

ISSN 2713-3192  
DOI 10.15622/ia.2024.23.4  
<http://ia.spcras.ru>

ТОМ 23 № 4  
90 лет  
главному  
редактору  
Юсупову Р.М.

**ИНФОРМАТИКА  
И АВТОМАТИЗАЦИЯ**

**INFORMATICS  
AND AUTOMATION**



**СПб ФИЦ РАН**

**Санкт-Петербург  
2024**



# INFORMATICS AND AUTOMATION

Volume 23 № 4, 2024

Scientific and educational journal primarily specialized in computer science, automation, robotics, applied mathematics, interdisciplinary research

Founded in 2002

---

## Founder and Publisher

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

---

## Editor-in-Chief

**R. M. Yusupov**, Prof., Dr. Sci., Corr. Member of RAS, St. Petersburg, Russia

---

## Editorial Council

<b>A. A. Ashimov</b>	Prof., Dr. Sci., Academician of the National Academy of Sciences of the Republic of Kazakhstan, Almaty, Kazakhstan
<b>I. A. Kalyaev</b>	Prof., Dr. Sci., Academician of RAS, Taganrog, Russia
<b>Yu. A. Merkuryev</b>	Prof., Dr. Sci., Academician of the Latvian Academy of Sciences, Riga, Latvia
<b>A. I. Rudskoi</b>	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
<b>V. Sgurev</b>	Prof., Dr. Sci., Academician of the Bulgarian Academy of Sciences, Sofia, Bulgaria
<b>B. Ya. Sovetov</b>	Prof., Dr. Sci., Academician of RAE, St. Petersburg, Russia
<b>V. A. Soyfer</b>	Prof., Dr. Sci., Academician of RAS, Samara, Russia

---

## Editorial Board

<b>O. Yu. Gusikhin</b>	Ph. D., Dearborn, USA
<b>V. Delic</b>	Prof., Dr. Sci., Novi Sad, Serbia
<b>A. Dolgui</b>	Prof., Dr. Sci., St. Etienne, France
<b>M. N. Favorskaya</b>	Prof., Dr. Sci., Krasnoyarsk, Russia
<b>M. Zelezny</b>	Assoc. Prof., Ph.D., Plzen, Czech Republic
<b>H. Kaya</b>	Assoc. Prof., Ph.D., Utrecht, Netherlands
<b>A. A. Karpov</b>	Assoc. Prof., Dr. Sci., St. Petersburg, Russia
<b>S. V. Kuleshov</b>	Dr. Sci., St. Petersburg, Russia
<b>A. D. Khomonenko</b>	Prof., Dr. Sci., St. Petersburg, Russia
<b>D. A. Ivanov</b>	Prof., Dr. Habil., Berlin, Germany
<b>K. P. Markov</b>	Assoc. Prof., Ph.D., Aizu, Japan
<b>R. V. Meshcheryakov</b>	Prof., Dr. Sci., Moscow, Russia
<b>N. A. Moldovian</b>	Prof., Dr. Sci., St. Petersburg, Russia
<b>V. V. Nikulin</b>	Prof., Ph.D., New York, United States
<b>V. Yu. Osipov</b>	Prof., Dr. Sci., St. Petersburg, Russia
<b>V. K. Pshikhopov</b>	Prof., Dr. Sci., Taganrog, Russia
<b>A. L. Ronzhin</b>	Prof., Dr. Sci., Deputy Editor-in-Chief, St. Petersburg, Russia
<b>H. Samani</b>	Assoc. Prof., Ph.D., Plymouth, UK
<b>A. V. Smirnov</b>	Prof., Dr. Sci., St. Petersburg, Russia
<b>B. V. Sokolov</b>	Prof., Dr. Sci., St. Petersburg, Russia
<b>L. V. Utkin</b>	Prof., Dr. Sci., St. Petersburg, Russia
<b>L. B. Sheremetov</b>	Assoc. Prof., Dr. Sci., Mexico, Mexico

---

**Editor:** A. S. Lopotova

**Interpreter:** Ya. N. Berezina

**Art editor:** N. A. Dormidontova

---

## Editorial office address

SPC RAS, 39 litera A , 14-th line V.O., St. Petersburg, 199178, Russia

e-mail: [ia@spcras.ru](mailto:ia@spcras.ru), web: <http://ia.spcras.ru>

## The journal is indexed in Scopus

The journal is published under the scientific-methodological supervision of Department for Nanotechnologies and Information Technologies of the Russian Academy of Sciences

© St. Petersburg Federal Research Center of the Russian Academy of Sciences, 2024

# ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

Том 23 № 4, 2024

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

Журнал основан в 2002 году

---

## Учредитель и издатель

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

---

## Главный редактор

**Р. М. Юсупов**, чл.-корр. РАН, д-р техн. наук, проф., Санкт-Петербург, РФ

---

## Редакционный совет

<b>А. А. Ашимов</b>	академик Национальной академии наук Республики Казахстан, д-р техн. наук, проф., Алматы, Казахстан
<b>И. А. Каляев</b>	академик РАН, д-р техн. наук, проф., Таганрог, РФ
<b>Ю. А. Меркурьев</b>	академик Латвийской академии наук, д-р, проф., Рига, Латвия
<b>А. И. Рудской</b>	академик РАН, д-р техн. наук, проф., Санкт-Петербург, РФ
<b>В. Сгурев</b>	академик Болгарской академии наук, д-р техн. наук, проф., София, Болгария
<b>Б. Я. Советов</b>	академик РАО, д-р техн. наук, проф., Санкт-Петербург, РФ
<b>В. А. Соيفер</b>	академик РАН, д-р техн. наук, проф., Самара, РФ

---

## Редакционная коллегия

<b>О. Ю. Гусихин</b>	д-р наук, Диаборн, США
<b>В. Делич</b>	д-р техн. наук, проф., Нови-Сад, Сербия
<b>А. Б. Долгий</b>	д-р наук, проф. Сент-Этьен, Франция
<b>М. Железны</b>	д-р наук, доцент, Пльзень, Чешская республика
<b>Д. А. Иванов</b>	д-р экон. наук, проф., Берлин, Германия
<b>Х. Каия</b>	д-р наук, доцент, Утрехт, Нидерланды
<b>А. А. Карпов</b>	д-р техн. наук, доцент, Санкт-Петербург, РФ
<b>С. В. Кулешов</b>	д-р техн. наук, Санкт-Петербург, РФ
<b>К. П. Марков</b>	д-р наук, доцент, Аizu, Япония
<b>Р. В. Мещеряков</b>	д-р техн. наук, проф., Москва, РФ
<b>Н. А. Молдовян</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>В. В. Никулин</b>	д-р наук, проф., Нью-Йорк, США
<b>В. Ю. Осипов</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>В. Х. Пшихопов</b>	д-р техн. наук, проф., Таганрог, РФ
<b>А. Л. Ронжин</b>	д-р техн. наук, проф., зам. главного редактора, Санкт-Петербург, РФ
<b>Х. Самани</b>	д-р наук, доцент, Плимут, Соединённое Королевство
<b>А. В. Смирнов</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>Б. В. Соколов</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>Л. В. Уткин</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>М. Н. Фаворская</b>	д-р техн. наук, проф., Красноярск, РФ
<b>А. Д. Хомоненко</b>	д-р техн. наук, проф., Санкт-Петербург, РФ
<b>Л. Б. Шереметов</b>	д-р техн. наук, Мехико, Мексика

---

**Выпускающий редактор:** А. С. Лопотова

**Переводчик:** Я. Н. Березина

**Художественный редактор:** Н. А. Дормидонтова

---

## Адрес редакции

14-я линия В.О., д. 39, лит. А, г. Санкт-Петербург, 199178, Россия

e-mail: [ia@spcras.ru](mailto:ia@spcras.ru), сайт: <http://ia.spcras.ru>

## 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», 2024  
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»

## CONTENTS

<b>Towards 90th Anniversary of Rafael M. Yusupov</b>	941
<b>Artificial Intelligence, Knowledge and Data Engineering</b>	
V. Soifer, V. Fursov, S. Kharitonov KALMAN FILTER FOR A PARTICULAR CLASS OF DYNAMIC OBJECT IMAGES	953
E. Shalfeeva, V. Gribova THE ISSUES OF CREATION OF MACHINE-UNDERSTANDABLE SMART STANDARDS BASED ON KNOWLEDGE GRAPHS	969
V. Manoilov, A. Borodinov, I. Zarutsky, A. Petrov, A. Saraev, V. Kurochkin ALGORITHMS FOR THE PRIMARY ANALYSIS OF LOCAL FLUORESCENCE OBJECTS IN THE DNA SEQUENCER «NANOFOR SPS»	989
K. Sobyenin, S. Kulikova UNET-BOOSTED CLASSIFIER – MULTI-TASK ARCHITECTURE FOR SMALL DATASETS APPLIED TO BRAIN MRI CLASSIFICATION	1022
M. Favorskaya, A. Pakhirka RESTORATION OF SEMANTIC-BASED SUPER-RESOLUTION AERIAL IMAGES	1047
V. Osipov INTELLIGENT NEURAL NETWORK MACHINE WITH THINKING FUNCTIONS	1077
V. Pimeshkov, M. Nikonorova, M. Shishaev A COMBINED TERM EXTRACTION METHOD FOR THE PROBLEM OF MONITORING THEMATIC DISCUSSIONS IN SOCIAL MEDIA	1110
A. Smirnov, A. Ponomarev, N. Shilov, T. Levashova, N. Teslya A CONCEPTION OF COLLABORATIVE DECISION SUPPORT SYSTEMS: APPROACH AND PLATFORM ARCHITECTURE	1139
A. Dvoynikova, I. Kagirov, A. Karpov A METHOD FOR RECOGNITION OF SENTIMENT AND EMOTIONS IN RUSSIAN SPEECH TRANSCRIPTS USING MACHINE TRANSLATION	1173
E. Vlasov, N. Krasnenko CASCADE CLASSIFIER FOR THE DETECTION AND IDENTIFICATION OF BIRDS IN A VIDEOSTREAM	1199
A. Verkhoturov, A. Stepanov, L. Illarionova USING SAR DATA FOR MONITORING OF AGRICULTURAL CROPS IN THE SOUTH OF THE RUSSIAN FAR EAST	1221
<b>Information Security</b>	
A. Moldovyan, D. Moldovyan, N. Moldovyan POST-QUANTUM PUBLIC-KEY CRYPTOSCHEMES ON FINITE ALGEBRAS	1246

## СОДЕРЖАНИЕ

<b>К 90-летию Юбилею Юсупова Рафаэля Мидхатовича</b>	941
<b>Искусственный интеллект, инженерия данных и знаний</b>	
В.А. Сойфер, В.А. Фурсов, С.И. Харитонов КАЛМАНОВСКАЯ ФИЛЬТРАЦИЯ ОДНОГО КЛАССА ИЗОБРАЖЕНИЙ	953
Е.А. Шалфеева, В.В. Грибова ВОПРОСЫ СОЗДАНИЯ МАШИНОПОНИМАЕМЫХ SMART- СТАНДАРТОВ НА ОСНОВЕ ГРАФОВ ЗНАНИЙ	969
В.В. Манойлов, А.Г. Бородинов, И.В. Заруцкий, А.И. Петров, А.С. Сараев, В.Е. Курочкин АЛГОРИТМЫ ПЕРВИЧНОГО АНАЛИЗА ЛОКАЛЬНЫХ ОБЪЕКТОВ ФЛУОРЕСЦЕНЦИИ В СЕКВЕНАТОРЕ ДНК «НАНОФОР СПС»	989
К.В. Собынин, С.П. Куликова UNET-BOOSTED CLASSIFIER – МУЛЬТИЗАДАЧНАЯ АРХИТЕКТУРА ДЛЯ МАЛЫХ ВЫБОРОК НА ПРИМЕРЕ КЛАССИФИКАЦИИ МРТ СНИМКОВ ГОЛОВНОГО МОЗГА	1022
М.Н. Фаворская, А.И. Пахирка ВОССТАНОВЛЕНИЕ АЭРОФОТОСНИМКОВ СВЕРХВЫСОКОГО РАЗРЕШЕНИЯ С УЧЕТОМ СЕМАНТИЧЕСКИХ ОСОБЕННОСТЕЙ	1047
В.Ю. Осипов ИНТЕЛЛЕКТУАЛЬНАЯ НЕЙРОСЕТЕВАЯ МАШИНА С ФУНКЦИЯМИ МЫШЛЕНИЯ	1077
В.К. Пимешков, М.Л. Никонорова, М.Г. Шишаев КОМБИНИРОВАННЫЙ МЕТОД ИЗВЛЕЧЕНИЯ ТЕРМИНОВ ДЛЯ ЗАДАЧИ МОНИТОРИНГА ТЕМАТИЧЕСКИХ ОБСУЖДЕНИЙ В СОЦИАЛЬНЫХ МЕДИА	1110
А.В. Смирнов, А.В. Пономарев, Н.Г. Шилов, Т.В. Левашова, Н.Н. Тесля КОНЦЕПЦИЯ ПОСТРОЕНИЯ КОЛЛАБОРАТИВНЫХ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ: ПОДХОД И АРХИТЕКТУРА ПЛАТФОРМЫ	1139
А.А. Двойникова, И.А. Кагиров, А.А. Карпов МЕТОД РАСПОЗНАВАНИЯ СЕНТИМЕНТА И ЭМОЦИЙ В ТРАНСКРИПЦИЯХ РУССКОЯЗЫЧНОЙ РЕЧИ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ПЕРЕВОДА	1173
Е.В. Власов, Н.П. Красенко КАСКАДНЫЙ КЛАССИФИКАТОР ДЛЯ ОБНАРУЖЕНИЯ И ИДЕНТИФИКАЦИИ ПТИЦ В ВИДЕОПОТОКЕ	1199
А.Л. Верхотуров, А.С. Степанов, Л.В. Илларионова ИСПОЛЬЗОВАНИЕ РАДИОЛОКАЦИОННЫХ ДАННЫХ ДЛЯ МОНИТОРИНГА СОСТОЯНИЯ ПОСЕВОВ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР НА ЮГЕ ДАЛЬНЕГО ВОСТОКА РОССИИ	1221
<b>Информационная безопасность</b>	
А.А. Молдовян, Д.Н. Молдовян, Н.А. Молдовян ПОСТКВАНТОВЫЕ ДВУХКЛЮЧЕВЫЕ КРИПТОСХЕМЫ НА КОНЕЧНЫХ АЛГЕБРАХ	1246

---

## К 90-ЛЕТНЕМУ ЮБИЛЕЮ ЮСУПОВА РАФАЭЛЯ МИДХАТОВИЧА



17 июля 2024 года исполнилось 90 лет со дня рождения Юсупова Рафаэля Мидхатовича, доктора технических наук, профессора, члена-корреспондента РАН, заслуженного деятеля науки и техники РСФСР, научного руководителя Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН).

Юсупов Рафаэль Мидхатович родился 17 июля 1934 года в г. Казани. После окончания в 1952 году с золотой медалью Казанской спецшколы ВВС Р.М. Юсупов был направлен в Ленинградскую военно-воздушную инженерную академию (ныне Военно-космическая академия (ВКА) имени А.Ф. Можайского), которую окончил с отличием в 1958 году по специальности «инженер-электрик». В 1964 году он окончил Ленинградский государственный университет по специальности «математика».

В 1958-1985 годах Рафаэль Мидхатович проходил службу в ВКА имени А.Ф. Можайского, где занимал должности инженера (1958-1959), старшего инженера (1959 г.), научного сотрудника (1959-1960), адъюнкта (1960-1962), старшего научного сотрудника (1962-1967), начальника научно-исследовательской лаборатории систем управления летательных аппаратов (1967-970), заместителя начальника кафедры систем управления ракет и космических аппаратов (1970-1971), начальника кафедры боевой эффективности применения ракетно-космической техники (1971-1976), начальника созданного с его участием, уникального в системе высшего военного образования, факультета сбора и обработки информации (1976-1985). В 1985 году Р.М. Юсупов назначается на должность начальника Направления моделирования стратегических операций Центра оперативно-стратегических исследований Генерального Штаба Вооруженных Сил СССР (ГШ ВС СССР). В 1986 году с целью улучшения оперативно-тактической подготовки выпускников ВКА имени А.Ф. Можайского для службы в космических войсках, учитывая опыт служебной деятельности Р.М. Юсупова

---

в ГШ ВС, он был назначен начальником кафедры оперативно-тактической подготовки ВКА имени А.Ф. Можайского.

В 1989 году Р.М. Юсупов назначен с оставлением в кадрах Вооруженных Сил заместителем директора по научной работе Ленинградского института информатики и автоматизации АН ССС (с 1991 года – Санкт-Петербургский институт информатики и автоматизации (СПИИРАН)), а в 1991 году был избран директором этого института и проработал в этой должности до 2018 года. С 2018 года Р.М. Юсупов – руководитель научного направления СПИИРАН.

Научно-педагогическая деятельность Рафаэля Мидхатовича началась в 1959 с должности научного сотрудника ВКА. Впоследствии он исполнял должности от начальника кафедр и факультетов до директора Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН), одного из ведущих институтов РАН в области информатики и автоматизации.

Р.М. Юсупов является организатором, руководителем и участником крупных фундаментальных и прикладных исследований в области кибернетики и информатики, результаты которых явились важным вкладом в отечественную и мировую науку, образование и экономику страны, ее безопасность и развитие международных научных связей.

В 1958-1976 годах Рафаэль Мидхатович проводил исследования в области теории управления полетом летательных аппаратов, теории самонастраивающихся (адаптивных) систем, теории идентификации и теории чувствительности динамических систем [1 – 3]. Он является одним из основоположников отечественной научной школы по теории чувствительности [4]. Благодаря работам Р.М. Юсупова, его коллег и учеников теория чувствительности вошла в число основных разделов общего курса теории автоматического управления. Анализ чувствительности стал обязательным этапом создания систем управления и их элементов. Его научные результаты в указанных областях используются при проектировании, настройке и испытаниях высокоточных приборов и систем.

Существенное влияние на становление Р.М. Юсупова как ученого в этот период оказал выдающийся ученый в области теории систем управления академик Евгений Павлович Попов. В то время член-корреспондент АН СССР Е.П. Попов был начальником кафедры

---

систем управления полетом ракет и космических аппаратов ВКА имени А.Ф. Можайского, а Р.М. Юсупов был старшим научным сотрудником этой кафедры. Именно по совету и при поддержке Е.П. Попова Р.М. Юсупов стал заниматься теорией и практикой самонастраивающихся систем управления. По этой тематике им в 1963 году была успешно защищена кандидатская диссертация и подготовлен совместно с Е.П. Поповым доклад для Второго конгресса ИФАК в Швейцарии. Развивая работу по данной тематике, Р.М. Юсупов получил пионерские мирового уровня результаты в теории чувствительности систем автоматического управления. Уровень и прикладное значение результатов позволили ему успешно защитить в 1968 году докторскую диссертацию.

Вспоминая эти годы Е.П. Попов в своих мемуарах (Попов Е.П. Воспоминания. М.: Изд. МГТУ им. Н.Э. Баумана. 1996) на стр. 131–132, после перечисления учеников, защитивших докторские диссертации В.М. Пономарева, А.М. Половко, И.П. Пальтова, Д.А. Башкирова, написал: «Наконец еще один мой ученик Р.М. Юсупов защитил в то время докторскую диссертацию по беспонсковым самонастраивающимся системам автоматического управления».

В 1981 году в издательстве «Наука» по теме диссертации совместно с Е.Н. Розенвассером опубликована монография «Чувствительность систем управления». Ее актуальность подтверждается публикацией в США английского перевода этой монографии [4, 5].

С 1976 года научные интересы Р.М. Юсупова связаны с проблемами сбора и обработки информации, геофизической кибернетики, математического моделирования, информатизации, телемедицины, информационной безопасности.

Термин геофизическая кибернетика был предложен Р.М. Юсуповым в 1977 году применительно к новому междисциплинарному научному направлению – теории и практике управления природными (геофизическими) объектами и процессами.

Под его руководством и с его участием в ВКА им. А.Ф. Можайского и в СПИИРАН выполнено более 90 научно-исследовательских работ. Исследования проводились по вопросам повышения эффективности военно-технических систем (ВТС), планирования их развития, создания математического

---

обеспечения, разработки информационно-расчетных систем оценки влияния на ВТС состояния природной среды, создания новых информационных технологий и их использования в различных сферах развития общества [6 – 9].

В области теории моделирования Р.М. Юсупов развивает новое научное направление – методы оценивания качества моделей (адекватность, чувствительность, сложность и т.д.), названное им квалиметрией моделей (моделеметрией). Им сформулированы концептуальные основы квалиметрии моделей, разработаны методы и алгоритмы оценивания адекватности и чувствительности моделей [10].

Избрание Р.М. Юсупова в 1991 году директором СПИИРАН совпало с массовым переходом во всем мире и в нашей стране на новое поколение вычислительной техники – персональные компьютеры. Персональные компьютеры приблизили мощные вычислительные ресурсы к непосредственному потребителю и проникли во все сферы человеческой деятельности. При этом стало ясно, что простая компьютеризация не может обеспечить обществу прорыва на новый технологический уровень [11]. Решением этой проблемы стал переход от простой компьютеризации к более широкой и глубокой информатизации общества.

Р.М. Юсупов является одним из инициаторов развития в России взамен исчерпавшего себя направления – компьютеризация нового весьма актуального научно-практического направления – информатизация общества. Важным вкладом Р.М. Юсупова в развитие этого направления является разработанная им универсальная структура концепции информатизации [12, 13].

Впервые эти концептуальные и научно-методологические основы информатизации были реализованы на практике в 1991-1993 гг. Под руководством Р.М. Юсупова и с его непосредственным участием созданы концепция информатизации Ленинградского экономического региона (1991 г.) и типовая концепция информатизации городского района (1992 г.). Он также является соавтором Концепции информатизации Санкт-Петербурга. С его участием предложена стратегия перехода Санкт-Петербурга в информационное общество, разработан ряд программ и приоритетных проектов для города в области телекоммуникаций, информационной безопасности, телемедицины, в том числе Целевая программа «Электронный Санкт-Петербург», Концепция

---

информационной безопасности исполнительных органов государственной власти Санкт-Петербурга и др.

В развитие этого направления Р.М. Юсуповым разработаны структурная и экономико-математическая модели информационного общества в предположении, что в последнем функционируют секторы материального и информационного (цифрового) производства [14]. С использованием этих моделей им получена новая параметрическая информационная модель развития науки, позволяющая исследовать влияние ряда параметров (объем финансирования, численность научных работников, старение знаний, «утечка умов» и технологий и т.д.) на эволюцию науки [15].

Следует также отметить работы Р.М. Юсупова по оценке влияния информатизации общества на его информационную и национальную безопасность такого общества, созданию и применению интегрированных информационных технологий и систем поддержки принятия решений [16 – 20].

Р.М. Юсупов – бессменный председатель оргкомитета конференций, проводимых в городе на регулярной основе при поддержке Правительства Санкт-Петербурга: «Региональная информатика» (1991-2024) и «Информационная безопасность регионов России» (1999-2023), которые способствуют формированию единого защищенного информационного пространства и информационно-образовательной среды города. Им был организован постоянно действующий семинар по разработке компьютерных моделей развития города и городской семинар при Научном совете по информатизации Санкт-Петербурга «Информатика и автоматизация». Благодаря его усилиям СПИИРАН стал по сути научно-методическим центром информатизации Санкт-Петербурга.

Свою высокоэффективную научную деятельность Р.М. Юсупов успешно сочетает с работой по подготовке высококвалифицированных инженерных и научных кадров. В течение 20 лет в ВКА имени А.Ф. Можайского на должностях начальника кафедры и начальника факультета он готовил высококвалифицированные инженерные кадры для Вооруженных Сил, многие из которых в настоящее время с успехом трудятся в различных структурах, в том числе и в РАН.

С 1991 года он является заведующим базовой кафедрой «Автоматизация научных исследований» Санкт-Петербургского

---

государственного электротехнического университета «ЛЭТИ», а с 2003 по 2016 годы заведовал базовой кафедрой «Прикладная информатика» Санкт-Петербургского государственного университета аэрокосмического приборостроения.

В 1996-1999 годах Р.М. Юсупов был профессором Санкт-Петербургского политехнического университета Петра Великого (СПбПУ), где при его непосредственном участии в 1998 году был создан факультет «Безопасность». Деканом этого факультета Р.М. Юсупов был с 1998 по 2001 гг. В 2009 году он избран заведующим базовой кафедрой «Распределённые интеллектуальные системы автоматизации» СПбПУ.

Профессор Р.М. Юсупов активно участвует в подготовке инженерных и научных кадров высшей квалификации. Он является инициатором создания шести базовых кафедр в ведущих университетах Санкт-Петербурга.

Р.М. Юсупов является почетным доктором Петрозаводского государственного университета, Санкт-Петербургского университета управления и экономики, почетным профессором Военно-космической академии имени А.Ф. Можайского, почетным академиком Академии наук Республики Татарстан.

С целью привития студентам и аспирантам навыков научной и прикладной деятельности в институте функционируют четыре научно-образовательных центра. Компьютерный научно-образовательный центр (с 1998 г.). Научно-образовательный центр «Технологии интеллектуального пространства» (с 2010 г.). Инновационно-образовательный Центр космических услуг (с 2011 г.). Учебный центр для подготовки сертифицированных специалистов в области обработки данных дистанционного зондирования Земли (с 2012 г.).

С этой же целью в институте созданы шесть базовых научно-исследовательских лабораторий при ряде университетов страны.

По инициативе Р.М. Юсупова в здании института созданы два музея. В 1995 году открылся Музей истории всемирно известной школы К. Мая, которая с 1910 по 1976 гг. размещалась в здании института. С 2008 года функционирует Музей истории СПИИРАН. Школа К. Мая образована в 1856 году. Среди выпускников этой школы более 35 членов Российской академии наук и Академии художеств, много выдающихся государственных и общественных

---

деятели России, космонавты дважды Герой Советского Союза Г.М. Гречко и Герой России А.И. Борисенко.

Ученые института на базе этих музеев ведут просветительскую и воспитательную работу со школьниками и студентами Санкт-Петербурга, пропагандируя лучшие научные, педагогические и культурно-нравственные традиции российской школы и науки. В Музее истории СПИИРАН выставлены образцы вычислительной техники прошлых времен.

Среди учеников Р.М. Юсупова 12 докторов наук и 46 кандидатов наук. Он является автором более 500 научных трудов, 40 монографий, учебников и учебных пособий, 19 изобретений.

Р.М. Юсупов ведет большую научно-организаторскую и общественную работу как член Президиума Санкт-Петербургского научного центра РАН (1992-2018); председатель Объединенного научного совета этого центра по информатике, телекоммуникациям и управлению (с 1992 г.), член Бюро Отделения нанотехнологий и информационных технологий РАН (до 2016 г.); член Научного совета при Совете Безопасности РФ (1999-2014); заместитель председателя Научного совета по информатизации Санкт-Петербурга (с 1994 г.); член Совета РАН «Научные телекоммуникации и информационная инфраструктура» (1998-2003); член Совета РАН «Высокопроизводительные вычислительные системы, научные телекоммуникации и информационная инфраструктура» (с 2003 г.); председатель Научно-технического совета по региональным проблемам национальной безопасности (1996-1998); заместитель председателя Научного совета при по информатизации Санкт-Петербурга при Правительстве города (1993-2024); член научных советов по государственным НТП «Перспективные информационные технологии» (1994-1998) и «Информатизация России» (с 1994 г.); член Северо-Западной секции содействия развитию экономической науки РАН (с 1998 г.); член комиссии при Губернаторе Санкт-Петербурга по реформированию научно-технической сферы (1998-1999); член правления Санкт-Петербургского отделения Ломоносовского фонда (с 2004 г.); член Общественного Совета Санкт-Петербурга (2002-2010); сопредседатель Координационного совета Партнерства для развития информационного общества на Северо-Западе России (ПРИОР Северо-Запад) (с 2002 г.); президент Национального общества имитационного моделирования (с 2011 г.); заместитель председателя Общественного

---

совета Комитета по информатизации и связи Правительства Санкт-Петербурга (с 2014 г.); вице-президент Санкт-Петербургского Общества информатики, вычислительной техники, систем связи и управления (с 1991 г.); член Объединённого совета по прикладным наукам и технологическому развитию промышленности Санкт-Петербургского отделения Российской академии наук (с 2024 г.).

Являясь членом Научного совета при Совете Безопасности Российской Федерации, Р.М. Юсупов принимал участие в разработке более десяти проектов государственных документов по обеспечению информационной безопасности, в которые вошли его предложения. В одном из последних документов совета «Основные направления государственной политики в области формирования культуры информационной безопасности личности до 2020 года» учтены два его предложения. Являясь заместителем председателя Научного совета по информатизации Санкт-Петербурга, Р.М. Юсупов, опираясь на разработанные им концептуальные и методологические основы информатизации и развития информационного общества, внес важный вклад в разработку документов, оказавших существенное положительное влияние на эти процессы в Санкт-Петербурге и в стране.

Р.М. Юсупов активно участвовал в работе редакционных советов журналов: «Информатика и ее применения», «Научно-технические ведомости СПбПУ» (председатель редакционного совета серии «Информатика. Телекоммуникации. Управление»), «Информация и космос», «Известия Петербургского университета путей сообщения», «Экономика и управление», «Прикаспийский журнал. Управление и высокие технологии», «Проблемы управления и информатики» (Украина), «Мехатроника, автоматизация и управление», «Информатизация и связь», «Телекоммуникации», «Journal of Intelligent Control» (США) др.

Он является председателем докторского диссертационного совета при СПИИРАН, а также главным редактором журнала «Информатика и автоматизация» («Труды СПИИРАН»). Благодаря его усилиям издаваемый институтом с 2001 г. журнал с 2011 г. включен в список ВАК. С 2014 г. журнал вошел в международную базу данных Scopus, в 2018 г. он включен в топ 100 российских журналов.

Международным признанием научных достижений Р.М. Юсупова являются присуждение ему ряда международных

---

премий, его многочисленные научные публикации (в том числе монографии) в зарубежных издательствах, приглашение в состав программных и организационных комитетов международных конференций (в том числе и в качестве докладчика), выступления с приглашенными докладами на международных конференциях, чтение лекций в зарубежных университетах. За работы в области обеспечения международной информационной безопасности он награжден орденом «Содружество» Межпарламентской ассамблеи государств-участников Содружества независимых государств (2013 г.), Почётным знаком МПА СНГ «За заслуги в области печати и информации» (2018 г.). В 2005 году ему присуждена Международная премия им. Н. Рериха за достижения в области педагогики и просветительства.

За заслуги в период военной службы Р.М. Юсупов награжден орденом «Красной звезды» (1978 г.) и тринадцатью медалями, ему присвоено воинское звание генерал-майор (1980 г.). За заслуги в развитии военной науки, вооружения и техники он избран действительным членом Академии военных наук России (1996 г.), награждён грамотой Военно-научного комитета Вооруженных Сил РФ «За большой вклад в развитие отечественной науки, вооружения, и военной техники» (2012 г.).

За научные достижения и педагогические заслуги Р.М. Юсупов награжден орденом «За заслуги перед Отечеством» IV степени (2005 г.), орденом «Почета» (1999 г.), медалью Совета Безопасности РФ «За заслуги в обеспечении национальной безопасности» (2009 г.; ему присвоено почетное звание «Заслуженный деятель науки и техники РФ» (1984 г.), «Почетный радист СССР» (1974 г.), присуждена ученая степень доктора технических наук (1968 г.), присвоено ученое звание профессора (1974 г.), он награжден Министерством высшего и среднего специального образования СССР Первой премией и медалью за лучшую научную работу (1983 г.), избран членом-корреспондентом Российской академии наук (2006 г.), награжден Почетной грамотой Президента РФ (2015 г.), удостоен звания «Почетный работник науки и высоких технологий Российской Федерации» за значительные заслуги в сфере науки и многолетний добросовестный труд (Приказ Минобрнауки № 38/к-н от 26 июня 2019 г.), за плодотворное сотрудничество с Республикой Татарстан

---

в области науки и образования, активную общественную деятельность награждён орденом «Дуслык» (Дружба) 15 апреля 2024 г.

Он удостоен премий Правительства РФ и Правительства Санкт-Петербурга в области образования (2009 г.), премии Правительства Санкт-Петербурга им. А.С. Попова в области электро- и радиотехники и информационных технологий, награжден Почетным знаком «За заслуги перед Санкт-Петербургом» (2018 г.), Знаком отличия «За заслуги перед Санкт-Петербургом» (2009 г.).

Р.М. Юсупов являлся руководителем работы, удостоенной премии Правительства Российской Федерации в области науки и техники (2022 г.) – за разработку и внедрение комплекса отечественных интеллектуальных наземных транспортно-технологических средств обслуживания судов гражданской авиации в едином цифровом пространстве аэропорта.

Р.М. Юсупов избран членом ряда общественных академий, в том числе Санкт-Петербургской инженерной академии, Академии военных наук, Академии навигации и управления, Российской академии естественных наук, которая присвоила ему за научные, педагогические и общественные заслуги почетное звание «Рыцарь науки и искусств» и наградила одноименным знаком (2004 г.).

За существенный вклад в научно-технологическое развитие Российской Федерации и содействие Российской академии наук в решении возложенных на нее задач Юсупов Р.М. награждён медалью 300 лет Российской академии наук (2024 г.), благодарственным письмом Президента Российской Федерации за вклад в развитие отечественной науки, многолетнюю плодотворную деятельность и в связи с 300-летием со дня основания Российской академии наук (2024 г.).

Благодаря высокому научному кругозору Р.М. Юсупова, его научной прозорливости, а также выдающимся организаторским способностям, в сложных условиях социально-экономических реформ и реформирования Российской академии наук удалось не только сохранить, но и развить и увеличить интеллектуальный потенциал СПИИРАН путём преобразования его в первый на Северо-Западе Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН), который приобрёл статус ведущего научного учреждения не только в области информатики и автоматизации, но и в области экологической и продовольственной

---

безопасности в Российской академии наук, в Санкт-Петербурге, в России и за ее пределами.

Сотрудники Санкт-Петербургского Федерального исследовательского центра Российской академии наук, ученики и коллеги, редакционная коллегия журнала Информатика и автоматизация от души поздравляют Рафаэля Мидхатовича Юсупова с юбилеем и желают ему крепкого здоровья, большого человеческого счастья и новых свершений во славу отечественной науки и на благо России!

### **Основные монографии**

1. Юсупов Р.М. Методы получения информации об управляемом процессе в самонастраивающихся системах. М: Энергия, 1966. 140 с.
2. Козлов Ю.М., Юсупов Р.М. Беспорядковые самонастраивающиеся системы. М.: Наука, 1969. 456 с.
3. Юсупов Р.М., Марков С.И. Функциональные устройства простейших беспорядковых самонастраивающихся систем. М.: Энергия, 1970. 104 с.
4. Розенwasser Е.Н., Юсупов Р.М., Чувствительность систем автоматического управления М. Энергия, 1969. 208 с.
5. Rozenwasser E., Yusupov R. Sensitivity of Automatic Control Systems. CRC Press, Boca Raton, London-New York-Washington, D.C., 1999. 456 p.
6. Юсупов Р.М., Городецкий В.И. и др. Элементы теории испытаний и контроля технических систем. М.: Энергия, 1977. 189 с.
7. Юсупов Р.М., Артемов В.М. и др. Справочник по теории автоматического управления / Под ред. А.А. Красовского. М.: Наука, 1987.
8. Юсупов Р.М., Гаскаров Д.В. и др. Введение в геофизическую кибернетику и экологический мониторинг / Под ред. Р.М. Юсупова СПб.:СПбГУВК, 1998. 166 с.
9. Юсупов Р.М., Городецкий В.И., Розенwasser Е.Н. и др. Методы теории чувствительности в автоматическом управлении. М.Энергия, 1971. 344 с.
10. Розенwasser Е.Н., Юсупов Р.М. Чувствительность систем управления. М.: Наука, 1981. 464 с.
11. Юсупов Р.М., Пальчун Б.П. Оценка надежности программного обеспечения. М.: Наука, 1994. 84 с.
12. Юсупов Р.М., Заболотский В.П. Научно-методологические основы информатизации. СПб.: Наука. 2001. 456 с.
13. Юсупов Р.М., Заболотский В.П. Концептуальные и научно-методологические основы информатизации. СПб.: Наука, 2009. 542 с.

- 
14. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов. М.: Наука, 2006. 410 с.
  15. История информатики и кибернетики в Санкт-Петербурге (Ленинграде). Вып. 1 – 2008, вып. 2 – 2009, вып. 3 – 2012, вып. 4 – 2014 / Под общей ред. и с участием Р.М. Юсупова. СПб.: Наука.
  16. Юсупов Р.М. Информационная безопасность и ее влияние на важнейшие компоненты национальной безопасности (глава 3) // Наука и безопасность России / Под.ред. А.Г. Назарова. М.: Наука, 2000.
  17. Юсупов Р.М. Наука и национальная безопасность. СПб.: Наука, 2006. 302 с.
  18. Юсупов Р.М. Наука и национальная безопасность. 2-е издание, переработанное и дополненное. СПб.: Наука, 2011. 376 с.
  19. Социоинженерные атаки. Проблемы анализа / Под ред. и с участием Р.М. Юсупова. СПб.: Наука, 2016. 349 с.
  20. Микони С.В., Соколов Б.В., Юсупов Р.М. Квалиметрия моделей и полимодельных комплексов. М.: РАН, 2018. 314 с.

В.А. СОЙФЕР, В.А. ФУРСОВ, С.И. ХАРИТОНОВ  
**КАЛМАНОВСКАЯ ФИЛЬТРАЦИЯ ОДНОГО КЛАССА  
ИЗОБРАЖЕНИЙ ДИНАМИЧЕСКИХ ОБЪЕКТОВ**

*Сойфер В.А., Фурсов В.А., Харитонов С.И. Калмановская фильтрация одного класса изображений динамических объектов.*

**Аннотация.** Рассматривается задача оценивания состояния динамического объекта по наблюдаемым изображениям, сформированным оптической системой. Цель исследования состоит в реализации нового подхода, обеспечивающего повышение точности автономного слежения за динамическим объектом по последовательности изображений. Используется векторная модель изображения объекта в виде ограниченного количества вершин (базовых точек). Предполагается, что в процессе регистрации объект удерживается в центральной области каждого кадра, поэтому параметры движения могут описываться в виде проекций на оси системы координат, связанной с оптической осью камеры. Новизна подхода состоит в том, что наблюдаемые параметры (расстояние вдоль оптической оси и угловое положение) объекта вычисляются по координатам заданных точек на изображении объекта. Для оценки состояний объекта строится фильтр Калмана-Бьюси в предположении, что движение динамического объекта описывается совокупностью уравнений поступательного движения центра масс вдоль оптической оси и изменений углового положения относительно плоскости изображения. Приведен пример оценивания углового положения объекта, иллюстрирующий работоспособность предложенного метода.

**Ключевые слова:** динамический объект, оценивание состояний, фильтр Калмана, векторная модель изображения.

**1. Введение.** Одной из важных составляющих обеспечения высокоточного наблюдения и управления динамическими объектами в системах реального времени является восстановление зашумленных данных. Наиболее эффективные известные подходы – это фильтры, минимизирующие среднеквадратичную ошибку и калмановская фильтрация [1, 2]. Фильтр Калмана используется для обработки изображений с начала 1970-х годов [3]. Например, в работе [4] он использовался для оценки позы в реальном времени, в [5] рассматривалась задача оценивания динамики объектов с помощью системы стереовидеокамеры. В работе [6] предложена модификация фильтра Калмана для повышения качества трекинга, в частности, для уменьшения количества срывов сопровождения объектов. В [7] исследовалась иерархическая структура, включающая модули адаптивной фильтрации Калмана, компенсации системных ошибок на основе нейронной сети и взвешенного сглаживания. В [8, 9, 10] предложены гибридные методы, в которых повышение качества оценок достигается совместным использованием нейронной сети и фильтра Калмана. В [11] предложен упрощенный фильтр Калмана для подавления гауссовского шума в системе позиционирования

мобильного устройства. Новая комбинация метода опорных векторов и фильтра Калмана рассматривалась в работе [13]. В работе [14] решались задачи идентификации параметров системы и последующего оценивания состояний с помощью фильтра Калмана. В [15] обсуждались проблемы оценивания параметров и состояний для космических систем с ограничениями параметров движения. В статье [16] рассматривалась проблема отслеживания траектории квадрокоптера в условиях неопределенности. В [17] изучалась возможность использования фильтра Калмана для фильтрации изображений, в частности, для удаления шума на полутонных изображениях. В работе [18] приведены сравнительные оценки производительности и согласованности дискретной и непрерывной реализаций фильтра Калмана. В [19] решалась задача попиксельной калмановской фильтрации изображений в реальном времени с представлением отсчетов изображения моделью авторегрессии первого порядка. В работе [20] описана технология высокоточного определения координат объектов звездного неба по текущим изображениям, регистрируемым в заданном направлении. Показано, что эта технология обеспечивает достаточно высокую точность, однако ее реализация требует значительных вычислительных ресурсов.

В настоящей работе предлагается новый метод оценки параметров движения динамического объекта с использованием векторной модели изображений. Полагаем, что геометрический объект, изображение которого наблюдается, состоит из примитивов: вершин, ребер и граней. Здесь мы ограничимся рассмотрением вершин. Несмотря на кажущуюся простоту, предлагаемая модель позволяет адекватно описать изображения широкого класса техногенных объектов. На рисунке 1 приведены примеры некоторых простейших изображений этого класса.

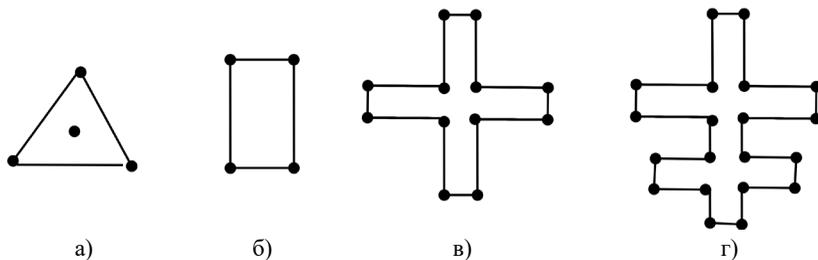


Рис. 1. Примеры векторных примитивов: а) три точки и центр; б) четыре точки; в) 12 точек; г) 20 точек

Здесь для простоты рассмотрим модель изображения в виде четырех точек (вершин). Обобщение решения на  $N$  точек может потребоваться для решения задач идентификации и/или управления динамическим объектом и в этой статье не рассматривается.

В данном случае рассматривается случай, когда наблюдаемый динамический объект описывается линейными дифференциальными уравнениями движения центра масс вдоль оптической оси камеры и углового движения относительно плоскости изображения. Текущие оценки углового положения и расстояния до объекта вдоль оптической оси вычисляются по вершинам наблюдаемого векторного изображения.

Определение координат вершин является самостоятельной задачей, которая успешно решается с использованием широко известных методов нахождения особых точек с использованием дескрипторов Harris, BRISK. MSER: SURF, FAST BRISK и др. Применение конкретного дескриптора связано с классом обрабатываемого изображения и здесь не рассматривается. В данном случае предполагается, что координаты точек на объекте заданы и отмечены каким-либо способом, например, маркерами.

Заметим также, что поскольку технология построена на использовании ограниченного числа (в настоящей задаче четырех) точек изображения это открывает возможность реализации калмановской фильтрации в реальном времени как в дискретном, так и в непрерывном варианте.

## **2. Связь параметров движения с точками на изображении.**

Мы развиваем метод оценивания параметров движения динамического объекта, который представляется в виде ограниченного числа точек на плоскости или сечении объекта плоскостью. Предполагается, что наблюдаемая плоскость изображения объекта удерживается в центре кадра, так что допустимо рассматривать лишь проекцию общего движения на оси системы координат связанной с оптической осью камеры. В частности, полагаем, что движение динамического объекта описывается в пространстве состояний совокупностью уравнений поступательного движения центра масс вдоль оптической оси и уравнений углового движения вокруг центра масс.

Идея метода состоит в том, чтобы формировать выходной вектор в уравнении наблюдения непосредственно по текущему изображению объекта. При этом для определения углового положения объекта используется связь координат точек, заданных на объекте, с зависящими от ориентации объекта координатами проекций этих точек на плоскость изображения камеры. Для определения расстояния

до объекта используется зависимость масштаба изображений от дальности наблюдения. Для описания связи параметров движения объекта, представляемых последовательностью координат точек, заданных на объекте, с последовательностью наблюдаемых координат этих точек на изображении используются следующие системы координат.

Система координат камеры (используется модель камеры обскуры) –  $OXYZ$ . Начало системы координат в точке схода лучей, а плоскость изображения  $хоу$  перпендикулярна оптической оси на фокусном расстоянии  $f$  от начала системы координат  $OXYZ$ . Прямоугольная система координат  $uov$ , связанная с плоскостью  $P_0$  объекта, начало которой является точкой пересечения этой плоскости с оптической осью камеры.

На плоскости  $P_0$  зададим объект в виде четырех точек по осям  $u, v$ , как показано на рисунке 2(а): точки 1, 3 на оси  $ov$  и точки 2, 4 на оси  $ou$ . Сначала рассмотрим решение задачи в предположении, что система координат  $OXYZ$  камеры ориентирована таким образом, что ось  $ou$ , заданная точками 2, 4, параллельна оси  $ох$  системы координат  $хоу$ , т.е. поворот плоскости объекта  $P_0$  совершается в плоскости  $OYZ$  системы координат камеры.

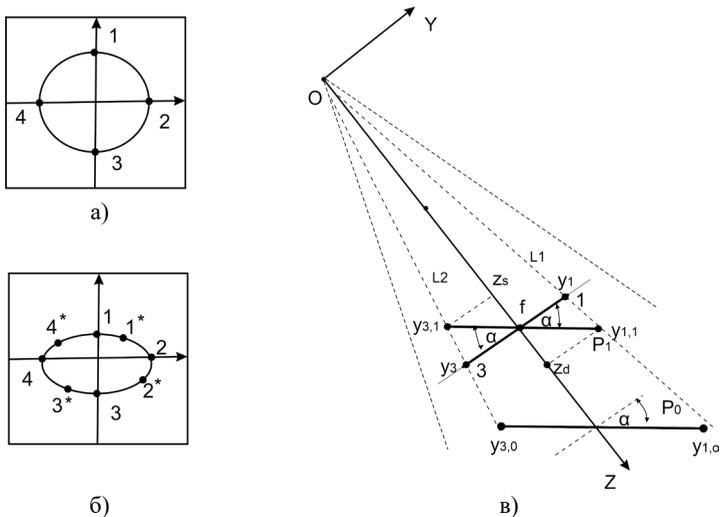


Рис. 2. Схематичное представление точечных изображений: а) фронтального; б) повернутого; в) в виде проекции на плоскость  $YOZ$

В системе координат камеры  $OXYZ$  введем вспомогательную плоскость  $P_1$ , параллельную плоскости  $P_0$  объекта (на рисунке 2(в)) показана проекция на плоскость  $OYZ$ ). В силу сделанных выше предположений точки  $y_{1,1}$  и  $y_{3,1}$  на плоскости  $P_1$  являются проекциями точек  $y_{1,0}$ ,  $y_{3,0}$  плоскости  $P_0$  объекта. На рисунке 2(а) эти точки обозначены цифрами 1 и 3 соответственно.

На рисунке 2(в) видно, что точки  $y_{1,1}$  и  $y_{3,1}$  на плоскости  $P_1$ , являются также соответственными для точек  $y_1, y_3$  на изображении, а также являются точками пересечения с линиями  $L1$  и  $L2$  соответственно. Запишем уравнения прямых  $L1$  и  $L2$  в плоскости  $YOZ$ , проходящих через точки  $(y_1, f)$  и  $(y_3, f)$ , и начало координат камеры  $(0, 0)$ :

$$y_* (Z - f) = f(Y - y_*), \quad (1)$$

где  $y_*$  – общее обозначение точек  $y_1, y_3$ .

Из рисунка 2(в) следует, что для точки  $y_1$   $Z = f + R \sin \alpha$ , а для точки  $y_3$   $Z = f - R \sin \alpha$ , где  $R$  – расстояние этих точек от центра на фронтальном изображении (рисунок 2(а)). Подставляя в (1) указанные значения  $Z$ , а также  $Y = y_{1,1} = R \cos \alpha$  и  $Y = y_{3,1} = R \cos \alpha$  после несложных преобразований получаем:

$$\begin{aligned} f R \cos \alpha &= (f + R \sin \alpha) y_1, \\ f R \cos \alpha &= (f - R \sin \alpha) y_3, \quad \alpha < \pi / 2. \end{aligned} \quad (2)$$

Исключив из системы (2)  $\cos \alpha$ , окончательно получаем:

$$\alpha = \arcsin \left[ \frac{f(y_3 - y_1)}{R(y_3 + y_1)} \right]. \quad (3)$$

Мы получили соотношение, позволяющее по координатам точек 1 и 3 на изображении вычислить угол между объектом и изображением в плоскости  $YOZ$  системы координат камеры.

Координаты точек  $y_1, y_3$  на изображении могут также использоваться для вычисления текущего расстояния от центра объекта до изображения. В частности, из подобия треугольников на рисунке 2(в) следует, что расстояние  $z$  от начала системы координат камеры до центра объекта можно определить как:

$$z = 2Rf(y_1 + y_3)^{-1}. \quad (4)$$

В случае, когда ось  $oi$  не параллельна оси  $ox$  необходимо вначале определить ориентацию плоскости поворота в системе координат камеры. Для этого на изображении необходимо определить координаты двух наиболее удаленных (на расстоянии  $2R$ ) точек, задающих направление большой оси полуэллипсов. Записав уравнение прямой, проходящей через эти точки, можно определить точки (на рисунке 2(б) они обозначены  $1^*$  и  $3^*$ ), принадлежащие разным полуэллипсам. Подставляя координаты  $x_1^*, y_1^*$  и  $x_3^*, y_3^*$  этих точек и длину большой оси  $a = 2R$  в уравнение эллипса, получаем значения малых осей полуэллипсов, которые и являются искомыми координатами  $y_1, y_3$  для определения угла  $\alpha$  и дальности  $z$  по формулам (3), (4).

Заметим, что уравнение большой оси полуэллипсов не обязано проходить через начало системы координат  $хоу$  (центр изображения). Расстояние между этой прямой и началом указанной системы координат можно использовать для коррекции отклонений движения центра масс объекта от оси  $OZ$ .

**3. Построение фильтра Калмана-Бьюси.** Пусть динамика объекта описывается системой дифференциальных уравнений углового движения в плоскости  $YOZ$  и движения центра масс вдоль оси ( $OZ$ ):

$$\begin{aligned} \ddot{\alpha} &= a_\alpha \dot{\alpha} + q_\alpha, \\ \ddot{z} &= a_z \dot{z} + q_z, \end{aligned} \quad (5)$$

где  $\alpha$  – угол между наблюдаемой плоскостью объекта и плоскостью изображения,  $z$  – расстояние от начала системы координат камеры до центра объекта вдоль оптической оси, а  $a_\alpha, q_\alpha, a_z, q_z$  – (по предположению известные) параметры объекта. Полагаем также, что параметры движения объекта  $\alpha$  и  $z$  доступны для «измерения» путем

их вычисления по соотношениям (3), (4) с использованием координат  $y_1, y_3$  на изображении.

Введем обозначения  $x_1 = \alpha$ ,  $x_2 = \dot{\alpha}$ ,  $x_3 = z$ ,  $x_4 = \dot{z}$ , тогда дополнив систему дифференциальных уравнений (5) уравнением измерений, можно записать:

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{Q}, \\ \mathbf{u} &= \mathbf{C}\mathbf{x} + \mathbf{v},\end{aligned}\quad (6)$$

где:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & a_\alpha & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & a_z \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} 0 \\ q_\alpha \\ 0 \\ q_z \end{bmatrix}, \mathbf{C}^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Здесь  $u_1 = \alpha$ ,  $u_2 = z$ ,  $\mathbf{v}$  – шум измерений и вычислений с известной ковариационной матрицей  $\mathbf{R}_v = M\{(\mathbf{v} - \bar{\mathbf{v}})(\mathbf{v} - \bar{\mathbf{v}})^T\} = \text{diag}(\sigma_\alpha^2, \sigma_z^2)$ , а  $q_\alpha$ ,  $q_z$  – параметры, описывающие управляющие воздействия. Легко проверить, что  $\text{Rank}[\mathbf{C}^T, \mathbf{A}^T \mathbf{C}^T, (\mathbf{A}^T)^2 \mathbf{C}^T, (\mathbf{A}^T)^3 \mathbf{C}^T] = 4$ , следовательно, система (6) наблюдаема.

