zhurnal – Ural Radio Engineering Journal*. 2020. vol. 4. no. 1. pp. 18–32. DOI: 10.15826/urej.2020.4.1.002. (In Russ.).
15. Li J., Li X., Li H., et al. Fabrication of the Cu/AgCuTi/Nb composite for superconducting radio-frequency material under extreme service conditions based on electroplating additive manufacturing. *Materials & Design*. 2024. vol. 244. 113220 p. DOI: 10.1016/j.matdes.2024.113220.
16. Belov A.M., Denisova A.Y. [Classification of remote sensing images using an algorithm for combining data from various sensors.]. *Komp'yuternaya optika – Computer Optics*. 2020. vol. 44. no. 4. pp. 627–635. DOI: 10.18287/2412-6179-CO-735. (In Russ.).

17. Vasilyeva D.V., Dvornikov S.V., Yakushenko S.A., Dvornikov S.S. [Automation of Forest Fire Detection Procedures Based on Video Processing Results]. *Vestnik Sankt-Peterburgskogo universiteta gosudarstvennoj protivopozharnoj sluzhby* MChS Rossii – Bulletin of the Saint Petersburg University of the State Fire Service of the Ministry of Emergency Situations of Russia. 2023. no. 4. pp. 47–58. DOI: 10.61260/2218-130X-2023-4-47-58. (In Russ.).
18. Surin V.A. [On the Processing of Noisy Contrast Images]. *Vestnik Yuzhno-Ural'skogo gosudarstvennogo universiteta* – Bulletin of the South Ural State University. 2021. vol. 13. no. 1. pp. 14–21. DOI: 10.14529/mmph210102. (In Russ.).
19. Veligosha A.V., Malyshev N.N., Strukov R.I. [Application of artificial neural networks to reduce the redundancy of the discrete wavelet transform of images.]. *Teoriya i tekhnika radiosvyazi* – Radio Communication Theory and Equipment. 2020. no. 2. pp. 5–12. (In Russ.).
20. Chebachiev V.O. [Signal Processing in Radio-Frequency Identification Systems Using Decorrelating Transformations]. *Voprosy`lektroniki* – Questions of radio electronics. 2016. no. 9. pp. 35–40. (In Russ.).
21. Bulatov Y.N. [Theorem on the norm of singular J-pseudodifferential operators by Kupriyanov–Katrakhov]. *Vestnik Voronezhskogo gosudarstvennogo universiteta* – Bulletin of Voronezh State University. 2025. no. 2. pp. 38–48. (In Russ.).
22. Osanov V.A. [Comparative analysis of network traffic decorrelation methods]. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta* – Vestnik of Ryazan State Radio Engineering University. 2025. no. 92. pp. 25–38. DOI: 10.21667/1995-4565-2025-92-25-38. (In Russ.).
23. Semenov V.I. [Comparison of the multi-scale image analysis algorithm in the frequency domain with Mallat's algorithm]. *Tomsk State University Journal of Control and Computer Science* – Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vy`chislitel'naya tekhnika i informatika. 2022. no. 59. pp. 73–82. DOI: 10.17223/19988605/59/8. (In Russ.).
24. Sheikh H.R., Bovik A.C. Image information and visual quality. *IEEE Transactions on Image Processing*. 2006. vol. 15. no. 2. pp. 430–444. DOI: 10.1109/tip.2005.859378.
25. Bisgard J. *Analysis and Linear Algebra: The Singular Value Decomposition and Applications* (Series: Student Mathematical Library). AMS. 2021. 217 p.
26. Dvornikov S.V., Dvornikov S.S., Konoplev M.A. [An algorithm for recognizing radio communication signals based on symmetric matrices.]. *Informacionny`e tekhnologii* – Information technology. 2010. no. 9. pp. 75–77. (In Russ.).
27. Harvey Mudd College. Karhunen-Loeve Transformation. Available at: <https://pages.hmc.edu/ruye/MachineLearning/lectures/ch8/node4.html> (accessed 02.10.2025). (In Russ.).
28. Ahmed N., Natarajan T., Rao K.R. Discrete Cosine Transform. *IEEE Transactions on Computers*. 1974. vol. 23. no. 1. pp. 90–93. DOI: 10.1109/T-C.1974.223784.
29. Loeffler C., Ligtenberg A., Moschytz G.S. Practical Fast 1-D DCT Algorithms with 11 Multiplications. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 1989. pp. 988–991.
30. Dvornikov S.V., Stepyanin D.V., Dvornikov A.S., Bukareva A.P. [Formation of signal feature vectors from wavelet coefficients of their frame transforms]. *Informacionny`e tekhnologii* – Information technology. 2013. no. 5. pp. 46–49. (In Russ.).
31. Flores-Soriano M. Solar radio bursts impact on the International GNSS Service Network during Solar Cycle 24. *Journal of Space Weather and Space Climate*. 2024. DOI: 10.1051/swsc/2024034.
32. Vasilyeva D.V. [Study of the sensitivity of feature vectors formed based on multi-scale transformations of the processed images.]. *Izvestiya vysshix uchebny`x zavedenij*