В рамках принятой модели движения угловое движение и движение центра масс можно рассматривать независимо. В частности, для углового движения матричное уравнение представляется в виде:

$$\begin{aligned}\dot{\mathbf{x}}_\alpha &= \mathbf{A}_\alpha \mathbf{x}_\alpha + \mathbf{Q}_\alpha, \\ \alpha &= \mathbf{C}_\alpha \mathbf{x} + v_\alpha,\end{aligned}\quad (7)$$

где:

$$\mathbf{x}_\alpha = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \mathbf{A}_\alpha = \begin{bmatrix} 0 & 1 \\ 0 & a_\alpha \end{bmatrix}, \mathbf{Q}_\alpha = \begin{bmatrix} 0 \\ q_\alpha \end{bmatrix}, \mathbf{C}_\alpha = [1 \ 0], v_\alpha = v_1.$$

Уравнение асимптотически устойчивого фильтра Калмана для канала углового движения, описываемого уравнениями (7), имеет вид:

$$\dot{\hat{\mathbf{x}}}_\alpha = \mathbf{A}_\alpha \hat{\mathbf{x}}_\alpha + \mathbf{K}_\alpha [\alpha^* - \mathbf{C}_\alpha \hat{\mathbf{x}}_\alpha], \quad (8)$$

где:

$$\mathbf{K}_\alpha = \mathbf{P}_\alpha \mathbf{C}_\alpha^T \sigma_{v_1}^{-2} \quad (9)$$

– матричный коэффициент (в данном случае вектор), а  $\alpha^*$  – наблюдаемое значение текущего угла. Фигурирующая здесь  $2 \times 2$  – матрица  $\mathbf{P}$  является решением алгебраического уравнения Риккати:

$$\mathbf{A}_\alpha \mathbf{P}_\alpha + \mathbf{P}_\alpha \mathbf{A}_\alpha^T - \mathbf{P}_\alpha \mathbf{C}_\alpha^T \sigma_{v_1}^{-2} \mathbf{C}_\alpha \mathbf{P}_\alpha = 0. \quad (10)$$

В силу симметрии матрицы  $\mathbf{P}$  решение этого уравнения сводится к решению системы трех уравнений с тремя неизвестными элементами матрицы  $\mathbf{P}$ .

По аналогии нетрудно записать уравнения системы и калмановской фильтрации в канале движения центра масс:

$$\begin{aligned} \dot{\mathbf{x}}_z &= \mathbf{A}_z \mathbf{x}_z + \mathbf{Q}_z, \\ z &= \mathbf{C}_z \mathbf{x}_z + v_z, \end{aligned} \quad (11)$$

$$\mathbf{x}_z = \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}, \quad \mathbf{A}_z = \begin{bmatrix} 0 & 1 \\ 0 & a_z \end{bmatrix}, \quad \mathbf{Q}_z = \begin{bmatrix} 0 \\ q_z \end{bmatrix}, \quad \mathbf{C}_z = [1 \quad 0], \quad v_z = v_2,$$

$$\dot{\hat{\mathbf{x}}}_z = \mathbf{A}_z \hat{\mathbf{x}}_z + \mathbf{K}_z [z^* - \mathbf{C}_z \hat{\mathbf{x}}_z]. \quad (12)$$

$$\mathbf{K}_z = \mathbf{P}_z \mathbf{C}_z^T \sigma_{v_2}^{-2}. \quad (13)$$

$$\mathbf{A}_z \mathbf{P}_z + \mathbf{P}_z \mathbf{A}_z^T - \mathbf{P}_z \mathbf{C}_z^T \sigma_{v_2}^{-2} \mathbf{C}_z \mathbf{P}_z = 0. \quad (14)$$

В данном случае матричный коэффициент  $\mathbf{K}_z$  также является вектором, а матрица  $\mathbf{P}$  в уравнении Риккати (14) –  $2 \times 2$  -матрицей.

Калмановская фильтрация в каналах углового движения и центра масс может осуществляться как отдельно, так и совместно.

**4. Пример.** Рассмотрим пример построения калмановских оценок параметров углового движения динамического объекта, описываемого уравнениями (7). Рассмотрим простой случай движения объекта в плоскости  $OYZ$  системы координат камеры, т.е. в данном примере использование точек 2 и 4 для определения направления оси поворота не понадобится. Будем моделировать собственное движение объекта для двух вариантов начального отклонения угла:  $\alpha_0 = \pi/6$  и  $\alpha_0 = \pi/3$ . Зададим параметры объекта и камеры:

$$\mathbf{A}_\alpha = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{Q}_\alpha = \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_\alpha, \quad \mathbf{C}_\alpha = [1 \quad 0], \quad f = 0.5 \text{ м.}, R = 0.5 \text{ м.}$$

Легко проверить, что система с указанными параметрами управляема и наблюдаема. Для асимптотической устойчивости фильтра Калмана система должна быть асимптотически устойчивой. Допустим сигнал управления  $u_\alpha$  в системе формируется по оценкам состояния в виде  $u_\alpha = \mathbf{L}\hat{\mathbf{x}}$ . Для оптимальной в среднеквадратичном смысле замкнутой системы получаем  $\mathbf{L} = [1, 0.73]$ . Тогда уравнение замкнутой системы в развернутом виде запишется так:

$$\dot{\mathbf{x}}_\alpha = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \mathbf{x}_\alpha - \begin{bmatrix} 0 \\ 1 \end{bmatrix} [1 \quad 0.73] \mathbf{x}_\alpha = \begin{bmatrix} 0 & 1 \\ -1 & -1.73 \end{bmatrix} \mathbf{x}_\alpha. \quad (15)$$

Характеристические числа матрицы замкнутой системы в (15)  $-0.865 \pm j0.502$ , т.е. система устойчива. Фильтр Калмана (8) для замкнутой устойчивой системы (также в развернутом виде):

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & -1.73 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} + \begin{bmatrix} k_\alpha \\ k_\alpha \end{bmatrix} (\alpha^* - \hat{\alpha}), \quad \hat{\alpha} = \mathbf{C}_\alpha \hat{\mathbf{x}} = [1 \quad 0] \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}, \quad (16)$$

где  $\alpha^*$  – текущее значение угла, вычисленное по формуле (3),  $\hat{\alpha}$  – калмановская оценка текущего угла.

Координаты  $y_1, y_3$ , необходимые для вычисления  $\alpha^*$  по формуле (3), здесь определялись по формулам (2), путем

подстановки в них текущих значений угла  $\alpha$ , которые генерировались в процессе моделирования динамики объекта по исходным дифференциальным уравнениям (7) с заданными выше параметрами. В реальном процессе определения координат этих точек по текущим изображениям неизбежны ошибки. Поэтому к вычисленным  $y_1$  и  $y_3$  добавлялся белый шум с дисперсией  $\sigma_{v_1}^2$  от 0.0005 до 0.0015.

Вектор  $K^T = [k_\alpha \quad k_\alpha]$  вычислялся по формуле (9), при этом для определения матрицы  $P$  в соответствии с (10) для каждого значения дисперсии ошибок измерений решалась система трех уравнений:

$$2p_{12} - p_{11}^2\sigma_{v_1}^{-2} = 0, \quad p_{22} - p_{12} - p_{11}p_{12}\sigma_{v_1}^{-2} = 0, \quad 2p_{22} + p_{12}^2\sigma_{v_1}^{-2} = 0. \quad (17)$$

Текущие оценки вектора состояний (16) вычислялись с шагом 0.05 с на интервале 2.5 с. При этом использовался простейший способ замены производных разностями первого порядка. В таблице 1 приведены значения среднеквадратичных ошибок (СКО) оценивания для различных значений отношения сигнал/шум (SNR) в диапазоне от 20 дБ до 30 дБ. Из таблицы видно, что независимо от начального значения угла  $\alpha_0$  ошибки оценивания снижаются с увеличением SNR. При увеличении интенсивности шума на порядок (SNR = 20.0 дБ) СКО оценивания параметра увеличивается в три раза, т.е. сохраняются приемлемые значения. На рисунке 3 в качестве иллюстрации приведены типичные графики процессов, полученных в ходе моделирования для  $\alpha_0 = \pi/3$  при значениях SNR = 30.0 дБ (рисунок 3 (а)) и SNR = 20.0 дБ (рисунок 3(б)).

Таблица 1. Среднеквадратические ошибки оценивания угла

Начальный угол $\alpha_0$	Отношение сигнал/шум (SNR)					
	20.0	22.0	24	26	28	30
$\alpha_0 = \pi/6$	0.067	0.056	0.051	0.047	0.028	0.022
$\alpha_0 = \pi/3$	0.034	0.028	0.021	0.018	0.015	0.011

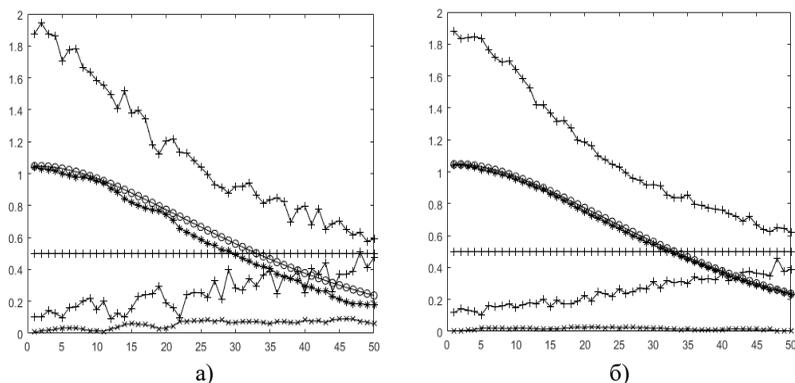


Рис. 3. Графики изменения угла (o), оценок угла (\*), координат  $y_1, y_3$  (+) и ошибок оценивания (x): (a) – SNR = 30.0 дБ, (б) – SNR = 20.0 дБ

Здесь по оси абсцисс маркером (+) отмечены графики изменения зашумленных координат  $y_1, y_3$ . В частности, верхний график изображает изменение координаты  $y_1$ , а нижний – координаты  $y_3$ , горизонтальная линия на уровне 0.5, также отмеченная маркером (+), соответствует случаю фронтального изображения, когда эти координаты совпадают. Маркером (o) отмечен график изменения текущего угла, полученный путем моделирования динамики объекта по исходным дифференциальным уравнениям. а маркером (\*) – калмановская оценка этого процесса. Маркером (x) отмечен график изменения ошибок оценивания. Видно, что метод сохраняет устойчивость к шумам в широком диапазоне значений SNR.

**5. Заключение.** Полученные результаты открывают новые возможности для построения систем обработки изображений и управления динамическим объектом. В частности, этот подход можно использовать для оценки ракурса объекта в технологиях распознавания для сокращения числа используемых эталонов и повышения достоверности принимаемых решений. Предложенный метод оценивания можно также использовать для решения различных задач обнаружения, удаленного управления движением и определения ориентации динамических объектов.

Необходимо заметить, что в настоящей статье пока рассмотрена достаточно простая модель объекта, который описывается совокупностью точек, принадлежащих плоскости. Представляет интерес обобщение этого подхода на случай трехмерных моделей, которые могут быть представлены совокупностью ограниченного

числа точек. При этом число возможных приложений к задачам обнаружения, автономного управления движением и распознавания существенно расширится. Эти направления являются предметом дальнейших исследований.

### Литература

1. Laplante P.A., Neill C.J. A class of Kalman filters for real-time image processing // Proceedings of the Real-Time Imaging VII conference. Santa Clara: SPIE, 2003. pp. 22–29.
2. Schneider F., Easterbrook S.M., Callahan J.R., Holzmann G.J. Validating requirements for fault tolerant systems using model checking // Proceedings of the Third IEEE International Symposium on Requirements Engineering: RE'98. 1998. pp. 4–13.
3. Biemond J., Riesek J., Gerbrands J. A fast Kalman filter for images degraded by both blur and noise // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1983. vol. 31. no. 5. pp. 1248–1256.
4. Xie X., Sudhakar R., Zhuang H. Real-time eye feature tracking from a video image sequence using Kalman Filter // IEEE Transactions on Systems, Man, and Cybernetics 1995. vol. 25. no. 12. pp. 1568–1577.
5. Lippiello V., Siciliano B., Villani L. A new method of image features pre-selection for real-time pose estimation based on Kalman filter // Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. 2002. vol. 1. pp. 372–377.
6. Горбачев В.А., Калугин В.Ф. Разработка алгоритма многообъектного трекинга с необучаемыми признаками сопоставления объектов // Компьютерная оптика. 2023. Т. 47. № 6. С. 1002–1010. DOI: 10.18287/2412-6179-CO-1275.
7. Cao K., Li J., Song R., Li Y. HELM-AD: Hierarchical and efficient attitude determination framework with adaptive error compensation module based on ELM network // ISPRS Journal of Photogrammetry and Remote Sensing. 2023. vol. 195. pp. 418–431. DOI: 10.1016/j.isprs.2022.12.010.
8. Al Bitar N., Gavrilov A. A new method for compensating the errors of integrated navigation systems using artificial neural networks // Measurement. 2021. vol. 168. no. 108391.
9. Chen C., Xiong R., Yang R., Shen W., Sun F. State-of-charge estimation of lithium-ion battery using an improved neural network model and extended Kalman filter. Journal of Cleaner Production. 2019. vol. 234. pp. 1153–1164. DOI: 10.1016/j.jclepro.2019.06.273.
10. Cui Z., Kang L., Li L., Wang L., Wang K. A combined state-of-charge estimation method for lithium-ion battery using an improved BGRU network and UKF // Energy. 2022. vol. 259. no. 124933.
11. Астафьев А.В., Титов А.В., Жизняков А.Л., Демидов А.А. Метод позиционирования мобильного устройства с использованием сенсорной сети BLE-маяков, аппроксимации значений уровней сигналов RSSI и искусственных нейронных сетей // Компьютерная оптика. 2021. Т. 45. № 2. С. 277–285. DOI: 10.18287/2412-6179-CO-826.
12. Baoa T., Zhao Y., Zaidia S.A.R., Xiea S., Yang P., Zhang Z.. A deep Kalman filter network for hand kinematics estimation using sEMG // Pattern Recognition Letters. 2021. vol. 143. pp. 88–94. DOI: 10.1016/j.patrec.2021.01.001.

13. Chen K., Yu J. Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach // *Applied Energy*. 2014. vol. 113. pp. 690–705. DOI: 10.1016/j.apenergy.2013.08.025.
14. Din F. Combined state and least squares parameter estimation algorithms for dynamic systems // *Applied Mathematical Modelling*. 2014. vol. 38. no. 1. pp. 403–412. DOI: 10.1016/j.apm.2013.06.007.
15. Ding F. State filtering and parameter estimation for state space systems with scarce measurements // *Signal Processing*. 2014. vol. 104. pp. 369–380. DOI: 10.1016/j.sigpro.2014.03.031.
16. Zhu X.-Z., Cabecinhas D., Xie W., Casau P., Silvestre C., Batista P., Oliveira P. Kalman–Bucy filter-based tracking controller design and experimental validations for a quadcopter with parametric uncertainties and disturbances // *International Journal of Systems Science*. 2023. vol. 54(1). pp. 17–41. DOI: 10.1080/00207721.2022.2096939.
17. Ionov I., Boldyrikhin N., Cherkesova L., Saveliev V. Filtering grayscale images using the Kalman filter // *E3S Web of Conferences*. 2022. vol. 363. DOI: 10.1051/e3sconf/202236303004.
18. Jwo D.-J., Biswal A. Implementation and Performance Analysis of Kalman Filters with Consistency Validation // *Mathematics*. 2023. vol. 11(3). no. 521. DOI: 10.3390/math11030521.
19. Piovoso M., Laplante P.A. Kalman filter recipes for real-time image processing // *Real-Time Imaging*. 2003. vol. 9. no. 6. pp. 433–439. DOI: 10.1016/j.rti.2003.09.005.
20. Васильюк Н.Н. Коррекция вращательного смаза в изображениях звёзд, наблюдаемых астроинерциальным датчиком ориентации на фоне дневного неба // *Компьютерная оптика*. 2023. Т. 47. № 1. С. 79–91. DOI: 10.18287/2412-6179-CO-1141.

**Соифер Виктор Александрович** — д-р техн. наук, профессор, академик Российской академии наук (РАН), лауреат государственной премии Российской Федерации в области науки и техники, президент, Самарский национальный исследовательский университет имени академика С.П. Королёва. Область научных интересов: компьютерная оптика, обработка изображений, распознавание образов. Число научных публикаций — более 700. soifer@ssau.ru; Московское шоссе, 34, 443086, Самара, Россия; р.т.: +7(846)332-2604.

**Фурсов Владимир Алексеевич** — д-р техн. наук, профессор, профессор кафедры, кафедра суперкомпьютеров и общей информатики, Самарский национальный исследовательский университет имени академика С.П. Королёва. Область научных интересов: теория оценивания, методы обработки изображений и распознавания образов. Число научных публикаций — 310. fursov@ssau.ru; Московское шоссе, 34, 443086, Самара, Россия; р.т.: +7(846)332-2604.

**Харитонов Сергей Иванович** — д-р физ.-мат. наук, доцент кафедры, кафедра наноинженерии, Самарский национальный исследовательский университет имени академика С.П. Королёва. Область научных интересов: дифракционная, квантовая оптика, волноводная оптика. Число научных публикаций — 87. prognoz2007@gmail.com; Московское шоссе, 34, 443086, Самара, Россия; р.т.: +7(846)332-2604.

**Поддержка исследований.** Работа выполнена при финансовой поддержке министерства науки и образования РФ в рамках госзадания Самарского университета (проект FSSS-2021-0016).

V. SOIFER, V. FURSOV, S. KHARITONOV  
**KALMAN FILTER FOR A PARTICULAR CLASS OF DYNAMIC  
OBJECT IMAGES**

*Soifer V., Fursov V., Kharitonov S.* **Kalman Filter for a Particular Class of Dynamic Object Images.**

**Abstract.** We discuss the problem of estimating the state of a dynamic object by using observed images generated by an optical system. The work aims to implement a novel approach that would ensure improved accuracy of dynamic object tracking using a sequence of images. We utilize a vector model that describes the object image as a limited number of vertexes (reference points). Upon imaging, the object of interest is assumed to be retained at the center of each frame, so that the motion parameters can be considered as projections onto the axes of a coordinate system matched with the camera's optical axis. The novelty of the approach is that the observed parameters (the distance along the optical axis and angular attitude) of the object are calculated using the coordinates of specified points in the object images. For estimating the object condition, a Kalman-Bucy filter is constructed on the assumption that the dynamic object motion is described by a set of equations for the translational motion of the center of mass along the optical axis and variations in the angular attitude relative to the image plane. The efficiency of the proposed method is illustrated by an example of estimating the object's angular attitude.

**Keywords:** dynamic object, state estimation, Kalman filter, four-point image.

### References

1. Laplante P.A., Neill C.J. A class of Kalman filters for real-time image processing. Proceedings of the Real-Time Imaging VII conference. Santa Clara: SPIE, 2003. pp. 22–29.
2. Schneider F., Easterbrook S.M., Callahan J.R., Holzmann G.J. Validating requirements for fault tolerant systems using model checking. Proceedings of the Third IEEE International Symposium on Requirements Engineering: RE'98. 1998. pp. 4–13.
3. Biemond J., Riesek J., Gerbrands J. A fast Kalman filter for images degraded by both blur and noise. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1983. vol. 31. no. 5. pp. 1248–1256.
4. Xie X., Sudhakar R., Zhuang H. Real-time eye feature tracking from a video image sequence using Kalman Filter. IEEE Transactions on Systems, Man, and Cybernetics 1995. vol. 25. no. 12. pp. 1568–1577.
5. Lippiello V., Siciliano B., Villani L. A new method of image features pre-selection for real-time pose estimation based on Kalman filter. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. 2002. vol. 1. pp. 372–377.
6. Gorbachev V.A., Kalugin V.F. [Development of a multi-object tracking algorithm with untrained features of object matching]. Computer Optics – Komp'yuternaja optika. 2023. vol. 47. no. 6. pp. 1002–1010. DOI: 10.18287/2412-6179-CO-1275. (In Russ.).
7. Cao K., Li J., Song R., Li Y. HELM-AD: Hierarchical and efficient attitude determination framework with adaptive error compensation module based on ELM network. ISPRS Journal of Photogrammetry and Remote Sensing. 2023. vol. 195. pp. 418–431. DOI: 10.1016/j.isprs.2022.12.010.

8. Al Bitar N., Gavrilov A. A new method for compensating the errors of integrated navigation systems using artificial neural networks. *Measurement*. 2021. vol. 168. no. 108391.
9. Chen C., Xiong R., Yang R., Shen W., Sun F. State-of-charge estimation of lithium-ion battery using an improved neural network model and extended Kalman filter. *Journal of Cleaner Production*. 2019. vol. 234. pp. 1153–1164. DOI: 10.1016/j.jclepro.2019.06.273.
10. Cui Z., Kang L., Li L., Wang L., Wang K. A combined state-of-charge estimation method for lithium-ion battery using an improved BGRU network and UKF. *Energy*. 2022. vol. 259. no. 124933.
11. Astafiev A.V., Titov D.V., Zhiznyakov A.L., Demidov A.A. A method for mobile device positioning using a sensor network of BLE beacons, approximation of the RSSI value and artificial neural networks. *Computer Optics – Komp'juternaja optika*. 2021. vol. 45. no. 2. pp. 277–285. DOI: 10.18287/2412-6179-CO-826. (In Russ.).
12. Baoa T., Zhao Y., Zaidia S.A.R., Xiea S., Yang P., Zhang Z. A deep Kalman filter network for hand kinematics estimation using sEMG. *Pattern Recognition Letters*. 2021. vol. 143. pp. 88–94. DOI: 10.1016/j.patrec.2021.01.001.
13. Chen K., Yu J. Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach. *Applied Energy*. 2014. vol. 113. pp. 690–705. DOI: 10.1016/j.apenergy.2013.08.025.
14. Din F. Combined state and least squares parameter estimation algorithms for dynamic systems. *Applied Mathematical Modelling*. 2014. vol. 38. no. 1. pp. 403–412. DOI: 10.1016/j.apm.2013.06.007.
15. Ding F. State filtering and parameter estimation for state space systems with scarce measurements. *Signal Processing*. 2014. vol. 104. pp. 369–380. DOI: 10.1016/j.sigpro.2014.03.031.
16. Zhu X.-Z., Cabecinhas D., Xie W., Casau P., Silvestre C., Batista P., Oliveira P. Kalman-Bucy filter-based tracking controller design and experimental validations for a quadcopter with parametric uncertainties and disturbances. *International Journal of Systems Science*. 2023. vol. 54(1). pp. 17–41. DOI: 10.1080/00207721.2022.2096939.
17. Ionov I., Boldyrikhin N., Cherckesova L., Saveliev V. Filtering grayscale images using the Kalman filter. *E3S Web of Conferences*. 2022. vol. 363. DOI: 10.1051/e3sconf/202236303004.
18. Jwo D.-J., Biswal A. Implementation and Performance Analysis of Kalman Filters with Consistency Validation. *Mathematics*. 2023. vol. 11(3). no. 521. DOI: 10.3390/math11030521.
19. Piovoso M., Laplante P.A. Kalman filter recipes for real-time image processing. *Real-Time Imaging*. 2003. vol. 9. no. 6. pp. 433–439. DOI: 10.1016/j.rti.2003.09.005.
20. Vasilyuk N.N. Correction of rotational blur in images of stars observed by an astroinertial attitude sensor against the background of the daytime sky. *Computer Optics – Komp'juternaja optika*. 2023. vol. 47. no. 1. pp. 79–91. DOI: 10.18287/2412-6179-CO-1141. (In Russ.).

**Soifer Victor** — Ph.D., Dr.Sci., Professor, Academician of RAS, Recipient of the Russian Federation State Prize in science and technology, President, S.P. Korolev Samara National Research University. Research interests: computer optics, image processing, pattern recognition. The number of publications — 60; see 700. soifer@ssau.ru; 34, Moscow Hgw., 443086, Samara, Russia; office phone: +7(846)332-2604.

**Fursov Vladimir** — Ph.D., Dr.Sci., Professor, Professor of the sub-department, Sub-department of supercomputers and informatics, S.P. Korolev Samara National Research University. Research interests: estimation theory, methods of image processing and pattern

recognition. The number of publications — 310. fursov@ssau.ru; 34, Moscow Hgw., 443086, Samara, Russia; office phone: +7(846)332-2604.

**Kharitonov Sergey** — Ph.D., Dr.Sci., Associate professor of the sub-department, Sub-department of nanoengineering, S.P. Korolev Samara National Research University. Research interests: diffractive and quantum optics, waveguide optics. The number of publications — 87. prognoz2007@gmail.com; 34, Moscow Hgw., 443086, Samara, Russia; office phone: +7(846)332-2604.

**Acknowledgements.** This work was supported by the Ministry of Science and Higher Education within the government project No. FSSS-2021-0016 of Samara University (Multi-spectral imaging).

Е.А. ШАЛФЕЕВА, В.В. ГРИБОВА  
**ВОПРОСЫ СОЗДАНИЯ МАШИНОПОНИМАЕМЫХ SMART-  
СТАНДАРТОВ НА ОСНОВЕ ГРАФОВ ЗНАНИЙ**

*Шалфеева Е.А., Грибова В.В.* **Вопросы создания машино понимаемых SMART-стандартов на основе графов знаний.**

**Аннотация.** Развитие цифровой трансформации требует широкого использования новых технологий в документах по стандартизации. Одной из задач является создание стандартов с машино понимаемым содержанием, которые позволяют использовать цифровые документы на различных этапах разработки и производства без необходимости участия человека-оператора. Целью данной работы является описание подхода для создания и перевода в машино понимаемое представление нормативных документов отрасли для дальнейшего их использования в программных сервисах и системах. Содержимое SMART-стандарта бывает трех видов: машино читаемое, машино интерпретируемое и машино понимаемое. Для формализации данных и знаний при решении различных задач активно используются графы знаний. Предложен новый двухуровневый подход для создания и перевода в машино понимаемое представление нормативных документов как графов знаний. Подход определяет два вида интерпретации такого документа (человекочитаемость и машино понимаемость) через два связанных формата: граф, каждый семантический узел которого представляет текст на естественном языке, и сеть понятий и строгих связей. Каждому узлу «человекочитаемого» графа соответствует (в общем случае) поддерево машино понимаемого графа знаний. В качестве основы для обеспечения преобразования одной формы представления SMART-стандарта в другую форму служат LLM модели, дополняемые специализированным адаптером, полученным в результате дообучения с помощью подхода Parameter-Efficient Fine-Tuning. Установлены требования к набору проблемно- и предметно-ориентированных инструментальных средств формирования графов знаний. Показана концептуальная архитектура системы поддержки решения комплекса задач на основе SMART-документов в виде графов, установлены принципы реализации программных компонентов, работающих со знаниями, для интеллектуальных программных сервисов.

**Ключевые слова:** SMART-стандарт, нормативный документ, машино понимаемое представление, граф знаний, двухуровневое представление, LLM модели.

**1. Введение.** В настоящее время профессиональная деятельность в социально-экономической, технико-технологической, правовой, медицинской и других сферах невозможна без использования различных типов нормативных документов (рекомендации, своды правил, ГОСТы, инструкции, технические задания, технические условия и др.). Они концентрируют накопленный опыт и знания, которые необходимы для работы и принятия решений; устанавливают требования к продукции, услугам, процессам и системам; помогают обеспечить качество работ, эффективность процессов, повысить безопасность и надежность. Число таких документов непрерывно растет, они представляются в различных

текстовых форматах (.doc, .txt, .pdf и др.), могут включать таблицы, рисунки, формулы, графики.

Традиционные методы работы с документами имеют ограниченные возможности автоматизации их обработки, управления, применения, и, как правило, не совместимы с современными ИТ-технологиями. Такие документы понятны специалистам предметной области, но не являются машинопонимаемыми или машиноисполняемыми при решении различных задач отрасли.

Именно поэтому актуальной является разработка так называемых SMART-стандартов (от англ. SMART: Standards Machine Applicable, Readable and Transferable) [1, 2, 3], которые являются неотъемлемой частью Индустрии 4.0 и соответствует четвертому уровню развития цифровизации в сфере стандартизации по классификации ИСО/МЭК. Главная особенность этого уровня – достижение такого качества стандартов в электронном представлении (изложении), которое делает их машинопонимаемыми и машиноисполняемыми, то есть, наряду с возможностью чтения человеком, дает возможность обработки и использования информационными и киберфизическими системами, минуя человека [4]. Понимая исключительную важность и актуальность создания таких стандартов, 23 октября 2023 года Росстандарт утвердил первый из серии предварительных национальных стандартов на умные стандарты (ПНСТ 864-2023. Наименование: Умные (SMART) стандарты. Общие положения. Статус: Принят. Дата введения: 01.02.2024) [4]. Этот нормативный документ существенно влияет на цифровизацию экономики в целом и определяет разработку таких стандартов в качестве одной из приоритетных задач для всех отраслей.

Развитие цифровой трансформации требует создания стандартов с машинопонимаемым содержанием, которые позволят использовать умные документы на различных этапах разработки и производства без необходимости участия человека-оператора. В работах [5, 6] отмечается, что, несмотря на высокую актуальность, исследования в области создания SMART-стандартов недостаточно освещены в научных публикациях и рассматриваются сообществом исследователей как весьма специфическая тема. Их создание является сложной и важной научной задачей.

Целью данной работы является описание подхода для создания и перевода в машинопонимаемое представление нормативных документов отрасли для дальнейшего их использования в прикладных программных системах.

**2. Материалы: SMART-стандарты.** Умные стандарты (в рамках использования цифровых технологий в документах по стандартизации) представляют собой взаимосвязанные информационные модели, пригодные к автоматизированной обработке.

Согласно ПНСТ 864-2023 «Умные (SMART) стандарты. Общие положения» содержимое SMART-стандарта бывает трех видов: машиночитаемое, машиноинтерпретируемое и машинопонимаемое.

Первый формат – машиночитаемый, он нужен для обработки автоматизированными средствами и для представления в воспринимаемой человеком форме, с его помощью специалист может ознакомиться с текстом документа и проанализировать его самостоятельно. Как правило, это документы в форматах DOCX, PDF, HTML.

Второй тип – машиноинтерпретируемый, позволяет создавать человеко-ориентированные аналитические сервисы, в частности выделять из документа отдельные формулировки-требования. Применяет их в таком случае по-прежнему человек, но SMART-стандарты и созданные на их базе SMART-сервисы позволяют сэкономить много ресурсов на поиске, анализе и подборе этих требований под конкретную задачу, а также снизить количество ошибок при передаче информации и адресно доставить те или иные формулировки к рабочему месту специалиста. Такие человекочитаемые требования и другие SMART-данные можно экспортировать во внешнее прикладное программное обеспечение и автоматизировать проверку их актуальности. Как правило, это документы в форматах XML, XHTML. Методы автоматической интерпретации разнообразны: от подключения лингвистических методов до семантико-графового моделирования [7, 8, 9].

Третий тип – машинопонимаемый, он позволяет передавать в другие информационные системы и напрямую на оборудование машинопонимаемые данные, а также создавать машиноориентированные SMART-сервисы. Однако для качественных изменений в производительности труда мало цифровизировать отдельные предприятия – нужно, чтобы они взаимодействовали друг с другом в информационной среде. И здесь на помощь приходит мощный инструмент кооперации: стандартизированные документы в SMART-формате позволяют разным организациям обмениваться более полными данными, в т.ч. через коммуникацию информационных систем без участия человека. Назначение SMART-документов – быть не просто контентом, а информационной базой и базовым

компонентом создаваемых умных человеко- и машиноориентированных сервисов и специализированных программных систем.

Стандарты нормируют деятельность специалистов всех областей, и их применяют для разных профессиональных задач. Спектр задач охватывает:

- технический и нормативный контроль объектов и условий их существования,
- диагностику, определение нарушений и их причин,
- определение критичности состояния (угроз),
- определение потенциала устранения, планов исправления и доработки,
- поиск сведений (в т.ч. с фильтрами) или ответов на вопросы,
- критику предлагаемых решений,
- прогноз изменения состояния, мониторинг,
- подбор средств воздействия, планов построения, обработки, возделывания, и многое другое.

Стандарт, в отличие от художественных произведений, призван быть однозначным:

- каждое понятие ассоциируется с его точным определением и, возможно, местом в иерархии от общих понятий; возможно, сопровождается распространенными синонимами;
- каждая формулировка включает нормирующую или объясняющую «связку» понятий, условия, параметры, степень строгости, обязательности.

В настоящее время различные отрасли только формируют подходы и форматы к созданию SMART-стандартов. Именно поэтому исследования в этом направлении [1, 3, 10, 11] являются исключительно важной научной задачей.

В работе [12] предложен модульный подход к формированию содержимого SMART-документа. На основе существующих подходов к модульности проектируется такая концепция для формул и, в частности, разработка модуля формул. Идентифицируются и структурируются описательные элементы формул. Такие модули служат шаблоном для будущей документации формул в стандартах XML и потенциально применимы в рамках предоставления стандартов на основе контента без необходимости предоставления целых документов.

**3. Методы: графы знаний.** Для формализации данных и знаний при решении различных задач активно используются графы знаний; они используются для генерации рекомендаций, аналитических сведений, поддержки принятия решений, предиктивной аналитики. Одним из ключевых преимуществ графов знаний, специфичных для предметной области, является их возможность описывать сложные семантические связи предметной области, и, таким образом, обеспечить возможность отвечать на специфические вопросы предметной области, выявлять скрытые связи между элементами, т.е. они обеспечивают представление семантического контента в машинопонимаемом формате.

Графы знаний могут быть «общими», а также ориентированными на конкретные сценарии использования или задачи, второе считается наиболее актуальным [11, 13, 14]. Примерами таких сфер приложения графов знаний являются: обнаружение и предотвращение финансового мошенничества; оценка финансовых рисков; управление на соответствие нормативным требованиям; улучшение обслуживания, поддержки клиентов и предоставления им точной и контекстно-зависимой помощи с интеграцией данных о клиентах, информации о продуктах и вспомогательной документации; оптимизация грузовых перевозок, управление транспортными средствами; анализ дорожного движения в режиме реального времени; анализ исторических закономерностей о погоде; помощи прокладки оптимальных маршрутов для транспортных средств и многие другие.

Графы знаний активно используются и в медицине, их построение является «горячей точкой» исследований в области искусственного интеллекта. Технология построения и использования графов знаний имеет широкие перспективы применения в данной области [14, 15].

Построение графов знаний неуклонно прогрессирует от ручного до полуавтоматического и далее – к автоматическому их построению [15, 16, 17]. Для этого активно применяются различные технологии и подходы обработки естественного языка. Отмечается, что автоматизация создания графов знаний имеет преимущества, среди которых: снижение трудозатрат на проведение всех этапов построения графов знаний, улучшение функциональной совместимости, поддержка принятия решений в режиме реального времени, масштабируемость решений. Также отмечается сложность автоматизации формирования экспертных знаний. При создании графа знаний для конкретной предметной области необходимо учитывать

сложные семантические связи разных типов между объектами. Недостаточно точное и полное представление или упрощение семантических связей может привести к неэффективному процессу поиска решения [17].

В настоящее время для построения графов знаний активно используются большие языковые модели (Large Language Models – LLM) [14, 18]. Однако их известной проблемой являются «галлюцинации». Языковая модель, обученная на существующих данных, может не иметь доступа к последним результатам текущих исследований. Отсутствие актуальной информации может препятствовать способности модели генерировать контекстуально релевантные ответы, что может привести к устаревшим или неполным выводам, невозможно закодировать все доступные знания в параметрах LLM для всестороннего понимания.

Гипотеза исследования состоит в целесообразности использования графов знаний для разработки машинопонимаемых документов и стандартов.

Их создание может осуществляться вручную, для этого привлекаются эксперты предметной области и применяются инструменты редактирования графов знаний, часто управляемые онтологиями знаний предметной области.

Для автоматизации построения графов знаний используют методы анализа естественного языка; в настоящее время актуально использование LLM. Однако, несмотря на некоторые положительные результаты, проблема далека от решения и требует создания новых моделей, методов и подходов к созданию SMART-стандартов.

LLM, обученные (на существующих текстах по нужной тематике) пониманию вопросов и формированию ответов на вопросы, дорогостоящи и ориентированы на интерактивное взаимодействие с пользователем. Альтернативой является метод двух связанных предметно-ориентированных форматов – связанных предметно-ориентированных языков. Один подобен структурированному, упорядоченному, размеченному тексту. Второй – детализирует сведения в текстах смысловых элементов-концептов и связей. Именно детализированный предметно-ориентированный формат позволит обращаться к нужным фрагментам, манипулировать ими при решении либо объяснении некоторой задачи. Смысловые элементы записываются как узлы графов знаний, а связи образуют именованные дуги (ребра).

**4. Методы: описание графового подхода к созданию SMART-стандартов.** Главными свойствами умных стандартов является их

понятность человеку (с соответствующими профессиональными знаниями), пригодность и достаточность для стандартизируемой информации, а также «понятность» для программных сервисов, используемых при решении профессиональных задач в предметной области.

Интерес в SMART-стандартах, прежде всего, представляют знания. Здесь графы могут обеспечить понятное и естественное представление знаний, например, с именованнием дуг-связей узлов-понятий, с иерархией вложенности понятий, что отличается от других моделей представления знаний (например, в виде триплетов или правил).

Наличие редакторов графов знаний, дающих возможность при порождении новых элементов знаний выбирать их смысл и тип их «предметной» связи (предопределяемых онтологией знаний) с уже порожденными элементами знаний, делает более строгим процесс создания и редактирования и упрощает чтение таких структур. Некоторые редакторы графов знаний (с представлением информации «сверху-вниз») дополняют именованные дуги спецификаторами и ограничителями (и могут связывать с правилами интерпретации).

Но очевидно, что все профессиональное сообщество не готово перейти к чтению и особенно созданию документов, представленных такими графами. Представление документа в таком формальном виде, где каждому узлу графов знаний соответствует смысловые элементы (как атомарные элементы), а «предметным» связям – именованные дуги/ребра этих графов, эксперту предметной области, не всегда удобно. Понятность человеку и «понятность программам» являются отчасти противоположными целями: максимально структурированная для программной обработки информация не удобна для восприятия человеком.

Поэтому предлагается обеспечить два вида интерпретации нормативного документа (человекочитаемость и машинопонимаемость) через два связанных формата (два предметно-ориентированных языка). Один формат (машинопонимаемый уровень) – сеть понятий и строгих формулировок (часть онтологии предметной области), другой (человекочитаемый уровень) – граф понятий и связей, каждый узел которого представляет текст на естественном языке, соответствующий типу узла-контейнера. «Человекочитаемое» представление – некий «компромисс» между строгостью, однозначностью и удобством чтения и создания человеком; этому графу ставится в соответствие детализированное и формальное представление, предназначенное для обработки программными

сервисами. Иными словами каждому узлу «человекочитаемого» графа соответствует (в общем случае) поддерево машинопонимаемого графа знаний.

Формирование графовых знаний каждого из этих двух «уровней» должно осуществляться по своей структуре, но с использованием единой отраслевой лексики, присущей действующим нормативно-правовым актам (ограниченный русский язык).

В традициях онтологического подхода (где под управлением одной онтологии создается любое число баз знаний) на предметно-ориентированном языке одного и второго уровня может быть написано произвольное множество (класс) документов и их версий.

Схема общего подхода приведена на рисунке 1.

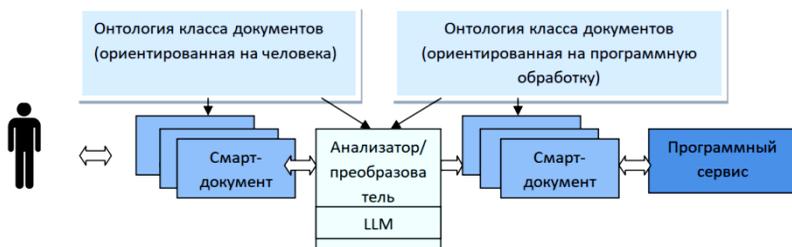


Рис. 1. Схематическое представление онтологического подхода к SMART-стандартам

**5. Результаты: общая архитектура программного комплекса.** Общая архитектура программного комплекса состоит из предметно и проблемно-независимых инструментальных средств, а также проблемно- и предметно-ориентированных. К таким инструментальным средствам относятся редактор онтологической (графовой) структуры, редакторы тезаурусов или баз терминов предметной области, редакторы графов знаний под управлением онтологии и с поддержкой лексикона из тезаурусов, визуализаторы человеко- и машинопонимаемых графовых документов отрасли в виде графов знаний.

В качестве инструментальной среды предлагается использовать платформу IACRaaS. Для формального представления онтологической структуры или структурных частей (модулей) онтологии предметной области применяется язык информации различных уровней общности (ИРУО, описан на <https://iacraas.dvo.ru/infores>) [19]. Онтологической структура строится как ориентированный граф (орграф) со

специальной разметкой, описывающей характеристики узлов (понятий) и отношений. Узлы орграфа делятся на два типа: нетерминалы и терминалы. Дуги орграфа описывают направленные отношения между парами понятий. Согласно нотации языка ИРУО отношения характеризуются набором спецификаторов, которые определяют правила порождения понятий и отношений в графах (графовых документах):  $\text{copy}(=)$  – "копия";  $\text{one}(!)$  – "в точности один";  $\text{set}(+)$  – "непустое множество";  $\text{seq}(\Delta)$  – "непущая последовательность";  $\text{copymm}(=[])$  – "возможное отсутствие";  $\text{onemmm}(!)$  – "ноль или один";  $\text{setmm}(=[+])$  – "возможно пустое множество";  $\text{seqmm}([\Delta])$  – "возможно пустая последовательность".

В качестве редакторов/просмотрщиков графов знаний могут выступать и другие редакторы, реализованные вне платформы IACPaaS, поскольку все графы знаний, созданные на платформе, экспортируются в Json-формат (и импортируются из него на платформу).

Предметно-ориентированными средствами являются терминологические базы знаний для каждой предметной области, для которой создаются умные стандарты (графовые документы отрасли), преобразователи человекопонимаемого стандарта в машинопонимаемый, все интерпретирующие и обрабатывающие их специализированные сервисы, размещаемые на специализированных порталах SMART-стандартов, в частности, на платформе IACPaaS.

В качестве основы для обеспечения преобразования одного формата представления SMART-стандарта в другой служат LLM модели. В качестве такой модели взята доступная модель Llama2. Для каждой предметной области LLM дополняется специализированным адаптером – проблемно-ориентированной надстройкой, полученной в результате дообучения LLM модели с помощью подхода PEFT (Parameter-Efficient Fine-Tuning), обеспечивающего тонкую настройку модели на требуемый класс задач на основе датасета из примеров, снижая при этом требования к памяти и вычислительным мощностям.

Анализатор текста, обращаясь к LLM модели с заданным адаптером, обеспечивает перевод текста из одного формата представления SMART-стандарта в другой. Результат формируется в формате JSON и экспортируется/импортируется на платформу IACPaaS. Удобные и понятные экспертам редакторы онтологий, средства просмотра и редактирования SMART-стандартов экспертами производятся на платформе IACPaaS, где создаются соответствующие порталы знаний (рисунок 2).

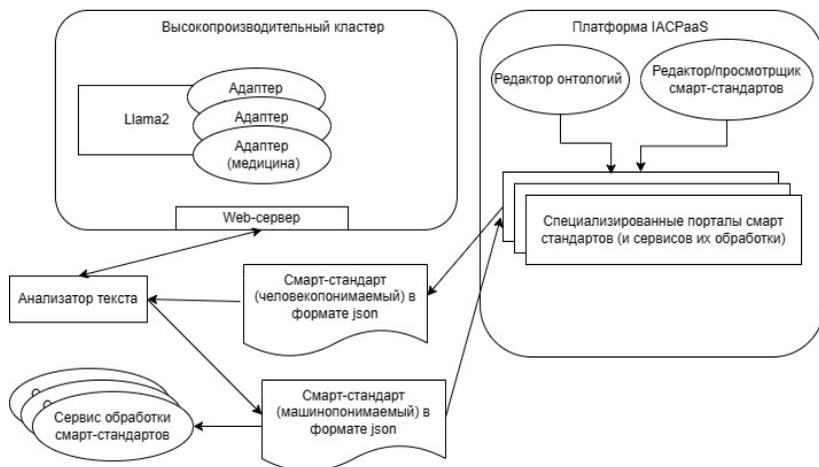


Рис. 2. Общая архитектура программного комплекса

Отдельной задачей является обеспечение качества преобразования человекопонимаемого SMART-стандарта в машинопонимаемый. В настоящее время данная задача исследуется. В качестве решений мы видим перманентное улучшение адаптера через расширение датасетов с примерами, использованием различных доступных LLM моделей, например, Mistral, Mixtral и др., и сравнение полученных результатов, а также настройкой различных гиперпараметров модели, например, получение одинакового результата при задании высокой «температуры» (температура – стандартный гиперпараметр LLM).

**6. Результаты: примеры человеко- и машинопонимаемых SMART-стандартов.** Рассмотрим примеры двухуровневого представления SMART-стандартов для предметной области «медицина» и «строительство».

Основным регламентирующим документом в медицине являются клинические рекомендации, которые определяют правила профилактики, диагностики, лечения, мониторинга и прогноза состояний пациента по различным заболеваниям [20, 21]. Каждая клиническая рекомендация представлена многостраничным документом с определенной в его оглавлении структурой. В общем случае – это текстовый документ, в некоторых случаях содержащий диаграммы и таблицы. Размеры этих документов варьируются в зависимости от заболевания и доходят (по одному заболеванию)

до 250 страниц текста (например, методические рекомендации по лечению covid-19, 18-ая версия [20]).

Рассмотрим фрагмент клинической SMART-рекомендации, вносимой через интерфейс семантического редактора, для описания этиологии и патогенеза (рисунок 3). Его компоненты требуются для решения разных задач (оценка риска, диагностика, прогноз, мониторинг) [21].

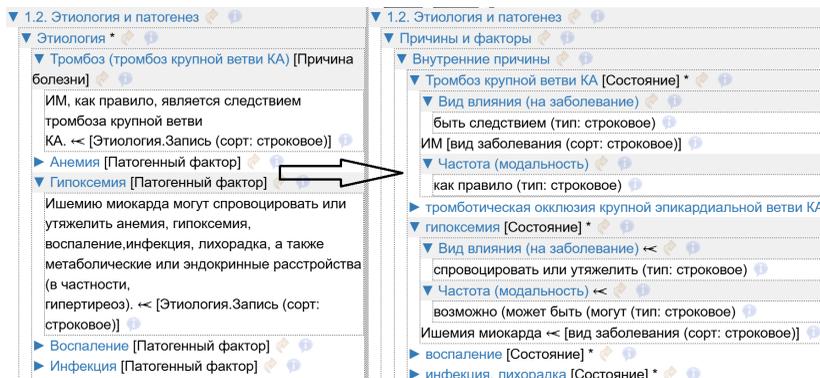


Рис. 3. Фрагмент двухуровневого представления клинической рекомендации

При обработке машинопонимаемой части стандарта (рисунок 3, справа) для потенциального запроса «Каковы наиболее частые внутренние причины этого заболевания?» ответ в терминах онтологии мог бы быть построен так:

найти в документе раздел «1.2. Этиология и патогенез»,  
в разделе – подсеть «Причины и факторы. Внутренние причины»,

в подсети – все узлы типа-[состояние] с внутренним узлом типа-[Вид влияния (на заболевание)] с внутренним значением «быть следствием», и с внутренним узлом типа-[частота (модальность)], внутри которого значение = «как правило»,

и вывести значения таких найденных узлов типа-[состояние].

Были проведены эксперименты с вносимой через семантический редактор документации от технического комитета по стандартизации в сфере строительства (рисунок 4).

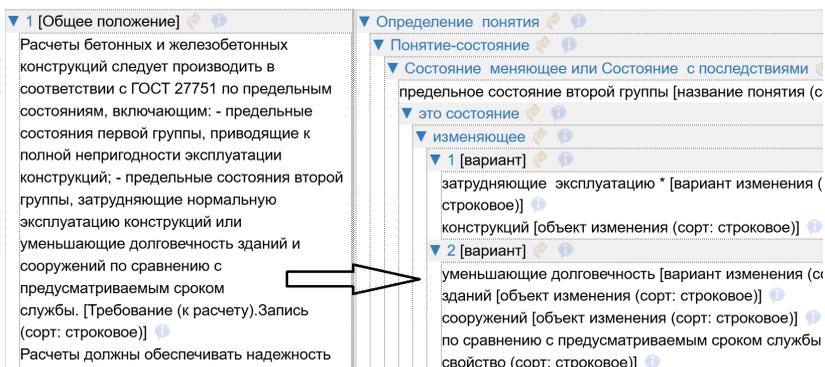


Рис. 4. Фрагмент двухуровневого представления SMART-свода правил

Пример потенциального запроса на способ расчета на изгиб бетонной конструкции без арматуры в терминах онтологии свода правил (рисунок 4, справа) может быть построен так: «Найти в документе узел [Требование] с внутренним узлом [Наименование объекта] = «бетонные конструкции» и с узлом [тип воздействия] = «изгиб»; найти в подсети найденного узла [Требование] узел [предельная величина], а в его подсети узел [формула] и показать его значение.

**7. Анализ текста с помощью LLM.** В общем случае процесс формирования графа знаний на основе текста является сложной и в целом пока неразрешимой проблемой, особенно для текстов, которые содержат сложные семантические связи. Примером является медицина. Вместе с тем, декомпозиция целостного текста на фрагменты, соответствующие элементам человекопонимаемого SMART-стандарта, упрощает процесс его семантического анализа. Каждому фрагменту человекочитаемого стандарта соответствует фрагмент онтологии машинопонимаемого стандарта.

Приведем пример разбора медицинского текста (из узла-контейнера в человеко-ориентированном графовом формате) с помощью обученной LLM. Возьмем фрагмент лечения COVID-19 с помощью препарата Фавипиравир [20]. Исходный текст:

«Механизм действия: Ингибирует РНК-зависимую РНК-полимеразу вируса SARS-CoV-2. Схема назначения. Таблетки. Для пациентов:

- с массой тела <75 кг: по 1600 мг 2 р/сут в 1-й день и далее по 600 мг 2 р/сут в 2-10-й дни;
- с массой тела 75 кг и более: по 1800 мг 2 раза/сут в 1-й день, далее по 800 мг 2 раза/сут со 2 по 10 день».

Результат обработки в формате JSON имеет следующий вид (Листинг 1):

```
{
  "name": "COVID-19",
  "meta": "Заболевание",
  "successors": [
    {
      "name": "Фавипиравир",
      "meta": "Рекомендуемый препарат",
      "successors": [
        {
          "name": "Механизм действия",
          "successors": [
            {
              "name": "Ингибирует РНК-зависимую РНК-полимеразу вируса SARS-CoV-2"
            }
          ]
        }
      ]
    }
  ],
  {
    "name": "Схема назначения",
    "successors": [
      {
        "name": "1",
        "meta": "Вариант назначения",
        "successors": [
          {
            "name": "Таблетки",
            "successors": [
              {
                "name": "Суточная дозировка",
                "successors": [
                  {
                    "name": "Условие для дозировки",
                    "successors": [
                      {
                        "name": "Тип условия",
                        "successors": [
                          {
                            "name": "Блок критериев",
                            "successors": [
                              {
                                "name": "Масса тела < 75 кг",
                                "meta": "Критерий",
                                "successors": [
                                  {

```

```
"name": "Числовые значения",
"successors": [
  {
    "name": "1-й день",
    "value": "1600 мг",
    "meta": "Разовая дозировка"
  },
  {
    "name": "2-10-й дни",
    "value": "600 мг",
    "meta": "Разовая дозировка"
  }
]
},
{
  "name": "Масса тела ≥ 75 кг",
  "meta": "Критерий",
  "successors": [
    {
      "name": "Числовые значения",
      "successors": [
        {
          "name": "1-й день",
          "value": "1800 мг",
          "meta": "Разовая дозировка"
        },
        {
          "name": "2-10-й дни",
          "value": "800 мг",
          "meta": "Разовая дозировка"
        }
      ]
    }
  ]
}
```

Листинг 1. Фрагмент экспортного представления детализированного графа для назначения лечения

Мы использовали версию LLM, которая прошла процесс fine-tuning, с целью более точного соответствия специфике анализа медицинских текстов.

В дополнение к fine-tuning, были разработаны и применены специализированные промпты для запросов к модели. Эти промпты были нацелены на определение и извлечение ключевых аспектов лечения, таких как названия препаратов, дозировки, формы выпуска и доказательства их эффективности. Использование настроенных

промттов позволило уточнить запросы к модели и получить более релевантные и точные результаты.

В результате обработки текста с использованием LLM с fine-tuning и специализированными промптами, мы получили структурированные данные в формате json, этот json-формат может использоваться для обработки различными сервисами, а может быть импортирован на платформу IASaaS, на которой в настоящее время функционирует программная оболочка для создания систем поддержки принятия решений по назначению медикаментозного лечения.

Для валидации графовых знаний, проверки свойств качества написанных SMART-стандартов и обеспечения доверия к ним будут создаваться эталонные задачи для конкретных ситуаций на основе ранее решенных в предметной области. Будут применяться средства проверки решений, вырабатываемых по формализованным графовым базам знаний. Признаком готовности (но не гарантией) становится 100%-ное соответствие вырабатываемых результатов или решений эталонным. Критерии соответствия будут разрабатываться с учетом решаемых задач. На сегодняшний день хорошо понятны критерии правильности решения для задач типа диагностики и распознавания класса ситуаций или сущностей.

**8. Заключение.** Предложен новый двухуровневый подход для создания и перевода в машинопонимаемое представление нормативных документов для дальнейшего их использования в прикладных системах и интеллектуальных сервисах. Показана концептуальная архитектура системы поддержки решения комплекса задач на основе графов знаний, установлены принципы реализации программных компонентов для подготовки к интерпретации графового документа со знаниями.

Авторы считают, что процесс проецирования графа понятий и строгих формулировок в детализированный граф понятий и связей (с помощью LLM с fine-tuning и специализированными промптами) и построение графо-интерпретатора (решателя профессиональных задач пользователя) эффективнее «обучения» LLM «решению» таких профессиональных задач (определения нарушений, поиска по критериям и т.д.), поскольку формализованный граф знаний может быть использован для широкого класса сервисов, в то время как LLM необходимо обучать для каждого класса задач и этот процесс является достаточно дорогим. Кроме того, интерпретатор вырабатывает предсказуемый результат, а обученная LLM «решает» задачу с правдоподобным, но непрозрачным и нестабильным результатом,

поэтому предложенный авторами подход делает вклад в доверие к результатам применения SMART-стандартов. Вместе с тем проблема все еще не «закрыта», авторы продолжают работы в данном направлении, связанные, прежде всего, с обеспечением качества предложенных решений.

### Литература

1. Smart Standards – From a market and industry perspective // Societal and technology trend report. URL: [https://www.iec.ch/system/files/2023-10/iec\\_str\\_smart\\_standards\\_en\\_lr\\_0.pdf](https://www.iec.ch/system/files/2023-10/iec_str_smart_standards_en_lr_0.pdf) (дата обращения: 18.03.2024).
2. Елагин Ф.Н. Цифровые технологии стандартизации // Инновации и инвестиции. 2023. № 8. С. 243–246.
3. Денисова О.А., Дмитриева С.Ю. SMART-стандарты: нормативные документы для цифровой экономики будущего // Стандарты и качество. 2023. № 6. С. 42–44.
4. Предварительный национальный стандарт российской федерации. ПНСТ 864-2023. Умные (SMART) стандарты. Общие положения. Издание официальное. Москва: Российский институт стандартизации, 2023. URL: <https://docs.cntd.ru/document/728306620> (дата обращения: 18.03.2024).
5. Van de Kaa G., Stoccutto S., Calderón C.V. A battle over smart standards: Compatibility, governance, and innovation in home energy management systems and smart meters in the Netherlands // Energy Research & Social Science. 2021. vol. 82.
6. Mutule A., Antoskova I., Papadimitriou C., Efthymiou V., Morch A. Development of Smart Grid Standards in View of Energy System Functionalities // 6th International Conference on Smart and Sustainable Technologies (SpliTech). IEEE, 2021. pp. 1–6. DOI: 10.23919/SpliTech52315.2021.9566337.
7. Peleg M. Computer-interpretable clinical guidelines: A methodological review // Journal of biomedical informatics. 2013. vol. 46. № 4. pp. 744–763.
8. Young O., Shahar Y., Liel Y., Lunenfeld E., Bar G., Shalom E., Martins S., Vaszar L., Marom T., Goldstein M.K. Runtime application of Hybrid-Asbru clinical guidelines // Journal of biomedical informatics. 2007. vol. 40. no. 5. pp. 507–526.
9. Novais P., Oliveira T., Satoh K., Neves J. The Role of Ontologies and Decision Frameworks in Computer-Interpretable Guideline Execution // Synergies between Knowledge Engineering and Software Engineering. 2018. vol. 626. pp. 197–216.
10. Головин С.А., Лоцманов А.Н., Тихомиров С.Г. Цифровая трансформация стандартизации требует системного подхода и практических действий // ИТ-Стандарт. 2023. № 3. С. 4–22.
11. Liu J., Peng G. Designing a Smart Standards Information Service: A Research Framework // International Conference on Human-Computer Interaction. Cham: Springer Nature Switzerland. 2023. pp. 348–365.
12. Luttmer J., Ehring D., Pluhnu R., Kocks C., Nagarajah A. SMART Standards: Modularization Approach for Engineering Standards // Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. 42nd Computers and Information in Engineering Conference. 2022. vol. 2(42). DOI: 10.1115/DETC2022-88206.
13. Zhong L., Wu J., Li Q., Peng H., Wu X. A Comprehensive Survey on Automatic Knowledge Graph Construction // ACM Computing Surveys. 2023. vol. 56. no. 4. pp. 1–62.
14. Qu J. A Review on the Application of Knowledge Graph Technology in the Medical Field // Scientific Programming. 2022. vol. 2022. 12 p.

15. Sezgin E., Hussain S.A., Rust S., Huang Y. Extracting medical information from free-text and unstructured patient-generated health data using natural language processing methods: feasibility study with real-world data // *JMIR Formative Research*. 2023. vol. 7.
16. Melnyk L., Dognin P., Das P. Knowledge graph generation from text // *arXiv preprint*. 2022. arXiv:2211.10 511v1. 13 p.
17. Ibáñez L-D., Domingue J., Kिरrane S., Seneviratne O., Third A., Vidal M-E. Trust, Accountability, and Autonomy in Knowledge Graph-Based AI for Self-Determination // *arXiv preprint*. 2023. arXiv:2310.19503. 33 p. DOI: 10.48550/arXiv.2310.19503.
18. Sajid H. Combining Large Language Models and Knowledge Graphs. URL: <https://www.wisecube.ai/blog/combining-large-language-models-and-knowledge-graphs/> (дата обращения: 15.03.2024).
19. Грибова В.В., Москаленко Ф.М., Тимченко В.А., Шалфеева Е.А. Платформа IACPaaS для разработки систем на основе онтологий: десятилетие использования // *Искусственный интеллект и принятие решений*. 2022. № 4. С 55–65.
20. Временные методические рекомендации. Профилактика, диагностика и лечение новой коронавирусной инфекции (COVID-19). Версия 18. URL: [https://static-0.minzdrav.gov.ru/system/attachments/attaches/000/064/610/original/BMP\\_COVID-19\\_V18.pdf](https://static-0.minzdrav.gov.ru/system/attachments/attaches/000/064/610/original/BMP_COVID-19_V18.pdf) (дата обращения: 15.03.2024).
21. Клинические рекомендации. Острый инфаркт миокарда с подъемом сегмента ST электрокардиограммы. 2020. 157 с. URL: [https://cardioweb.ru/files/glavny-kardiolog/rekomendation/%D0%9A%D0%BB%D0%B8%D0%BD\\_%D1%80%D0%B5%D0%BA%D0%BE%D0%BC%D0%B5%D0%BD%D0%B4%D0%B0%D1%86%D0%B8%D0%B8\\_%D0%9E%D0%9A%D0%A1\\_%D1%81\\_%D0%BF%D0%BE%D0%B4%D1%8A%D0%B5%D0%BC%D0%BE%D0%BC\\_ST\\_2020.pdf](https://cardioweb.ru/files/glavny-kardiolog/rekomendation/%D0%9A%D0%BB%D0%B8%D0%BD_%D1%80%D0%B5%D0%BA%D0%BE%D0%BC%D0%B5%D0%BD%D0%B4%D0%B0%D1%86%D0%B8%D0%B8_%D0%9E%D0%9A%D0%A1_%D1%81_%D0%BF%D0%BE%D0%B4%D1%8A%D0%B5%D0%BC%D0%BE%D0%BC_ST_2020.pdf) (дата обращения: 18.03.2024).

**Шалфеева Елена Арефьевна** — д-р техн. наук, ведущий научный сотрудник, лаборатория интеллектуальных систем, Институт автоматки и процессов управления Дальневосточного отделения РАН. Область научных интересов: онтологический инжиниринг, интерпретируемые клинические руководства, технология создания систем с декларативными знаниями, объяснительный искусственный интеллект, управление базами знаний. Число научных публикаций — 170. shalf@dvo.ru; улица Радио, 5, 690041, Владивосток, Россия; р.т.: +7(423)231-0424.

**Грибова Валерия Викторовна** — д-р техн. наук, член-корреспондент РАН, заместитель директора по научной работе, научный руководитель лаборатории, лаборатория интеллектуальных систем, Институт автоматки и процессов управления Дальневосточного отделения РАН. Область научных интересов: онтологии и базы знаний, прикладные и проблемно-ориентированные системы, основанные на знаниях, управление базами знаний. Число научных публикаций — 280. gribova@iacp.dvo.ru; улица Радио, 5, 690041, Владивосток, Россия; р.т.: +7(423)231-3999.

**Поддержка исследований.** Работа выполнена в рамках государственного задания ИАПУ ДВО РАН (тема FFW-2021-0004 – подход для создания нормативных документов как двухуровневых графов знаний), и ДВФУ (тема FZNS-2023-0010 – подход для перевода с помощью LLM моделей нормативных документов в машиночитаемое представление).

E. SHALFEEVA, V. GRIBOVA

**THE ISSUES OF CREATION OF MACHINE-UNDERSTANDABLE SMART STANDARDS BASED ON KNOWLEDGE GRAPHS**

*Shalfееva E., Gribova V. The Issues of Creation of Machine-Understandable Smart Standards Based on Knowledge Graphs.*

**Abstract.** The development of digital transformation requires the widespread use of digital technologies in standardization documents. One of the goals is to create standards with machine-understandable content that will allow the use of digital documents at various stages of development and production without the need for a human operator. The purpose of this work is to describe an approach for creating and translating industry normative documents into a machine-understandable representation for their further use in software services and systems. There are three types of SMART standard content: machine-readable, machine-interpretable, and machine-understandable. Knowledge graphs are actively used to formalize data and knowledge when solving various problems. The new two-level approach is proposed for the creation and translation into a machine-understandable representation of regulatory documents as knowledge graphs. The approach defines two types of interpretation of a smart document (human readability and machine understandability) through two related formats: a graph, each semantic node of which represents text in a natural language, and a network of concepts and strict connections. Each node of a human-readable graph corresponds (in general) to a subtree of a machine-readable knowledge graph. As the basis for ensuring the transformation of one form of smart standard representation into another form, LLM models are used, supplemented by a specialized adapter obtained as a result of additional training using the Parameter-Efficient Fine-Tuning approach. Requirements have been established for a set of problem- and subject-oriented tools for generating knowledge graphs. The conceptual architecture of the system for supporting the solution of a set of problems based on knowledge graphs is shown, and the principles for implementing software components that work with smart knowledge for intelligent software services are established.

**Keywords:** smart standard, regulatory document, machine-understandable representation, knowledge graph, two-level representation, LLM models.

**References**

1. Smart Standards – From a market and industry perspective // Societal and technology trend report. Available at: [https://www.iec.ch/system/files/2023-10/iec\\_str\\_smart\\_standards\\_en\\_lr\\_0.pdf](https://www.iec.ch/system/files/2023-10/iec_str_smart_standards_en_lr_0.pdf) (accessed: 18.03.2024).
2. Elagin F.N. [Digital standardization technologies]. Innovacii i investicii – Innovations and investments. 2023. no. 8. pp. 243–246. (In Russ.).
3. Denisova O.A., Dmitrieva S.Ju. [SMART standards: regulatory documents for the digital economy of the future]. Standarty i kachestvo – Standards and quality. 2023. no. 6. pp. 42–44. (In Russ.).
4. PNST 864-2023. Predvaritel'nyy natsional'nyy standart rossiyskoy federatsii. Umnyye (SMART) standarty. Obshchiye polozeniya [Preliminary national standard of the Russian Federation. Smart (SMART) standards. General provisions]. Moscow: Russian Institute of Standardization, 2023. Available at: <https://docs.cntd.ru/document/728306620> (accessed: 18.03.2024).
5. Van de Kaa G., Stoccutto S., Calderón C.V. A battle over smart standards: Compatibility, governance, and innovation in home energy management systems and smart meters in the Netherlands. Energy Research & Social Science. 2021. vol. 82.

6. Mutule A., Antoskova I., Papadimitriou C., Efthymiou V., Morch A. Development of Smart Grid Standards in View of Energy System Functionalities. 6th International Conference on Smart and Sustainable Technologies (SpliTech). IEEE, 2021. pp. 1–6. DOI: 10.23919/SpliTech52315.2021.9566337.
7. Peleg M. Computer-interpretable clinical guidelines: A methodological review. *Journal of biomedical informatics*. 2013. vol. 46. № 4. pp. 744–763.
8. Young O., Shahar Y., Liel Y., Lunenfeld E., Bar G., Shalom E., Martins S., Vaszar L., Marom T., Goldstein M.K. Runtime application of Hybrid-Asbru clinical guidelines. *Journal of biomedical informatics*. 2007. vol. 40. no. 5. pp. 507–526.
9. Novais P., Oliveira T., Satoh K., Neves J. The Role of Ontologies and Decision Frameworks in Computer-Interpretable Guideline Execution. Synergies between Knowledge Engineering and Software Engineering, 2018. vol. 626. pp. 197–216.
10. Golovin S.A., Locmanov A.N., Tihomirov S.G. [Digital transformation of standardization requires a systematic approach and practical actions]. *IT Standard – AT-Standart*. 2023. no. 3. pp. 4–22. (In Russ.).
11. Liu J., Peng G. Designing a Smart Standards Information Service: A Research Framework. *International Conference on Human-Computer Interaction*. Cham: Springer Nature Switzerland. 2023. pp. 348–365.
12. Luttmner J., Ehrling D., Pluhnu R., Kocks C., Nagarajah A. SMART Standards: Modularization Approach for Engineering Standards. *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. 42nd Computers and Information in Engineering Conference. 2022. vol. 2(42). DOI: 10.1115/DETC2022-88206.
13. Zhong L., Wu J., Li Q., Peng H., Wu X. A Comprehensive Survey on Automatic Knowledge Graph Construction. *ACM Computing Surveys*. 2023. vol. 56. no. 4. pp. 1–62.
14. Qu J. A Review on the Application of Knowledge Graph Technology in the Medical Field. *Scientific Programming*. 2022. vol. 2022. 12 p.
15. Sezgin E., Hussain S.A., Rust S., Huang Y. Extracting medical information from free-text and unstructured patient-generated health data using natural language processing methods: feasibility study with real-world data. *JMIR Formative Research*. 2023. vol. 7.
16. Melnyk I., Dognin P., Das P. Knowledge graph generation from text. *arXiv preprint*. 2022. arXiv:2211.10511v1. 13 p.
17. Ibáñez L-D., Domingue J., Kirrane S., Seneviratne O., Third A., Vidal M-E. Trust, Accountability, and Autonomy in Knowledge Graph-Based AI for Self-Determination. *arXiv preprint*. 2023. arXiv:2310.19503. 33 p. DOI: 10.48550/arXiv.2310.19503.
18. Sajid H. Combining Large Language Models and Knowledge Graphs. Available at: <https://www.wisecube.ai/blog/combining-large-language-models-and-knowledge-graphs/> (accessed: 15.03.2024).
19. Gribova V.V., Moskalenko F.M., Timchenko V.A., Shalfeeva E.A. [IACPaaS platform for ontology-based systems development: a decade of use]. *Iskusstvennyj intellekt i prinjatie reshenij – Artificial Intelligence and Decision Making*. 2022. no. 4. pp. 55–65. (In Russ.).
20. Vremennye metodicheskie rekomendacii. Profilaktika, diagnostika i lechenie novoj koronavirusnoj infekcii (COVID-19). Versija 18 (26.10.2023) [Temporary guidelines. Prevention, diagnosis and treatment of new coronavirus infection (COVID-19). Version 18]. Available at: [https://static-0.minzdrav.gov.ru/system/attachments/attaches/000/064/610/original/BMP\\_COVID-19\\_V18.pdf](https://static-0.minzdrav.gov.ru/system/attachments/attaches/000/064/610/original/BMP_COVID-19_V18.pdf) (accessed: 15.03.2024).

21. Klinicheskie rekomendacii. Ostryj infarkt miokarda s podemom segmenta ST jelektrokardiogrammy [Clinical recommendations. Acute myocardial infarction with ST segment elevation of the electrocardiogram]. 2020. 157 p. Available at: [https://cardioweb.ru/files/glavny-kardiolog/rekomendation/%D0%9A%D0%BB%D0%B8%D0%BD\\_%D1%80%D0%B5%D0%BA%D0%BE%D0%BC%D0%B5%D0%BD%D0%B4%D0%B0%D1%86%D0%B8%D0%B8\\_%D0%9E%D0%9A%D0%A1\\_%D1%81\\_%D0%BF%D0%BE%D0%B4%D1%8A%D0%B5%D0%BC%D0%BE%D0%BC\\_ST\\_2020.pdf](https://cardioweb.ru/files/glavny-kardiolog/rekomendation/%D0%9A%D0%BB%D0%B8%D0%BD_%D1%80%D0%B5%D0%BA%D0%BE%D0%BC%D0%B5%D0%BD%D0%B4%D0%B0%D1%86%D0%B8%D0%B8_%D0%9E%D0%9A%D0%A1_%D1%81_%D0%BF%D0%BE%D0%B4%D1%8A%D0%B5%D0%BC%D0%BE%D0%BC_ST_2020.pdf) (accessed: 18.03.2024).

**Shalfeeva Elena** — Ph.D., Dr.Sci., Leading researcher, Laboratory of intelligent systems, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences. Research interests: ontological engineering, interpretable clinical guidelines, technology for creating systems with declarative knowledge, explanatory artificial intelligence, knowledge base management. The number of publications — 170. shalf@dvo.ru; 5, Radio St., 690041, Vladivostok, Russia; office phone: +7(423)231-0424.

**Gribova Valeria** — Ph.D., Dr.Sci., Corresponding member of the Russian Academy of Sciences, Deputy director for research, scientific director of the laboratory, Laboratory of intelligent systems, Institute of Automation and Control Processes, Far Eastern Branch of the Russian Academy of Sciences. Research interests: ontologies and knowledge bases, applied and problem-oriented knowledge-based systems, knowledge base management. The number of publications — 280. gribova@iacp.dvo.ru; 5, Radio St., 690041, Vladivostok, Russia; office phone: +7(423)231-3999.

**Acknowledgements.** The research was carried out within the state assignment of IACP FEB RAS (Theme FFWF-2021-0004 – an approach for translating models of regulatory documents into a machine-understandable representation using LLM), and the Far Eastern Federal University (Theme FZNS-2023-0010 – an approach for creating regulatory documents as two-level graphs knowledge).

В.В. МАНОЙЛОВ, А.Г. БОРОДИНОВ, И.В. ЗАРУЦКИЙ, А.И. ПЕТРОВ,  
А.С. САРАЕВ, В.Е. КУРОЧКИН

## АЛГОРИТМЫ ПЕРВИЧНОГО АНАЛИЗА ЛОКАЛЬНЫХ ОБЪЕКТОВ ФЛУОРЕСЦЕНЦИИ В СЕКВЕНАТОРЕ ДНК «НАНОФОР СПС»

*Манойлов В.В., Бородинов А.Г., Заруцкий И.В., Петров А.И., Сараев А.С., Курочкин В.Е.*  
**Алгоритмы первичного анализа локальных объектов флуоресценции в секвенаторе ДНК «Нанофор СПС».**

**Аннотация.** В секвенаторе ДНК «Нанофор СПС», разработанном в Институте аналитического приборостроения РАН, реализован метод массового параллельного секвенирования для расшифровки последовательности нуклеиновых кислот. Этот метод позволяет определять последовательность нуклеотидов в ДНК или РНК, содержащих от нескольких сотен до сотен миллионов звеньев мономеров. Таким образом, имеется возможность получения подробной информации о геноме различных биологических объектов, в том числе человека, животных и растений. Важнейшей частью этого прибора является программное обеспечение, без которого невозможно решение задач по расшифровке генома. Выходными данными оптической детекции в секвенаторе являются набор изображений по четырем каналам, соответствующим типам нуклеотидов: А, С, G, Т. С помощью специального программного обеспечения определяется положение молекулярных кластеров и их интенсивностные характеристики вместе с параметрами окружающего фона. В ходе создания программного обеспечения прибора были разработаны алгоритмы и программы обработки сигналов флуоресценции, рассмотренные в работе. Также, для отладки и тестирования рабочих программ созданы модели построения изображений, аналогичных реальным данным, получаемым в ходе работы секвенатора. Данные модели позволили получить значительный массив информации без запуска дорогостоящих экспериментов. За последние годы достигнуты значительные успехи в области машинного обучения, в том числе и в области биоинформатики, что привело к реализации наиболее распространенных моделей и возможности их применения для практических задач. Однако, если на этапе вторичного анализа биоинформационных данных эти методы широко зарекомендовали себя, то их потенциал для первичного анализа остается недостаточно раскрытым. В данной работе особое внимание уделяется разработке и внедрению методов машинного обучения для первичного анализа оптических изображений сигналов флуоресценции в реакционных ячейках. Описаны методы кластеризации и их апробация на моделях и на изображениях, полученных на приборе. Цель этой статьи – продемонстрировать возможности алгоритмов первичного анализа сигналов флуоресценции, получающихся в процессе секвенирования на приборе «Нанофор СПС». В работе описаны основные задачи анализа сигналов флуоресценции и сравниваются традиционные методы их решения с использованием технологий машинного обучения.

**Ключевые слова:** секвенирование, нуклеиновая кислота, методы обработки сигналов флуоресценции ДНК и РНК, анализ изображений, машинное обучение.

**1. Введение.** В секвенаторе «Нанофор СПС» реализован метод массового параллельного секвенирования, который еще называют методом секвенирования нового поколения (NGS) [1]. Отличительной

особенностью технологии является возможность анализировать одновременно множество участков генома [2, 3]. В процессе секвенирования используется техника удлинения цепей отдельных частей ДНК и РНК кислот.

В секвенаторе «Нанофор СПС» флуоресцентное химическое соединение (флуорофор или краситель) «присоединяется» к нуклеотидам и может повторно излучать свет при его возбуждении, например, лазерным излучением. Каждый нуклеотид, помеченный красителем, излучает свет на длине волны, соответствующей его типу. После присоединения красителя к фрагментам нуклеиновой кислоты производится возбуждение красителя лазерным излучением. Полученный после возбуждения сигнал флуоресценции проходит через светофильтры разных длин волн. Длины волн светофильтров соответствуют длинам волн, которые излучают нуклеотиды, помеченные красителями. После прохождения через светофильтры сигнал флуоресценции регистрируется видеокамерами. В секвенаторе имеются четыре видеокамеры, каждая из которых фиксирует сигналы определенного типа нуклеотида (канала): аденин – «А», цитозин – «С», гуанин – «G» и тимин – «Т».

После регистрации видеокамерами изображений сигналов флуоресценции по всей длине реакционной ячейки происходит переход к следующему этапу. На этом этапе через камеру пропускают реактивы, которые отделяют краситель (флуорофор) и прекращают процесс синтеза. Затем добавляются другие реактивы, чтобы начать новый процесс синтеза – новый цикл.

Программное обеспечение (ПО) секвенатора «Нанофор СПС» решает следующие задачи обработки данных генетического анализа, полученных по результатам экспериментов:

- 1) Чтение изображений с видеокамер;
- 2) Фокусировка полученных изображений аппаратными и математическими методами;
- 3) Исключение фона в исходном изображении;
- 4) Распознавание и определение характеристик кластеров сигналов флуоресценции на реакционной ячейке;
- 5) Определение характеристик изображений «слипшихся» кластеров;
- 6) Исключение взаимовлияния флуоресценции в различных каналах;
- 7) Оценка качества результатов проведенного эксперимента после коррекции влияния химических процессов изменяющих

значения обрабатываемых сигналов: фазирование, перефазирование, затухание сигнала и др.

Решения задач обработки данных генетического анализа реализованы в программном обеспечении секвенатора «Нанофор СПС» и в нескольких зарубежных приборах, описанных в литературе [4, 5]. При разработке нижеперечисленных методов использовались алгоритмы из работ [6, 7, 8]. В последнее время технологии машинного обучения стали широко применяться для обработки биоизображений. Упомянутые методы были протестированы на изображениях, полученных с реальных приборов, а также на модельных изображениях. Методы построения моделей изображений описываются в настоящей статье.

**2. Алгоритм обнаружения сигналов флуоресценции на основе свертки.** В секвенаторе «Нанофор СПС» установлены четыре черно-белые видеокамеры, по одной на каждый канал. Эти камеры способны регистрировать изображения с 4096 оттенками серого. Изображения с камер передаются в компьютер в форме растровых массивов двоичных слов, где каждое слово содержит код яркости пикселя. На рисунке 1 представлен фрагмент изображения сигналов флуоресценции для канала «А» – аденин.

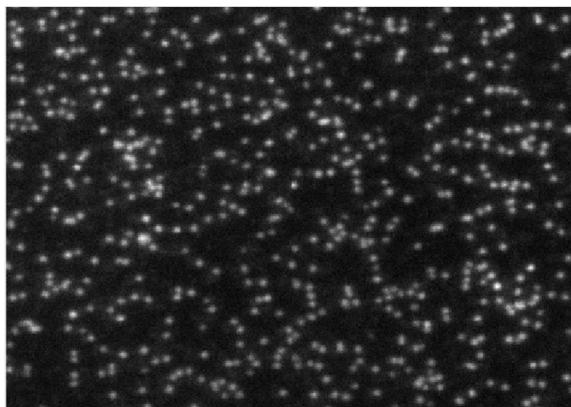


Рис. 1. Фрагмент изображения сигналов флуоресценции для канала «А» (аденин)

Для обеспечения высокого качества решения задач по обработке изображений сигналов флуоресценции, полученные из видеокамер цифровые данные подвергаются медианной фильтрации. Этот шаг помогает избежать негативного влияния «выбросов» регистрируемых сигналов и дефектных пикселей на видеокамерах, искажающих

изображения. Затем решается задача обнаружения кластеров нуклеиновых кислот. Задача обнаружения описывается как поиск точек (пикселей) на изображении, соответствующих центрам кластеров, что позволяет интерпретировать изображение, как результат искажения исходного сигнала различными причинами, например, плохой фокусировкой, шумами и другими причинами, которые могут испортить исходный сигнал [9 – 13]. Исходный сигнал представляет собой набор функций с координатами в центрах кластеров, а форма изображения кластера определяется функцией искажения. Задача обнаружения кластера сводится к задаче восстановления сигнала, решение которой является решением обратной задачи. Одним из методов решения обратных задач является деконволюция, для которой требуется знание формы функции искажения. Математически задача деконволюции решается с использованием прямого и обратного преобразования Фурье в двумерном пространстве частот с независимыми переменными [9 – 13].

Для обнаружения объектов, изображенных на рисунке 1, используется свертка с трехмерным образом второй производной гауссовой функции (1) с уменьшенной шириной, приблизительно равной половине средней ширины самих кластеров.

$$g(t) = A * \left( \frac{4t^2}{\mu^4} \exp \left[ -\left( \frac{t}{\mu^2} \right)^2 \right] - \frac{2}{\mu^2} \exp \left[ -\left( \frac{t}{\mu^2} \right)^2 \right] \right), \quad (1)$$

где  $t$  – независимая переменная,  $\mu$  – параметр ширины.

Формула (1) является результатом вычисления второй производной функции  $\exp \left[ -\left( \frac{t}{\mu^2} \right)^2 \right]$ .

Вторую производную гауссовой функции называют «мексиканской шляпой» (Mexican hat) и используют в программном обеспечении ряда приборов, в получаемых изображениях на которых необходимо обнаруживать сигналы флуоресценции, в том числе и в секвенаторе фирмы «Illumina» [4].

Получение трехмерного образа этой функции производится путем ее вращения по вертикальной оси, проходящей через максимум. Такое вращение формирует трехмерную функцию. После этого с помощью преобразования Фурье получается двумерный Фурье-образ в этой функции. Полученный Фурье-образ используется в алгоритме решения обратной задачи с помощью деконволюции. С помощью деконволюции происходит выделение полезного сигнала из шума,

«сужение» («обострение») обнаруженных сигналов и исключение влияния фоновой составляющей. «Обострение» обнаруженных сигналов необходимо для решения задачи разделения «наложившихся» кластеров.

Условия использования формулы (1) в рассматриваемой задаче следующие: соотношение сигнал шум находится в диапазоне: от 7 до 40, среднее квадратичное значение шума примерно 15 условных единиц, фоновая составляющая в приборе «Нанофор СПС», как правило, представляет собой нелинейную функцию со значениями интенсивностей от 100 до 200 единиц. Исходный размер изображения в приборе «Нанофор СПС» составляет 2000 x 2400 пикселей. В связи с тем, что ширина кластера на его полувысоте составляет от 6 до 15 пикселей, то фоновую составляющую под кластером можно считать линейной.

Величина амплитуды результирующего сигнала после свертки (деконволюции) зависит от значения параметра  $A$  в формуле (1). Параметр  $A$  подбирался опытным путем для обеспечения амплитуды результирующего сигнала равной или большей амплитуды исходного сигнала. В будущем планируется сделать адаптивный выбор параметра  $A$ , как функции экспозиции.

Для обеспечения высокого качества решения обратной задачи с помощью деконволюции к обрабатываемому изображению необходимо предварительно применить медианную фильтрацию. Этот шаг помогает избежать негативного влияния «выбросов» регистрируемых сигналов и дефектных пикселей на видеокамерах, искажающих изображения.

Качественное распознавание и определение характеристик кластеров на изображении затруднено из-за фоновой составляющей. Для решения этой проблемы используется алгоритм, описанный в работе [6]. В процессе вычислений могут возникать отрицательные значения яркости изображения, которые не играют ключевой роли. Фон обычно изменяется плавно и его можно вычислить, усредняя сигнал изображения в окне, скользящем по кадру. Ширина окна должна быть в два раза больше, нежели размер самого крупного кластера на изображении.

Вычисления по алгоритму деконволюции производятся следующим образом. Исходные данные видеoinформации  $S(x, y)$  в геометрическом пространстве  $(x, y)$  с помощью алгоритма быстрого преобразования Фурье преобразуются в данные видеoinформации в пространстве Фурье  $S(u, v)$ . Затем данные  $S(u, v)$  умножаются на Фурье-образ функции искажения  $Fl(u, v)$ . Полученное

произведение подвергается обратному преобразованию Фурье и получаются данные  $S_p(x, y)$ , которые затем сравниваются с порогом. В данном описании  $x, y$  – координаты исходного геометрического пространства,  $u, v$  – координаты пространства Фурье (частоты).

После выполнения операций по вычитанию фона и применению алгоритма деконволюции происходит решение задач по поиску и оценке положений координат кластеров флуоресценции (КФ). Задача поиска включает в себя обнаружение объектов и определение координат их центров. Обнаружение кластеров представляет собой процесс выделения областей на изображении, соответствующих искомому кластеру [13]. Для реализации этой операции важным моментом является определение порога, который позволит эффективно отделить «сигнал» (объект) от шума. Для определения порога используется метод гистограмм распределения интенсивности сигналов [10].

Величины интенсивностей сигналов различных нуклеотидов (А, С, G, Т) отличаются друг от друга, и поэтому значения порога для изображений сигналов флуоресценции каждого из нуклеотидов будут разные. Назовем изображения, полученные для сигналов флуоресценции нуклеотидов А, С, G, Т каналами А, С, G, Т, соответственно. Для определения значений порога для каждого из каналов строятся гистограммы распределения нормированных на максимальное значение интенсивностей сигналов в каждом пикселе.

Распределение интенсивностей сигналов флуоресценции представляет собой одномодальную асимметричную функцию. Величины интенсивностей отличные от шума «вносят вклад» в асимметричную (правую) часть функции. Определение порога с помощью гистограмм, имеющих асимметричную функцию распределения, немного сложнее, чем для гистограмм, имеющих двухмодальную функцию распределения, которая описывается в работе Отсу [14]. Для двухмодальных функций распределения порог определяется как среднее значение между двумя максимумами в функции распределения. В нашем случае с помощью гистограмм, соответствующих сигналам флуоресценции секвенатора, порог определялся следующим образом: производилась оценка среднеквадратичного значения (СКО) шума. Оценка СКО шума представляет собой значение полуширины на полувисоте пика гистограммы. Величина порога равна произведению коэффициента  $k$  на оценку СКО. Коэффициент  $k$  подбирался экспериментально путем обработки большого количества изображений сигналов

флуоресценции. Для большинства задач определения порога пригодным оказался коэффициент  $k=9$ .

Работа программы обнаружения заканчивается отметкой специальными маркерами (например, кружками) границ обнаруженных кластеров на исходном изображении, например, так, как показано на рисунке 2, а также построением матрицы, содержащей вертикальные ( $v$ ) и горизонтальные координаты ( $h$ ) обнаруженных кластеров.

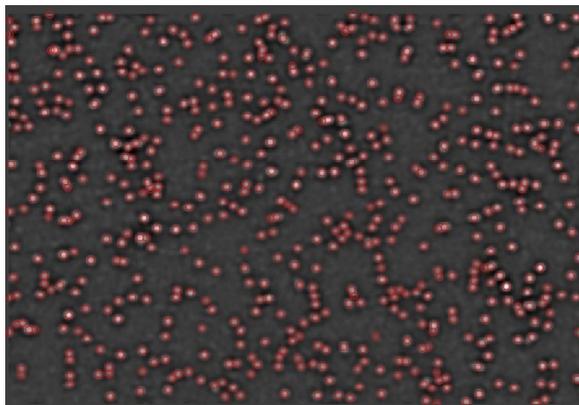


Рис. 2. Границы обнаруженных кластеров на исходном изображении, помеченные кружками

**3. Построение математических моделей изображений совокупностей кластеров сигналов флуоресценции для проверки алгоритмов обнаружения.** Для отладки и тестирования программного обеспечения обработки сигналов флуоресценции по рассмотренным выше алгоритмам были разработаны программы построения моделей изображений, практически соответствующих реальным изображениям, получаемым с прибора [15].

При построении программы моделирования сигналов флуоресценции приборов массового параллельного секвенирования выполняются следующие операции:

- 1) Генерация идеального одиночного объекта флуоресценции;
- 2) Генерация множества изображений одиночных объектов флуоресценции со случайными значениями амплитуды и ширины в соответствии с заданным количеством кластеров;
- 3) Генерация случайных значений координат  $x$  и  $y$  объектов флуоресценции на изображении, аналогично получаемому из видеокамеры;

4) Построение изображения на основе множества одиночных объектов флуоресценции со случайными значениями амплитуды и ширины для полученных в п. 3 значений координат  $x$  и  $y$ ;

5) Добавление к изображению, полученного в п. 4, случайного шума с заданными параметрами;

6) Добавление к изображению, полученного в п. 5, нелинейного фона с заданными параметрами.

Идеальный одиночный объект флуоресценции строится на основе трехмерной гауссовой функции в соответствии с формулой (2).

$$g(u, v) = A * \exp \left[ - \left( \frac{(u-x)^2}{2 \sigma_u^2} + \frac{(v-y)^2}{2 \sigma_v^2} \right) \right], \quad (2)$$

где  $A$  – амплитуда сигнала,  $x$  и  $y$  – координаты максимального значения,  $\sigma_u$  и  $\sigma_v$  – параметры ширины по координатам  $u$  и  $v$ .

На основе экспериментальных данных параметры  $\sigma_u$  и  $\sigma_v$  можно считать одинаковыми и равными  $\sigma$ . Параметры  $\sigma_u$  и  $\sigma_v$  отвечают за площадь кластера на изображении.

Амплитуда объектов флуоресценции в реальных экспериментах может меняться от 100 до 600 условных единиц. Площадь объектов флуоресценции на реальном изображении может меняться от 6 до 10 пикселей на уровне половины максимального значения. Для создания модели объектов флуоресценции изображений, используемых при отладке алгоритмов обработки информации, рассмотренных выше, необходимо сгенерировать  $N$  объектов флуоресценции с указанным диапазоном амплитуд  $A$  и шириной  $\sigma$  посредством использования метода Монте-Карло. Создание таких объектов осуществлялось при помощи генератора случайных чисел, имеющих плотность распределения вероятностей подчиняющейся нормальному закону.

Генерация координат  $x$  и  $y$  объектов флуоресценции осуществлялась с помощью генератора равномерно распределенных случайных чисел для заданного количества объектов  $N$  и заданного размера моделируемого изображения, например 320 x 456 пикселей. В результате получался набор из  $N$  пар чисел. Каждая пара чисел ( $x$  и  $y$ ) из этого набора соответствовала определенному кластеру с амплитудой  $A$  и шириной  $\sigma$  из указанных выше диапазонов. Значения координат  $x$ ,  $y$ , амплитуды  $A$  и ширины  $\sigma$  подставлялись в формулу (2). Таким образом, оказались сгенерированными  $N$  кластеров и получено изображение, содержащее модели сигналов, построенных на основе моделей идеального одиночного объекта флуоресценции.

Моделирование «слипшихся» объектов производится в той же самой программе, что и моделирование одиночных объектов. При этом для получения эффекта слипания увеличивается общее количество моделируемых объектов. Координаты объектов являются случайными числами и поэтому, чем большее количество объектов моделируется на изображении заданного размера, тем большая вероятность того, что часть объектов будет иметь близкие координаты – такие объекты окажутся наложенными друг на друга и станут «слипшимися». На реальных изображениях, получаемых в ходе эксперимента, часть объектов флуоресценции сливается друг с другом, т.к. технология их «выращивания» также является случайным процессом и локальные центры этих объектов могут быть близки. Количество синтезируемых в секвенаторе фрагментов нуклеиновых кислот, локальные центры кластеров которых близки друг к другу, зависит от концентрации анализируемой пробы. Чем выше концентрация, тем большее количество букв нуклеотидов может быть получено в результате эксперимента. Однако, слишком высокая концентрация анализируемой пробы может привести к ошибочным результатам из-за неправильной оценки параметров кластеров. Оптимальная концентрация анализируемой пробы подбирается опытным путем таким образом, чтобы не допустить, с одной стороны, большого количества «слипшихся» кластеров и, с другой стороны, получить желаемое количество букв нуклеотидов. Для оценки параметров «слипшихся» объектов применяются специальные алгоритмы. Один из таких алгоритмов описан в работе [7].

Для приближения полученного изображения, основанного на моделях идеального одиночного объекта флуоресценции, к изображению, получаемого с прибора массового параллельного секвенирования, необходимо добавить случайный шум и фон. Случайный шум создается с помощью генератора случайных чисел, подчиняющихся нормальному распределению с математическим ожиданием нуль и средним квадратичным отклонением, задаваемым оператором. Величина СКО составляет примерно 5...10 % от величины средней амплитуды сигнала флуоресценции кластера. Для генерации нелинейного фона использовалась модель двумерной гауссовой функции, аналогичная формуле (2), но с шириной, отличающейся от ширины сигналов флуоресценции кластеров примерно в 4000 раз. Центр такой двумерной гауссоиды моделировался в левом нижнем или правом верхнем углу изображения в зависимости от того, в каком месте находился источник света,

создающий флуоресценцию в реальном приборе. Значения ширины, центра и амплитуды функции фона задавались оператором.

На рисунке 3 представлены профили одиночного сигнала объекта в реальном приборе (пунктирная линия) и в модельном изображении (сплошная линия). Как видно из рисунка 3 профили сигналов объекта реального и моделируемого изображений близки. Численная оценка близости моделируемого и реального изображений выполнена в работе [16].

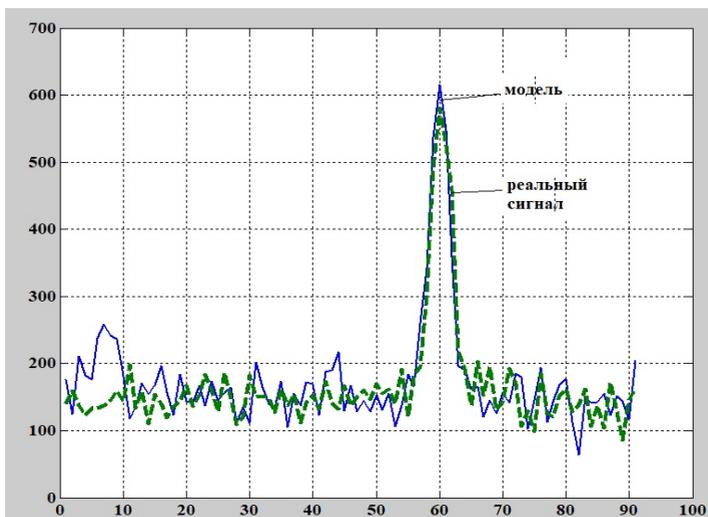


Рис. 3. Профили одиночного сигнала в реальном приборе (пунктирная линия) и модельном изображении (сплошная линия). По горизонтальной оси номер пикселя. По вертикальной оси величина интенсивности в условных единицах

Сравнение изображений, получаемых в ходе генерации модельных кластеров с изображениями кластеров, полученных в реальных экспериментах, показало их высокую идентичность. Подробнее описано в работе А.С. Сараева [16].

Разработанные программы моделирования кластеров молекул ДНК на приборах массового параллельного секвенирования позволили произвести отладку рабочих программ прибора «Нанофор СПС» без использования изображений сигналов флуоресценции, получаемых в ходе дорогостоящих экспериментов. Полученные результаты обеспечили возможность решения задач по обнаружению, оценке параметров кластеров в условиях шума, нелинейности фона и «наложения» кластеров при высокой концентрации объектов.

**4. Оценка амплитуды и ширины объекта флуоресценции по известным координатам методом наименьших квадратов на основе трехмерной гауссовой функции.** Одиночный объект флуоресценции (кластер) можно описать с помощью формулы (2). Предположим, что мы выделили из исходного изображения фрагмент размером  $9 \times 9$  пикселей с координатами  $x$  и  $y$ , например, как показано на рисунке 4(а). 3D изображение яркостей пикселей этого фрагмента, из которых вычтено значение «базовой» линии, показано на рисунке 4(б).

Предположим, что объект флуоресценции симметричный  $\sigma_u = \sigma_v = \sigma$ . Тогда значение яркости  $i$ -го пикселя можно записать следующим образом:

$$Z_i = A * \exp \left[ - \left( \frac{(u_i - x)^2}{2 \sigma^2} + \frac{(v_i - y)^2}{2 \sigma^2} \right) \right], \quad (3)$$

где  $u_i$  и  $v_i$  горизонтальная и вертикальная координата  $i$ -го пикселя внутри фрагмента. Произведем логарифмирование и сделаем следующие обозначения:

$$b_0 = \ln A \text{ и } b_1 = -\frac{1}{2\sigma}. \quad (4)$$

Теперь сумму  $Q$  квадратов отклонений логарифмов яркостей пикселей от аппроксимирующей функции, задаваемой по формуле (2) можно записать:

$$Q = \sum_{i=1}^{81} [b_0 + b_1(u_i - x)^2 + b_1(v_i - y)^2 - \ln Z_i]^2. \quad (5)$$

Для нахождения минимума суммы квадратов отклонений приравняем частные производные  $\frac{\partial Q}{\partial b_0}$  и  $\frac{\partial Q}{\partial b_1}$  к нулю и после алгебраических преобразований получаем систему из двух линейных уравнений с двумя неизвестными. Решая эту систему, получаем значения  $b_0$  и  $b_1$ . Учитывая обозначения (5), получаем формулы для определения амплитуды  $A$  и ширины  $\sigma$  пика кластера флуоресценции.

Найдя амплитуду и ширину, можно построить по формуле (2) фигуру, представленную на рисунке 4(в) и сравнить ее с фигурой, представленной на рисунке 4(б). Для оценки качества аппроксимации подсчет среднего квадратичного отклонения между исходными данными и данными, полученными по результатам аппроксимации по формуле (2) показал, что это среднее значение не превышает уровня шума в исходных данных. На рисунке 4(г) представлены профили

строк изображений, проходящих через максимум для исходных данных и для данных, полученных в результате аппроксимации. Из информации, представленной на рисунке 4(г) видно, что данные, полученные в результате аппроксимации и исходные данные, близки.

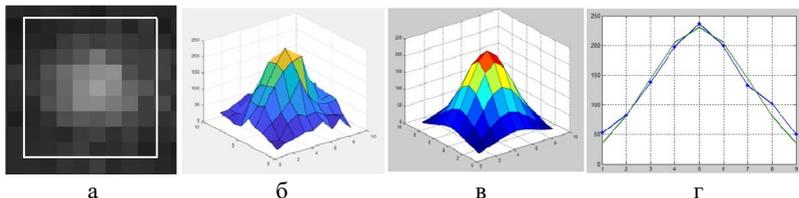


Рис. 4. а – фрагмент изображения размером 9х9, показанный белыми линиями; б – 3D изображение яркостей пикселей фрагмента изображения 9х9 пикселей, из которых вычтено значение «базовой» линии; в – 3D изображение яркостей пикселей фрагмента изображения 9х9 пикселей, полученного в результате аппроксимации с использованием формулы (3); г – профиль изображения яркостей пикселей фрагмента исходного изображения (сплошная линия) и изображения, полученного в результате аппроксимации (линия с «\*»)

**5. Методы машинного обучения для первичного анализа сигналов флуоресценции в технологии массового параллельного секвенирования.** В последние десятилетия искусственный интеллект (ИИ), (Artificial Intelligence, AI), машинное обучение (Machine Learning, ML) и глубокое обучение (Deep Learning, DL) стали высокоэффективными подходами, имеющими множество применений в сфере информатики.

В данной работе особое внимание уделяется разработке и внедрению методов машинного обучения для обработки и интерпретации оптических изображений сигналов флуоресценции в реакционных ячейках в приборе «Нанофор СПС». При этом если методам машинного обучения на этапе вторичного анализа биоинформационных данных секвенирования уделено уже достаточно много внимания в современных исследованиях [17, 18], то применение машинного обучения для первичного анализа только начало привлекать внимание исследователей [8]. Оптические изображения реакционных ячеек секвенаторов генерируют огромные объемы данных, но при этом демонстрируют высокую вариабельность из-за различий в подготовке образцов, методах визуализации, оборудовании и используемых параметрах сбора данных микроскопии. Присущая таким изображениям изменчивость создает проблемы для анализа классическими методами обработки изображений. Следовательно,

методы машинного обучения (ML), а в перспективе и методы глубокого обучения (DL) предлагают привлекательное решение для повышения скорости, точности, адаптируемости, воспроизводимости и эффективности анализа изображений.

**5.1. Методы ML.** Подходы ML/DL, используемые для анализа изображений в секвенаторе «Нанофор СПС», делятся на два типа (рисунок 5): обучение с учителем (Supervised learning) и обучение без учителя (Unsupervised learning). В обучении с учителем используется размеченный набор данных. Размеченный набор данных – это совокупность данных, разделенных на подмножества: обучающий набор (training set), тестовый набор (test set) и иногда проверочный набор данных (validation set). Методы обучения с учителем используются для прогнозного моделирования и подразделяются на задачи классификации и регрессии. В задаче обработки данных секвенатора типичной задачей классификации является задача base-calling. Base-calling – это процесс определения нуклеотидного основания, который генерирует соответствующее значение интенсивности в каналах флуоресценции для различных длин волн на кадрах изображения реакционной ячейки для различных циклов секвенирования методом синтеза [19]. Разнообразные алгоритмы такой классификации ML включают модель перцептрона, логистическую регрессию, деревья решений, метод опорных векторов (SVM), случайный лес и k-ближайшие соседи (KNN).

Проблема base-calling, сформулированная в виде обобщенной модели [20], сводится к последовательности задачи регрессии для оценки параметров этих моделей. Соответствующие алгоритмы стремятся подогнать точки данных посредством регрессионного анализа, например, логистической регрессии, полиномиальной регрессии и линейной регрессии. Тип используемой регрессии будет зависеть от количества независимых и зависимых переменных и формы распределения точек данных. Примеры такой реализации решения проблемы base-calling содержатся в работах [21].

Прогнозируемые выходные данные контролируемых моделей ML могут быть «переобучены» (overfitted) или недообучены в результате некорректного обучения, которое приводит к искаженным или смещенным результатам с высокой погрешностью. Переобучение является результатом слишком точного обучения модели к точкам данных тестового набора, при этом набор обучающих данных слишком мал для обучения или слишком зашумлен. Задача переобучения (overfitting) очень часто возникает при применении моделей машинного обучения в задачах регрессии. Здесь она решается с помощью

регуляризации соответствующих регрессионных моделей [22]. Также эффективным средством борьбы с такого рода переобучением являются методы снижения размерности модели. Различные методы снижения размерности для задачи base-calling рассмотрены в работе [23].



Рис. 5. Разные типы методов в машинном обучении, используемые при анализе изображений в биоинформатике. CNN – сверточная нейронная сеть; DBSCAN – основанная на плотности кластеризация для приложений с шумами; HDBSCAN – иерархическая версия DBSCAN; BIRCH – сбалансированное итеративное сокращение и кластеризация с помощью иерархий; GMM – модель гауссовой смеси; KNN – метод k-ближайших соседей; OPTICS – упорядочение точек для обнаружения кластерной структуры; PCA – анализ главных компонент; SVD – разложение по сингулярным значениям; SVM – метод опорных векторов; SVR – регрессия опорных векторов; t-SNE – стохастическое вложение соседей с t-распределением; UMAP – Uniform Manifold Approximation and Projection

Наконец, типичным классом задач обучения без учителя являются задачи кластеризации. При этом, не требуется наличие помеченного набора данных (labelled dataset). Набор данных не сгруппирован в обучающий и тестовый набор – весь набор данных вводится в модель машинного обучения для анализа данных. Кластерный анализ – это метод, который группирует схожие точки данных в кластеры в зависимости от их относительного сходства. В задачах обработки изображений сигналов флуоресценции задача кластеризации связана с обнаружением клональных кластеров, получаемых из фрагментов геномной библиотеки посредством мостиковой амплификации. Такая задача в области массового параллельного секвенирования сильно осложнена огромным количеством таких кластеров (сотни тысяч), небольшим количеством

пикселей изображений, приходящихся в среднем на каждый кластер, неопределенностью паттерна каждого отдельного кластера, плотной упаковкой кластеров в рамках одного изображения. Естественным образом, при анализе таких сложных физико-химических процессов необходимой задачей является определение зон чрезмерной плотности кластеров или аномальных зон оптической яркости для их последующей фильтрации в области обнаружения кластеров.

Таким образом, практически все типовые задачи обработки изображений (классификация, регрессия, кластеризация, снижение размерности, выявление аномалий) востребованы на стадии первичного анализа данных секвенирования.

**5.2. Процесс анализа изображений.** Типичный процесс анализа изображений состоит из нескольких этапов: предварительная обработка, детектирование объектов, сегментация объектов, извлечение признаков для интеллектуального анализа данных и других [24]. На рисунке 6 представлена структурная схема процесса анализа изображений в секвенаторах.



Рис.6. Общий рабочий процесс анализа изображений в секвенаторах

### 5.2.1. Предварительная обработка (Image preprocessing).

Процесс обычно начинается с предварительной обработки необработанных изображений после регистрации. Предварительная обработка включает в себя такие этапы, как контроль качества изображения и различные типы манипуляций с изображением (например, изменение размера и улучшение отношения сигнал/шум (SNR)). Этот шаг особенно полезен, например, при работе в условиях

низкой освещенности, что приводит к низкой интенсивности сигнала во время визуализации образца в реальном времени. Предварительная обработка может также включать удаление выбросов и шумоподавление в изображениях.

### **5.2.2. Детектирование объектов, сегментация и выделение признаков (object detection, object segmentation, feature extraction).**

Обнаружение объектов включает в себя локализацию и классификацию объектов на изображении. Его цель – идентифицировать конкретные объекты, представляющие интерес, и определить их ограничивающие рамки, что имеет решающее значение для таких задач, как отслеживание поведения объектов. Задача обработки изображений, полученных секвенатором, усложняется отсутствием образцов кластеров как таковых и небольшим количеством пикселей на каждый объект. Сегментация – это процесс разделения изображения или видео на значимые области для идентификации и дифференциации объектов. Она служит таким целям, как понимание границ объектов, извлечение детальной информации и обеспечение дальнейшего анализа. Одним из этапов сегментации является бинаризация изображений. Этот метод используется для распознавания объектов, т.к. позволяет отличить интересующий объект от фона, на котором он находится. Следующим этапом служит сегментация экземпляров (Instance Segmentation) [25]. Сегментация экземпляров связывает наборы пикселей с каждым отдельным экземпляром объекта и используется, когда необходимо различать уникальные экземпляры одной и той же категории объектов на изображении. Наконец, в процессе выделения признаков сегментированные объекты наделяются рядом признаков, которые служат в дальнейшем для последующего анализа. В частности, выделенные кластеры на изображениях флуоресцентных сигналов наделяются рядом характеристик, позволяющих, во-первых, отфильтровать низкоинформативные кластеры по признакам *chastity* и *purity*, во-вторых, эффективно провести классификацию *base-calling*.

### **5.2.3. Отслеживание объектов и их классификация.**

Алгоритмы машинного обучения могут сыграть заметную роль в задаче выравнивания изображений сигналов флуоресценции, полученных на различных циклах работы секвенатора. Наконец, финальной частью обработки изображения для секвенаторов является задача классификации по типу нуклеотида на каждом цикле секвенатора, что и является сутью *base-calling*.

**5.3. Алгоритмы кластеризации в обработке изображений секвенатора.** Чтобы не путать определения кластерного анализа, как одного из методов классификации в алгоритмах машинного обучения

и кластера в приборах для секвенирования нуклеиновых кислот дадим поясняющие определения.

**Кластерный анализ**, как алгоритм классификации – категория методов обучения без учителя, которая позволяет нам обнаруживать скрытые структуры в данных, где мы заранее не знаем правильного ответа. Цель кластеризации – найти естественную группировку данных, чтобы элементы в одном кластере были более похожи друг на друга, нежели на элементы из разных кластеров.