- Rossii. Radioelektronika – Journal of the Russian universities. Radioelectronics. 2024. vol. 27. no. 3. pp. 20–29. DOI: 10.32603/1993-8985-2024-27-3-20-29. (In Russ.).
33. Singh U., Dua A., Kumar N., Guizani M. QoS Aware Uplink Scheduling for M2M Communication in LTE/LTE-A Network: A Game Theoretic Approach. IEEE Transactions on Vehicular Technology. 2022. vol. 71. no. 4. pp. 4156–4170. DOI: 10.1109/TVT.2021.3132535.
34. Sai S.V., Zinkevich A.V., Fomina E.S. [Comparison of Discrete Cosine and Wavelet Transforms in RAW Image Compression Systems]. Computer Optics – Komp'yuternaya optika. 2022. vol. 46. no. 6. pp. 929–938. DOI: 10.18287/2412-6179-CO-1094. (In Russ.).
35. Veligosha A.V., Malyshko N.N., Strukov R.I. [Application of artificial neural networks to reduce the redundancy of the discrete wavelet transform of images.]. Teoriya i tekhnika radiosvyazi – Radio Communication Theory and Equipment. 2020. no. 2. pp. 5–12. (In Russ.).
36. Grishentsev A.Y., Korovkin N.V., Ostrovsky D.P. [Hermite–Gauss wavelets: synthesis of discrete forms and study of properties]. Nauchno-tekhnicheskij vestnik informacionny'x tekhnologij, mexaniki i optiki – Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2025. vol. 25. no. 4. pp. 789–796. DOI: 10.17586/2226-1494-2025-25-4-789-796. (In Russ.).
37. Chui C.K. An Introduction to Wavelets (Series: Wavelet Analysis & Its Applications). San Diego: Academic Press. 1992. 266 p.
38. Semenov V.I., Chumarov S.G. [From the design of wavelets based on the derivatives of the Gaussian function to the synthesis of filters with finite impulse response.]. Nauchno-tekhnicheskij vestnik informacionny'x tekhnologij, mexaniki i optiki – Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2024. vol. 24. no. 2. pp. 306–313. DOI: 10.17586/2226-1494-2024-24-2-306-313. (In Russ.).
39. Belyaev A.A., Kononov D.P., Kudrin A.M. [Research into the choice of the mother wavelet function in vibration diagnostics of diesel locomotive engines.]. Izvestiya Peterburgskogo universiteta putej soobshheniya – Proceedings of Petersburg Transport University. 2025. vol. 22. no. 2. pp. 290–299. DOI: 10.20295/1815-588X-2025-2-290-299. (In Russ.).
40. Vasilyeva D.V., Dvornikov S.S., Tolstukha Y.E., Obrezkov P.S., Dvornikov S.V. [Feature vector generation for video surveillance systems]. Voprosy` radioelektroniki – Questions of radio electronics. 2023. no. 4. pp. 62–68. (In Russ.).
41. Dvornikov S.S., Zheglov K.D., Dvornikov S.V. SSB signals with controlled pilot level. T-Comm. 2023. vol. 17. no. 3. pp. 41–47. DOI: 10.36724/2072-8735-2023-17-3-41-47.
42. Jiao Z., Jia G., Cai Y. A new approach to oil spill detection that combines deep learning with unmanned aerial vehicles. Computers & Industrial Engineering. 2019. vol. 135. pp. 1300–1311.
43. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778. DOI: 10.1109/CVPR.2016.90.
44. Howard A.G., Zhu M., Chen B., et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv preprint arXiv:1704.04861. 2017.
45. Venkatesan R., Li B. Convolutional Neural Networks in Visual Computing: A Concise Guide. Boca Raton: CRC Press. 2017. 186 p.
46. Orfanidis G., Ioannidis K., Aygerinakis K., et al. A Deep Neural Network for Oil Spill Semantic Segmentation in Sar Images. Proceedings of 25th IEEE International Conference on Image Processing (ICIP). 2018. pp. 3773–3777. DOI: 10.1109/ICIP.2018.8451113.