**Кластер в приборах для секвенирования** – клональная группа матричной ДНК, связанная с поверхностью реакционной ячейки. Каждый кластер засеивается одной цепью матричной ДНК и клонально амплифицируется посредством мостиковой амплификации до тех пор, пока кластер не наберет примерно 1000 копий исследуемой ДНК. Каждому кластеру реакционной ячейки в идеале соответствует одно прочтение (рид) исследуемой геномной последовательности.

Рассмотрим классификацию методов кластерного анализа, приведенного в новейшем обзоре [26], которые представлены на рисунке 7.

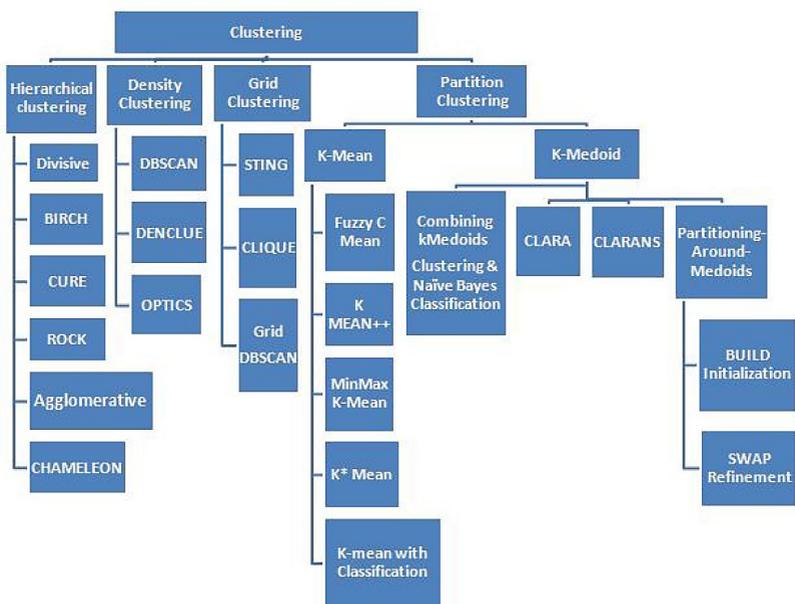


Рис. 7. Классификация методов кластерного анализа [44]

### **5.3.1. Иерархическая кластеризация (Hierarchical Clustering)**

– популярная стратегия обучения без учителя для группировки схожих элементов данных. Она создает кластер иерархической структуры путем многократного слияния и разделения кластеров на основе сходства и несходства. Основная концепция, лежащая в основе иерархической кластеризации, заключается в построении дендрограммы, которая представляет собой древовидную структуру, изображающую связи между точками данных и кластерами. Дендрограмма начинается с каждой точки данных как отдельного кластера и в конечном итоге объединяет связанные группы в соответствии со значением сходства и расстояния. Алгоритм оценивает близость кластеров и выбирает кластеры для объединения на каждом этапе. В зависимости от данных и проблемной области метрики сходства и расстояния, используемые в иерархической кластеризации, могут различаться. Обычно используемые меры расстояния включают евклидово расстояние, манхэттенское расстояние и корреляционное расстояние. Из методов, доступных в библиотеке Scikit-learn для анализа изображений в секвенаторе, были опробованы агломеративная иерархическая кластеризация [27], а также сбалансированное итеративное сокращение и кластеризация с иерархиями (BIRCH) [28].

**5.3.2. Кластеризация на основе плотности (Density-Based Clustering)** – это тип алгоритма кластеризации, который организует точки данных в пространстве объектов в зависимости от их плотности. Кластеризация на основе плотности – это метод кластеризации, который организует точки данных в пространстве данных на основе их плотности. Методы этого класса стремятся обнаруживать кластеры любой формы и размера путем распознавания мест со значительной плотностью данных, при этом не делается предположений относительно количества кластеров в наборе данных или предустановленных форм кластеров. Из методов, предоставляемых библиотекой Scikit-learn были испытаны методы DBSCAN и HDBSCAN.

Основными особенностями DBSCAN являются следующие [29]. Метод основан на поиске областей с высокой плотностью точек, разделенных областями с низкой плотностью. Такие плотные скопления и формируют кластеры. Метод может выявлять кластеры произвольной формы, в том числе неконвексные, не требует указания заранее известного числа кластеров, устойчив к выбросам и шумовым точкам, выделяет их в отдельные «кластеры». DBSCAN имеет два основных параметра: радиус окрестности эпсилон и минимальное

число точек в окрестности MinPts. Именно они определяют порог плотности для кластеризации. Таким образом, DBSCAN позволяет находить скопления объектов произвольной формы в данных без необходимости заранее задавать число кластеров. Это делает его гибким и удобным методом кластеризации во многих практических задачах.

Еще более применимым к обработке изображений в секвенаторе оказался метод HDBSCAN. Как и DBSCAN, метод HDBSCAN разработан для обнаружения кластеров различной формы в пространстве признаков, в отличие от методов K-средних или иерархической кластеризации. Вот основные различия между DBSCAN и HDBSCAN:

1. Параметры: DBSCAN требует два параметра: эpsilon (eps, радиус окружности для соседей) и минимальное количество точек (общее количество точек, которые должны находиться в пределах радиуса eps, чтобы образовать кластер). HDBSCAN требует только параметра minPts (минимальное количество точек), так как он динамически оценивает эpsilon на основе плотности данных.

2. Точки шума: DBSCAN отмечает все точки, лежащие вне minPts эpsilon-сферы окружности, как шум. HDBSCAN, тем не менее, имеет более сложный способ обработки шума и может назначить точки шума к кластерам на основе стойкости.

3. Гибкость: DBSCAN на самом деле сложно использовать на практике, так как он чувствителен к выбору параметра эpsilon. В свою очередь HDBSCAN более гибок, успевает работать со сложными структурами данных, способен обрабатывать и неравномерные плотности данных.

4. Количество кластеров: в отличие от алгоритма K-средних, число кластеров в DBSCAN и HDBSCAN не указывается заранее. HDBSCAN, как модифицированная версия DBSCAN, позволяет получить даже вложенную иерархию кластеров.

5. Иерархическая кластеризация: HDBSCAN выполняет иерархическую кластеризацию, создавая иерархию кластеров с разными уровнями гранулярности или уровня детализации, в то время как DBSCAN не создает явной структуры иерархии кластера.

**5.3.3. Методы кластеризации, основанные на сетке (Grid-Based Clustering).** Кластеризация на основе сетки – это подход к кластеризации, который использует поле данных, разделенное на ячейки сетки. Он может предоставить эффективные и масштабируемые методы кластеризации, которые чрезвычайно полезны для больших наборов данных [30]. В отличие

от традиционных алгоритмов кластеризации, таких как K-Means, которые работают непосредственно с точками данных в их исходном пространстве признаков, Grid-Based Clustering использует другой подход, заключающийся в нескольких этапах реализации:

1. Дискретизация: разделяет пространство данных на сетку ячеек.
2. Расчет плотности: рассчитывается плотность точек данных внутри каждой ячейки.
3. Формирование кластеров. Непрерывные ячейки с плотностью выше определенного порога группируются с образованием кластеров.

Данная работа ограничивается методами кластеризации, реализованными в рамках библиотеки Scikit-learn, которая на настоящее время не имеет специального модуля Grid-Based Clustering. В дальнейших планах авторов реализация такого метода на основе комбинации классов KBinsDiscretizer для создания разбиения пространства на ячейки сети, KernelDensity для расчета плотности точек и scipy.ndimage.label для идентификации кластеров. Огромное количество потенциально присутствующих кластеров на анализируемых изображениях при секвенировании служит веским основанием уделить этим методам особое внимание в дальнейших исследованиях.

**5.3.4. Разделительные методы кластеризации (Partitioning-Based Clustering).** Это тип алгоритмов кластеризации, который делит наборы данных на отдельные группы путем оптимизации целевой функции. Например, разбивает данные на заданное число кластеров, минимизируя расстояние внутри кластера и максимизируя расстояние между кластерами. При этом в ходе кластеризации алгоритм непрерывно распределяет точки данных по кластерам и сохраняет точки центраида кластера до тех пор, пока не произойдет сходимость по определенному критерию.

В нашей работе для обнаружения объектов флуоресценции был использован метод K-средних (K-means) из библиотеки Scikit-learn. Метод K-средних – один из основных алгоритмов кластеризации, целью которого является разделение набора наблюдений на K кластеров, в которых каждое наблюдение принадлежит кластеру с ближайшим средним значением, служащим прототипом кластера.

**6. Итоговый алгоритм обнаружения сигналов флуоресценции в секвенаторах.** Изображения, детектируемые

на протяжении циклов в процессе массового параллельного секвенирования, характеризуются рядом особенностей [31]:

- Огромное количество объектов (кластеров), требующих обнаружения (до нескольких сот тысяч объектов);
- Плотность объектов высока и неравномерна по кадрам, что требует процедур разделения кластеров и/или фильтрации плохо разрешимых участков изображения;
- Изображения получаются в четырех разных каналах флуоресценции и требуют аккуратного совмещения кадров по циклам и каналам (alignment);
- Уровень шумов на изображениях высок и требует эффективного отделения сигналов реальных кластеров от шумовой составляющей;
- Количество пикселей на каждый кластер достаточно мало и форма отдельных кластеров нерегулярна;
- Для оптимального функционирования дальнейших этапов требуется корректное определение центров кластеров;
- Производительность алгоритмов кластеризации крайне важна, поскольку лимитирована временем обработки каждого цикла секвенирования и последовательностью обработки различных участков изображения реакционной ячейки.

Исходя из вышеприведенных требований был предложен многостадийный этап кластеризации:

1) На первом этапе применен метод K-means++ с целью отделения полезного сигнала от фона. При этом число задаваемых кластеров варьировалось в диапазоне 3-5. По сути, на этом этапе мы проводим гибкую бинаризацию изображений для последующей обработки. Кластеризация происходит по значениям интенсивностей кластеров. Метод K-means++ обеспечивает высокую производительность при проведении этого этапа обработки. Главная цель при этом – снизить объем данных для последующей пространственной кластеризации.

2) На втором этапе из всего арсенала апробированных методов выбран BIRCH. BIRCH – это алгоритм иерархической кластеризации, разработанный непосредственно для сканирования больших наборов данных. Основные преимущества BIRCH заключаются в эффективности процесса кластеризации исходя из параметров близости пикселей на изображении и возможности выбирать пороговое значение и количество фрагментов в дереве BIRCH для оптимизации получаемой модели. Однако стоит учитывать, что BIRCH лучше всего подходит для равномерно распределенных

данных и может работать неудовлетворительно при кластеризации сложных структур, таких как кластеры с переменной плотностью или иррегулярной формы. В дальнейшем предлагается апробировать на этом этапе методы Density-Based Clustering, такие как DBSCAN и HDBSCAN с целью оптимального обнаружения кластеров нерегулярной формы.

3) На третьем этапе опционально может быть использована операция сегментации изображения, основанная на морфологических операциях библиотеки `skimage` [32]. Главная задача этого дополнительного этапа – получить большую информацию о кластерах в виде параметров их плотности и площади с целью отфильтровать плохо разрешимые кластеры и пренебречь случайными точками выбросов большой интенсивности на изображении.

Для пояснения работы предложенного комбинированного метода кластеризации приведены иллюстрации к каждому этапу. Изображение представляет собой моделированное распределение кластеров отдельных ридов в одном из каналов флуоресценции в одном из циклов секвенирования с плотностью, характерной для протокола работы секвенатора «Нанофор СПС». На рисунке 8 представлена стадия бинаризации изображения, т.е. его разделение на фоновую и сигнальную составляющую. В данном случае мы кластеризуем изображения по интенсивностям на 4 группы. Два класса самых интенсивных пикселей (зеленый цвет – самые интенсивные пиксели, желтый цвет – следующий класс пикселей по интенсивности) идут на следующий этап для пространственной кластеризации. На рисунке 9 показан результат пространственной кластеризации. Красными точками показаны истинные положения кластеров, взятые из программы моделирования. Голубая область вокруг них – пиксели, отнесенные кластеризацией BIRCH к данному кластеру. На следующем этапе (рисунок 10) изображение сегментируется (красные прямоугольники). Характеристики сегментов (плотность, площадь) позволяют применять фильтры, отсеивающие зоны наложения кластеров друг на друга. По результатам проведенного моделирования такой алгоритм кластеризации позволяет обнаруживать 85-90% процентов кластеров, при этом ошибка определения центра объекта для 80% процентов кластеров составляет меньше одного пикселя.

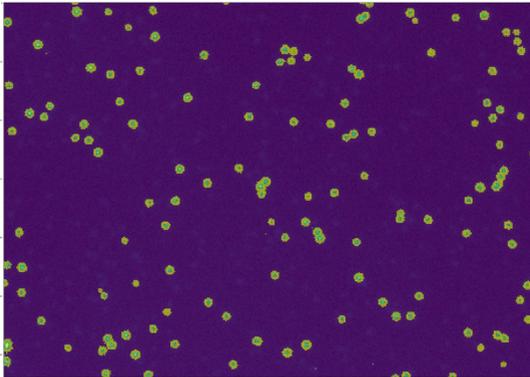


Рис. 8. Этап бинаризации

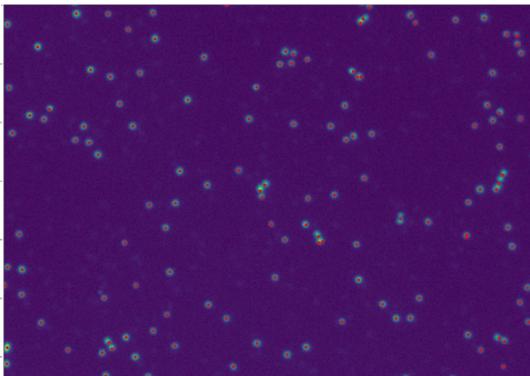


Рис. 9. Этап кластеризации методом Birch

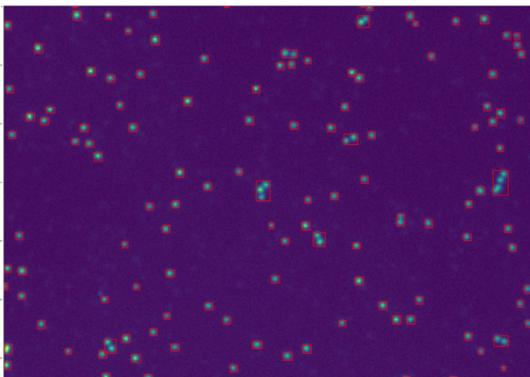


Рис. 10. Этап сегментации

Рассмотренный в данном разделе итоговый алгоритм обнаружения сигналов флуоресценции имеет два преимущества по сравнению с классическим алгоритмом обнаружения на основе свертки:

1. Для обнаружения сигналов не требуется предварительное знание функции формы обнаруживаемого кластера;

2. Этот алгоритм позволяет выделять «слипшиеся» кластеры, параметры которых, как правило, оцениваются не в режиме онлайн, а после окончания эксперимента по алгоритмам, описанным в работе [7].

**7. Сравнение программного обеспечения первичного анализа локальных объектов флуоресценции в секвенаторе «НАНОФОР СПС» с зарубежным аналогом.** Рассмотренные в настоящей работе алгоритмы первичного анализа локальных объектов флуоресценции в секвенаторе «НАНОФОР СПС» легли в основу разработанного в Институте аналитического приборостроения РАН программного обеспечения с предварительным названием PrImA. Данное программное обеспечение является первой отечественной разработкой для приборов, основанных на методе массового параллельного секвенирования.

Для сравнения этого программного обеспечения с его зарубежным аналогом RTA рассмотрим графики, полученные с помощью программного обеспечения UGENE. UGENE – это отечественное свободно распространяемое программное обеспечение для биоинформатики с открытым исходным кодом под системы Windows, macOS и Linux. UGENE собирает полную нуклеотидную последовательность из фрагментов, полученных по результатам первичного анализа локальных объектов флуоресценции в секвенаторе.

На рисунке 11 и 12 представлены распределения показателей качества отдельных букв нуклеотидов референсного генома бактериофага Phix174, для которых первичный анализ флуоресценции был выполнен с помощью программы RTA и программы PrImA. Референсный геном бактериофага Phix174 содержит 5386 нуклеотидов и в результате проведенного анализа на секвенаторе «Нанофор СПС» этот геном в среднем был покрыт примерно 100 раз.

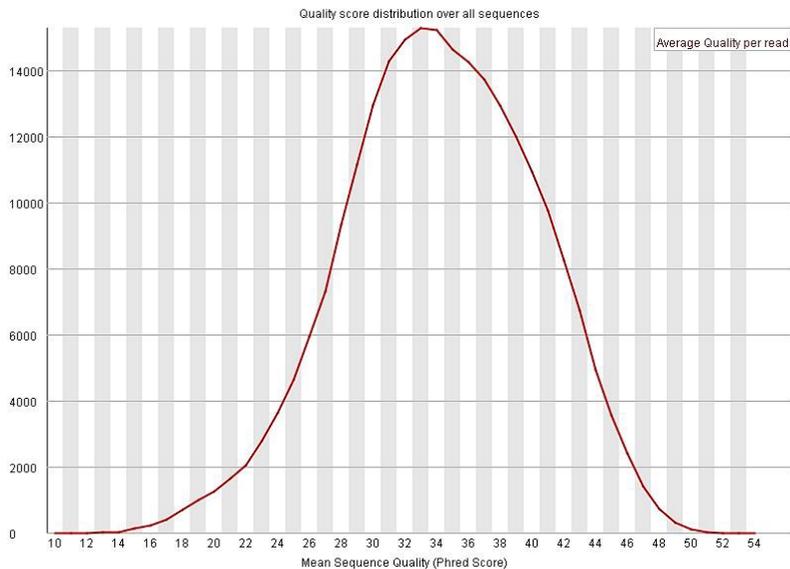


Рис. 11. Показатели качества последовательностей нуклеотидов, полученных с помощью программы RTA

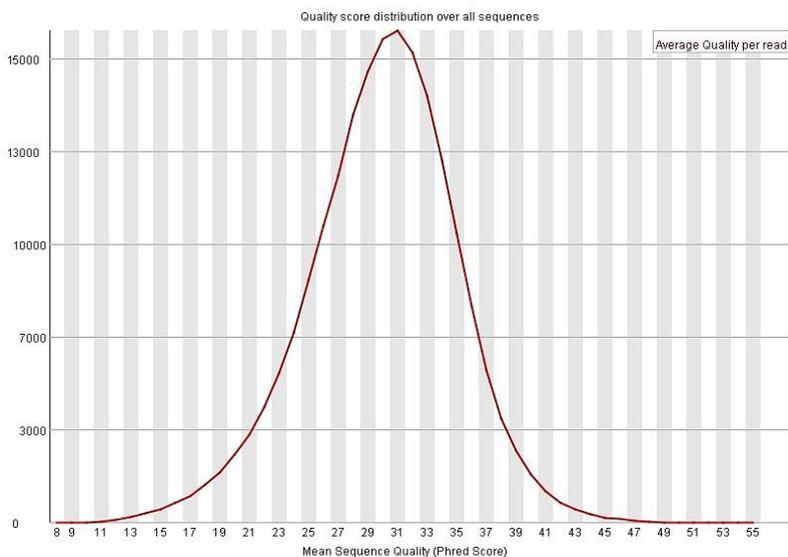


Рис. 12. Показатели качества последовательностей нуклеотидов, полученных с помощью программы PrImA

По горизонтальной оси на этих графиках отложены показатели качества по Phred Score, принятые в биоинформатике. Оценки качества Phred логарифмически связаны с вероятностью ошибок построения последовательности букв нуклеотидов и определяются как:

$$Q = -\log_{10} P.$$

Это соотношение можно записать как:

$$P = 10^{\frac{-Q}{10}}.$$

Например, Phred присваивает букве оценку качества, равную 30. Вероятность того, что эта буква в последовательности была названа неправильно, равна 1 к 1000, или вероятность правильности буквы равна 99.9%.

По вертикальной оси откладывается количество букв, оценка которых произведена с указанной вероятностью.

Сравнение графиков на рисунках 11 и 12 показывает, что программы RTA и PflmA генерируют примерно одинаковые результаты.

**8. Заключение.** Описанные алгоритмы являются результатом исследований, проведенных в процессе разработки секвенатора «Нанофор СПС». Созданное программное обеспечение на основе данных методик в значительной степени предопределило успешный запуск серийной партии отечественных приборов. Для проверки работоспособности алгоритмов были использованы изображения, полученные непосредственно в ходе экспериментов на приборе, а также математические модели изображений кластеров ДНК.

Использование технологий машинного обучения в программном обеспечении секвенатора ДНК «Нанофор СПС» позволило расширить диапазон методов первичного анализа сигналов флуоресценции и показало возможность их применения в обработке без априорного знания функции искажения, а также для автоматического обнаружения слипшихся кластеров.

Программное обеспечение, разработанное на основе рассмотренных алгоритмов, позволяет улучшить качество проводимых экспериментов в сфере геномных исследований и расширить области их применения.

**Литература**

1. Курочкин В.Е., Алексеев Я.И., Петров Д.Г., Евстапов А.А. Отечественные приборы для молекулярно-генетического анализа: разработки ИАП РАН и ООО «Синтол» // *Известия Российской Военно-медицинской академии*. 2021. Т. 40. № 3. С. 69–74. DOI: 10.17816/rmmar76918.
2. Ansoorge W.J. Next-generation DNA sequencing techniques // *Nature Biotechnology*. 2009. vol. 25. no. 4. pp. 195–203.
3. Bentley R.D. Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., et al. Accurate whole human genome sequencing using reversible terminator chemistry // *Nature*. 2008. vol. 456. no. 7216. pp. 53–59.
4. Whiteford N. The Solexa pipeline. 2012. URL: <http://4lj.com/blog/wp-content/uploads/2012/04/pipeline.pdf> (дата обращения: 20.02.2024).
5. Leshkowitz D. Introduction to Deep-Sequencing Data Analysis Illumina Primary Analysis Pipeline & Quality Control. 2017. URL: [http://dors.weizmann.ac.il/course/course2017/Dena\\_IlluminaPrimaryAnalysisPipeline-course2017.pdf](http://dors.weizmann.ac.il/course/course2017/Dena_IlluminaPrimaryAnalysisPipeline-course2017.pdf) (дата обращения: 20.02.2024).
6. Манойлов В.В., Бородинов А.Г., Сараев А.С., Петров А.И., Заруцкий И.В., Курочкин В.Е. Алгоритмы обработки изображений в секвенаторе ДНК НАНОФОР СПС // *Журнал технической физики*. 2022. Т. 92. № 7. С. 985–992. DOI: 10.21883/JTF.2022.07.52655.318-21.
7. Манойлов В.В., Бородинов А.Г., Заруцкий И.В., Петров А.И., Курочкин В.Е. Алгоритмы обработки сигналов флуоресценции массового параллельного секвенирования нуклеиновых кислот // *Труды СПИИРАН*. 2019. Т. 18. № 4. С. 1010–1036. DOI: 10.15622/sp.2019.18.4.1010-1036.
8. Бородинов А.Г., Манойлов В.В., Заруцкий И.В., Петров А.И., Курочкин В.Е., Сараев А.С. Машинное обучение в задачах base-calling для методов секвенирования нового поколения // *Информатика и автоматизация*. 2022. Т. 21. № 3. С. 572–603. DOI: 10.15622/ia.21.3.5.
9. Журавель И.М. Краткий курс теории обработки изображений. URL: <http://matlab.exponenta.ru/imageprocess/book2/49.php> (дата обращения: 26.10.2023).
10. Вудс Р., Гонсалес Р. Цифровая обработка изображений / 3-е изд. // М.: Техносфера. 2012. 1104 с.
11. Sizikov V.S. Spectral method for estimating the point-spread function in the task of eliminating image distortions // *Journal of Optical Technology*. 2017. vol. 84. no. 2. pp. 95–101.
12. Sizikov V.S., Stepanov A.V., Mezhenin A.V., Burlov D.I., Eksemplaryov R.A. Determining image-distortion parameters by spectral means when processing pictures of the earth's surface obtained from satellites and aircraft // *Journal of Optical Technology*. 2018. vol. 85. no. 4. pp. 203–110.
13. Бардин Б.В., Чубинский-Надеждин И.В. Обнаружение локальных объектов на цифровых микроскопических изображениях // *Научное приборостроение*. 2009. Т. 19. № 4. С. 96–102.
14. Otsu N. A Threshold Selection Method from Gray-Level Histograms // *IEEE Transactions on Systems, Man and Cybernetics*. 1979. vol. 9. pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.
15. Сараев А.С., Петров А.И., Манойлов В.В. Моделирование генерации кластеров молекул ДНК в приборах массового параллельного секвенирования // Тезисы докладов Четвертой международной конференции со школой молодых ученых «Физика – наукам о жизни» / СПб: ФТИ им. А.Ф. Иоффе. 2021. С. 153.

16. Сараев А.С. Научно-квалификационная работа «Разработка алгоритма распознавания кластеров нуклеиновых кислот в микрофлюидной ячейке секвенатора «Нанофор СПС». СПб: ИАП РАН. 2023. С. 16–22.
17. Schmidt B., Hildebrandt A. Deep learning in next-generation sequencing // *Drug discovery today*. 2021. vol. 26. no. 1. pp. 173–180.
18. Ozgur S., Orman M. Application of deep learning technique in next generation sequence experiments // *Journal of Big Data*. 2023. vol. 10. no. 1. DOI: 10.1186/s40537-023-00838-w.
19. Tegfalk E. Application of machine learning techniques to perform base-calling in next-generation DNA sequencing. 2020. 45 p.
20. Cacho A., Smirnova E., Huzurbazar S., Cui X. A comparison of base-calling algorithms for illumina sequencing technology // *Briefings in bioinformatics*. 2016. vol. 17. no. 5. pp. 786–795.
21. Kircher M., Stenzel U., Kelso J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies // *Genome biology*. 2009. vol. 10(8). DOI: 10.1186/gb-2009-10-8-r83.
22. Ghojogh B., Crowley M. The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial // *arXiv preprint arXiv:1905.12787*. 2019.
23. Бородинов А.Г., Ямановская А.Ю., Манойлов В.В., Петров А.И. Оптимальный выбор признаков для уменьшения размерности моделей машинного обучения в задаче base-calling // *Тезисы докладов Второй ежегодной всероссийской молодежной конференции по методам и приборам для анализа биологических объектов «АналитБиоПрибор-2023» (Санкт-Петербург, 23–24 ноября 2023 г.)*. Санкт-Петербург: Издательско-полиграфическая ассоциация высших учебных заведений, 2023. С. 135–138.
24. Whiteford N., Skelly T., Curtis C., Ritchie M.E., Lohr, A., Zaranek A.W., Abnizova I., Brown C. Swift: primary data analysis for the Illumina Solexa sequencing platform // *Bioinformatics*. 2009. vol. 25. no. 17. pp. 2194–2199.
25. Hafiz A.M., Bhat G.M. A survey on instance segmentation: state of the art // *International journal of multimedia information retrieval*. 2020. vol. 9. no. 3. pp. 171–189.
26. Chaudhry M., Shafi I., Mahnoor M., Vargas D.L.R., Thompson E.B., Ashraf I.A. Systematic literature review on identifying patterns using unsupervised clustering algorithms: a Data mining perspective // *Symmetry*. 2023. vol. 15. no. 1679. DOI: 10.3390/sym15091679.
27. Khandare A., Pawar R. Data clustering algorithms: experimentation and comparison // *Intelligent Computing and Networking: Proceedings of IC-ICN 2021*. 2022. pp. 86–99.
28. Sarang P. BIRCH: Divide and Conquer // *Thinking Data Science: A Data Science Practitioner's Guide*. Cham: Springer International Publishing. 2023. pp. 229–236.
29. Ester M., Kriegel H.P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise // *kdd*. 1996. vol. 96. no. 34. pp. 226–231.
30. Zhao Y., Cao J., Zhang C., Zhang S. Enhancing grid-density based clustering for high dimensional data // *Journal of Systems and Software*. 2011. vol. 84. no. 9. pp. 1524–1539.
31. Wolowski V.R. High-quality, high-throughput measurement of protein-DNA binding using HiTS-FLIP. Dissertation, LMU Munchen: Fakultat fur Chemie und Pharmazie 2016. DOI: 10.5282/edoc.19445.
32. Chityala R., Pudipeddi S. Image processing and acquisition using Python // *Chapman and Hall/CRC*. 2020. 452 p.

33. Kameshwaran K., Malarvizhi K. Survey on clustering techniques in data mining // International Journal of Computer Science and Information Technologies. 2014. vol. 5. no. 2. pp. 2272–2276.

**Манойлов Владимир Владимирович** — д-р техн. наук, доцент, заведующий лабораторией, лаборатория автоматизации измерений и цифровой обработки сигналов, ИАП РАН. Область научных интересов: представление и обработка сигналов и изображений в аналитических приборах. Число научных публикаций — 101. manoilov-vv@mail.ru; ул. Ивана Черных, 31-33А, 198095, Санкт-Петербург, Россия; р.т.: +7(812)363-0720.

**Бородинов Андрей Геннадьевич** — д-р физ.-мат. наук, старший научный сотрудник, лаборатория методов и приборов иммунного и генетического анализа, ИАП РАН. Область научных интересов: математическая статистика, проблемы анализа, обработки и представления данных, искусственный интеллект. Число научных публикаций — 15. borodinov@gmail.com; ул. Ивана Черных, 31-33А, 198095, Санкт-Петербург, Россия; р.т.: +7(812)363-0719.

**Заруцкий Игорь Вячеславович** — канд. техн. наук, старший научный сотрудник, лаборатория автоматизации измерений и цифровой обработки сигналов, ИАП РАН. Область научных интересов: представление и обработка сигналов и изображений в аналитических приборах. Число научных публикаций — 51. igorzv@yandex.ru; ул. Ивана Черных, 31-33А, 198095, Санкт-Петербург, Россия; р.т.: +7(812)363-0719.

**Петров Александр Иванович** — заведующий сектором электроники и программного обеспечения, лаборатория методов и приборов иммунного и генетического анализа, ИАП РАН. Область научных интересов: представление и обработка сигналов и изображений в аналитических приборах. Число научных публикаций — 21. fataip@mail.ru; ул. Ивана Черных, 31-33А, 198095, Санкт-Петербург, Россия; р.т.: +7(812)363-0719.

**Сараев Алексей Сергеевич** — инженер 1 категории, лаборатория методов и приборов иммунного и генетического анализа, ИАП РАН. Область научных интересов: представление и обработка сигналов и изображений в аналитических устройствах. Число научных публикаций — 4. alex.niispb@yandex.ru; ул. Ивана Черных, 31-33А, 198095, Санкт-Петербург, Россия; р.т.: +7(812)363-0719.

**Курочкин Владимир Ефимович** — д-р техн. наук, профессор, руководитель научного направления, заведующий лабораторией, лаборатория методов и приборов иммунного и генетического анализа, ИАП РАН. Область научных интересов: исследования и оптимизация электромиграционных методов анализа, развитие аналитических методик для капиллярного электрофореза, исследование оптических методов детектирования, разработка методов и приборов для ДНК анализа, разработка методик подготовки проб и специализированных реактивов. Число научных публикаций — 200. lavovas@yandex.ru; ул. Ивана Черных, 31-33А, 198095, Санкт-Петербург, Россия; р.т.: +7(812)363-0719.

**Поддержка исследований.** Работа выполнена в соответствии с Государственным заданием Министерства науки и высшего образования РФ № 075-01157-23-00 от 29.12.2022 г.

V. MANOILOV, A. BORODINOV, I. ZARUTSKY, A. PETROV, A. SARAEV,  
V. KUROCHKIN

**ALGORITHMS FOR THE PRIMARY ANALYSIS OF LOCAL  
FLUORESCENCE OBJECTS IN THE DNA SEQUENCER  
«NANOFOR SPS»**

*Manoilov V., Borodinov A., Zarutsky I., Petrov A., Saraev A., Kurochkin V.* **Algorithms for the Primary Analysis of Local Fluorescence Objects in the DNA Sequencer «Nanofor SPS».**

**Abstract.** The DNA sequencer "Nanofor SPS", developed at the Institute of Analytical Instrumentation of the Russian Academy of Sciences, implements a method for massively parallel sequencing to decrypt the sequence of nucleic acids. This method allows for the determination of the nucleotide sequence in DNA or RNA, containing from several hundred to hundreds of millions of bases. Thus, there is the opportunity to obtain detailed information about the genome of various biological entities, including humans, animals, and plants. A crucial part of this device is the software, without which it is impossible to solve genome decoding tasks. The output data of optical detection in the sequencer are a set of images over four channels, corresponding to nucleotide types: A, C, G, T. Through specialized software, the position of molecular clusters and their intensity characteristics, along with parameters of the surrounding background, are determined. Algorithms and programs for processing fluorescence signals, discussed in the paper, were developed during the creation of the device software. Also, to debug and test the working programs, models of image construction similar to real data obtained in the course of sequencer operation were created. These models made it possible to obtain a significant amount of information without running expensive experiments. Significant progress has been made in the field of machine learning in recent years, including in the field of bioinformatics, leading to the implementation of the most common models and their potential for practical tasks. However, while these methods have amply proven their worth in secondary bioinformatics data analysis, their potential for the primary analysis remains untapped. This paper focuses on the development and implementation of machine learning methods for primary analysis of optical images of fluorescence signals in reaction cells. The methods of clustering and their testing on models and images obtained from the device are described. The aim of this paper is to demonstrate the capabilities of algorithms for primary analysis of fluorescence signals that arise during sequencing in the «Nanofor SPS» device. The paper describes the main tasks of fluorescence signal analysis and compares traditional methods of solving them and solutions using machine learning technologies.

**Keywords:** sequencing, nucleic acid, processing DNA and RNA fluorescence signals methods, image analysis, machine learning.

## References

1. Kurochkin V.E., Alekseev Ya.I., Petrov D.G., Evstrapov A.A. [Domestic devices for molecular genetic analysis: developments of the IAP RAS and Syntol LLC]. *Izvestiya Rossijskoj Voenno-medicinskoj akademii – Proceedings of the Russian Military Medical Academy*. 2021. vol. 40. no. 3. pp. 69–74. DOI: 10.17816/rmmar76918. (In Russ.).
2. Ansorge W.J. Next-generation DNA sequencing techniques. *Nature Biotechnology*. 2009. vol. 25. no. 4. pp. 195–203.

3. Bentley R.D. Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008. vol. 456. no. 7216. pp. 53–59.
4. Whiteford N. The Solexa pipeline. 2012. Available at: <http://41j.com/blog/wp-content/uploads/2012/04/pipeline.pdf> (accessed 20.02.2024).
5. Leshkowitz D. Introduction to Deep-Sequencing Data Analysis Illumina Primary Analysis Pipeline & Quality Control. 2017. Available at: [http://dors.weizmann.ac.il/course/course2017/Dena\\_IlluminaPrimaryAnalysisPipeline-course2017.pdf](http://dors.weizmann.ac.il/course/course2017/Dena_IlluminaPrimaryAnalysisPipeline-course2017.pdf) (accessed 20.02.2024).
6. Manojlov V.V., Borodinov A.G., Saraev A.S., Petrov A.I., Zaruckij I.V., Kurochkin V.E. [Image processing algorithms in the DNA sequencer NANOPHORE SPS]. *Zhurnal tehnichej fiziki – Journal of Technical Physics*. 2022. vol. 92. no. 7. pp. 985–992. DOI: 10.21883/JTF.2022.07.52655.318-21. (In Russ.).
7. Manojlov V.V., Borodinov A.G., Zaruckij I.V., Petrov A.I., Kurochkin V.E. [Algorithms for processing fluorescence signals of mass parallel sequencing of nucleic acids]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2019. vol. 18. no. 4. pp. 1010–1036. DOI: 10.15622/sp.2019.18.4.1010-1036. (In Russ.).
8. Borodinov A.G., Manojlov V.V., Zaruckij I.V., Petrov A.I., Kurochkin V.E., Saraev A.S. Machine learning in base-calling tasks for next-generation sequencing methods. *Informatika i avtomatizacija – Informatics and Automation*. 2022. vol. 21. no. 3. pp. 572–603. DOI: 10.15622/ia.21.3.5. (In Russ.).
9. Zhuravel' I.M. Kratkij kurs teorii obrabotki izobrazhenij. Available at: <http://matlab.exponenta.ru/imageprocess/book2/49.php> (accessed 26.10.2023). (In Russ.).
10. Vuds R., Gonsales R. *Cifrovaya obrabotka izobrazhenij [Digital image processing]*. 3rd ed. Moscow: Tekhnosfera, 2012. 1104 p. (In Russ.).
11. Sizikov V.S. Spectral method for estimating the point-spread function in the task of eliminating image distortions. *Journal of Optical Technology*. 2017. vol. 84. no. 2. pp. 95–101.
12. Sizikov V.S., Stepanov A.V., Mezhenin A.V., Burlov D.I., Eksemplarov R.A. Determining image-distortion parameters by spectral means when processing pictures of the earth's surface obtained from satellites and aircraft. *Journal of Optical Technology*. 2018. vol. 85. no. 4. pp. 203–110.
13. Bardin B.V., Chubinskij-Nadezhdin I.V. [Local object detection in digital microscopic images]. *Nauchnoe priborostroenie – Nauchnoe Priborostroenie*. 2009. vol. 19. no. 4. pp. 96–102. (In Russ.).
14. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*. 1979. vol. 9. pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.
15. Saraev A.S., Petrov A.I., Manojlov V.V. [Modeling the generation of clusters of DNA molecules in massively parallel sequencing devices] *Tezisy dokladov Chetvertoj mezhdunarodnoj konferencii so shkoloj molodyh uchenyh «Fizika – naukam o zhizni» [Abstracts of reports of the Fourth International Conference with the School of Young Scientists «Physics – Life Sciences»]*. SPb: Ioffe Institute. 2021. pp. 153 (In Russ.).
16. Saraev A.S. *Nauchno-kvalifikacionnaya rabota «Razrabotka algoritma raspoznavaniya klasterov nukleinovyh kislot v mikroflyuidnoj yachejke sekvenatora «Nanofor SPS» [Scientific qualification work «Development of an algorithm for recognizing nucleic acid clusters in the microfluidic cell of the Nanofor SPS sequencer»]*. SPb: IAIRAS, 2023. pp. 16–22. (In Russ.).
17. Schmidt B., Hildebrandt A. Deep learning in next-generation sequencing. *Drug discovery today*. 2021. vol. 26. no. 1. pp. 173–180.

18. Ozgur S., Orman M. Application of deep learning technique in next generation sequence experiments. *Journal of Big Data*. 2023. vol. 10. no. 1. DOI: 10.1186/s40537-023-00838-w.
19. Tegfalk E. Application of machine learning techniques to perform base-calling in next-generation DNA sequencing. 2020. 45 p.
20. Cacho A., Smirnova E., Huzurbazar S., Cui X. A comparison of base-calling algorithms for illumina sequencing technology. *Briefings in bioinformatics*. 2016. vol. 17. no. 5. pp. 786–795.
21. Kircher M., Stenzel U., Kelso J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome biology*. 2009. vol. 10(8). DOI: 10.1186/gb-2009-10-8-r83.
22. Ghojogh B., Crowley M. The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. arXiv preprint arXiv:1905.12787. 2019.
23. Borodinov A.G., Yamanovskaya A.Yu., Manojlov V.V., Petrov A.I. [Optimal feature selection for reducing the dimensionality of machine learning models in the base-calling problem] *Tezisy dokladov Vtoroj ezhegodnoj vserossijskoj molodezhnoj konferencii po metodam i priboram dlja analiza biologicheskikh obektov «AnalitBioPribor-2023»* [Abstracts of reports of the Second Annual All-Russian Youth Conference on Methods and Instruments for the Analysis of Biological Objects «AnalitBioPribor-2023»]. St. Petersburg: Publishing and Printing Association of Higher Educational Institutions, 2023. pp. 135–138. (In Russ.).
24. Whiteford N., Skelly T., Curtis C., Ritchie M.E., Lohr, A., Zaranek A.W., Abnizova I., Brown C. Swift: primary data analysis for the Illumina Solexa sequencing platform. *Bioinformatics*. 2009. vol. 25. no. 17. pp. 2194–2199.
25. Hafiz A.M., Bhat G.M. A survey on instance segmentation: state of the art. *International journal of multimedia information retrieval*. 2020. vol. 9. no. 3. pp. 171–189.
26. Chaudhry M., Shafi I., Mahnoor M., Vargas D.L.R., Thompson E.B., Ashraf I.A. Systematic literature review on identifying patterns using unsupervised clustering algorithms: a Data mining perspective. *Symmetry*. 2023. vol. 15. no. 1679. DOI: 10.3390/sym15091679.
27. Khandare A., Pawar R. Data clustering algorithms: experimentation and comparison. *Intelligent Computing and Networking: Proceedings of IC-ICN 2021*. 2022. pp. 86–99.
28. Sarang P. BIRCH: Divide and Conquer. *Thinking Data Science: A Data Science Practitioner's Guide*. Cham: Springer International Publishing, 2023. pp. 229–236.
29. Ester M., Krieger H.P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. *kdd*. 1996. vol. 96. no. 34. pp. 226–231.
30. Zhao Y., Cao J., Zhang C., Zhang S. Enhancing grid-density based clustering for high dimensional data. *Journal of Systems and Software*. 2011. vol. 84. no. 9. pp. 1524–1539.
31. Wolowski V.R. High-quality, high-throughput measurement of protein-DNA binding using HiTS-FLIP. *Dissertation, LMU Munchen: Fakultat fur Chemie und Pharmazie* 2016. DOI: 10.5282/edoc.19445.
32. Chityala R., Pudipeddi S. Image processing and acquisition using Python. *Chapman and Hall/CRC*. 2020. 452 p.
33. Kameshwaran K., Malarvizhi K. Survey on clustering techniques in data mining. *International Journal of Computer Science and Information Technologies*. 2014. vol. 5. no. 2. pp. 2272–2276.

**Manoilov Vladimir** — Ph.D., Dr.Sci., Associate Professor, Head of the laboratory, Laboratory of automation of measurements and digital signal processing, IAI RAS. Research interests: representation and processing of signals and images in analytical devices. The number of publications — 101. manoilov-vv@mail.ru; 31-33A, Ivan Chernykh St., 198095, St. Petersburg, Russia; office phone: +7(812)363-0720.

**Borodinov Andrew** — Ph.D., Senior researcher, Laboratory of methods and instruments for immune and genetic analysis, IAI RAS. Research interests: mathematical statistics, problems of analysis, processing and presentation of data, artificial intelligence. The number of publications — 15. borodinov@gmail.com; 31-33A, Ivan Chernykh St., 198095, St. Petersburg, Russia; office phone: +7(812)363-0719.

**Zarutsky Igor** — Ph.D., Senior researcher, Laboratory of automation of measurements and digital signal processing, IAI RAS. Research interests: representation and processing of signals and images in analytical devices. The number of publications — 51. igorzv@yandex.ru; 31-33A, Ivan Chernykh St., 198095, St. Petersburg, Russia; office phone: +7(812)363-0719.

**Petrov Alexander** — Head of the sector of electronics and software, Laboratory of methods and instruments for immune and genetic analysis, IAI RAS. Research interests: representation and processing of signals and images in analytical devices. The number of publications — 21. fataip@mail.ru; 31-33A, Ivan Chernykh St., 198095, St. Petersburg, Russia; office phone: +7(812)363-0719.

**Saraev Alexey** — Engineer of the 1st category, Laboratory of methods and instruments for immune and genetic analysis, IAI RAS. Research interests: representation and processing of signals and images in analytical devices. The number of publications — 4. alex.niispb@yandex.ru; 31-33A, Ivan Chernykh St., 198095, St. Petersburg, Russia; office phone: +7(812)363-0719.

**Kurochkin Vladimir** — Ph.D., Dr.Sci., Professor, Head of the scientific direction, head of the laboratory, Laboratory of methods and instruments for immune and genetic analysis, IAI RAS. Research interests: research and optimization of electromigration analysis methods, development of analytical methods for capillary electrophoresis, study of optical methods of detection, development of methods and instruments for DNA analysis, development of methods for preparing samples and specialized reagents. The number of publications — 200. lavrovas@yandex.ru; 31-33A, Ivan Chernykh St., 198095, St. Petersburg, Russia; office phone: +7(812)363-0719.

**Acknowledgements.** This research was performed in accordance with the State Assignment of the Ministry of Science and Higher Education of the Russian Federation No. 075-01157-23-00 dated 12/29/2022.

К.В. СОБЯНИН, С.П. КУЛИКОВА  
**UNET-BOOSTED CLASSIFIER – МУЛЬТИЗАДАЧНАЯ  
АРХИТЕКТУРА ДЛЯ МАЛЫХ ВЫБОРОК НА ПРИМЕРЕ  
КЛАССИФИКАЦИИ МРТ СНИМКОВ ГОЛОВНОГО МОЗГА**

*Собянин К.В., Куликова С.П. Unet-boosted classifier – мультизадачная архитектура для малых выборок на примере классификации МРТ снимков головного мозга.*

**Аннотация.** Проблема обучения глубоких нейронных сетей на малых выборках особенно актуальна для медицинских задач. В работе рассматривается влияние попиксельной разметки значимых объектов на изображении, в дополнении к истинной метке класса, на качество решения задачи классификации. Для достижения лучших результатов классификации на малых выборках предлагается мультизадачная архитектура Unet-boosted classifier (UBC), обучаемая одновременно для решения задач классификации и семантической сегментации. В качестве исследуемого набора данных используются МРТ-снимки пациентов с доброкачественной глиомой и глиобластомой, взятые из открытого набора данных BraTS 2019. В качестве входа рассматривается один горизонтальный срез МРТ-изображения, содержащий глиому (всего 380 кадров в обучающей выборке), в качестве выхода – вероятность глиобластомы. В качестве базового решения используется ResNet34, обученный без аугментаций с функцией потерь на основе взаимной энтропии (CrossEntropyLoss). В качестве альтернативного решения используется UBC-ResNet34 – тот же ResNet34 усиленный декодером, построенным по принципу U-Net, и предсказывающим положение глиомы. В качестве дополнительной функции потерь используется сложенный коэффициент Соренсена-Дайса (DiceLoss). Результаты на тестовой выборке: доля правильных ответов (acc@casu) для базовой модели составила 0.71, для альтернативной – 0.81, коэффициент Дайса (Dice score) при этом составил 0.77. Таким образом, глубокую модель можно качественно обучить даже на небольшом наборе данных, используя предложенную архитектуру и добавив в разметку информацию о пораженных тканях в виде семантической маски. Предлагаемый подход потенциально может быть полезен и в любых других задачах классификации изображений с ограниченным набором данных.

**Ключевые слова:** классификация изображений, глубокое обучение, малый набор данных, семантическая сегментация, мультизадачная архитектура, церебральная патология, диагностика опухолей.

**1. Введение.** Разработка методов персонализированной медицины является одним из приоритетных направлений исследований в рамках Стратегии научно-технологического развития Российской Федерации. Это направление предоставляет обширные возможности для применения современных методов анализа данных и машинного обучения. В частности, одним из наиболее эффективных и перспективных методов анализа медицинских сигналов и изображений являются глубокие нейронные сети [1, 2]. Так, например, в обзоре [3] рассмотрено 49 научных исследований, посвященных диагностированию рака молочной железы, 39 из которых использовали глубокие сверточные сети. Обзор [4] описывает результаты 16 исследований, использующих глубокие сверточные сети

для диагностирования болезни Альцгеймера по МРТ-снимкам. Доля правильных ответов в этих исследованиях варьировалась от 0.79 до 0.99 на различных наборах данных. Также предыдущие исследования показывают, что глубокие сверточные сети успешно справляются с задачами классификации и локализации заболеваний органов грудной клетки [5] или выявления рака кожи на уровне компетентности специалистов [6].

Однако для обеспечения высокого качества работы таких моделей требуются значительные по объему массивы обучающих данных. Так, авторы в [6] использовали в своей работе по определению рака кожи модель архитектуры Inception v3 [7], предобученную на 1.41 млн изображений из набора ImageNet [8], которую они дообучили на наборе данных, содержащем 129 450 клинических изображений, размеченных экспертами-дерматологами. Для решения задачи классификации заболеваний органов грудной клетки (легких и сердца) авторы в [5] построили свою модель, объединив 4 другие модели, предобученные на ImageNet: AlexNet [9], GoogLeNet [10], VGGNet-16 [11] и ResNet-50 [12]. В этой работе для обучения классификатора и переходных слоев, объединяющих признаки всех четырех моделей, использовали набор данных «ChestX-ray8», собранный в 1992-2015 гг и содержащий 108 948 рентгеновских снимков грудной клетки 32 717 уникальных пациентов.

В то же время сбор данных в рамках исследования отдельной клинической задачи, как правило, является весьма трудоемким и зачастую продолжительным и дорогостоящим процессом. Стоит также отметить, что клинические данные нередко имеют пропущенные данные из-за невозможности прохождения пациентом тех или иных видов обследования, а сами данные могут варьироваться в зависимости от используемого клинического оборудования и протоколов обследований. Ввиду этих обстоятельств полноценно обследованная выборка даже из ста пациентов уже считается хорошим результатом. В связи с этим ограничением возникает потребность рассмотрения подходов, позволяющих повысить качество обучения модели на малых выборках.

Современные модели машинного обучения, сравнимые по качеству работы с человеком [13] часто содержат миллионы параметров [11] и обучаются на миллионах изображений [8]. Этот факт отсылает нас к проблеме «проклятия размерности», сформулированной в 1961 году американским математиком Ричардом Бэллманом [14], которую можно интерпретировать как «количество данных необходимое для оценки произвольной функции с заданным уровнем точности растет экспоненциально по отношению к числу входных данных» [15]. Таким

образом, число входных данных, то есть сам по себе размер изображения накладывает ограничение на задачу классификации. Для преодоления этой проблемы есть два принципиальных подхода: 1) увеличение количества обучающих данных и 2) регуляризация – ограничение количества степеней свободы модели, то есть фактически понижение размерности пространства входных данных.

В своем недавнем обзоре [16] авторы проанализировали 26 исследований, опубликованных за период с 2015 по 2021 г. и посвященных проблемам обучения моделей на малых наборах данных. В результате анализа авторы выделили 5 семейств подходов, используемых для обучения моделей на малых выборках:

– Модификации архитектуры – включают в себя различные модификации архитектуры на всех масштабах, регулирующие обучение: от дропаут-техник (англ. dropout) [17] и специальных «узких» блоков (англ. bottleneck block) [18] до различных встроенных фильтров, нацеленных на поиск определенных сигналов или геометрических примитивов, например гармонических блоков, которые обладают свойством инвариантности к вращению [19]. Также классическими примерами таких подходов являются дистилляция и ускорение обучения [20, 21]. Так, например, в [22] авторы использовали инвариантные к аффинному масштабированию сверточные блоки, использование которых позволило удвоить долю правильных ответов по сравнению с альтернативными моделями, таким как классические сверточные, гармонические [19] и капсульные нейронные сети [23], на ограниченной выборке (по 5 изображений на класс) из набора данных по распознаванию дорожных знаков Traffic Sign [24]. Еще одним примером использования подходов такого рода служит работа [25], в которой они использовали тангенсальные ядра (англ. Neural Tangent Kernel, NTK), эквивалентные линейным слоям неограниченной ширины, что позволило на 3% превзойти ResNet-34 [12] по результатам обучения на 640 изображениях из набора CIFAR-10 [26].

– Модификации функции потерь – обозначают подходы, ограничивающие обучение модели на уровне функции потерь. Подобным образом эффективно может быть использована косинусная мера близости [27], или в сочетании с обычной функцией потерь может использоваться метафункция потерь, обученная под конкретную задачу [28]. В частности, в [22] авторы использовали регуляризатор, инвариантный к вращению, который добавляется к функции потерь в качестве штрафного слагаемого. Еще одним примером подходов такого рода является работа [29], в которой авторы предложили геометрическую функцию потерь OLE loss (ang. Orthogonal Low-rank Embedding). Такая

функция потерь поощряет геометрическую разделимость между классами на уровне признаков, что позволило в сочетании с обычной функцией потерь достичь кратного прироста в качестве на выборках, содержащих менее 100 изображений из наборов MNIST [30] CIFAR-10, CIFAR-100 [26] и других.

– Аугментации данных – к этому семейству относятся все методы, увеличивающие размер набора обучающих данных [31], начиная от стандартных (кадрирование, смещение, поворот или масштабирование изображения и т.д.), до сложных методов генерации данных, например, с помощью генеративно-состязательных сетей (англ. Generative adversarial network, GAN) на основе исходного набора и бинарных масок [32]. Один из подходов такого рода представлен в работе [33], в которой была предложена соревновательная техника аугментации малых наборов данных DADA (англ. deep adversarial data augmentation) с добавлением объектов «фейковых» классов, находящихся в «спорных» точках признакового пространства с точки зрения меры близости объектов. Предложенный подход показал свои преимущества в том числе в задачах классификации электроэнцефалографических (ЭЭГ) сигналов и в определении рака груди на маммографических снимках.

– Аугментация признаков – это подходы, при которых к признакам добавляется шум с целью улучшения устойчивости обучения модели [34]. Одним из примеров такого подхода является предложенная в [34] состязательная аугментация признаков случайно выбранных скрытых слоев, что позволяет стабильно получать прирост в качестве 2-5% на небольших подвыборках MNIST [30] и CIFAR-10 [26].

– «Теплый старт» – это стратегия инициализации весов, при которой отдельные части модели предварительно обучаются на других данных, которые могут улучшить сходимость модели. Это могут быть другие реальные данные [35] или данные сгенерированные, например, с добавлением «фейковых» классов [33]. Так, например, в [36] авторы использовали обучение без учителя, чтобы обучить кодировщик MoCo v2 [37] для последующей дистилляции под задачу классификации.

Стоит отметить, что для малых выборок все ещё популярны классические модели, так как они зачастую содержат меньше параметров, и их довольно просто регуляризовать [38]. Так, например, Кристодулу и соавторы [39] отмечают в своем мета-анализе, что во многих клинических предсказательных задачах результаты работы моделей на основе методов машинного обучения зачастую сопоставимы с результатами моделей на основе обычной логистической регрессии. Тем не менее, авторы отмечают перспективность разработки подходов на основе машинного обучения.

Кроме выше перечисленных подходов, отдельно можно выделить семейство многозадачных архитектур, решающих одновременно несколько задач машинного обучения. Примером применения такого подхода является работа Мойаб и соавторов [40], в которой модель одновременно обучали решать задачу классификации и сегментации. Применение этого подхода позволило достичь прироста доли правильных ответов до 30% по сравнению с базовым решением в виде ResNet101 [12] на малых выборках из открытых баз данных MNIST [30], CIFAR-10 и CIFAR-100 [26], а также продемонстрировало свою работоспособность на клинической выборке, содержащей 45 изображений глазного дна, каждое из которых принадлежало одному из трех классов (без патологии, с признаками диабетической ретинопатии, наличие глаукомы).

Также хорошие результаты обучения на малых выборках для решения задачи семантической сегментации демонстрирует и архитектура U-Net [41], предложенная в 2015 году Роннебергом и соавт. В частности, авторам архитектуры удалось достичь качества сегментации клеток глиобластомы-астроцитомы U373 с коэффициентом Жаккара в 0.92 (англ. Intersection over Union, IoU), обучив модель на наборе данных PhC-U373, содержащем 32 частично размеченных снимка. Такие результаты обеспечиваются использованием меньшего количества параметров модели и большего объема полезной информации, извлекаемой из каждого отдельного пикселя разметки, по сравнению с архитектурами, содержащими полносвязные слои и обучающимися предсказывать только метку класса. Архитектура U-Net состоит из двух основных блоков: кодировщика (англ. encoder), который сжимает исходное изображение до признакововой карты (англ. feature map), и декодера (англ. decoder), который восстанавливает разметку. Поскольку кодировщик извлекает признаки из изображения и, по сути, является главным элементом в любой архитектуре, то, естественным образом, возникает желание переиспользовать его в задаче классификации.

В данной работе мы представляем подход, использующий дополнительную информацию об обучающей выборке в виде попиксельной разметки пораженных тканей, для лучшего обучения глубокой сверточной сети. Ключевой идеей подхода является обучение кодировщика одновременно внутри двух моделей – классической сверточной сети, решающей задачу классификации, и полносверточной архитектуры U-net, решающей задачу семантической сегментации и обладающей свойством обучаться даже на небольших наборах данных.

Качество работы предлагаемого подхода было протестировано на задаче классификации одиночных срезов МРТ-снимков на предмет

выявления злокачественной опухоли головного мозга. В результате обучения модели на снимках 76 пациентов предложенный подход показал прирост доли правильных ответов на 14% по сравнению с базовым решением.

Применение предлагаемого подхода может быть расширено на другие задачи анализа медицинских изображений или в других предметных областях для задач с малыми выборками.

**2. Данные.** В качестве исследуемых данных рассматривается открытый набор данных MICCAI по сегментации опухолей головного мозга (BraTS 2019) [42 – 44].

BraTS 2019 содержит данные 76 пациентов с диагнозами глиома (англ. lower grade glioma, LGG) и 259 пациентов с диагнозом глиобластома (англ. high grade glioma, HGG). Для каждого пациента представлены снимки 4 видов МРТ последовательностей:

- T1-взвешенное изображение (пропорционально времени продольной релаксации тканей);
  - T1-взвешенное изображение с контрастированием гадолинием;
  - T2-взвешенное изображение (пропорционально времени поперечной релаксации тканей);
  - T2 изображение инверсионного восстановления с подавлением сигнала от жидкости (англ. Fluid Attenuated Inversion Recovery, FLAIR).
- Указанные данные были записаны с использованием различных МРТ сканеров и клинических протоколов в 19 клинических центрах, но в дальнейшем были совмещены с единым анатомическим шаблоном [45] и передискретизированы к разрешению  $1 \times 1 \times 1$  мм.

Также для каждого пациента имеются аннотации, указывающие на тип ткани, верифицированные 4 клиническими экспертами [43]:

- Нормальная ткань;
- Перитуморальная отечная/инвазированная ткань;
- Неконтрастирующее ядро;
- Некротическое ядро, наполненное жидкостью;
- Контрастирующее ядро.

В качестве исследуемых данных были взяты снимки 76 пациентов с LGG и случайным образом выбранных 76 пациентов с HGG. Полученная выборка была разделена на две равные части с сохранением пропорции классов: обучающую и тестовую. Для каждого пациента были извлечены по 5 горизонтальных срезов FLAIR последовательности с наибольшей площадью поражения (рисунок 1) и соответствующие им бинарные маски (рисунок 2), являющиеся объединением всех пораженных тканей на срезе. В общей сложности, размер обучающей и тестовой выборки составил по

190 одноканальных изображений каждого класса с разрешением  $224 \times 224$  пикселей.

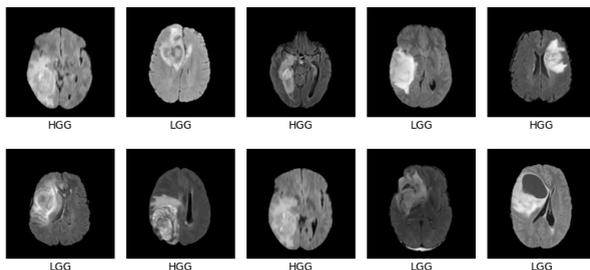


Рис. 1. Пример входных данных. На изображениях представлены МРТ-снимки FLAIR последовательности пациентов с диагнозами глиома (LGG) и глиобластома (HGG)

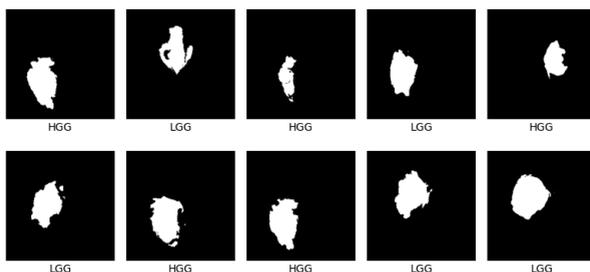


Рис. 2. Примеры бинарных масок новообразований головного мозга соответствующих МРТ-изображениям, представленным на рисунке 1

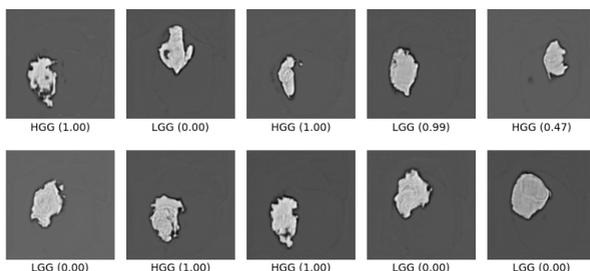


Рис. 3. Ответы UBC-resnet34 на изображения, представленные на рисунке 1. Представлена карта откликов до бинаризации, подпись указывает на истинный класс, число в скобках определяет уверенность модели в положительном классе – глиобластоме (HGG)

**3. Базовая модель.** Глубокие модели машинного обучения для решения задач классификации изображений обычно состоят из двух частей: кодировщика (англ. *encoder*), извлекающего признаки (англ. *feature map*) из изображения, и классификатора (англ. *classification head*), предсказывающего вероятность классов на основе признаков, извлеченных кодировщиком (рисунок 4). Благодаря иерархическому устройству сверточных сетей каждый последующий уровень модели извлекает все более сложные признаки, при этом на выходе из кодировщика объекты разных классов иногда даже становятся линейно разделимы в пространстве признакового описания. Поэтому в большинстве известных архитектур глубоких сверточных нейронных сетей, начиная с LeNet [46], классификатор представляет собой последовательность нескольких линейных полносвязных слоев с нелинейными функциями активации между ними.

Полносвязный слой суммирует все входные значения с некоторыми коэффициентами столько раз, сколько имеет выходов, и фактически решает задачу линейной регрессии для каждого выхода (1).

$$z_i(\mathbf{x}) = \mathbf{w}_i \cdot \mathbf{x} + b_i, \quad (1)$$

где  $z_i$  – значение  $i$ -го выхода полносвязного слоя,  $\mathbf{x}$  – входной вектор,  $\mathbf{w}_i$  – весовые коэффициенты  $i$ -го выхода,  $b_i$  – сдвиг  $i$ -го выхода (англ. *bias*).

Последовательность из нескольких полносвязных слоев очень быстро «раздувает» модель и во многих архитектурах большая часть параметров содержится именно в полносвязных слоях.

На малых выборках такие полносвязные сети имеют свойство очень быстро переобучаться – запоминать особенности выборки. Переобученная модель может хорошо описывать данные, на которых она обучалась, но её поведение непредсказуемо на новых данных, так как модель запомнила особенности именно тренировочного набора и не обладает обобщающей способностью. Поэтому, чтобы снизить эффект переобучения последних полносвязных слоев, в качестве базового решения была выбрана модель архитектуры ResNet34 [12] с однослойным линейным классификатором.

ResNet – это архитектура, решившая проблему плохого обучения глубоких моделей с помощью добавления специальных соединений (англ. *residual connections*), идущих в обход следующего блока и позволяющих градиенту во время обучения беспрепятственно распространяться через всю сеть.

ResNet34 состоит из 34 обучаемых слоев, 33 из которых содержатся в кодировщике и 1 в классификаторе. Кодировщик ResNet34 имеет входной блок, состоящий из сверточного слоя (Conv2d) с ядром  $7 \times 7$ , который расслаивает исходное изображение на 64 канала, не изменяя при этом его размер, нормализующего слоя (BatchNorm), функции активации (ReLU) и слоя понижающего дискретизацию признаков с сохранением наиболее сильных активаций (MaxPooling).

Далее устройство кодировщика имеет регулярную структуру, состоящую из 4-х уровней, каждый из которых работает со своим масштабом признаков и состоит из одинаковых базовых блоков (BasicBlock). Базовый блок представляет собой последовательность двух сверточных слоев (Conv2d) с ядром  $3 \times 3$  с последующей нормализацией (BatchNorm) и с активацией (ReLU) по середине. Вход каждого базового блока прибавляется к его выходу, усредняя тем самым исходное внутреннее представление и полученное в результате работы базового блока. Такой подход гарантирует прохождение градиента через всю сеть, даже если он занулится в каком-то конкретном слое. Переход к следующему масштабу обеспечивается дополнительным сверточным слоем в конце блока с ядром  $1 \times 1$  и с шагом 2 (англ. stride). В общей сложности, кодировщик ResNet34 содержит 21.3 млн обучаемых параметров и сжимает изображение разрешения  $224 \times 224$  до признакововой карты с разрешением  $512 \times 7 \times 7$  (таблица 1).

Таблица 1. Уровни кодировщика ResNet34

ResNet34 encoder levels and params					
name	B-blocks	convs	params	input shape	output shape
Input	0	1	3.26 K	$1 \times 224 \times 224$	$64 \times 112 \times 112$
level1	3	6	222 K	$64 \times 112 \times 112$	$64 \times 112 \times 112$
level2	4	8	1.12 M	$64 \times 112 \times 112$	$128 \times 56 \times 56$
level3	6	12	6.82 M	$128 \times 56 \times 56$	$256 \times 28 \times 28$
level4	3	6	13.1 M	$256 \times 28 \times 28$	$512 \times 7 \times 7$
total	16	33	21.3 M	$1 \times 224 \times 224$	$512 \times 7 \times 7$

Классификатор ResNet34 содержит 1026 параметров и состоит из AveragePooling слоя, который усредняет значения каналов, и одного линейного полносвязного слоя с одним выходом – вероятностью положительного класса.

**4. Unet-Boosted Classifier.** Предлагаемая в данной работе модель представляет собой гибридную мультизадачную архитектуру, одновременно решающую задачи классификации и семантической

сегментации (рисунок 5). Требование на уровне архитектуры кодирования и восстановления потенциально важных для классификации областей изображения позволяет обучить более глубокую сеть на тех же данных и повысить качество классификации. Поскольку такой подход нацелен на улучшение решения задачи классификации, мы назвали предложенную архитектуру Unet-Boosted Classifier (UBC), а модель, построенную по этому принципу на базе ResNet34 – UBC-ResNet34.

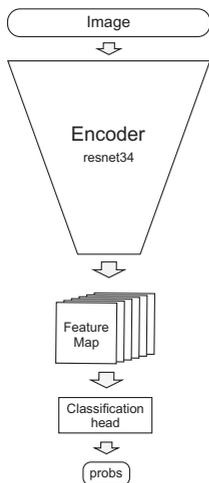


Рис. 4. Базовая модель:  
ResNet34

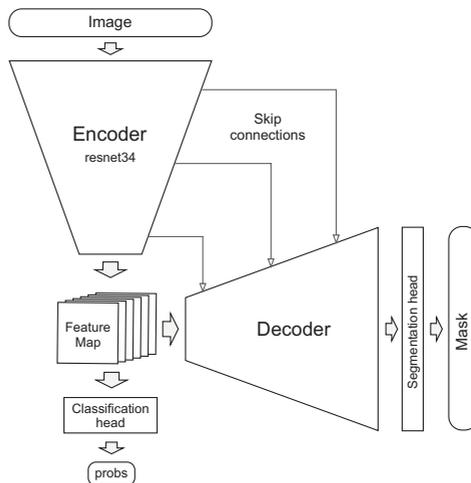


Рис. 5. Предлагаемая модель:  
UBC-ResNet34

UBC-ResNet34 состоит из тех же блоков, что и ResNet34: кодировщика и классификатора, но также содержит декодер (англ. decoder), восстанавливающий маску из признаков. Декодер UBC-ResNet34 построен по принципу U-Net и соединён с выходом кодировщика и его промежуточными результатами с помощью специальных связей (англ. skip connections), обеспечивающих лучшее распространение градиента и пропускание признаков от кодировщика к декодеру на разных масштабах. Декодер принимает на вход признаки всех масштабов и состоит из пяти уровней. Каждый уровень декодера смешивает признаки двух масштабов и отбирает наиболее значимые. Фактически представляет собой один повышающий дискретизацию слой (англ. up-convolution или ConvTranspose2d) для признаков меньшего масштаба с последующей конкатенацией с признаками следующего масштаба и последовательность двух сверточных слоев с ядром 3x3 (Conv2D). После каждого сверточного

слоя применяется нормализация (BatchNorm) для лучшей устойчивости обучения, а так же активация (ReLU). В общей сложности декодер содержит 3.15 млн. параметров и растягивает выход кодировщика в 32 раза до исходного размера изображения (таблица 2).

Таблица 2. U-Net декодер

U-Net decoder levels				
name	convs	params	input shape	output shape
level1	2	2.4 M	256×28×28 + 512×7×7	256×28×28
level2	2	590 K	128×56×56 + 256×28×28	128×56×56
level3	2	148 K	64×112×112 + 128×56×56	64×112×112
level4	2	46.2 K	64×112×112 + 64×112×112	32×112×112
level5	2	6.98 K	32×112×112	16×224×224
total	10	3.15 M	-	16×224×224

Далее результат работы декодера поступает на вход однослойной сегментирующей сети (англ. segmentation head), содержащей 145 параметров, которая восстанавливает маску изображения.

Таким образом, на входе модель также принимает одноканальное изображение, но имеет два выхода: выход классификатора – вероятность положительного класса, и выход сегментатора – маску, указывающую на поврежденные ткани.

**5. Обучение.** Для реализация моделей и их обучения использовались библиотеки pytorch и pytorch lightning. Реализация ResNet34 была взята из модуля torchvision.models, декодера Unet – из библиотеки segmentation models pytorch. Обучение моделей выполнялось методом обратного распространения ошибки (англ. backpropagation), с помощью стохастического градиентного спуска (англ. mini-batch gradient descent). Размер выборки итерации (англ. batch size) для обучения обеих моделей был одинаковый и составлял 50 обучающих примеров. Оптимизатор использовался адаптивный с инерцией Adam [47] с начальным шагом  $2 \cdot 10^5$  (англ. learning rate) и множителем угасания 0.9 (англ. learning rate annealing), применяющимся каждые 5 шагов.

Для оценки вклада предложенной архитектуры по сравнению с базовым решением инициализация параметров моделей происходила случайно, аугментации не использовались, а для вычисления ошибки классификации использовалась стандартная функция потерь на базе взаимной энтропии (CrossEntropyLoss) (2).

$$CE(\hat{p}, p) = \frac{1}{N} \sum_i^N \sum_{c=1}^C p_{ic} \log \hat{p}_{ic}, \quad (2)$$

где  $p$  – вектор разметки для выборки,  $\hat{p}$  – вектор вероятностей, полученный от классификатора,  $N$  – размер выборки,  $C$  – количество классов, в нашем случае 2.

Для вычисления ошибки сегментации использовалась функция потерь на основе сглаженного коэффициента Соренса-Дайса (DiceLoss) (3).

$$SD(\hat{m}, m) = 1 - 2 \frac{\sum (\hat{m} \circ m) + \epsilon}{\sum (\hat{m} + m) + \epsilon}, \quad (3)$$

где  $m$  – бинарная маска, фактически матрица со значениями 0 и 1;  $\hat{m}$  – предсказанная моделью маска;  $\circ$  – операция поэлементного умножения, фактически обнуляет все ложные срабатывания, оставляя только пересечение  $m$  и  $\hat{m}$ , после чего происходит суммирование одновременно по всем осям;  $\epsilon$  – сглаживающий параметр.

Для обучения мультизадачной модели ошибка определялась как сумма ошибок классификации и сегментации (4).

$$L(\hat{p}, \hat{m}, p, m) = \alpha \cdot CE(\hat{p}, p) + (1 - \alpha) \cdot SD(\hat{m}, m), \quad (4)$$

где  $\alpha$  – балансовый гиперпараметр, который в исследовании был принят равным 0.5.

Для более комплексного сравнения каждая из моделей была обучена на малых выборках разного объема: от 1 до 190 изображений на класс. Выборки размером от 1 до 38 изображений на класс представляют собой подвыборки обучающего набора данных, содержащие по 1 случайному изображению на пациента. Выборки из 76 и 190 изображений на класс содержали соответственно по 2 и 5 изображений от каждого пациента из обучающего набора данных.

Модели обучались на протяжении 200 эпох, значения метрик для сравнения выбирались при достижении плато и усреднялись по 5 значениям, вычисленным на тестовой выборке. Тестирование моделей, обученных на подвыборках разных объемов, выполнялось на одном и том же тестовом наборе данных, содержащем 380 изображений от 76 пациентов.

**6. Результаты и дискуссия.** Большая часть моделей, обучавшихся на выборках содержащих менее 50 изображений, достигла локального минимума функции потерь менее, чем за 100 шагов обучения, то есть, при batch size равным 50, менее чем за 100 эпох. ResNet34, обучавшийся на

полном наборе данных (380 изображений) – за 20 эпох, UBC-ResNet34 – за 60 эпох. Результаты обучения на выборках разной величины представлены на рисунке 6. Взаимосвязь качества классификации и сегментации моделей UBC-ResNet34 представлена на рисунке 7.

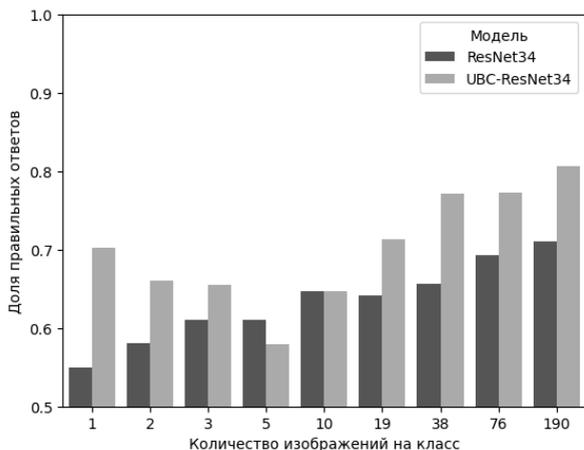


Рис. 6. Результаты обучения моделей на малых выборках

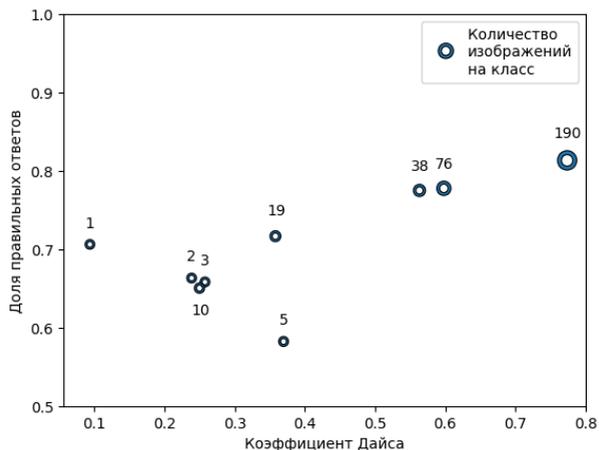


Рис. 7. Взаимосвязь доли правильных ответов с коэффициентом Дайса

Высокое значение доли правильных ответов модели UBC-ResNet34 обученной на выборке из двух изображений можно интерпретировать как выброс, по скольку согласно точечной диаграмме на рисунке 7,

наблюдается положительная зависимость между качеством сегментации и классификации. Мы полагаем, что два данных изображения, попавших в обучающую выборку, наиболее характерны для всего представленного набора данных, поскольку добавление последующих восьми изображений снижает качество классификации на каждом этапе. Так, интерпретируя глубокую просадку качества модели UBC-ResNet34 и снижение темпа роста качества базовой модели – ResNet34, можно прийти к выводу что выборка, состоящая из 10 изображений (по 5 на класс), содержит нехарактерные изображения для своего класса.

Результаты обучения на полных данных представлены в таблице 3. На тестовых данных из 380 изображений, ResNet34 достиг доли правильных ответов ( $accuracy = (TP + TN)/n$ ) в  $0.71 \pm 0.05$ , что значительно отличается от тривиального решения (англ. dummy classifier)  $\hat{y}(x) = 1$  с результатом в  $0.50 \pm 0.05$ . В случае использования предложенной модели UBC-ResNet34 доля правильных ответов составила  $0.81 \pm 0.04$ , коэффициент Дайса (Dice score) при этом составил 0.77. Интервальные оценки получены методом Уилсона для 95% уровня доверия. Примеры ответов модели представлены на рисунке 3.

Таблица 3. Результаты обучения ResNet34 и UBC-ResNet34 на 380 изображениях

Результаты на тестовой выборке из 380 изображений				
Модель	Параметры	Метки	Доля правильных ответов, 95% ДИ	Коэффициент Дайса
dummy	-	-	$0.50 \pm 0.05$	-
ResNet34	21.3 М	190	$0.71 \pm 0.05$	-
UBC-ResNet34	24.4 М	0.90 М	$0.81 \pm 0.04$	0.77

Полученные результаты можно объяснить одновременным влиянием как минимум двух факторов. Во-первых, количество обучающей информации увеличилось соизмеримо с количеством пикселей в изображении, так как в среднем по обучающей выборке, площадь поражения составляет 5% от площади снимка. Таким образом, количество обучаемых параметров выросло всего на 15% (столбец «Параметры» в таблице 3), а количество положительных меток в обучающей выборке выросло более чем на три порядка (столбец «Метки» в таблице 3), что в свою очередь повлияло на качество обучения кодировщика. Во-вторых, есть основания полагать, что требование описывать положение пораженных тканей на уровне признаков позволило кодировщику научиться извлекать важные признаки для детектирования злокачественной опухоли.

Из второго фактора следует важное потенциальное преимущество предложенной архитектуры – возможность продолжать обучать кодировщик на данных пациентов с неустановленным диагнозом. Такие обучающие примеры можно использовать в одном цикле с полностью размеченными данными и обновлять веса кодировщика, используя градиент, полученный только от декодера.

Так же второй фактор представляется наиболее интересным с точки зрения стратегии разметки изображений. Развивая мысль о взаимосвязи качества классификации с качеством семантической маски, мы полагаем, что потенциальный прирост может дать обучение с многоклассовой семантической маской, показывающей, например, принадлежность объекта на изображении к определенной анатомической области. Наличие данной информации может также потенциально улучшить возможность клинической интерпретации результатов работы модели.

В то же время добавление в аннотацию информации об анатомической принадлежности того или иного участка изображения само по себе является непростой задачей. Золотым стандартом для анатомического аннотирования изображений по-прежнему является ручная разметка, выполняемая клиническим специалистом (радиологом, неврологом, патологоанатомом), которая требует значительных временных затрат, и качество которой зависит от опыта специалиста.

Наиболее распространенные автоматические подходы для анатомической сегментации МРТ-изображений головного мозга, реализованные в таких программных продуктах как FreeSurfer, BrainVisa, volBrain, FSL, SPM, BrainSuite, Cat12 и т.д., основываются на использовании существующих, вручную созданных, анатомических атласов головного мозга и позволяют определить принадлежность пикселей изображений к одной из предопределенных в атласе области интереса. Однако качество работы таких алгоритмов сильно зависит от качества совмещения МРТ-снимка с шаблоном атласа, что в свою очередь, также зависит от качества самого МРТ-снимка. При этом стоит отметить значительную индивидуальную вариабельность анатомического строения головного мозга, что также усложняет задачу аннотирования.

Кроме этого, неизбежно возникают сложности, связанные с проблемой соотнесения области пораженной ткани с соответствующей её анатомической областью атласа, так как атласы создаются на основе МРТ-изображений здоровых взрослых людей, а области поражения отличаются изменениями интенсивности изображения в одной или нескольких модальностях. В результате этих обстоятельств, ни один

из выше указанных подходов не обладают анатомической точностью сопоставимой с результатами аннотирования «вручную». Существует надежда, что современные методы машинного обучения смогут справиться и с этой задачей [51 – 55]. Однако здесь возникает некий «порочный круг», так как обучение этих методов само по себе требует больших объемов анатомически аннотированных данных, а использование в качестве обучающих данных результатов наивных автоматизированных систем сегментации, часто содержащих неточные нейроанатомические описания, очень вероятно приведут к переносу этих нейроанатомических неточностей и в модели машинного обучения.

Поскольку подобные аннотации создаются вручную, стоит так же упомянуть проблему меж-экспертной вариабельности при разметке медицинских изображения (англ. *inter-expert variations*), в том числе МРТ-снимков, что потенциально может негативно сказаться на качестве результатов классификации. Однако эта проблема может быть частично решена с использованием полуавтоматических подходов, например, таких, когда эксперт сначала выделяет область возможного новообразования, а затем алгоритм сравнивает области здорового и пораженного мозга и уточняет границу опухоли. Такой подход позволяет достичь среднего коэффициента Дайса в 90%, что соизмеримо с экспертной оценкой [48]. Кроме того, существуют подходы, позволяющие уменьшить меж-экспертные различия уже существующей разметки с помощью вариационных автоэнкодеров [49].

Также не исключено, что качество разметки пораженных тканей может не сильно влиять на качество итоговой классификации, так как классификатор применяет к признаковой карте поканальное усреднение (*Average Pooling*), т.е. фактически никак не учитывает локализацию поражения. Возможно, для качества модели гораздо важнее зафиксировать сам факт наличия признаков, описывающих пораженные структуры, и для достижения максимального эффекта достаточно предсказывать только ограничивающую рамку (англ. *bounding box*).

Развивая идею важности наличия признаков, можно предположить, что часть полезных признаков при описанном в работе способе обучения может скрываться на других масштабах кодировщика. Поэтому дополнительный прирост качества классификации может быть обеспечен требованием восстановления семантической маски на всех масштабах декодера по аналогии с подходом «пирамида признаков» (англ. *feature pyramid*) [50].

Ещё одним важным направлением будущих исследований предложенной архитектуры является изучение вопроса устойчивости её

работы к наличию шумов и артефактов на изображениях. В настоящей работе, как и в большинстве существующих, посвященных вопросам сегментации МРТ изображений, и обучение модели, и её тестирование проводилось на относительно качественных, специально отобранных и предобработанных, данных с хорошим пространственным разрешением из базы данных BraTS. Однако, в реальной клинической практике качество изображений сильно зависит от многих факторов таких, как используемый сканер (в том числе качество магнитной катушки, её сила), параметры протокола сканирования (в том числе пространственное разрешение), интенсивность движений пациента в ходе сканирования (вызывают разнообразные артефакты движения) и применяемых методов для пост-коррекции артефактов [56]. Поэтому в дальнейших исследованиях для проверки возможностей практического применения предлагаемой архитектуры в реальных задачах клинической практики, необходимо провести тестирование устойчивости работы разработанного подхода в зависимости от всех выше перечисленных параметров.

**7. Заключение.** В работе рассмотрено сравнение моделей ResNet34 и UBC-ResNet34 – мультизадачной архитектуры Unet-Boosted Classifier (UBC) на базе ResNet34 на малой выборке клинических МРТ-изображений головного мозга. Использование предложенной архитектуры позволяет достичь существенно лучших результатов на тех же данных с использованием расширенных аннотаций.

Как видно из результатов исследования, предложенный подход позволяет эффективно обучить даже такую объемную модель, как ResNet34, содержащую 21.3 млн параметров, всего лишь на 380 снимках, без использования каких либо других приемов. В сочетании с правильно подобранными аугментациями, использованием предобученного компактного кодировщика и другими популярными приемами этот подход может стать серьезным инструментом в руках исследователей на пути к достижению лучших метрик на малых обучающих выборках и представляет собой перспективное направление для дальнейших исследований.

В частности, с практической точки зрения, описанный подход может быть применен во многих задачах классификации медицинских изображений, в том числе к задаче определения этиологии криптогенного ишемического инсульта [57], установление точной причины которого является трудоемкой и социально значимой задачей.

Однако для обучения модели архитектуры UBC требуется дополнительная аннотация изображений в виде семантической маски, отмечающей важные элементы изображения. Поскольку предлагаемая

архитектура нацелена на решение медицинских задач с ограниченными данными, разметка имеющихся изображений может оказаться существенно более доступным способом повышения качества модели, чем сбор большего количества клинических случаев.

## Литература

1. Li Q., Cai W., Wang X., Zhou Y., Feng D.D., Chen M. Medical image classification with convolutional neural network // 13th international conference on control automation robotics vision (ICARCV). IEEE. 2014. pp. 844–848.
2. Cai L., Gao J., Zhao D. A review of the application of deep learning in medical image classification and segmentation // *Annals of translational medicine*. 2020. vol. 8. no. 11.
3. Murtaza G., Shuib L., Abdul Wahab A.W., Mujtaba G., Mujtaba G., Nweke H.F., Al-garadi M.A., Zulfiqar F., Raza G., Azmi N.A. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges // *Artificial Intelligence Review*. 2020. vol. 53. pp. 1655–1720.
4. Yamanakkanavar N., Choi J.Y., Lee B. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer’s disease: a survey // *Sensors*. 2020. vol. 20. no. 11. pp. 3243.
5. Wang X., Peng Y., Lu L., Lu Z., Bagheri M., Summers R.M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases // *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. pp. 2097–2106.
6. Esteva A., Kuprel B., Novoa R.A., Ko J., Swetter S.M., Blau H.M., Thrun S. Dermatologist-level classification of skin cancer with deep neural networks // *nature*. 2017. vol. 542. no. 7639. pp. 115–118.
7. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision // *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. pp. 2818–2826.
8. Ridnik T., Ben-Baruch E., Noy A., Zelnik-Manor L. Imagenet-21k pretraining for the masses // *arXiv preprint arXiv:2104.10972*. 2021.
9. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks // *Advances in neural information processing systems*. 2012. vol. 25.
10. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions // *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. pp. 1–9.
11. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // *arXiv preprint arXiv:1409.1556*. 2014.
12. 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*. 2016. pp. 770–778.
13. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A., Fei-Fei L. Imagenet large scale visual recognition challenge // *International journal of computer vision*. 2015. vol. 115. pp. 211–252.
14. Hammer P. Adaptive control processes: a guided tour (R. Bellman). 1962. 163 p.
15. Chen L. Curse of Dimensionality // *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009. pp. 545–546.
16. Brigato L., Barz B., Iocchi L., Denzler J. Image classification with small datasets: Overview and benchmark // *IEEE Access*. 2022. vol. 10. pp. 49233–49250.

17. Hinton G.E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R.R. Improving neural networks by preventing co-adaptation of feature detectors // arXiv preprint arXiv:1207.0580. 2012.
18. Howard A., Sandler M., Chu G., Chen L.-C., Chen B., Tan M., Wang W., Zhu Y., Pang R., Vasudevan V., Le Q., Adam H. Searching for mobilenetv3 // Proceedings of the IEEE/CVF international conference on computer vision. 2019. pp. 1314–1324.
19. Kim J., Jung W., Kim H., Lee J. CyCNN: A rotation invariant CNN using polar mapping and cylindrical convolution layers // arXiv preprint arXiv:2007.10588. 2020.
20. Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks // International conference on machine learning. PMLR. 2019. pp. 6105–6114.
21. Tan M., Le Q. Efficientnetv2: Smaller models and faster training // International conference on machine learning. PMLR. 2021. pp. 10096–10106.
22. Xu W., Wang G., Sullivan A., Zhang Z. Towards learning affine-invariant representations via data-efficient cnns // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020. pp. 904–913.
23. Sabour S., Frosst N., Hinton G.E. Dynamic routing between capsules // Advances in neural information processing systems. 2017. vol. 30.
24. Stallkamp J., Schlipsing M., Salmen J., Igel C. The German traffic sign recognition benchmark: a multi-class classification competition // 2011 international joint conference on neural networks. IEEE. 2011. pp. 1453–1460.
25. Arora S., Du S.S., Li Z., Salakhutdinov R., Wang R., Yu D. Harnessing the power of infinitely wide deep nets on small-data tasks // arXiv preprint arXiv:1910.01663. 2019.
26. Krizhevsky A., Hinton G. et al. Learning multiple layers of features from tiny images. 2009. 60 p.
27. Barz B., Denzler J. Deep learning on small datasets without pre-training using cosine loss // Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020. pp. 1371–1380.
28. Hai Z., Liu X. Evolving parametrized Loss for Image Classification Learning on Small Datasets // arXiv preprint arXiv:2103.08249. 2021.
29. Lezama J., Qiu Q., Mus e P., Sapiro G. Ole: Orthogonal low-rank embedding-a plug and play geometric loss for deep learning // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 8109–8118.
30. LeCun Y., Cortes C., Burges C. MNIST handwritten digit database. 2010. vol. 2. 18 p.
31. Shorten C., Khoshgoftaar T.M. A survey on image data augmentation for deep learning // Journal of big data. 2019. vol. 6. no. 1. pp. 1–48.
32. Reinders C., Schubert F., Rosenhahn B. Chimeramix: Image classification on small datasets via masked feature mixing // arXiv preprint arXiv:2202.11616. 2022.
33. Zhang X., Wang Z., Liu D., Lin Q., Ling Q. Deep adversarial data augmentation for extremely low data regimes // IEEE Transactions on Circuits and Systems for Video Technology. 2020. vol. 31. no. 1. pp. 15–28.
34. Ishii M., Sato A. Training deep neural networks with adversarially augmented features for small-scale training datasets // 2019 International Joint Conference on Neural Networks (IJCNN). IEEE. 2019. pp. 1–8.
35. Agarwal N., Sondhi A., Chopra K., Singh G. Transfer learning: Survey and classification // Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS 2020. 2021. pp. 145–155.
36. Zhao B., Wen X. Distilling visual priors from self-supervised learning // Computer Vision – ECCV 2020 Workshops. Springer. 2020. pp. 422–429.

37. Chen X., Fan H., Girshick R., He K. Improved baselines with momentum contrastive learning // arXiv preprint arXiv:2003.04297. 2020.
38. Miranda G., Rubio-Manzano C. Image Classification Using Deep and Classical Machine Learning on Small Datasets: A Complete Comparative. 2022. DOI: 10.20944/preprints202201.0367.v1.
39. Christodoulou E., Ma J., Collins G.S., Steyerberg E.W., Verbakel J.Y., Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models // Journal of clinical epidemiology. 2019. vol. 110. pp. 12–22.
40. Mojab N., Yu P.S., Hallak J.A., Yi D. Cvs: Classification via segmentation for small datasets // arXiv preprint arXiv:2111.00042. 2021.
41. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th international conference. Springer International Publishing, 2015. pp. 234–241.
42. Baid U., Ghodasara S., Mohan S., Bilello M., Calabrese E., Colak E., Farahani K., Kalpathy-Cramer J., Kitamura F.C., Pati S. et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification // arXiv preprint arXiv:2107.02314. 2021.
43. Menze B.H., Jakab A., Bauer S., Kalpathy-Cramer J., Farahani K., Kirby J., Burren Y., Porz N., Slotboom J., Wiest R., et al. The multimodal brain tumor image segmentation benchmark (BRATS) // IEEE transactions on medical imaging. 2014. vol. 34. no. 10. pp. 1993–2024.
44. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby J.S., Freymann J.B., Farahani K., Davatzikos C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features // Scientific data. 2017. vol. 4. no. 1. pp. 1–13.
45. Rohlfing T., Zahr N.M., Sullivan E.V., Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure // Human brain mapping. 2010. vol. 31. no. 5. pp. 798–819.
46. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition // Proceedings of the IEEE. 1998. vol. 86. no. 11. pp. 2278–2324.
47. Kingma D.P., Ba J. Adam A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014.
48. Akkus Z., Sedlar J., Coufalova L., Korfiatis P., Kline T.L., Warner J.D., Agrawal J., Erickson B.J. Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging // Cancer Imaging. 2015. vol. 15. pp. 1–10.
49. Liu X., Xing F., Marin T., El Fakhri G., Woo J. Variational inference for quantifying inter-observer variability in segmentation of anatomical structures // Medical Imaging 2022: Image Processing. 2022. vol. 12032. pp. 438–443.
50. Lin T.-Y., Dollar P., Girshick R., He K., Hariharan B., Belongie S. Feature pyramid networks for object detection // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. pp. 2117–2125.
51. Huo Y., Xu Z., Xiong Y., Aboud K., Parvathaneni P., Bao S., Bermudez C., Resnick S., Cutting L., Landman B. 3D whole brain segmentation using spatially localized atlas network tiles // NeuroImage. 2019. vol. 194. pp. 105–119.
52. Paschali M., Gasperini S., Roy A.G., Fang M.Y.S., Navab N. 3DQ: Compact quantized neural networks for volumetric whole brain segmentation // Medical Image Computing and Computer Assisted Intervention: 22nd International Conference. Springer International Publishing, 2019. pp. 438–446.

53. Roy A.G., Conjeti S., Navab N., Wachinger C. Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control // *NeuroImage*. 2019. vol. 195. pp. 11–22.
54. Henschel L, Conjeti S., Estrada S., Diers K., Fischl B., Reuter, M. Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline // *NeuroImage*. 2020. vol. 219. pp. 117012.
55. Coupe P., Mansencal, B., Clement, M., Giraud, R., de Senneville B.D., Ta V.T., Lepetit V., Manjon J.V. AssemblyNet: A large ensemble of CNNs for 3D whole brain MRI segmentation // *NeuroImage*. 2020. vol. 219(5). pp. 117026.
56. Dubois J., Alison M., Counsell S.J., Hertz-Pannier L., Huppi P.S., Benders M.J. MRI of the neonatal brain: a review of methodological challenges and neuroscientific advances // *Journal of Magnetic Resonance Imaging*. 2021. vol. 53. no. 5. pp. 1318–1343.
57. Ntaios G. Embolic stroke of undetermined source: JACC review topic of the week // *Journal of the American College of Cardiology*. 2020. vol. 75. no. 3. pp. 333–340.

**Собынин Кирилл Валентинович** — младший научный сотрудник, центр когнитивных нейронаук, Национальный исследовательский университет «Высшая школа экономики». Область научных интересов: машинное обучение, применение методов математики и информатики в неврологии и нейронауках. Число научных публикаций — 9. k.sobyenin@gmail.com; бульвар Гагарина, 37а, 614060, Пермь, Россия; р.т.: +7(342)200-9555.

**Куликова Софья Петровна** — Ph.D., доцент, старший научный сотрудник, центр когнитивных нейронаук, Национальный исследовательский университет «Высшая школа экономики». Область научных интересов: моделирование когнитивных процессов, анализ МРТ снимков, развитие головного мозга. Число научных публикаций — 19. spkulikova@hse.ru; бульвар Гагарина, 37а, 614060, Пермь, Россия; р.т.: +7(342)200-9555.

**Поддержка исследований.** Публикация подготовлена в ходе проведения исследования № 23-00-026 «Разработка автоматических подходов для определения этиологии криптогенного инсульта с целью профилактики вторичных острых нарушений мозгового кровообращения» в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)».

K. SOBYANIN, S. KULIKOVA  
**UNET-BOOSTED CLASSIFIER – MULTI-TASK ARCHITECTURE  
FOR SMALL DATASETS APPLIED TO BRAIN MRI  
CLASSIFICATION**

*Sobyanin K., Kulokova S. Unet-boosted Classifier – Multi-Task Architecture for Small Datasets Applied to Brain MRI Classification.*

**Abstract.** The problem of training deep neural networks on small samples is especially relevant for medical issues. The paper examines the impact of pixel-wise marking of significant objects in the image, over the true class label, on the quality of the classification. To achieve better classification results on small samples, we propose a multitasking architecture – Unet-boosted classifier (UBC), that is trained simultaneously to solve classification and semantic segmentation problems. As the exploratory dataset, MRI images of patients with benign glioma and glioblastoma taken from the BRaTS 2019 data set are used. One horizontal slice of the MRI image containing a glioma is considered as the input (a total of 380 frames in the training set), and the probability of glioblastoma – as the output. Resnet34 was chosen as the baseline, trained without augmentations with a loss function based on cross-entropy. As an alternative solution UBC-resnet34 is used – the same resnet34, boosted by a decoder built on the U-Net principle and predicting the pixels with glioma. The smoothed Sorensen-Dice coefficient (DiceLoss) is used as a decoder loss function. Results on the test sample: accuracy for the baseline reached 0.71, for the proposed model – 0.81, and the Dice score – 0.77. Thus, a deep model can be well trained even on a small data set, using the proposed architecture, provided that marking of the affected tissues in the form of a semantic mask is provided.

**Keywords:** image classification, deep learning, small dataset, semantic segmentation, multi-task architecture, cerebral pathology, tumor diagnosis.

## References

1. Li Q., Cai W., Wang X., Zhou Y., Feng D.D., Chen M. Medical image classification with convolutional neural network. 13th international conference on control automation robotics vision (ICARCV). IEEE. 2014. pp. 844–848.
2. Cai L., Gao J., Zhao D. A review of the application of deep learning in medical image classification and segmentation. *Annals of translational medicine*. 2020. vol. 8. no. 11.
3. Murtaza G., Shuib L., Abdul Wahab A.W., Mujtaba G., Mujtaba G., Nweke H.F., Al-garadi M.A., Zulfiqar F., Raza G., Azmi N.A. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artificial Intelligence Review*. 2020. vol. 53. pp. 1655–1720.
4. Yamanakkanavar N., Choi J.Y., Lee B. MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: a survey. *Sensors*. 2020. vol. 20. no. 11. pp. 3243.
5. Wang X., Peng Y., Lu L., Lu Z., Bagheri M., Summers R.M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. pp. 2097–2106.
6. Esteva A., Kuprel B., Novoa R.A., Ko J., Swetter S.M., Blau H.M., Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *nature*. 2017. vol. 542. no. 7639. pp. 115–118.

7. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. pp. 2818–2826.
8. Ridnik T., Ben-Baruch E., Noy A., Zelnik-Manor L. Imagenet-21k pretraining for the masses. arXiv preprint arXiv:2104.10972. 2021.
9. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012. vol. 25.
10. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. pp. 1–9.
11. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
12. 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. 2016. pp. 770–778.
13. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A., Fei-Fei L. Imagenet large scale visual recognition challenge. International journal of computer vision. 2015. vol. 115. pp. 211–252.
14. Hammer P. Adaptive control processes: a guided tour (R. Bellman). 1962. 163 p.
15. Chen L. Curse of Dimensionality. Encyclopedia of Database Systems. Boston, MA: Springer US, 2009. pp. 545–546.
16. Brigato L., Barz B., Iocchi L., Denzler J. Image classification with small datasets: Overview and benchmark. IEEE Access. 2022. vol. 10. pp. 49233–49250.
17. Hinton G.E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R.R. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580. 2012.
18. Howard A., Sandler M., Chu G., Chen L.-C., Chen B., Tan M., Wang W., Zhu Y., Pang R., Vasudevan V., Le Q., Adam H. Searching for mobilenetv3. Proceedings of the IEEE/CVF international conference on computer vision. 2019. pp. 1314–1324.
19. Kim J., Jung W., Kim H., Lee J. CyCNN: A rotation invariant CNN using polar mapping and cylindrical convolution layers. arXiv preprint arXiv:2007.10588. 2020.
20. Tan M., Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning. PMLR. 2019. pp. 6105–6114.
21. Tan M., Le Q. Efficientnetv2: Smaller models and faster training. International conference on machine learning. PMLR. 2021. pp. 10096–10106.
22. Xu W., Wang G., Sullivan A., Zhang Z. Towards learning affine-invariant representations via data-efficient cnns. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020. pp. 904–913.
23. Sabour S., Frosst N., Hinton G.E. Dynamic routing between capsules. Advances in neural information processing systems. 2017. vol. 30.
24. Stallkamp J., Schlipsing M., Salmen J., Igel C. The German traffic sign recognition benchmark: a multi-class classification competition. 2011 international joint conference on neural networks. IEEE. 2011. pp. 1453–1460.
25. Arora S., Du S.S., Li Z., Salakhutdinov R., Wang R., Yu D. Harnessing the power of infinitely wide deep nets on small-data tasks. arXiv preprint arXiv:1910.01663. 2019.
26. Krizhevsky A., Hinton G. et al. Learning multiple layers of features from tiny images. 2009. 60 p.
27. Barz B., Denzler J. Deep learning on small datasets without pre-training using cosine loss. Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2020. pp. 1371–1380.

28. Hai Z., Liu X. Evolving parametrized Loss for Image Classification Learning on Small Datasets. arXiv preprint arXiv:2103.08249. 2021.
29. Lezama J., Qiu Q., Mus e P., Sapiro G. Ole: Orthogonal low-rank embedding-a plug and play geometric loss for deep learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 8109–8118.
30. LeCun Y., Cortes C., Burges C. MNIST handwritten digit database. 2010. vol. 2. 18 p.
31. Shorten C., Khoshgoftaar T.M. A survey on image data augmentation for deep learning. Journal of big data. 2019. vol. 6. no. 1. pp. 1–48.
32. Reinders C., Schubert F., Rosenhahn B. Chimeramix: Image classification on small datasets via masked feature mixing. arXiv preprint arXiv:2202.11616. 2022.
33. Zhang X., Wang Z., Liu D., Lin Q., Ling Q. Deep adversarial data augmentation for extremely low data regimes. IEEE Transactions on Circuits and Systems for Video Technology. 2020. vol. 31. no. 1. pp. 15–28.
34. Ishii M., Sato A. Training deep neural networks with adversarially augmented features for small-scale training datasets. 2019 International Joint Conference on Neural Networks (IJCNN). IEEE. 2019. pp. 1–8.
35. Agarwal N., Sondhi A., Chopra K., Singh G. Transfer learning: Survey and classification. Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS 2020. 2021. pp. 145–155.
36. Zhao B., Wen X. Distilling visual priors from self-supervised learning. Computer Vision – ECCV 2020 Workshops. Springer. 2020. pp. 422–429.
37. Chen X., Fan H., Girshick R., He K. Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297. 2020.
38. Miranda G., Rubio-Manzano C. Image Classification Using Deep and Classical Machine Learning on Small Datasets: A Complete Comparative. 2022. DOI: 10.20944/preprints202201.0367.v1.
39. Christodoulou E., Ma J., Collins G.S., Steyerberg E.W., Verbakel J.Y., Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. Journal of clinical epidemiology. 2019. vol. 110. pp. 12–22.
40. Mojab N., Yu P.S., Hallak J.A., Yi D. Cvs: Classification via segmentation for small datasets. arXiv preprint arXiv:2111.00042. 2021.
41. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th international conference. Springer International Publishing, 2015. pp. 234–241.
42. Baid U., Ghodasara S., Mohan S., Bilello M., Calabrese E., Colak E., Farahani K., Kalpathy-Cramer J., Kitamura F.C., Pati S. et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint arXiv:2107.02314. 2021.
43. Menze B.H., Jakab A., Bauer S., Kalpathy-Cramer J., Farahani K., Kirby J., Burren Y., Porz N., Slotboom J., Wiest R., et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging. 2014. vol. 34. no. 10. pp. 1993–2024.
44. Bakas S., Akbari H., Sotiras A., Bilello M., Rozycki M., Kirby J.S., Freymann J.B., Farahani K., Davatzikos C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific data. 2017. vol. 4. no. 1. pp. 1–13. Rohlfing T., Zahr N.M., Sullivan E.V., Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. Human brain mapping. 2010. vol. 31. no. 5. pp. 798–819.

45. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 1998. vol. 86. no. 11. pp. 2278–2324.
46. Kingma D.P., Ba J. Adam A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 2014.
47. Akkus Z., Sedlar J., Coufalova L., Korfiatis P., Kline T.L., Warner J.D., Agrawal J., Erickson B.J. Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging. *Cancer Imaging*. 2015. vol. 15. pp. 1–10.
48. Liu X., Xing F., Marin T., El Fakhri G., Woo J. Variational inference for quantifying inter-observer variability in segmentation of anatomical structures. *Medical Imaging 2022: Image Processing*. 2022. vol. 12032. pp. 438–443.
49. Lin T.-Y., Dollar P., Girshick R., He K., Hariharan B., Belongie S. Feature pyramid networks for object detection. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. pp. 2117–2125.
50. Huo Y., Xu Z., Xiong Y., Aboud K., Parvathaneni P., Bao S., Bermudez C., Resnick S., Cutting L., Landman B. 3D whole brain segmentation using spatially localized atlas network tiles. *NeuroImage*. 2019. vol. 194. pp. 105–119.
51. Paschali M., Gasperini S., Roy A.G., Fang M.Y.S., Navab N. 3DQ: Compact quantized neural networks for volumetric whole brain segmentation. *Medical Image Computing and Computer Assisted Intervention: 22nd International Conference*. Springer International Publishing, 2019. pp. 438–446.
52. Roy A.G., Conjeti S., Navab N., Wachinger C. Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control. *NeuroImage*. 2019. vol. 195. pp. 11–22.
53. Henschel L., Conjeti S., Estrada S., Diers K., Fischl B., Reuter, M. Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. *NeuroImage*. 2020. vol. 219. pp. 117012.
54. Coupe P., Mansencal, B., Clement, M., Giraud, R., de Senneville B.D., Ta V.T., Lepetit V., Manjon J.V. AssemblyNet: A large ensemble of CNNs for 3D whole brain MRI segmentation. *NeuroImage*. 2020. vol. 219(5). pp. 117026.
55. Dubois J., Alison M., Counsell S.J., Hertz-Pannier L., Huppi P.S., Benders M.J. MRI of the neonatal brain: a review of methodological challenges and neuroscientific advances. *Journal of Magnetic Resonance Imaging*. 2021. vol. 53. no. 5. pp. 1318–1343.
56. Ntaios G. Embolic stroke of undetermined source: JACC review topic of the week. *Journal of the American College of Cardiology*. 2020. vol. 75. no. 3. pp. 333–340.

**Sobyanin Kirill** — Junior researcher, Center for cognitive neuroscience, National Research University Higher School of Economics. Research interests: machine learning, application of mathematics and computer science in neurology and neuroscience. The number of publications — 9. k.sobyanin@gmail.com; 37a, Gagarin Blvd, 614060, Perm, Russia; office phone: +7(342)200-9555.

**Kulikova Sofya** — Ph.D., Associate Professor, Senior researcher, Center for cognitive neuroscience, National Research University Higher School of Economics. Research interests: modeling cognitive processes, MRI analysis, brain development. The number of publications — 19. spkulikova@hse.ru; 37a, Gagarin Blvd., 614060, Perm, Russia; office phone: +7(342)200-9555.

**Acknowledgements.** The publication was prepared within the framework of the Academic Fund Program at HSE University (grant № 23-00-026 "Development of automatic approaches to determine the etiology of cryptogenic stroke for the purpose of preventing secondary acute cerebrovascular accidents").