Dvornikov Sergei — Ph.D., Dr.Sci., Professor, Department of Design and Technology of Electronic and Laser Facilities of the Institute of Radio Engineering and Infocommunication Technologies, Federal State Autonomous Educational Institution of Higher Education Saint Petersburg University State University of Aerospace Instrumentation (GUAP Federal State Educational Institution of Higher Education); Professor, Department of Radio Communication, The Military Academy of the Orders of Zhukov and Lenin, the Red Banner Academy of Communications named after Marshal of the Soviet Union S.M. Budyonny (S.M. Budyonny Military Academy of Communications). Research interests: construction of interference-resistant radio communication systems, formation and processing of complex signal structures. The number of publications — 700. practicdsv@yandex.ru; 67, Bolshaya Morskaya St., 190000, St. Petersburg, Russia; office phone: +7(812)247-9400.

Vasilieva Dina — Senior lecturer, Department of Radio Engineering Systems at the Institute of Radio Engineering and Information and Communication Technologies, Federal State Autonomous Educational Institution of Higher Education Saint Petersburg University State University of Aerospace Instrumentation (GUAP Federal State Educational Institution of Higher Education). Research interests: digital signal processing. The number of publications — 61. dolli.dina@mail.ru; 67, Bolshaya Morskaya St., 190000, St. Petersburg, Russia; office phone: +7(931)385-9181.

Acknowledgements. This study was supported by a grant from the Russian Science Foundation (project no. 24-79-10259).

Руководство для авторов

Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Информатика и автоматизация» <http://ia.spcras.ru/>. При регистрации авторам рекомендуется заполнить все предложенные поля данных. Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше или LaTeX. Объем основного текста (до раздела Литература) - от 20 до 30 страниц включительно. Переносы запрещены. Номера страниц не проставляются. Основная часть текста статьи разбивается на разделы, среди которых являются обязательными: введение, хотя бы один «содержательный» раздел и заключение. Допускается также мотивированное содержанием и структурой материала выделение подразделов. В основную часть допускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно изложены на нашем сайте в разделе «Руководство для авторов».

Author guidelines

Interaction between authors and the Editorial board is carried out via the personal account on the website of the journal "Informatics and Automation" <http://ia.spcras.ru/>. Upon registration, authors are requested to fill in all the provided data fields in the registration form. Manuscripts should be prepared using MS Word 2007 or later, or LaTeX. The length of the main text (up to the References section) must be between 20 and 30 pages, inclusive. Hyphenation is not allowed. Page numbers should not be inserted. The main body of the manuscript should be divided into sections. The following sections are mandatory: an Introduction, at least one "substantive" section, and a Conclusion. The use of subsections is permitted if justified by the content and structure of the material. Figures, tables, listings, and formulas may be included in the main body. Detailed rules for their formatting are available on our website in the "Author Guidelines" section.

Подписано к печати 30.09.2026.

Формат 60×90 1/16. Усл. печ. л. 23,1. Заказ № 325. Тираж 300 экз.

Дата выхода в свет: 02.10.2026.

Отпечатано в Редакционно-издательском центре ГУАП.

Адрес типографии: Россия, 190000, г. Санкт-Петербург, ул. Большая Морская, д. 67, лит. А
Подписной индекс П5513 по каталогу «Почта России»

Цена свободная

Signed to print 30.09.2026

Format 60×90 1/16. Cond. printed sheets 23.1.

Order No. 325. Circulation 300 copies.

Passed for print 02.10.2026.

Printed in Publishing center GUAP.

Address: 67 Bolshaya Morskaya Street., St.Petersburg, 190000, Russia
Subscription Index П5513, Russian Post Catalog

Price not fixed