М.Н. ФАВОРСКАЯ, А.И. ПАХИРКА  
**ВОССТАНОВЛЕНИЕ АЭРОФОТОСНИМКОВ  
СВЕРХВЫСОКОГО РАЗРЕШЕНИЯ С УЧЕТОМ  
СЕМАНТИЧЕСКИХ ОСОБЕННОСТЕЙ**

*Фаворская М.Н., Пахирка А.И. Восстановление аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей.*

**Аннотация.** В настоящее время происходит активное развитие технологий обработки изображений дистанционного зондирования, включающих как спутниковые снимки, так и аэрофотоснимки, полученные от видеокамер беспилотных летательных аппаратов. Зачастую такие снимки имеют артефакты, связанные с низким разрешением, размытостью фрагментов изображения, наличием шумов и т.д. Одним из способов преодоления таких ограничений является применение современных технологий для восстановления снимков сверхвысокого разрешения на основе методов глубокого обучения. Особенностью аэрофотоснимков является представление текстуры и структурных элементов более высокого разрешения, чем на спутниковых снимках, что объективно способствует лучшим результатам восстановления. В статье приводится классификация методов сверхвысокого разрешения с учетом основных архитектур глубоких нейронных сетей, а именно сверточных нейронных сетей, визуальных трансформеров и генеративно-состязательных сетей. В статье предлагается метод восстановления аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей SemESRGAN за счет использования на этапе обучения дополнительной глубокой сети для семантической сегментации. При этом минимизируется общая функция потерь, включающая состязательные потери, потери на уровне пикселей и потери восприятия (сходства признаков). Для экспериментов использовались шесть наборов аннотированных аэрофотоснимков и спутниковых снимков CLCD, DOTA, LEVIR-CD, UAVid, AAD и AID. Было выполнено сравнение результатов восстановления изображений предложенным методом SemESRGAN с базовыми архитектурами сверточных нейронных сетей, визуальных трансформеров и генеративно-состязательных сетей. Получены сравнительные результаты восстановления изображений с применением объективных метрик PSNR и SSIM, что позволило оценить качество восстановления с использованием различных моделей глубоких сетей.

**Ключевые слова:** аэрофотоснимки, сверхвысокое разрешение, семантическая сегментация, сверточные нейронные сети, визуальные трансформеры, генеративно-состязательные сети.

**1. Введение.** Задача восстановления изображений сверхвысокого разрешения относится к методам предварительной обработки изображений. До появления методов глубокого обучения задача решалась традиционными методами интерполяции, включая билинейную интерполяцию, бикубическую интерполяцию, интерполяцию Ланцоша (Lanczos) (для снимков дистанционного зондирования Земли) и т.д. Развитие методов глубокого обучения существенно повлияло на качество восстановления изображений сверхвысокого разрешения.

Существуют два подхода к восстановлению снимков сверхвысокого разрешения: подход на основе одного исходного изображения (Single Image Super-Resolution, SISR) и подход с использованием нескольких исходных изображений или кадров видеопоследовательности (Multi Image Super-Resolution, MISR). Преимущественно используется первый подход (SISR), не требующий одновременного получения нескольких исходных снимков низкого разрешения (Low Resolution, LR) при разных ракурсах съемки. Следует отметить, что обе постановки обратных задач являются некорректными с математической точки зрения, поскольку существует множество способов восстановления снимка сверхвысокого разрешения, близкого по объективным метрикам к реальному снимку высокого разрешения (High Resolution, HR) [1].

Известны четыре категории SISR методов: на основе интерполяции, реконструкции, обучения и преобразований [2]. В настоящее время методы на основе интерполяции используются в качестве предварительной обработки при восстановлении SR снимков (при необходимости). Основное развитие получили методы реконструкции, обучения и преобразований. Методы реконструкции основаны на применении априорной информации. Они широко используются в медицинских приложениях. Методы на основе обучения анализируют взаимосвязи между снимками высокого и низкого разрешения (так называемые парные снимки) из предварительно подготовленного набора данных. При этом методы глубокого обучения считаются гибридными методами, основанными на реконструкции и обучении. Алгоритмы на основе преобразований, использующих сети-трансформеры, имеют более сложную структуру, включая модули самокалибровки, внимания, иерархического разделения фрагментов и т.д.

Для решения задачи восстановления SR снимков подходят два типа архитектур глубоких нейронных сетей – сверточные нейронные сети (СНС) и генеративно-состязательные сети (ГСС). Они основаны на разных принципах обучения и, соответственно, восстановления LR снимков. Исторически первыми появились СНС, и их использование для решения задачи SISR превалировало до 2021 года. В свою очередь первая архитектура ГСС были разработана в 2014 году [3], и в настоящее время ГСС демонстрируют лучшие результаты восстановления за счет более сложных, но и более эффективных стратегий обучений [4]. Тем не менее, применяются оба подхода. Сети-трансформеры были разработаны в 2017 году как относительно простой способ улучшения результатов в языковом переводе.

Визуальные трансформеры для SISR задач позволили переосмыслить архитектуру классических СНС. Одним из последних достижений является разработка SwinV2 трансформера для восстановления SR изображений [5].

SISR технологии применяются во многих областях, включая удаленное зондирование Земли [6], медицинскую диагностику [7], биометрию [8], видеонаблюдение [9], метеорологию [10] и т.д. Каждая область применения предъявляет свои требования к результатам визуальной восстановления, и поскольку выполнить их одновременно не представляется возможным, формируются модификации SISR методов, учитывающие основные особенности той или иной сферы применения.

В данном исследовании предлагается метод восстановления аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей на основе усложненного обучения нейронных сетей различных видов. Помимо базовой архитектуры дополнительно применяется глубокая сеть для семантической сегментации. При этом минимизируется общая функция потерь, включающая составительные потери, потери на уровне пикселей и потери восприятия (сходства признаков). Были исследованы различные базовые архитектуры, в частности, модели SRCNN, ESRGAN, а также модели СНС для семантической сегментации. Для экспериментов использовались шесть наборов аннотированных аэрофотоснимков и спутниковых снимков CLCD, DOTA, LEVIR-CD, UAVid, AAD и AID, что позволило провести анализ полученных результатов на основе объективных метрик оценки восстановленных изображений.

Снимки дистанционного зондирования характеризуются следующими особенностями: они получены с большого расстояния, изображения содержат множество малоразмерных объектов, изображения сцены зависят от времени года, различных атмосферных условий и разной геометрии обзора датчиков. Спутниковые снимки, как правило, являются мультиспектральными или гиперспектральными в отличие от аэрофотоснимков, которые преимущественно формируются в оптическом диапазоне. Поэтому можно говорить о разных методах восстановления спутниковых снимков и аэрофотоснимков. Далее приведем краткий обзор существующих в настоящее время SISR методов для восстановления аэрофотоснимков сверхвысокого разрешения.

**2. Обзор SISR методов для восстановления аэрофотоснимков сверхвысокого разрешения.** В настоящее время в реализации SISR методов для задач дистанционного зондирования преобладают методы

на основе глубокого обучения. Однако следует отметить, что существуют и менее распространенные методы обучения, такие как встраивание соседей (neighbor embedding) или разреженное кодирование (sparse coding). Методы глубокого обучения для восстановления снимков сверхвысокого разрешения можно разделить на девять категорий по типам связей:

- линейные связи;
- остаточные связи;
- рекурсивные связи;
- связи, основанные на внимании;
- многопоточные соединения;
- соединения высокой плотности;
- связи, обрабатывающие множественные искажения, на основе так называемой технологии обучения с нулевым выстрелом (zero-shot learning);
- связи, используемые в ГСС;
- связи прогрессивной реконструкции.

Каждая из категорий представлена несколькими моделями глубоких сетей, среди которых имеются как широко известные модели, так и редко применяемые.

Помимо связей важны виды глубокого обучения (с учителем или без учителя) и типы архитектуры. В настоящее время лучшие результаты демонстрируют архитектуры, основанные на использовании наборов данных, состоящих из парных LR–HR снимков, а методы на основе накапливаемой при обучении статистики, преобразовании фрагментов (визуальные трансформеры) или словарей фрагментов требуют дальнейшего развития.

Рассмотрим более подробно тенденции развития архитектур СНС и ГСС как наиболее часто используемых для восстановления аэрофотоснимков сверхвысокого разрешения. В работе [11] используется мультимасштабное представление аэрофотоснимков на основе вейвлет-анализа. Каждое кратномасштабное LR представление (прямое вейвлет-преобразование) обрабатывается своей предварительно обученной СНС, а затем для получения SR представления выполняется вейвлет-синтез (обратное вейвлет-преобразование). Отметим, что восстановление SR снимков с использованием частотных преобразований (обычно вейвлет-преобразований) широко применяется для восстановления контуров малоразмерных объектов. Для обнаружения малоразмерных объектов была спроектирована нейронная сеть с глубокой памятью, имеющая архитектуру, подобную U-Net модели (Deep Memory Connected

Network) [12]. Улучшенная глубокая рекурсивная остаточная сеть (Improved Deep Recursive Residual Network) была представлена в работе [13]. Основная идея состояла в уменьшении сложности обучения базовой модели ResNet на основе глобального остаточного обучения и локального остаточного обучения с использованием рекурсивного блока. Точность восстановления существенно повышалась за счет подключения нескольких вторичных фильтров с адаптацией для параллельной обработки. Восстановление аэрофотоснимков проводилось в масштабах 2×, 3× и 4×. Однако значения пикового отношения сигнал–шум не превышали 30-35 дБ. Оригинальная архитектура в виде двухуровневой взаимно дополняемой аффинной СНС предложена в работе [14]. С точки зрения авторов взаимная дополняемость между уровнями позволяет получать более информативные признаки, используя стратегию адаптивного множественного внимания к разномасштабным визуальным объектам. Таким образом, реализуется эффективное объединение низкоуровневых и высокоуровневых признаков с помощью операции взаимной аффинной свертки.

Одним из направлений исследований SISR задачи является создание моделей с меньшим количеством параметров, не уступающих, а иногда и превосходящих полные модели СНС. Так, например, в работе [15] предложен новый сверточный слой, названный слоем контекстного преобразования (contextual transformation layer), который с одной стороны упрощает традиционный сверточный слой 3×3, а с другой стороны извлекает эффективные контекстные функции на разных иерархических уровнях обработки. Однако более существенным улучшением традиционных моделей СНС для задачи восстановления снимков дистанционного зондирования сверхвысокого разрешения является применение визуальных трансформеров.

Модели СНС имеют два существенных недостатка. Во-первых, они используют одно и то же ядро свертки для обработки разномасштабных областей изображения. Во-вторых, модели СНС имеют ограниченное поле восприятия, восстановление которого зависит только от локальной информации. В то же время сходные по контексту фрагменты могут предоставить дополнительную информацию для восстановления текущего фрагмента. Архитектура визуальных трансформеров, возникшая на основе моделей СНС, обеспечивает механизм самовнимания (self-attention mechanism) для сбора глобальной информации и использует свойство самоподобия изображения. В настоящее время семейство визуальных

трансформеров расширяется, причем, не только для решения рассматриваемой задачи.

Большинство моделей СНС использует слои повышающей дискретизации для получения SR представления, которые игнорируют извлечение признаков в многомерном пространстве, что влияет на качество выходных данных. Для устранения этой проблемы предложена улучшенная сеть на основе трансформеров (Transformer-based Enhancement Network, TransENet) [16]. Ядром модели TransENet является иерархическая структура, объединяющая традиционные SR структуры с мультимасштабными функциями высокой или низкой размерности. Гибридная иерархическая сеть-трансформер (Hybrid-Scale Hierarchical Transformer Network, HSTNet) [17] решает аналогичную задачу за счет использования трансформера кросс-масштабного улучшения, позволяющего улавливать долгосрочные зависимости в визуальных данных.

Контекстно-зависимая облегченная сеть (Context-Aware Lightweight Super-Resolution Network, CALSRN) для восстановления изображений дистанционного зондирования сверхвысокого разрешения была предложена в работе [18]. Сеть, имеющая U-Net архитектуру, состоит из последовательных контекстно-зависимых блоков-трансформеров, которые извлекают как локальный контекст, так и глобальный контекст. Ветвь генерации динамических весов позволяет динамически регулировать процесс агрегации локальных и глобальных функций.

Мотивацией применения ГСС для задачи восстановления изображений дистанционного зондирования сверхвысокого разрешения является множество наземных сцен с разномасштабными объектами, характеризующимися различными спектральными характеристиками. Модели СНС, как правило, игнорируют такие особенности. Лучшим вариантом является применение моделей ГСС. Модель ГСС, основанная на механизмах локального и глобального внимания (Attention-based Generative Adversarial Network), была представлена в работе [19]. Механизм локального внимания позволяет сосредоточиться на структурных компонентах земной поверхности, а механизм глобального внимания используется для выявления долгосрочных пространственных взаимозависимостей. Используемый в любых моделях ГСС процесс состязательного обучения позволяет улучшить дискриминационную способность и применять градиентный штраф к комплексной функции потерь, включающей потери пикселей, потери восприятия и состязательные потери.

Восстановление аэрофотоснимков сверхвысокого разрешения зачастую приносит искажения в текстурные области. Для устранения подобных артефактов была разработана плотная генеративно-сопоставительная сеть (Novel Dense Generative Adversarial Network for real aerial imagery Super-Resolution reconstruction, NDSRGAN), которая объединяла многоуровневую плотную сеть и матричный дискриминатор, учитывающий средние значения пикселей [20]. Для обучения были созданы наборы данных с парными реальными изображениями дистанционного зондирования высокого и низкого разрешения. Для ускорения сходимости модели вместо функции потерь  $L1$  была применена функция потерь  $smoothL1$  для лучшего визуального восприятия текстур.

Было замечено, что обучение на «идеальных» наборах данных приводит к резкому падению производительности модели на реальных снимках дистанционного зондирования, поскольку качество реальных LR снимков зависит от множества факторов, таких как освещение, состояние атмосферы, используемые датчики и т.д. В работе [21] подобные искажения моделировались с помощью ядер размытия и шумов. В качестве генератора была разработана сеть остаточного сбалансированного внимания (Residual Balanced Attention Network) для оценки результатов сверхвысокого разрешения на основе входных LR снимков. В качестве сопоставительного обучения был применен дискриминатор на основе U-Net модели для генерации более реалистичных текстур.

Методология трансферного обучения для восстановления SR аэрофотоснимков была применена в работе [22]. В качестве базовой архитектуры использовалась GCG сверхвысокого разрешения (Super Resolution Generative Adversarial Network). При этом набор данных DIV2K применялся для предварительного обучения генеративной модели, а затем метод трансферного обучения использовался для обучения отдельных моделей на наборах проверочных данных xView и DOTA. Судя по значениям индекса восприятия и среднеквадратической ошибки, метод трансферного обучения показал хорошие результаты восстановления аэрофотоснимков. В дальнейшем такая технология была применена теми же авторами для улучшения обнаружения объектов на аэрофотоснимках [23].

Краткий обзор показал, что восстановление аэрофотоснимков сверхвысокого разрешения на основе методов глубокого обучения является актуальной задачей, требующей дальнейшего развития и поиска неординарных решений в связи со сложностью и многообразием реальных визуальных объектов.

**3. Постановка задачи.** Пусть  $I_{HR} \in R^{H \times W \times C}$  и  $I_{SR} \in R^{H \times W \times C}$  – исходный снимок высокого разрешения  $I_{HR}$  и восстановленный снимок сверхвысокого разрешения  $I_{SR}$ , которые представлены в многомерном пространстве  $R$ , имеющем размерности высоты  $H$ , ширины  $W$  и цветовых каналов  $C$ . Для моделирования снимка низкого разрешения  $I_{LR}$  строится модель искажений  $\Psi$  на основе снимка высокого разрешения  $I_{HR}$ , которая имеет вид:

$$I_{LR} = \Psi(I_{HR}, \theta_\eta),$$

где  $\theta_\eta$  – параметры искажений, например, коэффициент масштабирования, шум, размытие и т.д.

В простейшем случае доступны наборы данных с парными снимками  $I_{HR}$  и  $I_{LR}$ , а целью является получение параметров искажений  $\theta_\eta$ . Таким образом, задача восстановления снимка сверхвысокого разрешения заключается в устранении искажений и восстановления снимка, похожего на исходный снимок высокого разрешения  $I_{HR}$ .

$$I_{SR} = \Psi^{-1}(I_{LR}, \theta_\xi),$$

где  $\theta_\xi$  – параметры модели сверхвысокого разрешения  $\Psi^{-1}$ .

В случае неизвестных параметров искажений задача восстановления SR изображений усложняется. При этом на процесс восстановления, как правило, влияют несколько факторов: шум (белый гауссов шум), размытие (расфокусировка, движение), сжатие и другие артефакты. Процесс обучения заключается в оптимизации параметров  $\theta_\xi$  для модели  $\Psi^{-1}$ :

$$\hat{\theta}_\xi = \arg \min(Loss(I_{SR}, I_{HR})),$$

где  $Loss$  – функция потерь.

В силу сложности SR задача для изображений дистанционного зондирования обычно формулируется как задача обучения с учителем. При наличии парных HR–LR снимков обучение упрощается, однако HR снимки не всегда доступны. Если исходные снимки можно считать HR изображениями, то LR изображения в простейшем случае генерируются следующим образом [24]:

$$I_{LR} = \downarrow_s(I_{HR}),$$

где  $\downarrow_s$  – понижающая выборка с коэффициентом масштабирования  $s$ .

Однако на практике предпочитают использовать усложненную модель с ядрами размытия  $k$  и наложенным шумом  $n$ :

$$I_{LR} = \downarrow_s (I_{HR} \otimes k) + n,$$

где символ  $\otimes$  означает оператор свертки.

Если исходные снимки имеют низкое разрешение, то при применении СНС из них синтезируются LR изображения еще более низкого разрешения (методом понижающей дискретизации), а при применении ГСС они сравниваются с восстановленными SR снимками, также подвергнутыми понижающей дискретизации. Данный случай является самым сложным, демонстрирующим наихудшие результаты.

Таким образом, целью восстановления SR снимка методами обучения является минимизация функции потерь  $Loss(\cdot)$ :

$$Loss(I_{SR}, \theta_\eta, k) = \|(I_{HR} \otimes k) - I_{LR}\| + \alpha \Psi(I_{HR}, \theta_\eta),$$

где  $\alpha$  – параметр регуляризации. При этом первое слагаемое определяет точность модели, а второе слагаемое отвечает за регуляризацию.

**4. Восстановление аэрофотоснимков сверхвысокого разрешения с применением СНС.** Базовой моделью восстановления изображений сверхвысокого разрешения на основе СНС считается модель SRCNN [25], разработанная в 2014 году. Она использовала бикубическую интерполяцию для повышения дискретизации входного LR изображения до целевого SR изображения и существенно улучшила качество восстановления по сравнению с традиционными методами. Далее для улучшения базовой модели исследовались различные концепции, например, концепции субпиксельных сверточных слоев, остаточных блоков, плотных блоков, механизмов рекурсии и внимания, пирамидальной обработки, каскадной обработки и т.д. Более того, известны модели, которые совмещают сразу несколько механизмов улучшения. Тем не менее, основной концепцией является последовательное применение трех операций:

– извлечение и представление фрагментов (patches) исходного LR изображения, когда каждый фрагмент преобразуется в многомерный вектор;

– нелинейное отображение, когда многомерный вектор нелинейно отображается на другой многомерный вектор большей размерности;

– реконструкция, когда выполняется объединение фрагментов сверхвысокого разрешения, преобразованных из многомерных векторов большей размерности, для генерации выходного изображения сверхвысокого разрешения, похожего на реальное HR изображение.

В данном исследовании был проведен подробный анализ моделей СНС для восстановления аэрофотоснимков. Для демонстрации преимущества учета семантических особенностей была выбрана одна из последних моделей СНС для восстановления снимков дистанционного зондирования, а именно, гибридная U-образная сеть, основанная на внимании HAUNet (Hybrid Attention-based U-shaped Network) [26]. Модель HAUNet, имеющая достаточно сложную структуру, извлекает и адаптивно агрегирует мультимасштабную информацию с помощью двух модулей извлечения одномасштабных признаков на основе сверточного внимания (Single-scale feature Extraction Modules, SEM): модуль извлечения глобального пространственного контекста (Spatial-aware Context feature Extraction Module, S-SEM) и модуль извлечения абстрактного контента (Content feature Extraction Module, CEM). Модуль межмасштабного взаимодействия (Cross-scale Interaction Module, CIM), расположенный между энкодерами и декодерами на разных уровнях масштабирования, предназначен для устранения семантических разрывов на одном уровне масштабирования, а также разрывов в разрешении между различными уровнями масштабирования. Энкодеры и декодеры состоят из модулей S-SEM и CEM. Исходное изображение низкого разрешения  $I_{HR} \in R^{H \times W \times 3}$  поступает на сверточный слой  $3 \times 3$  для преобразования данных из RGB цветового пространства в представление низкоуровневых признаков  $F_0 \in R^{H \times W \times C}$ , где  $H \times W$  – пространственная размерность, а  $C$  – количество каналов. Затем представление  $F_0$  пропускается через три модуля, последовательно уменьшающих пространственный размер до  $C \times H \times W$  ( $F_1$ ),  $C \times H/2 \times W/2$  ( $F_2$ ) и  $C \times H/4 \times W/4$  ( $F_3$ ):

$$\begin{cases} F_1 = Enc_1(F_0) \\ F_i = Enc_i(Conv_{2 \times 2}^\downarrow(F_{i-1})) \quad i = 2, 3 \end{cases}$$

где  $Conv_{2 \times 2}^\downarrow$  означает сверточный слой  $2 \times 2$  с шагом 2 для понижения частоты дискретизации.

После вычисления представления  $F_3$  модуль CIM адаптивно объединяет и устраняет разрывы между тремя разномасштабными представлениями  $[F_1, F_2, F_3]$ , получая выходные данные  $[O_1, O_2, O_3]$ :

$$[O_1, O_2, O_3] = CIM[F_1, F_2, F_3].$$

Выходные данные самого низкого разрешения поступают на вход декодера последнего уровня  $Dec_3$ , формирующего выход  $P_1$ . Аналогично происходит повышающая дискретизация на 2-ом и 1-ом уровнях с постепенным обогащением представлений  $P_2$  и  $P_3$ :

$$\begin{cases} P_1 = DeConv_{2 \times 2}^\uparrow(Dec_3(O_3)) \\ P_2 = DeConv_{2 \times 2}^\uparrow(Dec_2(O_2 + P_1)), \\ P_3 = Dec_1(O_1 + P_2) \end{cases}$$

где  $DeConv_{2 \times 2}^\uparrow$  означает транспонированный сверточный слой  $2 \times 2$  с шагом 2 для повышения частоты дискретизации.

Далее выходные данные декодера  $P_3$  реконструируются и подвергаются повышающей дискретизации с помощью сверточного слоя  $3 \times 3$  и операций перемешивания пикселей для получения окончательных результатов сверхвысокого разрешения. Сформированные таким образом контекстно-зависимые данные суммируются с данными после бикубической интерполяции исходного снимка, в результате чего и происходит формирование снимка  $I_{SR}$ . При обучении сети HAUNet используется функция потерь  $L1$ :

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|I_{HR}^{(i)} - I_{SR}^{(i)}\|,$$

где  $\theta$  – параметры модели,  $N$  – количество обучающих экземпляров.

Модель HAUNet способна увеличивать масштаб в 2, 3 и 4 раза. Обучение выполняется по методу обучения с учителем. Входными изображениями являются случайные фрагменты (patches) LR изображений размерностью  $48 \times 48$  пикселей. Результат восстановления оценивается с помощью соответствующих доступных HR фрагментов изображений по известным метрикам PSRN и SSIM.

Интересно, что настройка сети в процессе обучения выполняется с использованием диагностической технологии LAM (Local Attribution Maps) [27].

Однако ограниченное поле восприятия СНС не способствует эффективному обнаружению малоразмерных объектов, а также ограничивает производительность моделей и возможность их установки на терминальные устройства из-за их высокой вычислительной сложности и большого количества параметров. Для решения подобных проблем были разработаны так называемые визуальные трансформеры, которые в настоящее время являются наиболее интересной и востребованной модификацией СНС. Визуальные трансформеры используются в разных задачах машинного зрения и, в частности, при восстановлении снимков дистанционного зондирования сверхвысокого разрешения. В данном исследовании для сравнительного анализа была выбрана контекстно-зависимая облегченная сеть сверхвысокого разрешения (Context-Aware Lightweight Super-Resolution Network, CALSRN) [18]. Модель CALSRN, в основном, состоит из блоков контекстно-зависимых преобразователей (Context-Aware Transformer Block, CATB), предназначенных как для извлечения локального контекста (Local Context Extraction Branch, LCEB), так и глобального контекста (Global Context Extraction Branch, GCEB) изображений. При этом модуль LCEB использует механизм перекрестного внимания на основе СНС для извлечения локальной информации, а модуль GCEB имеет структуру Swin Transformer для получения глобальной информации. Генерация весов агрегации модулей LCEB и GCEB осуществляется модулем генерации динамических весов (Dynamic Weight Generation Branch, DWGB), что повышает качество восстановления изображений сверхвысокого разрешения. Усложненная архитектура глубокой сети CALSRN требует меньшего количества параметров и меньшей вычислительной сложности по сравнению с существующими методами.

Архитектура модели CALSRN состоит из двух базовых ветвей: ветвь билинейной интерполяции и ветвь реконструкции. В свою очередь ветвь реконструкции состоит из трех основных частей: модуля извлечения низкоуровневых признаков, модуля извлечения высокоуровневых признаков и слоя реконструкции. Модуль извлечения низкоуровневых признаков использует сверточный слой  $3 \times 3$  и параметрическую функцию активации PReLU (Parametric Rectification Linear Unit). Таким образом, на выходе данного модуля формируются признаки  $F_0 \in \mathbb{R}^{H \times W \times C}$  в соответствии с выражением:

$$F_0 = \text{PRELU}(\text{conv}_{3 \times 3}(I_{LR})),$$

где  $\text{conv}_{3 \times 3}$  – сверточный слой с ядром  $3 \times 3$ .

Высокоуровневые функции извлекаются с помощью  $n$  каскадных блоков САТВ. На вход первого каскада поступают низкоуровневые признаки  $F_0$ , а выходные признаки  $n$ -го каскада  $F_n \in \mathbb{R}^{H \times W \times C}$  вычисляются следующим образом:

$$F_n = f_{\text{САТВ}}^n \left( f_{\text{САТВ}}^{n-1} \dots \left( f_{\text{САТВ}}^1 (F_0) \right) \right),$$

где  $f_{\text{САТВ}}^n$  – функция  $n$ -го блока САТВ.

Выходы всех блоков САТВ объединяются и последовательно пропускаются через два сверточных слоя  $\text{conv}_{1 \times 1}$  и  $\text{conv}_{3 \times 3}$  для извлечения высокоуровневых признаков. На выходе модуля извлечения высокоуровневых признаков выполняется суммирование низкоуровневых и высокоуровневых признаков, в результате чего формируются признаки  $F_{\text{add}} \in \mathbb{R}^{H \times W \times C}$ :

$$F_{\text{add}} = \text{conv}_{3 \times 3} \left( \text{conv}_{1 \times 1} ([F_1, F_2, \dots, F_n]) \right) + F_0,$$

где символ  $[\cdot, \cdot]$  обозначает операцию конкатенации.

Таким образом, выходные функции блоков САТВ разных уровней объединяются для получения изображения  $I_{\text{SR}}$ . Каждый блок САТВ имеет сложную структуру, включающую небольшую СНС с механизмом внимания и трансформер с возможностью генерации динамических весов модулем DWGB.

Слой реконструкции состоит из сверточного слоя  $3 \times 3$  с выходной функцией  $F_{\text{add}}$  и оператора перемешивания пикселей (Pixel Shuffle). Далее выходные функции двух базовых ветвей (ветвь билинейной интерполяции и ветвь реконструкции) суммируются, образуя изображение сверхвысокого разрешения  $I_{\text{SR}}$ :

$$I_{\text{SR}} = H_P(\text{conv}_{3 \times 3}(F_{\text{add}})) + H_B(I_{\text{LR}}),$$

где  $H_P$  – оператор перемешивания пикселей,  $H_B$  – оператор билинейной интерполяции.

Интересно отметить, что применение структуры смещенных окон (Swin Transformer) в модуле GCEB позволяет лучше

анализировать пространственные контексты, одновременно улучшая восприятие локальных пространственных особенностей и снижая вычислительную сложность.

При обучении сети CALSRN используется функция потерь  $L1$ , аналогичная функции потерь сети HAUNet. Модель CALSRN также способна увеличивать масштаб в 2, 3 и 4 раза.

**5. Восстановление аэрофотоснимков сверхвысокого разрешения с применением ГСС.** Для учета семантических особенностей аэрофотоснимков разработана новая модель ГСС (SemESRGAN), состоящая из базовой сети на основе известной модели восстановления изображений сверхвысокого разрешения ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) [28] и сети сегментации изображений, настроенной на конкретную прикладную задачу. Архитектура предлагаемой модели ГСС на этапе обучения представлена на рисунке 1. Генератор модели ESRGAN основан на архитектуре плотной сети «остаток в остатке» (Residual-in-Residual Dense Network, RRDN), сочетающей многоуровневую остаточную сеть и плотные соединения. Для повышения производительности и снижения сложности вычислений были удалены все слои пакетной нормализации (batch normalization layers). Дискриминатор модели ESRGAN предсказывает вероятность того, что реальное изображение более реалистично, чем поддельное изображение (в данном случае  $I_{SR}$ ). Сеть сегментации была предварительно обучена для генерации семантических масок таких типичных классов, встречающихся на аэрофотоснимках, как автомобили, дороги, растительность, крыши и т.д., с использованием общедоступных наборов данных CLCD, DOTA LEVIR-CD и UAVid. Сеть сегментации представляет собой СНС с архитектурой энкодер-декодер, на выходе которой формируется изображение в виде псевдоцветов. На этапе обучения требуются две идентичные сети сегментации (по сути, сиамская архитектура). На вход одной сети подается изображение из обучающего набора данных  $I_{HR}$ , в то время, как на вход другой сети поступает реконструированное генератором изображение  $I_{SR}$ . Такое решение позволяет оценить сходство признаков как часть общей функции потерь и, следовательно, улучшить обучение генератор в семантическом аспекте.

Функция состязательных потерь ГСС  $L_{GAN}$  формулируется как задача min-max оптимизации, когда генератор обучен минимизировать потери, а дискриминатор обучен их максимизировать:

$$\min_G \max_D L_{GAN}(G, D) = E_{I_{HR}} \left[ \log D_{\theta_D}(I_{HR}) \right] + E_{I_{LR}} \left[ \log \left( 1 - D_{\theta_D}(G_{\theta_G}(I_{LR})) \right) \right],$$

где  $\theta_G$  и  $\theta_D$  – параметры генератора  $G$  и дискриминатора  $D$  соответственно.

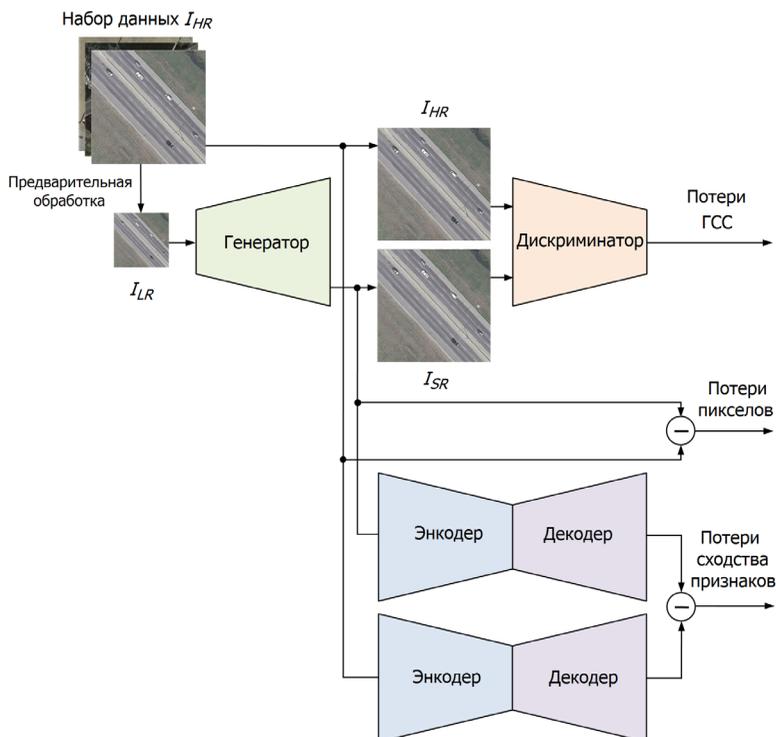


Рис. 1. Архитектура предложенной модели ГСС на этапе обучения

Обучение ГСС останавливается, когда дискриминатор достигает значение вероятности 0,5 для всех сгенерированных изображений  $\{I_{SR}\}$ . Следует отметить, что после обучения дискриминатор удаляется из архитектуры ГСС.

Как показано в работе [29], при восстановлении изображений сверхвысокого разрешения средняя абсолютная ошибка (Mean Absolute Error, MAE) выше среднеквадратической ошибки (Mean Square Error, MSE) с точки зрения различий на уровне пикселей. Применим функцию потерь пикселей вида MAE:

$$L_{MAE} = \frac{1}{HWC} \|I_{HR} - I_{SR}\|_1,$$

где  $H$ ,  $W$ ,  $C$  – высота, ширина и количество каналов изображения.

Недостаток функции потерь пикселей заключается в том, что эти потери рассчитываются на уровне пикселей при условии минимизации средней абсолютной ошибки и, следовательно, создают размытые фрагменты изображения. Поэтому функцию потерь  $L_{MAE}$  необходимо применять с понижающим коэффициентом в выражении для функции общих потерь.

Потери сходства признаков, известные как потери восприятия, показывают различия семантических особенностей, вычисленных энкодерами-декодерами СНС:

$$L_{feat} = \frac{1}{HWC} \left\| M(I_{HR}) - M(I_{SR}) \right\|_2^2,$$

где  $M(I_{HR})$  и  $M(I_{SR})$  – маски, сгенерированные СНС, предварительно обученными на аэрофотоснимках высокого разрешения.

Таким образом, общая функция потерь имеет вид:

$$L = \alpha L_{GAN} + \beta L_{MAE} + \gamma L_{feat},$$

где  $\alpha$ ,  $\beta$  и  $\gamma$  – эмпирически подобранные коэффициенты.

**6. Экспериментальные результаты.** Для обучения и тестирования предложенной модели глубокой сети (SemESRGAN) использовались шесть открытых наборов данных (рисунк 2):

- CropLand Change Detection (CLCD) [30] состоит из 600 изображений (512×512 пикселей) пахотных земель, собранных при помощи спутника Gaofen-2 в Китае, с пространственным разрешением от 0,5 до 2 м;

- Dataset of Object deTaction in Aerial (DOTA) [31] включает коллекцию изображений с разрешением от 800×800 до 20000×20000 пикселей собранных из различных источников (сервис Google Earth, спутник GF-2, другие аэрофотоснимки). Содержит 18 категорий объектов и состоит из 11268 изображений;

- LEVIR building Change Detection (LEVIR-CD) [32] состоит из 637 изображений, взятых из сервиса Google Earth высокого разрешения (50 см/пиксел) размером 1024×1024 пикселей. Набор данных охватывает различные типы зданий жилого частного сектора;

- UAVid [33] содержит изображения уличных сцен высокого разрешения (3840×2160 пикселей) полученных с применением БПЛА. В общей сложности набор состоит из 300 изображений, на которых

размечены 8 классов объектов: здания, дороги, деревья, растительность, движущиеся автомобили, статичные автомобили, люди, фон;

– Airbus Aircraft Detection (AAD) [34] содержит 103 изображения с разрешением  $2560 \times 2560$  пикселей (пространственное разрешение 50 см). Набор данных включает изображения аэропортов по всему миру, на некоторых изображениях присутствуют туман или облака;

– Aerial Image Dataset (AID) [35] включает 10000 изображений разрешением  $600 \times 600$  пикселей, принадлежащих 30 классам (аэропорты, стадионы, поля, пляжи, мосты, коммерческая застройка, пустыня, посевные площади, лес, горы, парки, стоянки, детские площадки, порты, железнодорожные станции, реки, школы, жилые массивы, площади, виадуки и т.д.).

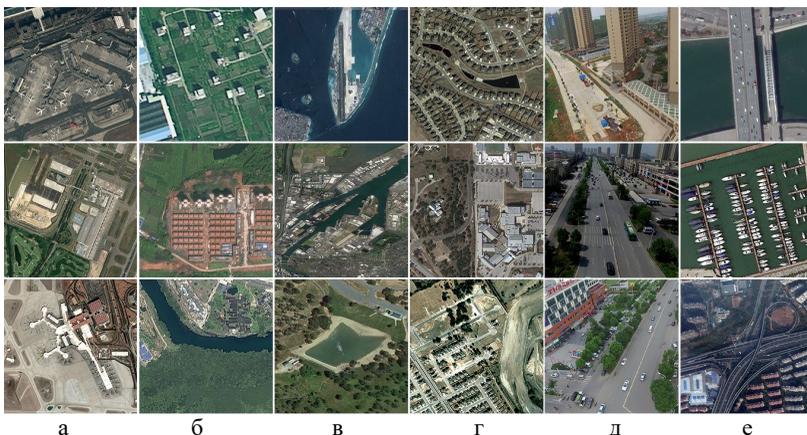


Рис. 2. Примеры изображений из используемых наборов данных:  
а) AAD, б) CLCD, в) DOTA, г) LEVIR-CD, д) UAVid, е) AID

Каждый набор данных, за исключением наборов данных AAD и AID, был разделен на обучающую, проверочную и тестовую выборки в соотношении 70/20/10 соответственно. В качестве аугментации применялось вертикальное и горизонтальное отражение изображений. Предложенная модель SemESRGAN была реализована на языке Python с использованием Pytorch repository. В экспериментах использовались графические процессоры NVIDIA Geforce RTX 2080 Ti (11 Гб), операционная система – MS Windows 10.

На рисунке 3 показаны графики функции потерь и валидации в процессе обучения предложенной модели сети SemESRGAN для получения изображений в масштабах  $\times 2$ ,  $\times 3$  и  $\times 4$ .

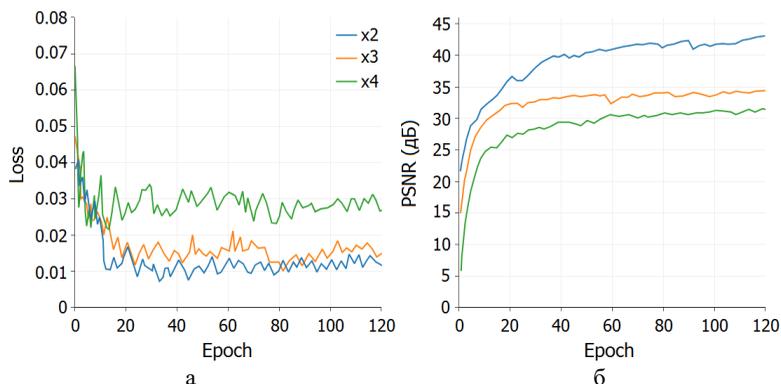


Рис. 3. Результаты обучения предложенной модели SemESRGAN: а) графики потерь, б) графики валидации

Для оценки качества восстановления изображений применялись две метрики: пиковое отношение сигнала к шуму (PSNR, Peak Signal-to-Noise Ratio) и индекс структурного сходства (SSIM, Structure Similarity).

В таблице 1 показаны средние значения метрик PSNR и SSIM для тестовых изображений из наборов данных CLCD, DOTA, LEVIR-CD и UAVid, а также сравнительные показатели предложенного метода SemESRGAN с другими методами (HAUNet, CALSRN и ESRGAN). При этом эксперименты проводились для 2-, 3- и 4-кратного увеличения исходных LR изображений.

Наборы данных AAD и AID не участвовали в обучении, а применились для тестирования разработанной модели на обобщаемость (таблица 2). Полученные оценки свидетельствуют о том, что предложенная модель SemESRGAN способна обеспечить качественную генерацию изображений сверхвысокого разрешения. В ходе экспериментов были выявлены некоторые артефакты при восстановлении SR изображений, связанные с нарушением текстуры при сильном размытии исходных LR изображений. Однако модель SemESRGAN лучше других моделей восстанавливает контуры объектов, что видно из PSNR показателей в таблицах 2 и 3, а также на рисунках 4-9.

Таблица 1. Средние значения PSNR (дБ)/SSIM

Метод	CLCD	DOTA	LEVIR-CD	UAVid
Увеличение ×2				
HAUNet [26]	32,15/0,895	33,04/0,914	25,78/0,714	28,34/0,833
CALSRN [18]	39,12/0,948	35,98/0,925	26,77/0,740	31,71/0,897
ESRGAN [28]	41,14/0,934	39,28/0,929	29,88/0,841	35,38/0,935
SemESRGAN	44,42/0,978	40,64/0,962	30,62/0,848	37,54/0,958
Увеличение ×3				
HAUNet [26]	30,91/0,839	28,63/0,824	23,88/0,678	26,30/0,716
CALSRN [18]	30,22/0,865	33,82/0,892	25,27/0,692	28,41/0,779
ESRGAN [28]	38,10/0,897	32,21/0,874	25,98/0,684	30,98/0,843
SemESRGAN	40,24/0,942	34,56/0,902	26,65/0,693	32,30/0,868
Увеличение ×4				
HAUNet [26]	29,36/0,783	30,45/0,847	22,98/0,528	25,08/0,632
CALSRN [18]	27,58/0,793	31,79/0,844	24,91/0,592	26,94/0,696
ESRGAN [28]	31,25/0,822	32,93/0,884	25,12/0,638	28,27/0,734
SemESRGAN	34,00/0,868	34,87/0,895	26,66/0,653	29,60/0,774

Таблица 2. Средние значения PSNR (дБ)/SSIM

Метод	AAD	AID
Увеличение ×2		
HAUNet [26]	29,76/0,892	27,12/0,720
CALSRN [18]	33,21/0,932	29,52/0,860
ESRGAN [28]	35,55/0,932	34,94/0,920
SemESRGAN	37,81/0,967	36,44/0,953
Увеличение ×3		
HAUNet [26]	27,53/0,811	26,57/0,708
CALSRN [18]	29,26/0,849	28,59/0,750
ESRGAN [28]	30,73/0,870	32,26/0,901
SemESRGAN	32,17/0,899	33,52/0,916
Увеличение ×4		
HAUNet [26]	25,88/0,697	26,22/0,698
CALSRN [18]	27,07/0,768	27,93/0,737
ESRGAN [28]	27,27/0,798	30,12/0,850
SemESRGAN	29,07/0,817	31,36/0,872

Дополнительно для оценки визуального качества изображений использовалась метрика LPIPS (Learned Perceptual Image Patch Similarity). Метрика LPIPS [36] применяется для измерения сходства восприятия между изображениями, созданными нейросетевыми моделями. В таблицах 3-4 показаны среднее значение метрики LPIPS для различных наборов данных. Более высокие значения означают большие различия в изображениях, а более низкие значения – большее

сходство оригинальных и восстановленных изображений. Если оценка по метрике LPIPS принимает нулевое значение, то изображения идентичны с точки зрения восприятия человеком.

Таблица 3. Средние значения LPIPS

Метод	CLCD	DOTA	LEVIR-CD	UAVid
Увеличение ×2				
HAUNet [26]	0,07011	0,05234	0,02916	0,01781
CALSRN [18]	0,06591	0,04879	0,03672	0,03314
ESRGAN [28]	0,00487	0,00523	0,00846	0,00440
SemESRGAN	0,00162	0,00146	0,00088	0,00043
Увеличение ×3				
HAUNet [26]	0,16280	0,06919	0,04812	0,02732
CALSRN [18]	0,10083	0,05515	0,04331	0,03752
ESRGAN [28]	0,02496	0,00566	0,00764	0,00583
SemESRGAN	0,01978	0,00266	0,00183	0,00086
Увеличение ×4				
HAUNet [26]	0,24689	0,09020	0,08371	0,04773
CALSRN [18]	0,15654	0,06528	0,05594	0,04411
ESRGAN [28]	0,08494	0,00956	0,01101	0,00932
SemESRGAN	0,07709	0,00627	0,00620	0,00395

Таблица 4. Средние значения LPIPS

Метод	AAD	AID
Увеличение ×2		
HAUNet [26]	0,01430	0,03062
CALSRN [18]	0,01993	0,03619
ESRGAN [28]	0,00240	0,00443
SemESRGAN	0,00033	0,00110
Увеличение ×3		
HAUNet [26]	0,02614	0,05373
CALSRN [18]	0,02554	0,04700
ESRGAN [28]	0,00255	0,00807
SemESRGAN	0,00062	0,00381
Увеличение ×4		
HAUNet [26]	0,04701	0,08912
CALSRN [18]	0,03474	0,06234
ESRGAN [28]	0,00454	0,03341
SemESRGAN	0,00267	0,01420

На рисунках 4-9 показаны примеры применения SR методов для 4-кратного увеличения изображения. Поскольку оригинальные HR изображения из разных наборов данных имеют разное разрешение, они были предварительно приведены к единому масштабу.



Рис. 4. Примеры восстановления SR фрагментов для набора данных CLCD в четырехкратном увеличении, файл – 00526.png: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN



Рис. 5. Примеры восстановления SR фрагментов для набора данных DOTA в четырехкратном увеличении, файл – P4219.png: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

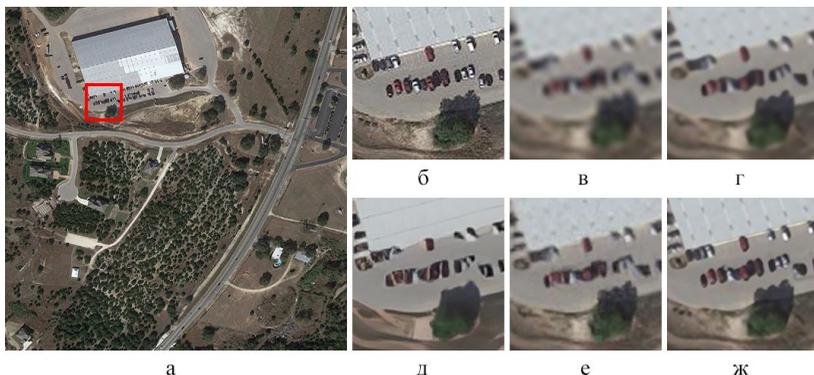


Рис. 6. Примеры восстановления SR фрагментов для набора данных LEVIRCD в четырехкратном увеличении, файл – test\_16.png: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

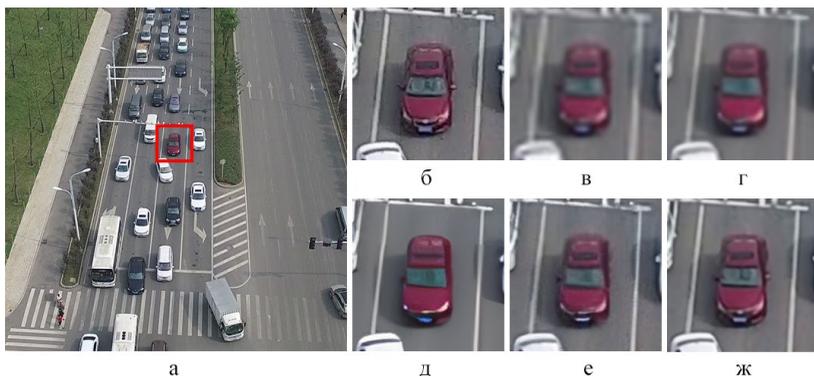


Рис. 7. Примеры восстановления SR фрагментов для набора данных UAVid в четырехкратном увеличении, файл – file15-1.png: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

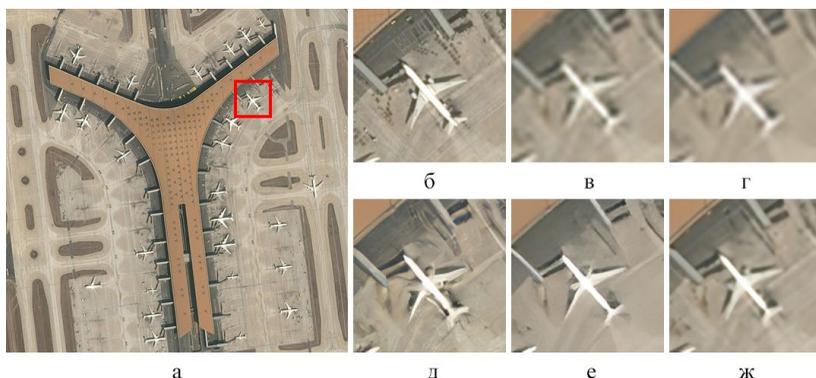


Рис. 8. Примеры восстановления SR фрагментов для набора данных AAD в четырехкратном увеличении, файл – 12210ad7-83f8-4b54-bb4b-e93f8ff6ac1f.jpg:  
 а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

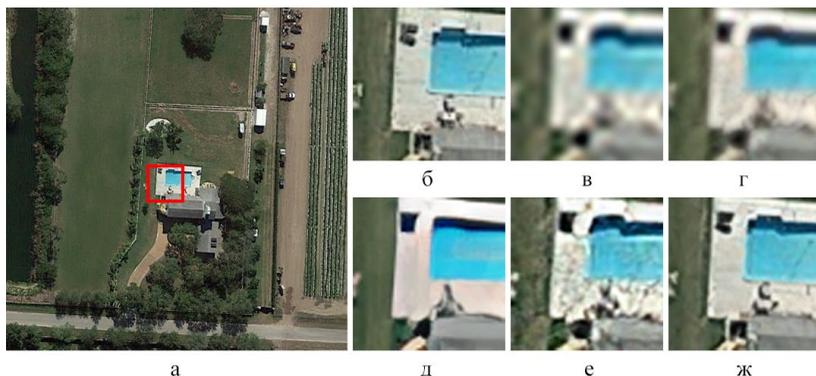


Рис. 9. Примеры восстановления SR фрагментов для набора данных AID в четырехкратном увеличении, файл sparseresidential\_51.jpg: а) входное HR изображение; б) фрагмент оригинального HR изображения; в) бикубическая интерполяция; г) модель HAUNet; д) модель CALSRN; е) модель ESRGAN; ж) предложенная модель SemESRGAN

**7. Заключение.** Предложенная нейросетевая модель SemESRGAN на основе ГСС позволяет генерировать достаточно реалистичные снимки дистанционного зондирования. Данный подход характеризуется повышенной чувствительностью к деталям изображения, которую обычно не могут обеспечить нейросетевые модели на основе СНС, что приводит к чрезмерной размытости

и исчезновению текстурных особенностей. Отметим, что потеря текстурных особенностей является общей проблемой любых методов восстановления снимков сверхвысокого разрешения, а качество восстановления сильно зависит от параметров исходных изображений, входящих в обучающие наборы данных.

### Литература

1. Фаворская М.Н. Аналитическое исследование моделей глубокого обучения для создания снимков ДЗЗ сверхвысокого разрешения // *Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов (SDM-2023)*: Сб. тр. Всероссийской конф. с междунар. участ. 2023. С. 17–25.
2. Lepcha D.C., Goyal B., Dogra A., Goyal V. Image super-resolution: A comprehensive review, recent trends, challenges and applications // *Information Fusion*. 2023. vol. 91. pp. 230–260.
3. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial nets. *Advances in Neural Information Processing Systems (NIPS 2014)*. 2014. vol. 27. pp. 1–9.
4. Фаворская М.Н., Пахирка А.И. Улучшение разрешения снимков ДЗЗ на основе глубоких генеративно-сопоставительных сетей // *Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов (SDM-2023)*: Сб. тр. Всероссийской конф. с междунар. участ. 2023. С. 163–168.
5. Conde M.V., Choi U.J., Burchi M., Timofte R. Swin2SR: SwinV2 transformer for compressed image super-resolution and restoration // *Computer Vision – ECCV 2022 Workshops*. LNCS. Springer, Cham. 2023. vol. 13802. pp. 669–687.
6. Wang P., Bayram B., Sertel E. A comprehensive review on deep learning based remote sensing image super-resolution methods // *Earth-Science Reviews*. 2022. vol. 232(15). DOI: 10.1016/j.earscirev.2022.104110.
7. Qiu D., Cheng Y., Wang X. Medical image super-resolution reconstruction algorithms based on deep learning: A survey // *Computer Methods and Programs in Biomedicine*. 2023. vol. 238. DOI: 10.1016/j.cmpb.2023.107590.
8. Jiang J., Wang C., Liu X., Ma J. Deep learning-based face super-resolution: A survey // *ACM Computing Surveys*. 2021. vol. 55. no. 1. pp. 1–36.
9. Liu H., Ruan Z., Zhao P., Dong C., Shang F., Liu Y., Yang L., Timofte R. Video super-resolution based on deep learning: A comprehensive survey // *Artificial Intelligence Review*. 2022. vol. 55. no. 8. pp. 5981–6035.
10. Sun Y., Deng K., Ren K., Liu J., Deng C., Jin Y. Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2024. vol. 208. pp. 14–38.
11. Wang T., Sun W., Qi H., Ren P. Aerial image super resolution via wavelet multiscale convolutional neural networks // *IEEE Geoscience and Remote Sensing Letters*. 2018. vol. 15. no. 5. pp. 769–773.
12. Xu W.-J., Xu G.-L., Wang Y., Sun X., Lin D.-Y., Wu Y.-R. High quality remote sensing image super-resolution using deep memory connected network. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018)*. 2018. pp. 8889–8892.
13. Tang J., Zhang J., Chen D., Al-Nabhan N., Huang C. Single-frame super-resolution for remote sensing images based on improved deep recursive residual network // *EURASIP J Image Video Proc*. 2021. vol. 2021. DOI: 10.1186/s13640-021-00560-8.

14. Tang S., Liu J., Xie X., Yang S., Zeng W., Wang X. A stage-mutual-affine network for single remote sensing image super-resolution // Chinese Conference on Pattern Recognition and Computer Vision (PRCV). 2022. pp. 249–261.
15. Wang S., Zhou T., Lu Y., Di H. Contextual transformation network for lightweight remote-sensing image super-resolution // IEEE Transactions on Geoscience and Remote Sensing. 2022. vol. 60. pp. 1–13. DOI: 10.1109/TGRS.2021.3132093.
16. Lei S., Shi Z., Mo W. Transformer-based multistage enhancement for remote sensing image super-resolution // IEEE Transactions on Geoscience and Remote Sensing. 2022. vol. 60. pp. 1–11. DOI: 10.1109/TGRS.2021.3136190.
17. Shang J., Gao M., Li Q., Pan J., Zou G., Jeon G. Hybrid-scale hierarchical transformer for remote sensing image super-resolution // Remote Sens. 2023. vol. 15. no. 13. pp. 1–20.
18. Peng G., Xie M., Fang L. Context-aware lightweight remote-sensing image super-resolution network // Frontiers in Neurorobotics. 2023. vol. 17. DOI: 10.3389/fnbot.2023.1220166.
19. Li Y., Mavromatis S., Zhang F., Du Z., Sequeira J., Wang Z., Zhao X., Liu R. Single-image super-resolution for remote sensing images using a deep generative adversarial network with local and global attention mechanisms // IEEE Transactions on Geoscience and Remote Sensing. 2021. vol. 60. pp. 1–24. DOI: 10.1109/TGRS.2021.3093043.
20. Guo M., Zhang Z., Liu H., Huang Y. NDSRGAN: A novel dense generative adversarial network for real aerial imagery super-resolution reconstruction // Remote Sens. 2022. vol. 14. no. 7. pp. 1–23. DOI: 10.3390/rs14071574.
21. Zhang J., Xu T., Li J., Jiang S., Zhang Y. Single-image super resolution of remote sensing images with real-world degradation modeling // Remote Sens. 2022. vol. 14. no. 12. pp. 1–22. DOI: 10.3390/rs14122895.
22. Haykir A.A., Oksuz I. Transfer learning based super resolution of aerial images // 2022 30th Signal Processing and Communications Applications Conference (SIU). 2022. pp. 1–4.
23. Haykir A.A., Öksüz I. Super-resolution with generative adversarial networks for improved object detection in aerial images // Information Discovery and Delivery. 2023. vol. 51. no. 4. pp. 349–357.
24. Tuna C., Unal G., Sertel E. Single-frame super resolution of remote-sensing images by convolutional neural networks // Int. J. Remote Sens. 2018. vol. 39. no. 8. pp. 2463–2479.
25. Dong C., Loy C.C., He K., Tang, X. Learning a deep convolutional network for image super-resolution // Computer Vision – ECCV 2014: 13th European Conference. 2014. pp. 184–199.
26. Wang J., Wang B., Wang X., Zhao Y., Long T. Hybrid attention-based U-shaped network for remote sensing image super-resolution // IEEE Transactions on Geoscience and Remote Sensing. 2023. vol. 61. pp. 1–15.
27. Gu J., Dong C. Interpreting super-resolution networks with local attribution maps // Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. pp. 9199–9208.
28. Wang X., Yu K., Wu S., Gu J., Liu Y., Dong C., Qiao Y., Loy C.C. ESRGAN: Enhanced super-resolution generative adversarial networks // Computer Vision – ECCV 2018 Workshops. 2019. pp. 63–79.
29. Johnson J., Alahi A., Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution // Computer Vision – ECCV 2016: 14th European Conference. 2016. pp. 694–711.
30. Liu M., Chai Z., Deng H., Liu R. A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection // IEEE Journal of Selected

- Topics in Applied Earth Observations and Remote Sensing. 2022. vol. 15. pp. 4297–4306.
31. Xia G., Bai X., Ding J., Zhu Z., Belongie S., Luo J., Datcu M., Pelillo M., Zhang L. DOTA: A large-scale dataset for object detection in aerial images // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. pp. 3974–3983.
  32. Chen H., Shi Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection // Remote Sens. 2020. vol. 12. no. 10. DOI: 10.3390/rs12101662.
  33. Lyu Y., Vosselman G., Xia G-S., Yilmaz A., Yang M.Y. UAVid: A semantic segmentation dataset for UAV imagery // ISPRS Journal of Photogrammetry and Remote Sensing. 2020. vol. 165. pp. 108–119.
  34. Airbus Aircraft Detection. URL: [www.kaggle.com/datasets/airbusgeo/airbus-aircrafts-sample-dataset](http://www.kaggle.com/datasets/airbusgeo/airbus-aircrafts-sample-dataset) (дата обращения: 04.03.2024).
  35. Xia G.-S., Hu J., Hu F., Shi B., Bai X., Zhong Y., Zhang L. AID: A benchmark dataset for performance evaluation of aerial scene classification // IEEE Transactions on Geoscience and Remote Sensing. 2017. vol. 55. no. 7. pp. 3965–3981.
  36. Zhang R., Isola P., Efros A.A., Shechtman E., Wang O. The unreasonable effectiveness of deep features as a perceptual metric // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: Salt Lake City, UT, USA. 2018. pp. 586–595.

**Фаворская Маргарита Николаевна** — д-р техн. наук, профессор, заведующий кафедрой, кафедра информатики и вычислительной техники, Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева (СибГУ им. М.Ф. Решетнева). Область научных интересов: компьютерное зрение, обработка изображений и видеопоследовательностей, машинное обучение, глубокое обучение, распознавание образов. Число научных публикаций — 300. [favorskaya@sibsau.ru](mailto:favorskaya@sibsau.ru); проспект им. газеты Красноярский Рабочий, 31, 660037, Красноярск, Россия; р.т.: +7(391)213-9622.

**Пахирка Андрей Иванович** — канд. техн. наук, доцент, кафедра информатики и вычислительной техники, Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева (СибГУ им. М.Ф. Решетнева). Область научных интересов: компьютерное зрение, обработка изображений и видеопоследовательностей, машинное обучение, глубокое обучение, распознавание образов. Число научных публикаций — 50. [pahirka@sibsau.ru](mailto:pahirka@sibsau.ru); проспект им. газеты Красноярский Рабочий, 31, 660037, Красноярск, Россия; р.т.: +7(391)213-9622.

M. FAVORSKAYA, A. PAKHIRKA  
**RESTORATION OF SEMANTIC-BASED SUPER-RESOLUTION  
AERIAL IMAGES**

*Favorskaya M., Pakhirka A.* **Restoration of Semantic-Based Super-Resolution Aerial Images.**

**Abstract.** Currently, technologies for remote sensing image processing are actively developing, including both satellite images and aerial images obtained from video cameras of unmanned aerial vehicles. Often such images have artifacts such as low resolution, blurred image fragments, noise, etc. One way to overcome such limitations is to use modern technologies to restore super-resolution images based on deep learning methods. The specificity of aerial images is the presentation of texture and structural elements in a higher resolution than in satellite images, which objectively contributes to better results of restoration. The article provides a classification of super-resolution methods based on the main architectures of deep neural networks, namely convolutional neural networks, visual transformers and generative adversarial networks. The article proposes a method for reconstructing super-resolution aerial images SemESRGAN taking into account semantic features by using an additional deep network for semantic segmentation during the training stage. The total loss function, including adversarial losses, pixel-level losses, and perception losses (feature similarity), is minimized. Six annotated aerial and satellite image datasets CLCD, DOTA, LEVIR-CD, UAVid, AAD, and AID were used for the experiments. The results of image restoration using the proposed SemESRGAN method were compared with the basic architectures of convolutional neural networks, visual transformers and generative adversarial networks. Comparative results of image restoration were obtained using objective metrics PSNR and SSIM, which made it possible to evaluate the quality of restoration using various deep network models.

**Keywords:** aerial images, super-resolution, semantic segmentation, convolutional neural networks, visual transformers, generative adversarial networks.

## References

1. Favorskaya M.N. [Analytical study of deep learning models for the problem of remote sensing single image super resolution] *Obrabotka prostranstvennykh dannykh v zadachax monitoringa pripodnykh i antropogennykh processov (SDM-2023)*: Sb. tr. Vserossijskoj nkonf. s mezhdunar. Uchast. [Processing of spatial data in tasks of monitoring natural and anthropogenic processes: Collected papers]. 2023. pp. 17–25. (In Russ.).
2. Lepcha D.C., Goyal B., Dogra A., Goyal V. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Information Fusion*. 2023. vol. 91. pp. 230–260.
3. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial nets. *Advances in Neural Information Processing Systems (NIPS 2014)*. 2014. vol. 27. pp. 1–9.
4. Favorskaya M.N., Pakhirka A.I. [GAN-based remote sensing single-image resolution improvement] *Obrabotka prostranstvennykh dannykh v zadachax monitoringa pripodnykh i antropogennykh processov (SDM-2023)*: Sb. tr. Vserossijskoj nkonf. s mezhdunar. Uchast. [Processing of spatial data in tasks of monitoring natural and anthropogenic processes: Collected papers]. 2023. pp. 163–168. (In Russ.).

5. Conde M.V., Choi U.J., Burchi M., Timofte R. Swin2SR: SwinV2 transformer for compressed image super-resolution and restoration. *Computer Vision – ECCV 2022 Workshops*. LNCS. Springer, Cham. 2023. vol. 13802. pp. 669–687.
6. Wang P., Bayram B., Sertel E. A comprehensive review on deep learning based remote sensing image super-resolution methods. *Earth-Science Reviews*. 2022. vol. 232(15). DOI: 10.1016/j.earscirev.2022.104110.
7. Qiu D., Cheng Y., Wang X. Medical image super-resolution reconstruction algorithms based on deep learning: A survey. *Computer Methods and Programs in Biomedicine*. 2023. vol. 238. DOI: 10.1016/j.cmpb.2023.107590.
8. Jiang J., Wang C., Liu X., Ma J. Deep learning-based face super-resolution: A survey. *ACM Computing Surveys*. 2021. vol. 55. no. 1. pp. 1–36.
9. Liu H., Ruan Z., Zhao P., Dong C., Shang F., Liu Y., Yang L., Timofte R. Video super-resolution based on deep learning: A comprehensive survey. *Artificial Intelligence Review*. 2022. vol. 55. no. 8. pp. 5981–6035.
10. Sun Y., Deng K., Ren K., Liu J., Deng C., Jin Y. Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2024. vol. 208. pp. 14–38.
11. Wang T., Sun W., Qi H., Ren P. Aerial image super resolution via wavelet multiscale convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*. 2018. vol. 15. no. 5. pp. 769–773.
12. Xu W.-J., Xu G.-L., Wang Y., Sun X., Lin D.-Y., Wu Y.-R. High quality remote sensing image super-resolution using deep memory connected network. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2018)*. 2018. pp. 8889–8892.
13. Tang J., Zhang J., Chen D., Al-Nabhan N., Huang C. Single-frame super-resolution for remote sensing images based on improved deep recursive residual network. *EURASIP J Image Video Proc*. 2021. vol. 2021. DOI: 10.1186/s13640-021-00560-8.
14. Tang S., Liu J., Xie X., Yang S., Wang S., Zeng W., Wang X. A stage-mutual-affine network for single remote sensing image super-resolution. *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*. 2022. pp. 249–261.
15. Wang S., Zhou T., Lu Y., Di H. Contextual transformation network for lightweight remote-sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*. 2022. vol. 60. pp. 1–13. DOI: 10.1109/TGRS.2021.3132093.
16. Lei S., Shi Z., Mo W. Transformer-based multistage enhancement for remote sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*. 2022. vol. 60. pp. 1–11. DOI: 10.1109/TGRS.2021.3136190.
17. Shang J., Gao M., Li Q., Pan J., Zou G., Jeon G. Hybrid-scale hierarchical transformer for remote sensing image super-resolution. *Remote Sens*. 2023. vol. 15. no. 13. pp. 1–20.
18. Peng G., Xie M., Fang L. Context-aware lightweight remote-sensing image super-resolution network. *Frontiers in Neuroinformatics*. 2023. vol. 17. DOI: 10.3389/fnbot.2023.1220166.
19. Li Y., Mavromatis S., Zhang F., Du Z., Sequeira J., Wang Z., Zhao X., Liu R. Single-image super-resolution for remote sensing images using a deep generative adversarial network with local and global attention mechanisms. *IEEE Transactions on Geoscience and Remote Sensing*. 2021. vol. 60. pp. 1–24. DOI: 10.1109/TGRS.2021.3093043.
20. Guo M., Zhang Z., Liu H., Huang Y. NDSRGAN: A novel dense generative adversarial network for real aerial imagery super-resolution reconstruction. *Remote Sens*. 2022. vol. 14. no. 7. pp. 1–23. DOI: 10.3390/rs14071574.

21. Zhang J., Xu T., Li J., Jiang S., Zhang Y. Single-image super resolution of remote sensing images with real-world degradation modeling. *Remote Sens.* 2022. vol. 14. no. 12. pp. 1–22. DOI: 10.3390/rs14122895.
22. Haykir A.A., Oksuz I. Transfer learning based super resolution of aerial images. 2022 30th Signal Processing and Communications Applications Conference (SIU). 2022. pp. 1–4.
23. Haykir A.A., Öksüz I. Super-resolution with generative adversarial networks for improved object detection in aerial images. *Information Discovery and Delivery.* 2023. vol. 51. no. 4. pp. 349–357.
24. Tuna C., Unal G., Sertel E. Single-frame super resolution of remote-sensing images by convolutional neural networks. *Int. J. Remote Sens.* 2018. vol. 39. no. 8. pp. 2463–2479.
25. Dong C., Loy C.C., He K., Tang, X. Learning a deep convolutional network for image super-resolution // *Computer Vision – ECCV 2014: 13th European Conference.* 2014. pp. 184–199.
26. Wang J., Wang B., Wang X., Zhao Y., Long T. Hybrid attention-based U-shaped network for remote sensing image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing.* 2023. vol. 61. pp. 1–15.
27. Gu J., Dong C. Interpreting super-resolution networks with local attribution maps. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 2021. pp. 9199–9208.
28. Wang X., Yu K., Wu S., Gu J., Liu Y., Dong C., Qiao Y., Loy C.C. ESRGAN: Enhanced super-resolution generative adversarial networks. *Computer Vision – ECCV 2018 Workshops.* 2019. pp. 63–79.
29. Johnson J., Alahi A., Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. *Computer Vision – ECCV 2016: 14th European Conference.* 2016. pp. 694–711.
30. Liu M., Chai Z., Deng H., Liu R. A CNN-transformer network with multiscale context aggregation for fine-grained cropland change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.* 2022. vol. 15. pp. 4297–4306.
31. Xia G., Bai X., Ding J., Zhu Z., Belongie S., Luo J., Datcu M., Pelillo M., Zhang L. DOTA: A large-scale dataset for object detection in aerial images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2018. pp. 3974–3983.
32. Chen H., Shi Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.* 2020. vol. 12. no. 10. DOI: 10.3390/rs12101662.
33. Lyu Y., Vosselman G., Xia G-S., Yilmaz A., Yang M.Y. UAVid: A semantic segmentation dataset for UAV imagery. *ISPRS Journal of Photogrammetry and Remote Sensing.* 2020. vol. 165. pp. 108–119.
34. Airbus Aircraft Detection. Available at: [www.kaggle.com/datasets/airbusgeo/airbus-aircrafts-sample-dataset](http://www.kaggle.com/datasets/airbusgeo/airbus-aircrafts-sample-dataset) (accessed 04.03.2024).
35. Xia G-S., Hu J., Hu F., Shi B., Bai X., Zhong Y., Zhang L. AID: A benchmark dataset for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing.* 2017. vol. 55. no. 7. pp. 3965–3981.
36. Zhang R., Isola P., Efros A.A., Shechtman E., Wang O. The unreasonable effectiveness of deep features as a perceptual metric. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE: Salt Lake City, UT, USA. 2018. pp. 586–595.

**Favorskaya Margarita** — Ph.D., Dr.Sci., Professor, Head of the department, Department of informatics and computer techniques, Reshetnev Siberian State University of Science and Technology (Reshetnev University). Research interests: computer vision, image and video sequence processing, machine learning, deep learning, pattern recognition. The number of publications — 300. favorskaya@sibsau.ru; 31, Krasnoyarsky Rabochy Ave., 660037, Krasnoyarsk, Russia; office phone: +7(391)213-9622.

**Pakhirka Andrey** — Ph.D., Associate professor, Department of informatics and computer techniques, Reshetnev Siberian State University of Science and Technology (Reshetnev University). Research interests: computer vision, image and video sequence processing, machine learning, deep learning, pattern recognition. The number of publications — 50. pahirka@sibsau.ru; 31, Krasnoyarsky Rabochy Ave., 660037, Krasnoyarsk, Russia; office phone: +7(391)213-9622.

В.Ю. ОСИПОВ  
**ИНТЕЛЛЕКТУАЛЬНАЯ НЕЙРОСЕТЕВАЯ МАШИНА  
С ФУНКЦИЯМИ МЫШЛЕНИЯ**

*Осипов В.Ю. Интеллектуальная нейросетевая машина с функциями мышления.*

**Аннотация.** В последние годы существенно вырос интерес к искусственному интеллекту на основе нейросетевых подходов. Получен ряд значимых научных результатов, которые нашли широкое применение на практике. Большое внимание привлекли генеративно – состязательные нейросетевые модели, нейросетевые трансформеры и другие решения. Достигнут очевидный прогресс в нейросетевом распознавании и генерации образов, обработке текстов и речи, прогнозировании событий, управлении трудно формализуемыми процессами. Однако пока не удалось наделять нейросетевые машины мышлением. Все получаемые с использованием нейросетевых машин результаты можно отнести к решениям на основе различных видов связывания сигналов без полноценного управления процессами их обработки. Типичными представителями таких машин выступают ChatGPT. Возможности по интеллектуальному оперированию различными сигналами в известных нейросетевых машинах очень ограничены. Среди основных причин таких ограничений следует выделить несовершенство используемых базовых принципов нейросетевой обработки информации. Свойства нейронов длительное время рассматривались упрощенно. Обуславливалось это, как пробелами в области биологических исследованиях, так и отсутствием возможностей построения больших нейронных сетей на сложных моделях нейронов. В последние годы ситуация изменилась. Появились новые способы реализации больших нейронных сетей. Также установлено, что даже отдельные нейроны могут обладать обширной внутренней памятью и реализовывать различные функции. Однако до сих пор многие механизмы функционирования нейронов и их взаимодействия остаются не раскрытыми. Мало исследованы вопросы управляемого ассоциативного обращения к внутренней памяти нейронов. Эти недостатки существенно сдерживает создание мыслящих нейросетевых машин. Объектом исследования в статье выступает процесс интеллектуальной нейросетевой обработки информации. Предмет исследования: принципы, модели и методы такой обработки. Преследуется цель расширения функциональных возможностей нейросетевых машин по решению трудно формализуемых творческих задач за счет разработки новых принципов, моделей и методов интеллектуальной обработки информации. В интересах достижения этой цели уточняются принципы функционирования интеллектуальных нейросетевых машин, предлагаются новые модели и методы нейросетевой обработки информации. Раскрывается новая модель импульсного нейрона, как базового элемента таких машин. Искусственный мозг нейросетевых машин рекомендуется формировать в виде многослойных нейронных сетей, наделенных логическими структурами, с различными по параметрам нейронами. Предлагается новый метод многоуровневой интеллектуальной обработки информации в нейросетевых машинах на основе умных импульсных нейронов. Поясняются механизмы мышления нейросетевых машин, лежащие в их основе функции интеллектуального оперирования образами и понятиями в нейросетевой памяти. Приводятся результаты моделирования, подтверждающие справедливость предложенных решений.

**Ключевые слова:** нейросетевая машина, интеллектуальность, функции мышления, умные нейроны, трансформация сигналов.

**1. Введение.** Впервые в широкой постановке вопрос о мышлении машин был поставлен Аланом Тьюрингом [1]. Однако до сих пор идут споры: может ли машина мыслить? Известно много определений понятия «мышление» [2]. Применительно к человеку это высшая ступень познания, процесс отражения в мозгу окружающего реального мира. Он основан на психофизических механизмах обучения и непрерывного пополнения запаса понятий, представлений и вывода новых суждений и умозаключений. В другой трактовке мышление можно рассматривать, как способность отражать в мозгу реальный мир и взаимодействие с ним, а также производить интеллектуальные операции с сигналами, образами и понятиями. Применительно к интеллектуальным нейросетевым машинам предлагается придерживаться последнего определения. Мышление машин согласно [3] относится к способности их имитировать когнитивные процессы человека, включая восприятие, рассуждения, принятие решений и решение проблем. Если можно смоделировать эти процессы, то допустимо говорить о наделении интеллектуальных нейросетевых машин функциями мышления (операциями различных видов анализа и синтеза сигналов). Под такими машинами будем понимать, прежде всего, абстрактные конструкции, которые могут реализовываться программными и аппаратными способами. Допустим, что отражение реального мира в искусственном мозгу машины осуществимо путем ее непрерывного обучения. Однако это обучение должно обладать широкими возможностями, включая кодирование и запоминание больших объемов информации с реализацией последующего извлечения из памяти интересующих событий. Известные решения [4, 5] пока не обеспечивают результатов, сравнимых с биологическими системами. Нет однозначного взгляда на то, как должна кодироваться информация в памяти интеллектуальных нейросетевых машин. Количество запоминаемой информации существующими нейронными сетями невелико [6]. Что касается осуществления в нейросетевой машине интеллектуальных операций, то здесь много темных пятен. Разработан ряд нейросетевых машин [3, 7, 8], реализующих интеллектуальные операции на основе заложенных в них правил. Однако их можно считать интеллектуальными лишь в узком плане [9, 10]. Известные нейросетевые машины ориентированы на решение только отдельных задач. Среди них фильтрация сигналов, распознавание, прогнозирование и восстановление событий, управление объектами и другие [6]. Ни одна из существующих нейросетевых машин без изменения структуры не способна полноценно обрабатывать широкий

спектр творческих задач. Одна из причин такого состояния – это несовершенство принципов функционирования нейросетевых машин. Они далеки от принципов, свойственных биологическим объектам [4, 5]. Положенные в их основу нейросетевые решения имеют существенные ограничения, как по одноуровневой, так и многоуровневой интеллектуальной обработке сигналов. Во многих случаях применяемые в интеллектуальных машинах нейронные сети рассматривают как черные ящики. Лишь в последние годы удалось наделять нейронные сети прозрачными логическими структурами [11]. Несмотря на это, проблемными остаются аспекты многоуровневой ассоциативной обработки потоков сигналов в рекуррентных нейронных сетях (РНС), составляющих основу интеллектуальных нейросетевых машин. Без такой обработки нет возможности переходить в нейросетевой машине от элементарных сигналов к образам и понятиям, а также возвращаться назад. К известным моделям и методам многоуровневой ассоциативной обработки сигналов в РНС можно отнести:

- иерархическую временную память (ИВП) и корковые алгоритмы обучения;
- рекуррентные сверточные нейронные сети и родственные методы;
- иерархические рекуррентные гибридные нейросетевые модели;
- многоуровневые потоковые РНС с управляемыми элементами и методы обработки информации в них.

Наиболее полный обзор и развитие идей ИВП был представлен в [12]. Основными функциями ИВП являются обучение, распознавание и прогнозирование. Особенности ИВП заключаются в наличии регионов, организованных по иерархическому принципу. Нейроны (клетки) каждой области объединены в двумерный (2D) массив столбцов. Каждый нейрон характеризуется прямыми и горизонтальными дендритами, что обеспечивает взаимодействие нейронов как внутри областей, так и между различными областями. Недостатки ИВП и связанных с ним методов обработки сигналов заключаются в следующем. Модель не подходит для обработки непрерывных потоков сигналов. Сеть ИВП представляет собой самоорганизующуюся структуру и требует достаточно много времени в части обработки входных сигналов. Функции и возможности сети и ее элементов по интеллектуальной обработке информации весьма ограничены. Не предусмотрено управление обработкой сигналов как

внутри регионов, так и между ними. Синапсы имеют только двоичный вес.

Общее описание рекуррентных сверточных нейронных сетей и методов обработки информации в них можно найти в [13, 14]. Эти сети могут использовать разные архитектуры, типичные как для традиционных сверточных сетей [13], так и для капсульных моделей [14]. У этих архитектур есть общая особенность, например, наличие хотя бы одного сверточного слоя. В общем, рекуррентные сверточные нейронные сети включают восходящие, нисходящие и горизонтальные связи. К недостаткам рекуррентных сверточных нейронных сетей с учетом ряда снятых ограничений капсульными моделями можно отнести их узкую специализацию, направленную исключительно на распознавание образов, а также трудности реализации и обучения.

Иерархические рекуррентные гибридные модели нейронных сетей представляют собой комбинированные структуры нейронных сетей с восходящей, нисходящей и продольной обработкой сигналов. Как правило, эти модели предназначены для распознавания, машинного перевода речи и прогнозирования событий. Они состоят из рекуррентных и нейронных сетей прямого распространения. Долгая краткосрочная память [15–17], самоорганизующиеся карты [6], структуры теории адаптивного резонанса [18] и другие нейронные сети могут служить конструкциями для этих моделей [19–21]. Проводятся исследования, направленные на придание комбинированным структурам нейронных сетей большей динамики [22, 23]. К недостаткам иерархических рекуррентных гибридных моделей нейронных сетей относятся: ограниченный функционал, несовершенство правил ассоциативной многоуровневой обработки сигналов; малый объем памяти; проблемы с обучением.

Сети с управляемыми элементами можно выделить в отдельный вид многоуровневых рекуррентных нейронных сетей [11], на основе которых осуществимо создание интеллектуальных нейросетевых машин. Эти потоковые сети применимы как для оперативной, так и для глубокой обработки информации, генерации новых сигналов [24, 25, 26]. В зависимости от вариантов ввода сигналов в РНС можно решать различные творческие задачи. Многоуровневая ассоциативная обработка сигналов в таких сетях реализуется по слоям. Эти РНС основаны на идеях, отраженных в [27]. В РНС используются импульсные нейроны. Время рефрактерности после их возбуждения превышает время задержки сигналов в образуемых двухслойных контурах сети. За счет приоритета сильных связей в РНС

обеспечивается однозначное соответствие между элементами входов и выходов сети. Эти РНС способны управлять ассоциативным взаимодействием сигналов в зависимости от текущих состояний сетей. К недостаткам потоковых РНС и методов обработки сигналов относятся следующие. Это низкоуровневая обработка сигналов в каждой из сигнальных систем, что приводит к увеличению объема памяти и затратам времени. Имеются трудности ассоциативной обработки в РНС информации о времени возникновения событий различного масштаба. Как и в других РНС существенно ограничены возможности базовых элементов – нейронов. На низком уровне обработки все прозрачно, можно оперировать в РНС прошлым, настоящим и будущим временем. Однако для перехода к логическим уровням обработки различных временных масштабов, интервалов и нечетких значений времени необходимо дальнейшее развитие теории и методов связывания сигналов в РНС.

В работе [28] сформулирована принципиальная идея расширенной многоуровневой обработки сигналов в интеллектуальных нейросетевых машинах на основе РНС с управляемыми элементами. Однако она требует дальнейшего развития. Идея обращена к поиску подходящих методов управляемого связывания пространственно-временных сигналов в РНС на различных уровнях иерархии. Подобные методы должны не только расширить память и сократить затраты времени на обработку сигналов в РНС, но и повысить возможности моделирования творческой деятельности человека. В интересах этого необходимо развить принципы интеллектуальной нейросетевой обработки информации и пересмотреть взгляды на функционирование искусственных нейронов и их взаимодействие.

В последние годы уделяется повышенное внимание к устранению несовершенства моделей нейронов, используемых в нейросетевых машинах [29, 30]. Исследователи стремятся, чтобы нейроны более полно учитывали информацию о пространственно-временных характеристиках обрабатываемых сигналов. Были предприняты некоторые усилия по изучению передачи пространственной информации о входных сигналах на выходы нейронов [29 – 33]. Предложены решения [30], позволяющие с помощью векторных нейронов реализовывать вращения обрабатываемых образов, но без управляемого ассоциативного вызова сигналов из памяти нейронных сетей. В [32] рассмотрены нейроны с большим объемом внутренней памяти на микротрубочках, формализуемых мемристорами. Однако пока не взглянули на нейроны,

как на полноценные ассоциативные процессоры. Несовершенство базовых элементов (нейронов) нейросетевых машин не позволяет сделать прорыв в наделении этих машин полноценным искусственным интеллектом. Требуется поиск новых идей, моделей, методов и технологий их реализации.

**2. Принципы функционирования ИНМ.** Опираясь на результаты биологических исследований [3, 4, 29, 32 – 34], предлагаются следующие принципы функционирования интеллектуальной нейросетевой машины (ИНМ). На входы искусственного мозга ИНМ должны подаваться одновременно не только сигналы внешнего мира, но и формируемые управляющие воздействия. Это позволяет связывать в мозгу ИНМ все эти сигналы, и обеспечивать формирование в нем модели окружающего мира и взаимодействие с ним самой машины. Перед подачей всех сигналов на вход искусственного мозга ИНМ они должны быть приведены к «языку», воспринимаемому его нейронной сетью. Эти сигналы представимы в виде последовательных совокупностей единичных импульсов (СЕИ), несущих всю информацию об исходных воздействиях. В интересах этого сигналы могут раскладываться на пространственно-частотные, пространственно-химические, пространственно-тактильные и другие составляющие. Каждая составляющая может преобразовываться в последовательность единичных импульсов с частотой и фазой, как функциями от амплитуды и фазы составляющей. В дискретном времени эти последовательности могут представляться в виде следующих друг за другом совокупностей единичных импульсов.

Искусственный мозг ИНМ должен быть биоподобным, устойчиво функционировать в условиях ограничений потребляемой энергии. Он должен формироваться с использованием многослойных управляемых рекуррентных нейронных сетей с прозрачной логической организацией, с возможностью реализации многоуровневой ассоциативной обработки информации. Обрабатываемые в РНС СЕИ должны продвигаться как между слоями, так и вдоль их. В искусственном мозгу ИНМ должна присутствовать возможность перехода от обработки СЕИ на нижних уровнях к обработке на верхних уровнях и наоборот.

ИНМ должна мысленно оперировать различными сигналами, образами и понятиями, а также временем. Эти операции должны отражать возможные действия над реальными объектами. К таким действиям можно отнести их перемещения (сдвиги), повороты, приближения и удаления, перестановку местами, дополнение

конструктивными элементами и другие. Оперирование временем предусматривает сдвиги точек отсчета, ускорение, замедление, изменение масштаба, запоминание и сравнение интервалов, другие операции.

В искусственном мозгу ИНМ должны присутствовать механизмы управления его деятельностью, обусловленные некоторыми мотивами. Основой их на самом нижнем уровне могут выступать минимальные расходы энергии при устранении раздражителей, а на верхнем уровне – минимизация информационных противоречий. Заметим, что информационные противоречия в РНС на физическом уровне проявляются в конфликтах между нейронами.

Для обеспечения всех этих возможностей нейроны искусственного мозга ИНМ должны обладать существенно расширенными функциями, чем известные модели. Они должны в широких пределах трансформировать различные распределения сигналов на входах в другие распределения на выходах с учетом управляющих воздействий и состояния внутренней памяти. При этом внутренняя память нейронов должна быть существенно больше памяти на их синапсах [35].

**3. Умные нейроны как базовые элементы ИНМ.** Поясним принципы функционирования ИНМ, начиная с рассмотрения новых моделей импульсных нейронов, как базовых элементов искусственного мозга рассматриваемой машины. В отличие от известных моделей в умном импульсном нейроне предусматривается учет, откуда и какие сигналы поступили на его входы. Полагается, что каждое индивидуальное воздействие на нейрон поступает на отдельный пространственно соотнесенный вход. Такая пространственная развязка в биологических нейронах реализуема, прежде всего, дендритами. При подаче сигнала на отдельный вход токи распределяются по поверхности нейрона с ослаблением. В случае нескольких сигналов, воздействующих на привязанные к ним пространственно соотнесенные входы нейрона, на его поверхности формируются совместные распределения токов. В зависимости от характеристик этих сигналов и пространственных параметров задействованных входов нейрона, эти распределения будут иметь различные пространственно соотнесенные максимальные значения. Таким образом, для каждого варианта входных сигналов можно поставить в соответствие некоторую пространственную точку максимального приложения сигналов к нейрону. В случае превышения током в этой точке порога возбуждения нейрона формируется единичный заряд, который распределяется по синапсам с учетом точки

начала возбуждения и внутренней памяти нейрона. Полагается, что от каждой пространственной точки начала возбуждения нейрона через его тело и аксон до синапсов существуют свои пути переноса зарядов. Этим путям, микротрубочкам [32], свойственны свои проводимости, изменяемые в зависимости от перенесенных по ним зарядов. Опираясь на эти положения, можно утверждать, что каждому варианту входных сигналов будет соответствовать на выходе обученного умного нейрона свой вариант сигналов. Таким образом, даже отдельно взятый умный нейрон может однозначно трансформировать одни входные сигналы в другие. В частности, каждый умный нейрон может успешно реализовывать не только операции, связанные со сдвигами, поворотами и другими действиями над сигналами, но и с операциями, эквивалентными математическим действиям (сложением, умножением, делением, возведением в степень, логарифмированием и другими). В рассматриваемом случае происходят трансформации одних сигналов в другие в зависимости от управляющих воздействий. Заметим, что входные сигналы умного нейрона можно разделить на информационные и управляющие воздействия (рисунок 1(а, б)).

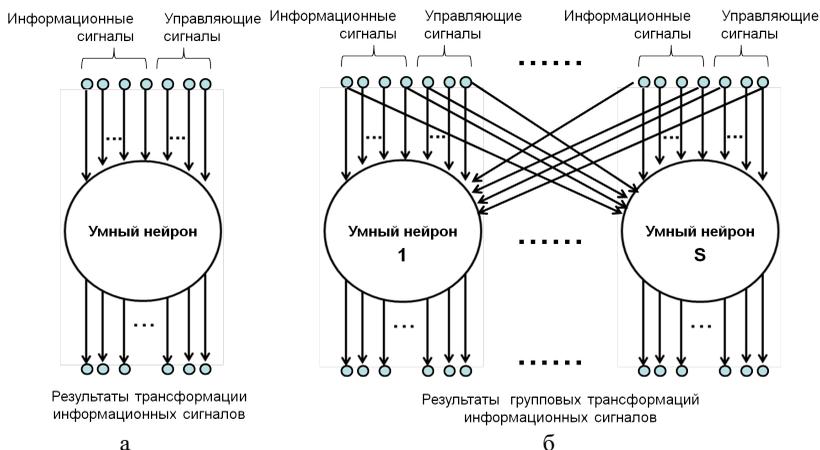


Рис. 1. Представление умных нейронов с: а) индивидуальной и б) групповой трансформацией сигналов

Изменяя только управляющие воздействия, можно одни и те же информационные сигналы преобразовывать в другие сигналы и запоминать результаты обработки. Эти умные нейроны можно рассматривать как полноценные ассоциативные процессоры с внутренней памятью, способные обращаться к ней по содержанию

входных сигналов новым способом. Число вариантов трансформации сигналов у каждого такого процессора может быть свое в зависимости от полученного опыта, результатов обучения. Группа параллельно функционирующих нейронов может реализовывать различные групповые трансформации (рисунок 1(б)), результаты их в дальнейшем могут подвергаться другим преобразованиям, в том числе рекуррентным трансформациям. Полноценное обращение к памяти по содержанию входных сигналов позволяет переходить к осмысленной обработке информации, в том числе оперировать смыслами и модельным временем.

С формальной точки зрения возбуждение  $j$ -го «умного» нейрона на момент времени  $t$  осуществимо, если ток  $i_j(x_o, y_o, t)$ , приложенный к точке его «тела» с координатами  $(x_o, y_o)$  больше или равен пороговому току  $I_{jd}(x_o, y_o, t)$  возбуждения,  $i_j(x_o, y_o, t) \geq I_{jd}(x_o, y_o, t)$ , и время  $t_j(t)$  после предыдущего возбуждения больше или равно времени  $T_r$  невосприимчивости нейрона после срабатывания,  $t_j(t) \geq T_r$ . Значение  $i_j(x_o, y_o, t)$  и координаты  $(x_o, y_o)$  предлагается определять согласно (1):

$$i_j(x_o, y_o, t) = \max_{x, y \in \Omega} \sum_{z=1}^N i_{zj}(x, y, m_{x_{zj}}, m_{y_{zj}}, t), \quad (1)$$

где  $N$  – число воздействующих на  $j$ -й нейрон  $z$ -х нейронов, равное числу входных сигналов;  $i_{zj}(x, y, m_{x_{zj}}, m_{y_{zj}}, t)$  – значение тока от  $z$ -го нейрона на поверхности  $j$ -го нейрона в точке с координатами  $x, y$ ;  $m_{x_{zj}}, m_{y_{zj}}$  координаты точки приложения сигнала от  $z$ -го нейрона к  $j$ -му нейрону;  $\Omega$  – множество возможных точек приложения сигналов к нейрону.

При возбуждении умного нейрона относительно точки  $(x_o, y_o)$  осуществляется свое распределение сгенерированного заряда по его внутренним каналам с учетом их проводимостей и проводимостей синапсов, а также состояний принимающих нейронов. Формируется вектор выходных сигналов. В зависимости от перенесенных по внутренним каналам и синапсам зарядов изменяются их проводимости, запоминаются результаты обработки информации нейроном.

При реализации памяти нейронов на мемристорах проводимости (веса)  $w_{km}(t) = w_{km}(x_o, y_o, t)$  цепей передачи сигналов с  $k$ -го входа памяти на  $m$ -е выходы можно рассчитать как (2):

$$w_{km}(t) = 1 / (R_{ON_{km}}(t) + (R_{OFF_{km}}(t) - R_{ON_{km}}(t)) \cdot \exp(-A \cdot (q_{km}(t - \Delta t) + q_{km}(\Delta t))), \quad (2)$$

где  $R_{ON_{km}}(t)$ ,  $R_{OFF_{km}}(t)$  – наименьшее и наибольшее сопротивления мемристорной  $km$  цепи,  $R_{ON_{km}}(t) = R_{ON} / \psi_{km}(t)$ ,  $R_{OFF_{km}}(t) = R_{OFF} / \psi_{km}(t)$ ;  $A$  – коэффициент размерностью 1/кулон;  $q_{km}(t - \Delta t)$  – заряд, переданный через  $km$  цепь на момент  $t - \Delta t$ ;  $q_{km}(\Delta t)$  – заряд, переданный за интервал времени  $\Delta t$ . Функция  $\psi_{km}(t)$ , от которой зависят  $R_{ON_{km}}(t)$  и  $R_{OFF_{km}}(t)$ , может определяться согласно [11].

Значения переносимых по цепям зарядов вычисляются как (3):

$$q_{km}(t) = \int_{t_0}^t u_k(\tau) \cdot (1 - u_m(\tau) / u_k(\tau)) \cdot w_{km}(\tau) d\tau, \quad (3)$$

$u_k(\tau)$  – напряжение на выходе возбужденного  $k$ -го нейрона в момент времени  $\tau$ ;  $u_m(\tau)$  – текущее напряжение на входе  $m$ -го нейрона;  $w_{km}(\tau)$  – проводимость (вес)  $km$  цепи.

Заметим, что при выделении только одной пространственно соотнесенной точки начала возбуждения нейрона, удовлетворяющей условию (1) и  $i_j(x_o, y_o, t) \geq I_{jd}(x_o, y_o, t)$ , в зависимости от ее координат возможно  $M$  вариантов распределения сигналов на выходе при одном и том же состоянии внутренней памяти. В случаях, когда может выделяться не одна, а до  $K$  локальных точек со своими максимумами токов, способных возбудить нейрон, ситуация иная. Число возможных вариантов распределения сигналов на выходе существенно возрастает и равно (4):

$$C_{\Sigma} = \sum_{k=1}^K \frac{M!}{k!(M-k)!}, \quad (4)$$

где  $M$  – число всех возможных точек возбуждения нейрона;  $!$  – знак факториала;  $k$  – число точек, одновременно удовлетворяющих

условиям возбуждения нейрона,  $k \leq K \leq M$ . Это хорошо согласуется с результатами исследований [36], согласно которым биологический нейрон функционирует как множество независимых пороговых элементов.

Дополнительно отметим следующее. При обработке сигналов сетями умных нейронов также могут реализовываться пространственные сдвиги сигналов при передаче их от слоя к слою. За счет этого исключаются конфликты между возбужденными нейронами и нейронами, находящимися в состояниях невосприимчивости. За счет этих сдвигов нейронные сети могут наделяться логическими структурами. Особенности функционирования таких нейронных сетей рассмотрим в следующем разделе.

**4. Структура искусственного мозга ИНМ и метод многоуровневой обработки информации.** В качестве базовой рекуррентной нейронной сети для построения искусственного мозга ИНМ применима двухслойная сеть с управляемыми элементами и обратными связями, замыкающими контуры со временем задержки единичных импульсов меньше времени невосприимчивости нейронов сети после их возбуждения [11]. Обобщенная схема этой РНС приведена на рисунке 2(а). Эта РНС является своеобразным конструктором, позволяющим формировать различные варианты других структур нейронных сетей. В РНС предусматриваются пространственные сдвиги СЕИ при передаче от слоя к слою. За счет этих сдвигов РНС может наделяться различными одно- и многоуровневыми логическими структурами (линейными, спиральными, петлевыми и другими), а СЕИ продвигаются не только между слоями, но и вдоль них. Один из простых примеров одноуровневой логической структуры РНС приведен на рисунке 2(б). В рассматриваемом случае на вход этой сети подаются СЕИ, несущие информацию как о внешнем мире, так об управляющих воздействиях. Если после обучения сети со структурой на рисунке 2(б) усилить ассоциативный вызов сигналов в направлении ее входа, она будет последовательно вызывать из памяти связанные события.

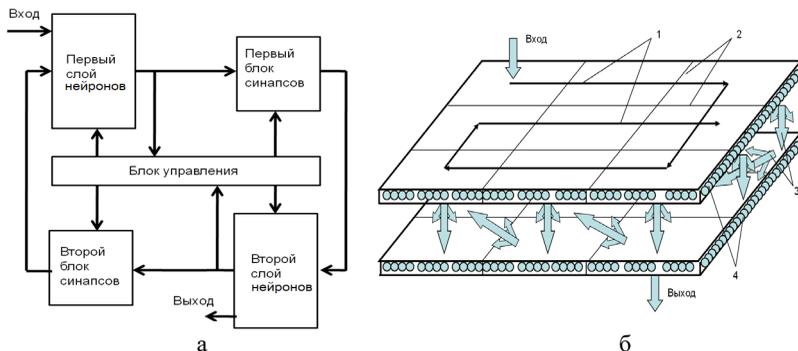


Рис. 2. Рекуррентная нейронная сеть с управляемыми элементами:  
 а) обобщенная схема; б) пример логической структуры (1, 3 – направления продвижения СЕИ вдоль и между слоями; 2 – линии разбивки слоев на логические поля за счет реализуемых пространственных сдвигов СЕИ; 4 – нейроны)

Заметим, что логическую структуру РНС на рисунке 2(б) можно представить также в однослойном варианте (рисунок 3(а)), а с помощью его осуществимо формирование многослойной РНС (искусственного мозга) с многоуровневой обработкой информации (рисунок 3(б)). В соответствии с рисунком 3(б) в РНС выделяются «кортикальные колонки». Они формируются, прежде всего, за счет многоуровневых пространственных сдвигов сигналов вдоль слоев.

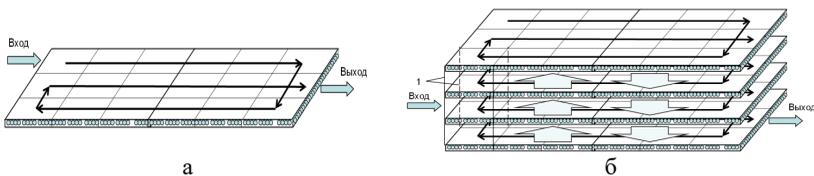


Рис. 3. Варианты логических структур РНС с а) одноуровневой и б) многоуровневой обработкой сигналов: 1 – границы «кортикальных колонок»

С учетом этих сдвигов многоуровневая обработка сигналов возможна согласно методу, основу которого составляет способ [28]:

А.1. Сигнал поступает в РНС, где на каждом  $k + 1$  ( $k = 0, 1, \dots, K-1$ ) уровне обработки пороги возбуждения нейронов и амплитуды генерируемых ими единичных импульсов больше их значения на  $k$ -м уровне.

А.2. При обработке единичных импульсов каждый нейрон определяет пространственно-энергетические характеристики этих импульсов.

А.3. Когда нейрон возбужден, информация об этих характеристиках сохраняется. Она учитывается при распределении нейроном сгенерированного импульса по синапсам, соединяющим его с принимающими нейронами.

А.4. В РНС пересекающиеся выборки из последовательности СЕИ  $k$ -го уровня преобразуются в СЕИ  $k + 1$  уровня. Длина этих выборок на каждом  $k$ -м уровне выбирается больше, чем реализованные пространственные сдвиги СЕИ при передаче с уровня на уровень.

А.5. Каждая СЕИ  $k + 1$  уровня связывается с соответствующей выборкой  $k$ -го уровня.

А.6. Дополнительно сформированные СЕИ на уровне  $k+1$  обрабатываются аналогично обработке СЕИ на  $k$ -м уровне и связываются в пространстве и времени посредством запоминания связей на элементах сети.

А.7. Эти связи используются для извлечения сигналов из памяти РНС.

А.8. Обработанные в нейронной сети СЕИ уровня  $k+1$  преобразуются обратно в соответствующие объединенные выборки СЕИ  $k$ -го уровня.

А.9. Эти выборки используются для формирования результатов обработки сигналов в РНС.

Согласно А.1, за счет повышенных порогов возбуждения нейронов на уровне  $k+1$  по сравнению с  $k$ -м уровнем физически можно осуществлять пространственно-временное компрессионное кодирование выборок из последовательностей СЕИ. СЕИ, сформированные на уровне  $k+1$ , продвигаются по сети синхронно с СЕИ  $k$ -го уровня, между ними формируются и сохраняются обратные связи. Однако для того, чтобы СЕИ  $k+1$  уровней смогли восстановить соответствующие выборки на  $k$ -м уровне, необходимо выполнение следующего условия. Требуется наличие нейронов уровня  $k+1$  с более высокими амплитудами генерируемых выходных импульсов, чем у нейронов уровня  $k$ . Это вполне согласуется с наличием в биологическом мозгу слоев нейронов со своими характеристиками порогов возбуждения и формируемых импульсов.

В целом поданные в РНС последовательные СЕИ представляются в виде других последовательных СЕИ в соответствии с предварительно заданными правилами связывания и трансформации с учетом обратных

результатов. Результаты связывания и трансформации запоминаются на элементах сети. При этом в РНС преобразование одних последовательных СЕИ в другие осуществляется путем трансформации их нейронами в зависимости от результатов связывания входных единичных образов и состояний нейронов. В качестве результатов обработки используются управляющие СЕИ на выходе РНС после обратного преобразования в соответствующие им исходные сигналы. Предлагаемый метод можно пояснить схемой обработки сигналов, приведенной на рисунке 4. Согласно этой схеме за счет реализуемых в РНС пространственных сдвигов сигналов при передаче от слоя к слою СЕИ на всех уровнях продвигаются синхронно слева направо. При восходящем преобразовании СЕИ формируются также связи, обеспечивающие возможность обратной трансформации. Как в РНС с обычными импульсными нейронами, так и в РНС с умными нейронами, учитывается приоритетность сильных связей, сохраняющих логическую структуру сетей. Однако при использовании умных нейронов их состояния на каждый момент времени характеризуются не скалярами, а векторами. Кроме этого в РНС на умных нейронах помимо расширенных возможностей по трансформации и запоминанию сигналов, присутствуют возможности по управлению переходами с одного уровня на другой в зависимости от содержания обрабатываемой информации.

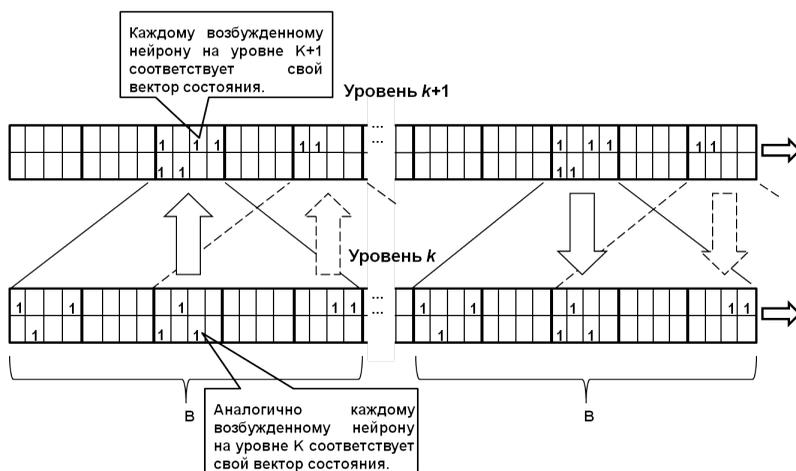


Рис. 4. Схема преобразования в РНС пересекающихся выборок из последовательности СЕИ  $k$ -го уровня в последовательность СЕИ  $k+1$  уровня и наоборот: В – длина сворачиваемых (разворачиваемых) выборок; 1 – признак возбужденного умного нейрона

Присутствует также возможность временного отрыва высокоуровневой обработки сигналов от обработки на нижних уровнях. Это позволяет на высоких уровнях решать творческие задачи без прерывания обработки на нижних уровнях.

**5. Механизмы мышления ИНМ.** Поясним возможные механизмы мышления ИНМ на примере ее обобщенной структуры, показанной на рисунке 5, где  $D1/\mathcal{E}1, \dots, Dn/\mathcal{E}n$  – управляемые датчики (Д) информации, совмещенные с эффекторами (Э);  $1, \dots, 2n-1$  – информационные сигналы с датчиков;  $2, \dots, 2n$  – сигналы о состоянии датчиков, совмещенных с эффекторами.

Согласно рисунку 5 предусматривается преобразование различных видов сигналов внешнего мира в информационные последовательные совокупности единичных импульсов (СЕИ) с учетом управляющих СЕИ с выхода сети. Такое преобразование реализуется в блоках Д/Э. Также в этих блоках сигналы об их состоянии перед подачей в РНС преобразуются в дополнительные последовательные СЕИ. В зависимости от управляющих воздействий на блоки Д/Э могут восприниматься различные информационные сигналы.

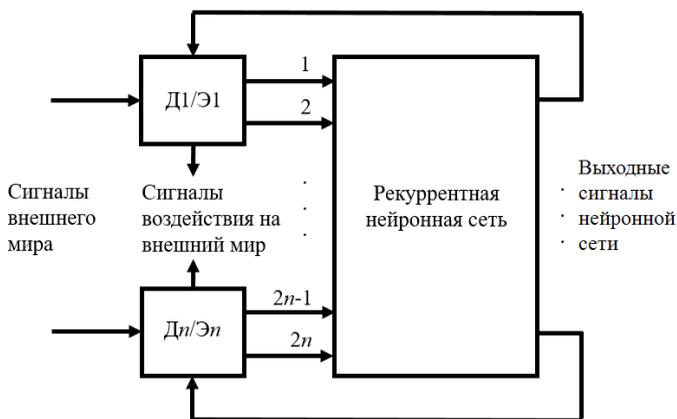


Рис. 5. Обобщенная структура интеллектуальной нейросетевой машины

Сигналы внешнего мира относительно ИНМ могут сдвигаться в пространстве и во времени, поворачиваться, изменяться по частотам и фазам, другим параметрам. От управляющих воздействий может изменяться положение и характеристики самих датчиков, совмещенных с эффекторами, а также самой ИНМ. Управление датчиками информации может быть осуществимо сначала путем

слежения за сигналами внешнего мира, их поведением, а в дальнейшем, согласно правилам сканирования пространства, в интересах преследуемых целей.

Для дальнейших пояснений принципов мышления ИНМ рассмотрим с учетом рисунков 2(б) и 5 одноуровневую обработку пары последовательностей СЕИ на канальном уровне РНС (рисунок 6).

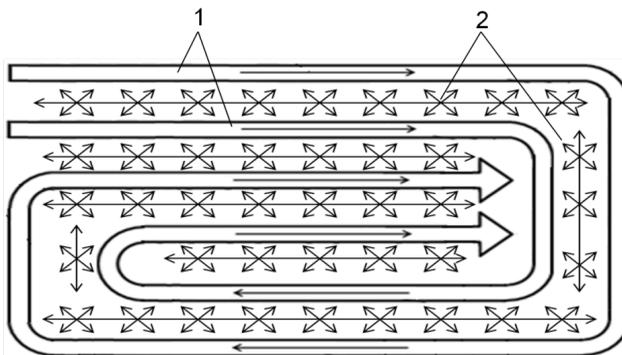


Рис. 6. Вариант логической структуры РНС на канальном уровне: 1 – каналы продвижения по сети СЕИ; 2 – управляемые ассоциативные взаимодействия СЕИ и отдельных импульсов

Обозначим поступающие в РНС информационные последовательные СЕИ как  $SSP^a$ , а последовательные СЕИ, отражающие результаты управляющих воздействий на датчики, совмещенные с эффекторами, в виде  $SSP^c$ . Связывание этих двух последовательностей СЕИ в РНС можно представить в виде (5):

$$\left\{ \begin{array}{l} SSP^c \\ \otimes \\ SSP^a \end{array} \right\} = \left\{ \begin{array}{l} SSP_1^c, SSP_2^c, \dots, SSP_n^c \\ \otimes \otimes \otimes \otimes \otimes \\ SSP_1^a, SSP_2^a, \dots, SSP_n^a \end{array} \right\}, \quad (5)$$

где  $\otimes$  – знаки связывания обрабатываемых совокупностей единичных импульсов. При простом обучении  $SSP_1^a, SSP_2^a, \dots, SSP_n^a$  и  $SSP_1^c, SSP_2^c, \dots, SSP_n^c$  связываются друг с другом при продвижении по

сети. Совокупностям  $SSP_1^c, SSP_2^c, \dots, SSP_n^c$  единичных импульсов в сети ставятся в соответствие информационные СЕИ, обозначенные как  $SSP_1^a, SSP_2^a, \dots, SSP_n^a$ . Заметим, что на начальном этапе формирование управляющих воздействий на блоки Д/Э нейросетевой машины осуществляется из информационных СЕИ. Нейросетевая машина учится отслеживать изменения сигналов внешнего мира. Например, при смещении или повороте наблюдаемого машиной объекта для отслеживания его действий изменяются положение датчиков, с помощью которых объект воспринимается. Устанавливается однозначное соответствие между изменениями сигналов внешнего мира и воздействиями на блоки Д/Э. После такого обучения, если кратковременно прервать поступление в РНС информационных СЕИ, но оставить воздействия  $SSP_1^c, SSP_2^c, \dots, SSP_i^c, SSP_{i+1}^c, SSP_{i+2}^c, \dots, SSP_n^c$ , то состояние сети до момента ассоциативного вызова из ассоциативной памяти недостающих информационных СЕИ представимо в виде (6):

$$\left\{ \begin{array}{l} SSP_1^c, SSP_2^c, \dots, SSP_i^c, SSP_{i+1}^c, SSP_{i+2}^c, \dots, SSP_n^c \\ \otimes \\ \emptyset, \emptyset, \dots, \emptyset, SSP_{i+1}^a, SSP_{i+2}^a, \dots, SSP_n^a \end{array} \right\}, \quad (6)$$

$\emptyset$  – знаки пустых множеств (пустых СЕИ).

После ассоциативного вызова сигналов из памяти РНС с трансформацией из одной формы в другую, в зависимости от управляющих  $SSP_1^c, SSP_2^c, \dots, SSP_i^c$ , получаем (7):

$$\left\{ \begin{array}{l} SSP_1^c, SSP_2^c, \dots, SSP_i^c, SSP_{i+1}^c, SSP_{i+2}^c, \dots, SSP_n^c \\ \otimes \\ SSP_1^b, SSP_2^b, \dots, SSP_i^b, SSP_{i+1}^a, SSP_{i+2}^a, \dots, SSP_n^a \end{array} \right\}, \quad (7)$$

где  $SSP_1^b, SSP_2^b, \dots, SSP_i^b$  – результат обработки. Формируя разные виды управляющих  $SSP_1^c, SSP_2^c, \dots, SSP_i^c$ , при обрабатываемых  $SSP_{i+1}^a, SSP_{i+2}^a, \dots, SSP_n^a$ , можно не только вызывать из памяти связанные сигналы, но и осуществлять интеллектуальное

оперирование ими в памяти, эквивалентное обработке в управляемых датчиках, совмещенных с эффекторами, и реальным действиям. Следует отметить, что управляющие воздействия могут формироваться также с преследованием других целей обработки информации в РНС. На самом низком уровне в качестве таких целей может выступать минимизация расходов энергии на поддержание устойчивого функционирования сети при внешних воздействиях.

Для случая, когда все входы РНС закрыты, но в сети обрабатываются поступившие сигналы, имеет место состояние (8):

$$\left\{ \begin{array}{l} \emptyset, \emptyset, \dots, \emptyset, \quad SSP_{i+1}^c, SSP_{i+2}^c, \dots, SSP_n^c \\ \otimes \\ \emptyset, \emptyset, \dots, \emptyset, \quad SSP_{i+1}^a, SSP_{i+2}^a, \dots, SSP_n^a \end{array} \right\}. \quad (8)$$

Относительно этого состояния в обученной сети могут вызываться из памяти как информационные, так и СЕИ, отражающие результаты управляющих воздействий на датчики, совмещенные с эффекторами. Они могут продолжать последовательности  $SSP_{i+1}^c, SSP_{i+2}^c, \dots, SSP_n^c$  и  $SSP_{i+1}^a, SSP_{i+2}^a, \dots, SSP_n^a$ . Аналогичным образом в памяти РНС осуществимо воспроизведение предшествующих сигналов с возможностью интеллектуального оперирования ими в зависимости от управляющих воздействий.

Рассмотрим теперь механизмы мышления искусственного мозга, связанные с операциями над временем. Ответим на вопросы, как можно ускорять, или замедлять время обработки событий в искусственном мозгу при условии, что передача сигналов внутри РНС осуществляется с одной и той же скоростью.

Чтобы ответить на эти вопросы вернемся к совместной обработке сигналов внешнего мира от датчиков и о состояниях последних, отражающих результаты управляющих воздействий. При отслеживании сигналов внешнего мира, изменяющихся с различной скоростью, можно получать соответствующие им результаты управляющих воздействий. При последующем вызове ими с различной скоростью информационных сигналов (в виде СЕИ) из ассоциативной памяти РНС осуществимо ускорение или замедление обработки событий. Изменение скорости такого вызова эквивалентно сжатию или растяжению последовательностей управляющих СЕИ. Формируя управляющие воздействия в сжатом виде, и осуществляя ими

ассоциативный вызов сигналов из памяти в направлении входа или выхода сети, можно оперативно обращаться к будущему или прошлому.

Учитывая это, можно утверждать, что предлагаемая интеллектуальная нейросетевая машина обладает смоделированными функциями мышления. Она способна реализовывать интеллектуальные операции в своей памяти над различными СЕИ, образами и понятиями, а также над временем.

Для того чтобы эта машина могла успешно проявлять свои интеллектуальные возможности необходимо следующее. В составе ИНМ должны присутствовать подсистемы управления датчиками, совмещенными с эффекторами, отслеживающие (повторяющие) поведение воспринимаемых сигналов (их источников). Такая машина должна связывать на элементах РНС сигналы о внешнем мире с выходах блоков Д/Э с сигналами об состоянии этих блоков, несущими информацию о реализованных преобразованиях над сигналами внешнего мира. В этом случае между сигналами внешнего мира и реализуемыми над ними преобразованиями устанавливается однозначное соответствие. РНС должна обеспечить полноценное связывание и трансформирование обрабатываемых сигналов без потерь информации о соответствующих исходных воздействиях. Это достижимо построением РНС на основе рассмотренных выше новых моделей умных импульсных нейронов с обширной внутренней памятью и новым способом ассоциативного обращения к памяти.

**6. Схемные решения и результаты моделирования.** Для подтверждения результативности и преимуществ предложенных моделей и методов интеллектуальной нейросетевой обработки информации проводилось их моделирование.

**6.1. Схема и результаты моделирования умного нейрона.** Учитывая предложенные теоретические решения, схему функционирования умного нейрона можно представить в виде рисунка 7. Согласно рисунку 7 эта схема состоит из пяти блоков. Для проверки условия (1) в первом блоке в качестве моделируемой поверхности нейрона можно использовать сетку из обычных резисторов (рисунок 8).

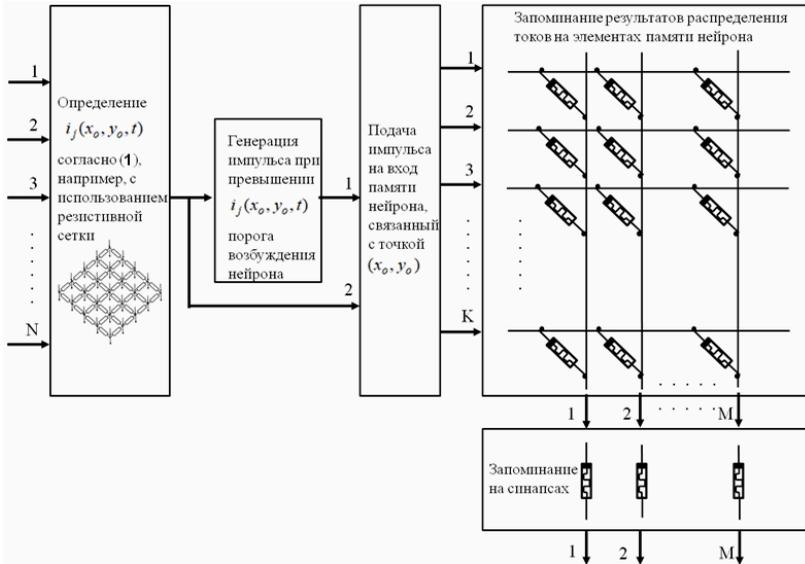


Рис. 7. Схема функционирования «умного» нейрона с реализацией внутренней памяти на мемристорах

На входы этой сетки поступают сигналы от  $z$ -х нейронов. В качестве интересующих выходов сетки используют один или несколько выходов, удовлетворяющий условию (1).

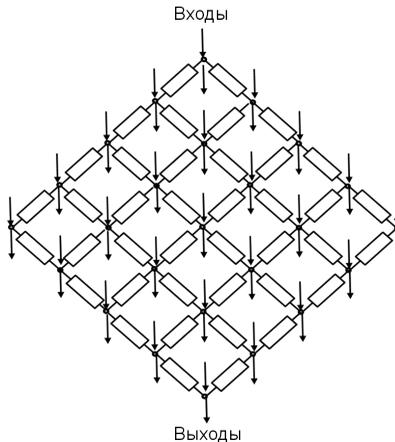


Рис. 8. Резистивная сетка, моделирующая принимающую поверхность умного нейрона

Для подтверждения справедливости сформулированных научных положений проводилось компьютерное моделирование. Сначала проверялась возможность умных нейронов реализовывать различные функции с использованием предложенного нового метода ассоциативного обращения к их внутренней памяти. В интересах этого распределение токов от отдельных воздействий на поверхности умных нейронов аппроксимировали двумерным нормальным законом. Рассматривался случай поступления на отдельно взятый нейрон трех воздействий, соотнесенных со своими точками приложения к поверхности нейрона. Результаты суммарных распределений токов по поверхности нейрона в соответствии с (1) отражены на рисунке 9.

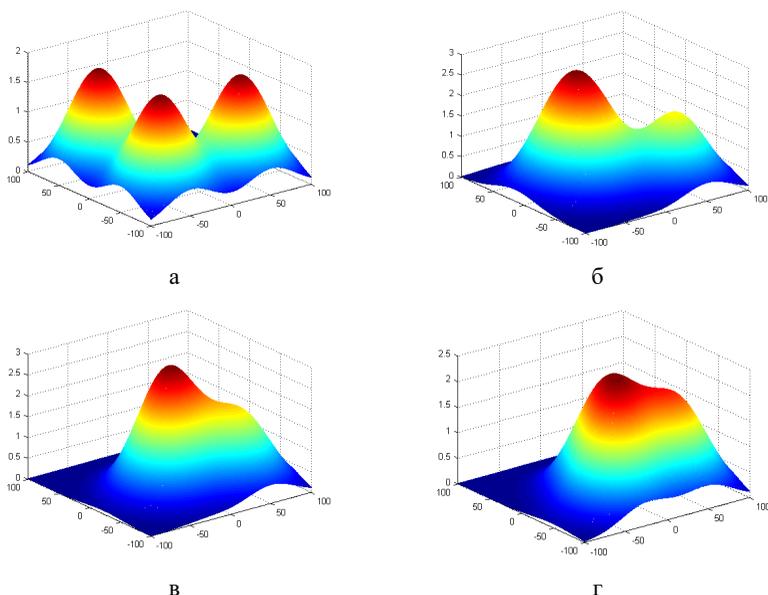


Рис. 9. Примеры распределения суммарных токов на поверхности принимающего нейрона при возбуждающих воздействиях, приложенных к трем разным точкам: а, б, в, г) распределения при различных координатах этих точек приложения

Согласно рисунку 9 получаемые распределения существенно зависят, как от точек приложения воздействий на нейрон, так и уровней этих воздействий. В случае большого разноса точек приложения воздействий (рисунок 9(а)) в явном виде просматриваются три максимума. При сближении этих точек (рисунок 9(б, в, г)) картина

меняется. Однако во всех случаях согласно (1) можно выделить точку с наибольшим максимумом со своими координатами. Причем координаты этого максимума могут не совпадать с координатами ни с одной из исходных точек. При превышении этого максимума порога возбуждения нейрона, он срабатывает. Относительно точки срабатывания формируется и передается к выходам нейрона по своим внутренним путям единичный заряд. На выходах нейрона получаем конкретное распределение токов с учетом пространственно-энергетических характеристик воздействий на нейрон и внутренней памяти.

Поясним это на простом примере. Допустим, что воздействия на нейрон в двух точках отражает поступление на него сигналов, соотнесенных, например, с числами  $A$  и  $B$ . Третье воздействие  $F$  со своей точкой приращения несет информацию о реализуемой с этими числами операции. Тогда при одних и тех же  $A$  и  $B$  в зависимости от управления  $F$  на выходах умного нейрона будут формироваться разные распределения токов. Этим распределениям могут быть поставлены в соответствие конкретные результаты выполняемых операций. В частности если  $A = 8$ , а  $B = 2$  и  $F = A + B$ , то получаемое распределение токов на выходе нейрона соответствует числу 10.

В случае  $F = \log_B A$  имеем распределение, соотнесенное с числом 3. Аналогично могут выполняться не только арифметические, но и другие различные по содержанию операции над сигналами. Таким образом, обученный нейрон в зависимости от управляющих воздействий может реализовывать различные трансформации над обрабатываемыми информационными сигналами, а группы параллельных нейронов – различные групповые трансформации, в том числе связанные с поворотами сигналов, их сдвигами, сжатием, расширением и другими действиями. Последовательности групп параллельных нейронов, в том числе с обратными связями, могут реализовывать последовательные трансформации над сигналами в соответствии управляющими воздействиями. Это, в некоторой мере, эквивалентно преобразованиям сигналов по программе.

**6.2. Результаты ассоциативного вызова сигналов из памяти РНС с умными и традиционными импульсными нейронами.** Сравнивались возможности рекуррентных импульсных нейронных сетей с умными нейронами и традиционными их моделями по запоминанию сигналов и ассоциативному вызову их из памяти. Рассматривался случай обработки в РНС сигналов с высокой корреляцией. Особенность такой обработки при использовании

традиционных нейронов в том, что в РНС формируется много альтернативных связей, мешающих как запоминанию, так и корректному извлечению сигналов из памяти. В качестве РНС выступала двухслойная сеть со структурой, подобной рисунку 2(б). Число нейронов в каждом слое сети равнялось 1890 единиц. Слои РНС за счет пространственных сдвигов СЕИ при передаче от слоя к слою разбивались на логические поля размером  $6 \times 7$  нейронов. Используемые умные нейроны, в отличие от традиционных импульсных моделей, были способны различать по четыре типа пространственно-энергетических воздействий (условно имели по четыре пространственно разнесенных точки приложения входных воздействий). Обе РНС одинаково обучались. На входы РНС последовательно подавались одинаковые последовательности СЕИ в виде изображений букв. Буквы представлялись совокупностями единичных импульсов (образов) с количеством от 8 до 12 единиц. Изображения различных букв имели пересечения по единичным образам. Обучение сводилось к прогону по сети этих последовательностей и запоминание их. Проводимости (веса) цепей передачи сигналов в РНС определялись согласно (2), (3). После этого оценивалась эффективность ассоциативного вызова сигналов из памяти РНС с умными и традиционными нейронами после одного и того же числа тактов работы. Полученные результаты сведены в таблицы 1, 2.

Таблица 1. Эффективность ассоциативного вызова сигналов из памяти РНС с одинаковым уровнем усиления вызывающих сигналов

Тип РНС и приращения показателей	Эксперимент 1		Эксперимент 2		Эксперимент 3		Эксперимент 4	
	N/S	S/(S+N)	N/S	S/(S+N)	N/S	S/(S+N)	N/S	S/(S+N)
РНС с традиционными импульсными нейронами	0.258	0.795	0.267	0.789	0.388	0.720	0.519	0.659
РНС с умными нейронами	0.217	0.821	0.143	0.875	0.167	0.857	0.171	0.854
Приращения (I)	-0.041	0.026	-0.124	0.086	-0.221	0.137	-0.348	0.195
I%	-15.891	3.270	-46.442	10.900	-56.959	19.028	-67.052	29.590

Таблица 2. Эффективность ассоциативного вызова сигналов из памяти РНС с разными уровнями усиления вызывающих сигналов

Тип РНС и приращения показателей	Эксперимент 5		Эксперимент 6		Эксперимент 7		Эксперимент 8	
	N/S	S/(S+N)	N/S	S/(S+N)	N/S	S/(S+N)	N/S	S/(S+N)
РНС с традиционными импульсными нейронами	0.310	0.760	0.310	0.760	0.419	0.705	0.419	0.705
РНС с умными нейронами	0.190	0.840	0.107	0.903	0.190	0.840	0.107	0.903
Приращения (I)	-0.120	0.080	-0.203	0.143	-0.229	0.135	-0.312	0.198
%	-38.709	10.526	-65.484	18.816	-54.654	19.149	-74.463	28.085

Эксперименты различались количеством тактов работы РНС, на которых анализировались состояния нейронов. В таблицах приняты обозначения:  $N$ ,  $S$  – число ложных и истинных импульсов (единичных образов), вызываемых из ассоциативной памяти РНС. Из анализа таблиц 1, 2 видно, что РНС с умными нейронами существенно превосходят сети с традиционными импульсными нейронами. При этом усиление ассоциативного вызова сигналов из памяти РНС позволяет получать более высокие оценки.

**6.3. Результаты восстановления обрабатываемых сигналов РНС с умными нейронами, представляемыми группами обычных импульсных моделей.** Проводились также другие эксперименты по ассоциативной обработке сигналов в РНС, подтверждающие реализуемость предложенных решений с умными нейронами. Рассматривался случай, когда умные нейроны представлялись в слоях РНС группами обычных импульсных нейронов. В РНС с числом нейронов в слоях по 3150 единиц каждый умный нейрон моделировался 42 обычными импульсными нейронами. Умный нейрон был эквивалентен одному логическому полю ( $6 \times 7$  нейронов), на которые разбивались слои РНС. За счет этого каждый слой РНС сводился к 75 умным нейронам. В каждой группе обычных импульсных нейронов, моделирующих умный нейрон, одновременно мог возбуждаться только один нейрон. Такая схема также предусматривала соответствующее кодирование входных сигналов. Для рассматриваемого случая примеры состояния первого слоя РНС при решении задачи восстановления СЕИ, отсутствующих в середине обрабатываемой последовательности, приведены на рисунке 10.

Заметим, что результат восстановления на рисунке 10(б) полностью соответствует эталонному варианту (рисунок 10(в)), за исключением разницы в пространственных сдвигах СЕИ при продвижении вдоль слоев.

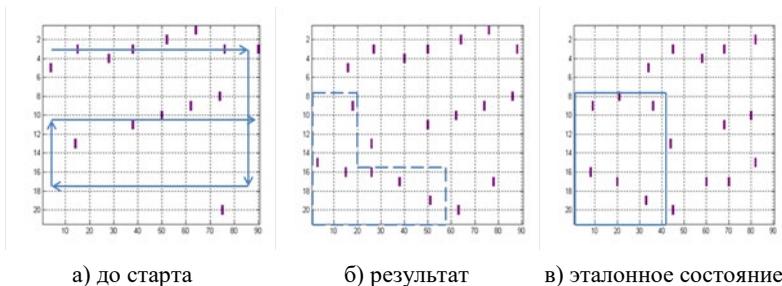


Рис. 10. Примеры состояния первого слоя РНС при восстановлении сигналов

Стрелки на рисунке 10(а) отражают схему продвижения СЕИ вдоль слоев. В данном примере за каждым пространственно соотнесенным единичным импульсом закреплялось свое событие. В рассматриваемом примере такими событиями выступали закодированные последовательные слова обрабатываемых текстов.

**6.4. Результаты моделирования сложных преобразований сигналов в памяти РНС с умными нейронами.** Проводились эксперименты по реализации интеллектуальных операций над сигналами в памяти РНС. Использовалась двухслойная РНС с умными нейронами с той же схемой продвижения СЕИ вдоль слоев, что и на рисунке 2(б). Число умных нейронов в каждом слое сети составляло 3150 единиц. За счет пространственных сдвигов СЕИ каждый слой сети разделялся на 75 логических полей с размерами 6x7 нейронов. Эти поля дополнительно делились на два подполя по 3x3 нейронов и подполя 6x4 нейронов. В эксперименте задействовались только подполя 3x3 умных нейронов, различающих по четыре типа пространственно-энергетических воздействий (имеющих по четыре пространственно разнесенных точки приложения входных сигналов). Во время обучения РНС закодированные информационные сигналы и соответствующие им управляющие воздействия подавались на свои входные подполя. При прохождении по сети информационных сигналов и управляющих воздействий между ними устанавливались связи. После обучения в нейронную сеть вводились управляющие воздействия, которые успешно вызывали связанные с ними информационные сигналы. Заметим, что каждому управляющему воздействию соответствует своя операция действий над обрабатываемыми сигналами. Пример такого вызова информации приведен на рисунках 11 и 12.

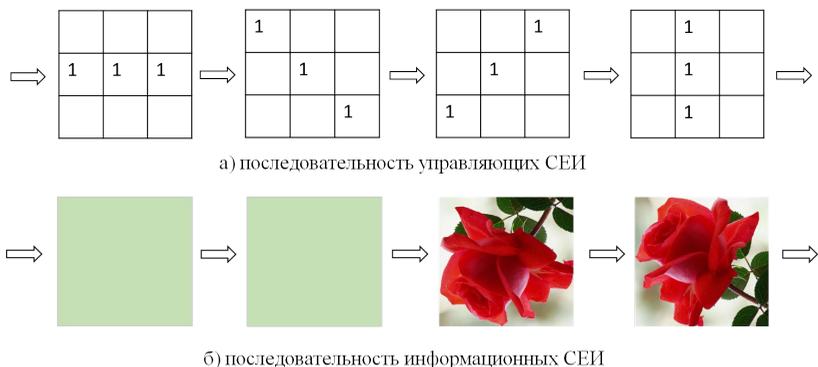


Рис. 11. Примеры последовательностей управляющих и информационных СЕИ, обрабатываемых в РНС до начала вызова информации из памяти

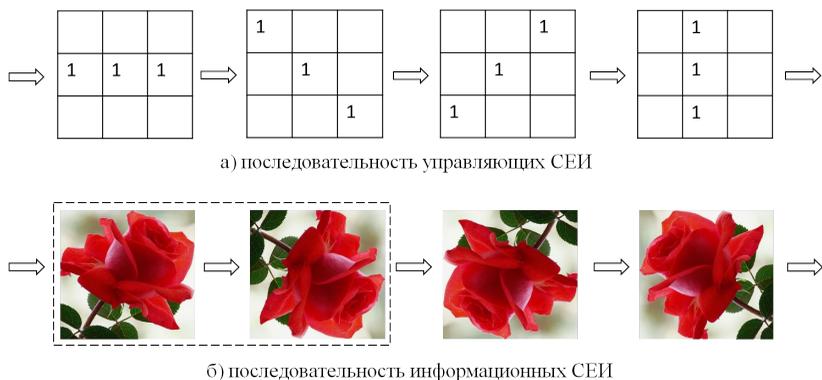


Рис. 12. Результат ассоциативного вызова информации из памяти РНС в соответствии с управляющими воздействиями

Для наглядности отражения результатов на рисунках 11 и 12 информационные СЕИ заменены на соответствующие им исходные образы. Это могут быть и другие образы, например, изображения букв, лиц, птиц. Промежуточные результаты работы сети при передаче СЕИ от слоя к слою здесь не показаны. Из рисунков 11, 12 видно, что каждой управляющей СЕИ соответствует своя операция действий над обрабатываемыми образами. В соответствии с рисунком 11(б) в РНС введены два образа розы (справа), а за ними – два пустых образа, которые формируются при прерывании воздействия информационных сигналов на сеть. Однако на интервале такого прерывания в сеть поступили две управляющих СЕИ (слева на рисунке 11(а)), которые ассоциативно вместе с присутствующими информационными СЕИ

вызывают из памяти РНС на рисунке 12(б) (слева) новые образы розы. Таким образом, формируя различные управляющие последовательности СЕИ внутри РНС, можно оперировать информационными образами, которые в ней активны, и вызывать из памяти новые связанные с ними сигналы. В этом примере рассмотрена реализация внутри сети операций вращения. Также внутри РНС согласно предложенным в разделе 5 механизмам мышления ИНМ могут реализовываться пространственные сдвиги и другие операции. Заметим, что человек, наблюдая за любым материальным объектом, может мысленно его переместить, повернуть, наклонить. Это также может делать ИНМ.

Для обеспечения устойчивого извлечения сигналов из памяти РНС производилось усиление вызывающих сигналов в направлении входа сети с учетом числа возбужденных нейронов. Этот случай вполне адекватно отражает возможности различных интеллектуальных операций над сигналами, если установлены соответствующие ассоциативные связи при обработке сигналов внешнего мира согласно схеме на рисунке 5.

**7. Заключение.** Предложены новые принципы функционирования ИНМ с функциями мышления. Согласно этим принципам ИНМ, чтобы научиться реализовывать интеллектуальные операции с сигналами, образами и понятиями в искусственном мозгу, должна уметь подражать событиям во внешнем мире. Кроме этого в искусственном мозгу необходимо иметь механизмы управления его деятельностью, обусловленные некоторыми мотивами. Искусственный мозг ИНМ должен обладать широкими возможностями по связыванию, трансформированию и запоминанию сигналов. Внутренняя память умных нейронов искусственного мозга должна превышать память на синапсах.

Разработана модель умного импульсного нейрона, как базового элемента перспективных ИНМ, с существенно расширенными функциями по интеллектуальной обработке информации. Эту модель можно рассматривать как полноценный ассоциативный процессор с внутренней памятью с новым методом обращения к ней по содержанию обрабатываемой информации.

Предложено формировать искусственный мозг ИНМ на основе многоуровневых РНС, наделенных логическими структурами и использующих в качестве базовых элементов умные импульсные нейроны.

Раскрыт метод, позволяющий осуществлять многоуровневую ассоциативную обработку информации в РНС с возможностью

восстановления результатов обработки в соответствующей им исходной форме.

Выявлены механизмы, позволяющие наделять ИНМ функциями мышления. В искусственном мозгу должны связываться и запоминаться сигналы внешнего мира и воздействия, отражающие операции над этими сигналами. В этом случае ИНМ с использованием управляющих сигналов может ассоциативно вызывать из памяти информационные сигналы с их различными трансформациями (преобразованиями). Также информационными сигналами, в зависимости от их содержания, можно вызывать из ассоциативной памяти ИНМ соответствующие им управляющие воздействия на внешний мир с различной скоростью.

Результаты компьютерного моделирования подтверждают результативность и преимущества предложенных решений по сравнению с известными подходами.

Предложенные решения позволяют устранить разрыв между интеллектуальной обработкой физических сигналов и символических представлений. Эти решения могут быть востребованы при создании перспективных программно и аппаратно реализуемых ассоциативных процессоров, мыслящих нейросетевых машин, умных роботов для решения широкого спектра трудно формализуемых творческих прикладных задач.

### Литература

1. Тьюринг А. Может ли машина мыслить? С приложением статьи Дж. фон Неймана «Общая и логическая теория автоматов». Перевод с английского Ю.А. Данилова. М.: Физ.-Мат. Лит., 1960. 112 С.
2. Мышление – Большой энциклопедический словарь. URL: <https://gufo.me/dict/bes/МЫШЛЕНИЕ> (дата доступа 05.04.2024).
3. Velankar M.R., Mahalle P.N., Shinde G.R. Machine Thinking: New Paradigm Shift. In: Cognitive Computing for Machine Thinking. Innovations in Sustainable Technologies and Computing. 2024. pp. 43–53.
4. Malsburg C. Toward understanding the neural code of the brain. Biological Cybernetics. 2021. vol. 115. no. 5. pp. 439–449.
5. Yamakawa H. The whole brain architecture approach: accelerating the development of artificial general intelligence by referring to the brain. Neural Networks. 2021. vol. 144. pp. 478–495.
6. Haykin S. Neural Networks and Learning Machines, third ed., Prentice Hall, New York. 2008. URL: <http://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf> (дата доступа 24.04.2024).
7. Kotseruba I., Tsotsos J. 40 years of cognitive architectures: core cognitive abilities and practical applications. Artificial Intelligence Review. 2020. vol. 53. no. 1. pp. 17–94.
8. Dormehl L. Thinking machine: The Quest for Artificial Intelligence – and Where It's Taking Us Next. Penguin, 2017. 209 p.

9. Takano S. *Thinking Machines. Machine Learning and Its Hardware Implementation.* Academic Press, 2021. 306 p.
10. Hawkins J., Blakeslee S. *On intelligence.* Brown Walker, 2006. 174 p.
11. Osipov V., Osipova M. Space-time signal binding in recurrent neural networks with controlled elements. *Neurocomputing.* 2018. vol. 308. pp. 194–204.
12. Hawkins J., Ahmad S. Hierarchical temporal memory including HTM cortical learning algorithms. Hosted at Numenta.org. 2011. 68 p.
13. Spoerer C.J., McClure, P., Kriegeskorte, N., 2017. Recurrent convolutional neural networks: a better model of biological object recognition. *Frontiers in psychology.* 2017. vol. 8. DOI: 10.3389/fpsyg.2017.01551.
14. Patrick M., Adekoya A., Mighty A., Edward B. Capsule networks – a survey. *Journal of King Saud University – Computer and Information Sciences.* 2022. vol. 34(1). pp. 1295–1310.
15. Yang G., Ding F. Associative memory optimized method on deep neural networks for image classification. *Information Sciences.* 2020. vol. 533. pp. 108–119.
16. Yang J., Zhang L., Chen C., Li Y., Li R., Wang G., Jiang S., Zeng Z. A hierarchical deep convolutional neural network and gated recurrent unit framework for structural damage detection. *Information Sciences.* 2020. vol. 540. pp. 117–130.
17. Ma T., Lv S., Huang L., Hu S. HiAM: A hierarchical attention based model for knowledge graph multi-hop reasoning. *Neural Networks.* 2021. vol. 143. pp. 261–270.
18. Grossberg S. Adaptive resonance theory: how a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks.* 2013. vol. 37. pp. 1–47.
19. Khawaja S., Lee S.L. Hybrid and hierarchical fusion networks: a deep cross-modal learning architecture for action recognition. *Neural Computing and Applications.* 2020. vol. 32. no. 14. pp. 10423–10434.
20. Saha S., Gan Z., Cheng L., Gao J., Kafka O., Xie X., Li H., Tajdari M., Kim H., Liu W. Hierarchical deep learning neural network (HiDeNN): an artificial intelligence (AI) framework for computational science and engineering. *Computer Methods in Applied Mechanics and Engineering.* 2021. vol. 373. DOI: 10.1016/j.cma.2020.113452.
21. Yang M., Chen L., Lyu Z., Liu J., Shen Y., Wu Q. Hierarchical fusion of common sense knowledge and classifier decisions for answer selection in community question answering. *Neural Networks.* 2020. vol. 132. pp. 53–65.
22. Wolfgram P., Wolff C., Lucke J., Malsburg C. A recurrent dynamic model for correspondence-based face recognition. *Journal of Vision.* 2008, vol. 8(7). no. 34. pp. 1–18. DOI: 10.1167/8.7.34.
23. Han Y., Huang G., Song S., Yang L., Wang H., Wang Y. Dynamic neural networks: a survey. *arXiv:2102.04906v4.* 2021. pp. 1–20.
24. Osipov V., Nikiforov V., Zhukova N., Miloserdov D. Urban traffic flows forecasting by recurrent neural networks with spiral structures of layers. *Neural Computing and Applications.* 2020. vol. 32. no. 18. pp. 14885–14897.
25. Osipov V., Kuleshov S., Zaytseva A., Levonevskiy D., Miloserdov D. Neural network forecasting of news feeds. *Expert systems with applications.* 2021. vol. 169. DOI: 10.1016/j.eswa.2020.114521.
26. Osipov V., Kuleshov S., Miloserdov D., Zaytseva A., Aksenov A. Recurrent Neural Networks with Continuous Learning in Problems of News Streams Multifunctional Processing. *Informatics and Automation.* 2022. vol. 21. no. 6. pp. 1145–1168.
27. Osipov V., Osipova M. Method and device of intellectual processing of information in neural network, Patent RU2413304. 2011.
28. Osipov V. Method for intelligent multi-level information processing in neural network, Patent RU2737227. 2020.

29. He J., Yang H., He L., Zhao L. Neural networks based on vectorized neurons. *Neurocomputing*. 2021. vol. 465. pp. 63–70.
30. Deng C., Litany O., Duan Y., Poulenard A., Tagliasacchi A., Guibas L. Vector neurons: a general framework for SO(3)-Equivariant networks. *arXiv:2104.12229v1*. 2021. pp. 1–12.
31. Kryzhanovsky B., Litinskii L., Mikaelian A. Vector-Neuron Model of Associative Memory. *IEEE International Joint Conference on Neural Networks*. 2004. vol. 2. pp. 909–914.
32. Tuszyński J.A., Friesen D.E., Freedman H., Sbitnev V.I., Kim H., Santelices L., Kalra A., Patel S., Shankar K., Chua L.O. Microtubules as Sub-Cellular Memristors. *Scientific Reports*. 2020. vol. 10(1). DOI: 10.1038/s41598-020-58820-y.
33. Bicanski A., Burgess N. Neural vector coding in spatial cognition. *Nature Reviews Neuroscience*. 2020. vol. 21. pp. 453–470.
34. Rvachev M. V. Neuron as a reward-modulated combinatorial switch and a model of learning behavior. *Neural Networks*. 2013. vol. 46. pp. 62–74.
35. Осипов В.Ю. Векторные свойства и память нейронов. Сборник тезисов XXIV съезда физиологического общества им. И.П. Павлова. 2023. С. 586–587.
36. Sardi S., Vardi R., Sheinin A., Goldental A., Kanter I. New types of experiments reveal that a neuron functions as multiple independent threshold units. *Scientific Reports*. 2017. vol. 7(1). DOI: 10.1038/s41598-017-18363-1.

**Осипов Василий Юрьевич** — директор, Санкт-Петербургский институт информатики и автоматизации Российской академии наук, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук. Область научных интересов: математическое моделирование, интеллектуальные системы, нейронные сети, информационная безопасность. Число научных публикаций — 170. [osipov\\_vasily@mail.ru](mailto:osipov_vasily@mail.ru); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)323-0366.

**Поддержка исследований.** грант Российского научного фонда № 23 19 20081 и Санкт-Петербургского научного фонда.

V. OSIPOV  
**INTELLIGENT NEURAL NETWORK MACHINE WITH  
THINKING FUNCTIONS**

---

*Osipov V. Intelligent Neural Network Machine with Thinking Functions.*

**Abstract.** In recent years, interest in artificial intelligence based on neural network approaches has grown significantly. A number of significant scientific results have been obtained that have found wide application in practice. Generative adversarial neural network models, neural network transformers, and other solutions have attracted much attention. Obvious progress has been achieved in neural network recognition and image generation, text and speech processing, event forecasting, and control of processes that are difficult to formalize. However, it has not yet been possible to endow neural network machines with thinking. All results obtained using neural network machines can be attributed to solutions based on various types of signal binding without full control of their processing processes. Typical representatives of such machines are ChatGPT. The capabilities for intelligently operating various signals in known neural network machines are very limited. Among the main reasons for such limitations, one should highlight the imperfection of the basic principles of neural network information processing used. The properties of neurons have long been considered in a simplified manner. This was due to both gaps in the field of biological research and the lack of opportunities to build large neural networks on complex neuron models. In recent years the situation has changed. New ways to implement large neural networks have emerged. It has also been established that even individual neurons can have extensive internal memory and implement various functions. However, many mechanisms of neuron functioning and their interactions still remain unclear. The issues of controlled associative access to the internal memory of neurons have been little studied. These shortcomings significantly hinder the creation of thinking neural network machines. The object of research in the article is the process of intelligent neural network information processing. The subject of research: principles, models, and methods of such processing. The goal is to expand the functionality of neural network machines to solve difficult-to-formalize creative problems through the development of new principles, models, and methods of intelligent information processing. In the interests of achieving this goal, the operating principles of intelligent neural network machines are clarified, and new models and methods of neural network information processing are proposed. A new model of a pulse neuron is revealed as a basic element of such machines. It is recommended to form the artificial brain of neural network machines in the form of multilayer neural networks endowed with logical structures with neurons of different parameters. A new method of multi-level intelligent information processing in neural network machines based on smart impulse neurons is proposed. The mechanisms of thinking of neural network machines, and the underlying functions of intellectual operation of images and concepts in neural network memory are explained. Simulation results are presented that confirm the validity of the proposed solutions.

**Keywords:** neural network machine, intelligence, thinking functions, smart neurons, signal transformation.

---

## References

1. Turing A. Can a Machine Think? With the attached article by J. Von Neumann. General and Logical Theory of Automaton. Moscow: State Publishing House Physical and Mathematical Literature, 1960. 112 p.

2. Myshlenie – Bol'shoj enciklopedicheskij slovar' [Thinking is a great encyclopedic dictionary]. Available at: <https://gufo.me/dict/bes/МЫШЛЕНИЕ> (accessed 05.04.2024).
3. Velankar M.R., Mahalle P.N., Shinde G.R. Machine Thinking: New Paradigm Shift. In: Cognitive Computing for Machine Thinking. Innovations in Sustainable Technologies and Computing. 2024. pp. 43–53.
4. Malsburg C. Toward understanding the neural code of the brain. *Biological Cybernetics*. 2021. vol. 115. no. 5. pp. 439–449.
5. Yamakawa H. The whole brain architecture approach: accelerating the development of artificial general intelligence by referring to the brain. *Neural Networks*. 2021. vol. 144. pp. 478–495.
6. Haykin S. *Neural Networks and Learning Machines*, third ed., Prentice Hall, New York. 2008. Available at: <http://dai.fmph.uniba.sk/courses/NN/haykin.neural-networks.3ed.2009.pdf> (accessed 24.04.2024).
7. Kotseruba I., Tsotsos J. 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*. 2020. vol. 53. no. 1. pp. 17–94.
8. Dormehl L. *Thinking machine: The Quest for Artificial Intelligence – and Where It's Taking Us Next*. Penguin, 2017. 209 p.
9. Takano S. *Thinking Machines. Machine Learning and Its Hardware Implementation*. Academic Press, 2021. 306 p.
10. Hawkins J., Blakeslee S. *On intelligence*. Brown Walker, 2006. 174 p.
11. Osipov V., Osipova M. Space-time signal binding in recurrent neural networks with controlled elements. *Neurocomputing*. 2018. vol. 308. pp. 194–204.
12. Hawkins J., Ahmad S. Hierarchical temporal memory including HTM cortical learning algorithms. Hosted at Numenta.org. 2011. 68 p.
13. Spoerer C.J., McClure, P., Kriegeskorte, N., 2017. Recurrent convolutional neural networks: a better model of biological object recognition. *Frontiers in psychology*. 2017. vol. 8. DOI: 10.3389/fpsyg.2017.01551.
14. Patrick M., Adekoya A., Mighty A., Edward B. Capsule networks – a survey. *Journal of King Saud University – Computer and Information Sciences*. 2022. vol. 34(1). pp. 1295–1310.
15. Yang G., Ding F. Associative memory optimized method on deep neural networks for image classification. *Information Sciences*. 2020. vol. 533. pp. 108–119.
16. Yang J., Zhang L., Chen C., Li Y., Li R., Wang G., Jiang S., Zeng Z. A hierarchical deep convolutional neural network and gated recurrent unit framework for structural damage detection. *Information Sciences*. 2020. vol. 540. pp. 117–130.
17. Ma T., Lv S., Huang L., Hu S. HiAM: A hierarchical attention based model for knowledge graph multi-hop reasoning. *Neural Networks*. 2021. vol. 143. pp. 261–270.
18. Grossberg S. Adaptive resonance theory: how a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*. 2013. vol. 37. pp. 1–47.
19. Khowaja S., Lee S.L. Hybrid and hierarchical fusion networks: a deep cross-modal learning architecture for action recognition. *Neural Computing and Applications*. 2020. vol. 32. no. 14. pp. 10423–10434.
20. Saha S., Gan Z., Cheng L., Gao J., Kafka O., Xie X., Li H., Tajdari M., Kim H., Liu W. Hierarchical deep learning neural network (HiDeNN): an artificial intelligence (AI) framework for computational science and engineering. *Computer Methods in Applied Mechanics and Engineering*. 2021. vol. 373. DOI: 10.1016/j.cma.2020.113452.
21. Yang M., Chen L., Lyu Z., Liu J., Shen Y., Wu Q. Hierarchical fusion of common sense knowledge and classifier decisions for answer selection in community question answering. *Neural Networks*. 2020. vol. 132. pp. 53–65.

22. Wolfrum P., Wolff C., Lucke J., Malsburg C. A recurrent dynamic model for correspondence-based face recognition. *Journal of Vision*. 2008, vol. 8(7). no. 34. pp. 1–18. DOI: 10.1167/8.7.34.
23. Han Y., Huang G., Song S., Yang L., Wang H., Wang Y. Dynamic neural networks: a survey. arXiv:2102.04906v4. 2021. pp. 1–20.
24. Osipov V., Nikiforov V., Zhukova N., Miloserdov D. Urban traffic flows forecasting by recurrent neural networks with spiral structures of layers. *Neural Computing and Applications*. 2020. vol. 32. no. 18. pp. 14885–14897.
25. Osipov V., Kuleshov S., Zaytseva A., Levonevskiy D., Miloserdov D. Neural network forecasting of news feeds. *Expert systems with applications*. 2021. vol. 169. DOI: 10.1016/j.eswa.2020.114521.
26. Osipov V., Kuleshov S., Miloserdov D., Zaytseva A., Aksenov A. Recurrent Neural Networks with Continuous Learning in Problems of News Streams Multifunctional Processing. *Informatics and Automation*. 2022. vol. 21. no. 6. pp. 1145–1168.
27. Osipov V., Osipova M. Method and device of intellectual processing of information in neural network, Patent RU2413304. 2011.
28. Osipov V. Method for intelligent multi-level information processing in neural network, Patent RU2737227. 2020.
29. He J., Yang H., He L., Zhao L. Neural networks based on vectorized neurons. *Neurocomputing*. 2021. vol. 465. pp. 63–70.
30. Deng C., Litany O., Duan Y., Poulencard A., Tagliasacchi A., Guibas L. Vector neurons: a general framework for SO(3)-Equivariant networks. arXiv:2104.12229v1. 2021. pp. 1–12.
31. Kryzhanovskiy B., Litinskii L., Mikaelian A. Vector-Neuron Model of Associative Memory. *IEEE International Joint Conference on Neural Networks*. 2004. vol. 2. pp. 909–914.
32. Tuszyński J.A., Friesen D.E., Freedman H., Sbitnev V.I., Kim H., Santelices L., Kalra A., Patel S., Shankar K., Chua L.O. Microtubules as Sub-Cellular Memristors. *Scientific Reports*. 2020. vol. 10(1). DOI: 10.1038/s41598-020-58820-y.
33. Bicanski A., Burgess N. Neural vector coding in spatial cognition. *Nature Reviews Neuroscience*. 2020. vol. 21. pp. 453–470.
34. Rvachev M. V. Neuron as a reward-modulated combinatorial switch and a model of learning behavior. *Neural Networks*. 2013. vol. 46. pp. 62–74.
35. Osipov V.Yu. [Vector properties and memory of neurons] Vektornye svoystva i pamjat' neyronov. *Sbornik tezisev XXIV s'ezda fiziologicheskogo obshchestva im. I.P. Pavlova Collection of abstracts of the XXIV Congress of the Physiological Society named after. I.P. Pavlova*. 2023. pp. 586–587. (In Russ.).
36. Sardi S., Vardi R., Sheinin A., Goldental A., Kanter I. New types of experiments reveal that a neuron functions as multiple independent threshold units. *Scientific Reports*. 2017. vol. 7(1). DOI: 10.1038/s41598-017-18363-1.

**Osipov Vasily** — Director, St. Petersburg Institute of Informatics and Automation of the Russian Academy of Sciences, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: mathematical modeling, intelligent systems, neural networks, information security. The number of publications — 170. osipov\_vasily@mail.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)323-0366.

**Acknowledgements.** This research is supported by the Russian Science Foundation No. 231920081 and the St. Petersburg Science Foundation.

В.К. ПИМЕШКОВ, М.Л. НИКОНОВА, М.Г. ШИШАЕВ  
**КОМБИНИРОВАННЫЙ МЕТОД ИЗВЛЕЧЕНИЯ ТЕРМИНОВ  
ДЛЯ ЗАДАЧИ МОНИТОРИНГА ТЕМАТИЧЕСКИХ  
ОБСУЖДЕНИЙ В СОЦИАЛЬНЫХ МЕДИА**

*Пимешков В.К., Никонорова М.Л., Шишаев М.Г. Комбинированный метод извлечения терминов для задачи мониторинга тематических обсуждений в социальных медиа.*

**Аннотация.** Извлечение терминов является важным этапом автоматизированного построения систем знаний на основе естественно-языковых текстов, поскольку обеспечивает формирование базовой системы понятий, используемой затем в прикладных задачах интеллектуальной обработки информации. В статье рассмотрена проблема автоматизированного извлечения терминов из естественно-языковых текстов с целью их дальнейшего использования при построении формализованных систем знаний (онтологий, тезаурусов, графов знаний) в рамках задачи мониторинга тематических обсуждений в социальных медиа. Данная задача характеризуется необходимостью включения в формируемую систему знаний как понятий из нескольких различных предметных областей, так и некоторых общеупотребительных понятий, используемых аудиторией социальных медиа в рамках тематических обсуждений. Кроме того, формируемая система знаний является динамичной как с точки зрения состава охватываемых ею предметных областей, так и состава релевантных понятий, подлежащих включению в систему. Применение существующих классических методов извлечения терминов в данном случае затруднительно, поскольку они ориентированы на извлечение терминов в рамках одной предметной области. Исходя из этого, для решения рассматриваемой задачи предложен комбинированный метод, совмещающий в себе подходы на основе внешних источников знаний, инструментов NER и правил. Результаты проведенных экспериментов демонстрируют эффективность предложенной комбинации подходов к извлечению терминов для задачи мониторинга и анализа тематических обсуждений в социальных медиа. Разработанный метод значительно превосходит по точности существующие инструменты извлечения терминов. В качестве дальнейшего направления исследования рассмотрена возможность развития метода для решения задачи выделения вложенных терминов или сущностей.

**Ключевые слова:** интеллектуальный анализ текстов, извлечение терминов, социальные медиа, извлечение знаний.

**1. Введение.** Задача извлечения терминов (term extraction) представляет собой процесс автоматического выделения ключевых терминов или слов из текста или набора текстов с целью определения наиболее репрезентативных слов в контексте рассматриваемой прикладной задачи. В данной работе рассматривается специфический класс подобных прикладных задач, характеризуемый множественностью предметных областей, с одной стороны, и важностью учета общеупотребительной лексики – с другой. Примером такой задачи является мониторинг и анализ социально-политической ситуации на некоторой ограниченной территории (город, регион и т.п.). Использование для ее решения данных

социальных медиа сопряжено с проблемой извлечения структурированной информации из огромного объема постоянно генерируемых сырых данных, составляющих содержание тематических обсуждений в социальных медиа. Базовым компонентом такой информации являются термины, являющиеся лексическими обозначениями понятий, используемых участниками дискуссий. В то же время, специфика как лексических особенностей коммуникаций в социальных медиа в целом, так и рассматриваемой прикладной задачи, ограничивает возможность применения для извлечения терминов известных готовых технологий и методов.

Извлечение терминов является критически важным этапом для задачи извлечения знаний – концептов (единиц мысли) и связей между ними. Процесс извлечения знаний из текстов заключается в выделении всех релевантных (обозначающих значимые для последующего использования концепты) терминов и в дальнейшем определении прямых или опосредованных связей между ними. При этом релевантность, чаще всего, определяется принадлежностью термина некоторой предметной области. В свою очередь, предметная область задается языковым ресурсом – коллекцией текстов соответствующей тематики, словарем, тезаурусом и т.п.

Как правило, общение в социальных медиа хоть и представлено в виде текста, но имеет присущие живой разговорной речи особенности, такие как свернутость, спонтанность, экспрессивность, высокая доля лексики с разговорной окраской, обилие специфических разговорных конструкций и т.д. [1]. Помимо этого, для социальных медиа характерна еще одна особенность – мультипредметность: наличие/употребление терминов из многих предметных областей. При этом мы различаем предметные термины (специфичные для некоторой предметной области) и общеупотребительные термины, в равной степени широко используемые в коммуникациях любой тематики. Важной особенностью социальных медиа как источника релевантных терминов является отсутствие каких-либо ограничений (присущих профессиональной, предметной, коммуникации) в использовании терминологии. Как следствие, интересующие нас понятия могут обозначаться как предметными терминами, так и терминами общей лексики.

Таким образом, в отличие от классической задачи извлечения терминов, где требуется извлечение ограниченного набора релевантных терминов в рамках одной предметной области, в задаче мониторинга тематических обсуждений нас интересуют не только предметные, но и общеупотребительные термины. Кроме того,

предметные термины в нашем случае могут принадлежать нескольким различным предметным областям, лексический состав которых может изменяться, если мониторинг осуществляется достаточно длительное время. В таких условиях вполне ожидаемо, что готовые словари или тезаурусы с подобными терминами в открытом доступе отсутствуют. Данная особенность рассматриваемой в работе задачи иллюстрируется рисунком 1.



Рис. 1. Структура лексики коммуникаций в рамках рассматриваемой прикладной задачи

В результате, применение только классических методов извлечения терминов не покрывает все интересующие нас термины и, следовательно, не обеспечивает достаточной полноты в рамках решаемой задачи. Это обуславливает необходимость разработки специального метода, учитывающего динамичность тематических обсуждений, выражающуюся в изменении набора интересующих нас предметных областей или набора интересующих нас терминов в рамках одной предметной области. В данной работе предложен подобный метод, основанный на комбинировании различных подходов к извлечению терминов, и проведена оценка его эффективности на рассматриваемом классе прикладных задач. Особенностью метода является учет структуры коммуникации и пользовательских реакций при принятии решения о включении некоторого термина в число значимых. В процессе разработки и валидации предложенного метода использовались данные, аккумулированные ранее в рамках мониторинга социально-политической обстановки в Кировско-Апатитском регионе Мурманской области: размеченный датасет, сформированный из сообщений сети ВКонтакте, а также коллекция соответствующих региональных нормативно-правовых документов.

## 2. Методы извлечения терминов для создания формализованных систем знаний

**2.1. Краткая характеристика существующих («классических») методов извлечения терминов.** Классические методы извлечения терминов можно разделить на три категории: методы на основе правил, методы, использующие словари или внешние источники знаний, а также статистические методы. Вне зависимости от метода задача заключается в поиске *значимых (релевантных)* терминов – терминов, важных в контексте рассматриваемой прикладной задачи.

Методы на основе правил используют лексические или морфосинтаксические шаблоны для извлечения значимых в рамках задачи терминов. Значимость термина в данной категории методов определяется правилами, задаваемыми экспертно или с помощью автоматизированных методов. Формально данный метод извлечения можно представить следующим образом:

*Дано:*

$D = \{d_1, \dots, d_N\}$  – множество (коллекция) документов различных тематик, при этом  $D = D^S \cup D^G$ , где  $D^S$  – множество документов предметной области,  $D^G$  – множество документов общей тематики;

$P = \{p_1, \dots, p_K\}$  – множество правил извлечения терминов, где:

$p_i: D \rightarrow B(T)$ ,  $B(\cdot)$  – булеан множества,  $T$  – множество всевозможных терминов.

*Найти:*

$T(D) = \cup_i T_i$  – множество соответствующих правилам терминов из коллекции документов  $D$ , где  $T_i = \cup_{j=1}^N p_i(d_j)$  – множество терминов, соответствующих правилу  $p_i$ .

Например, в [2] используются морфологические правила, определенные на основе анализа предметных текстов, для извлечения типизированных структур данных. В [3] используют лингвистические шаблоны, дополненные контекстными правилами, принимающие во внимание фразы с предлогами и причастиями, для повышения эффективности извлечения терминов. В рамках такого метода сначала извлекаются структуры, содержащие определенные предлоги. Полученные кандидаты фильтруются от терминов неверной структурной формы (т.е. ошибки токенизации и частеречевой разметки) и от семантически бедных терминов (грамматически правильных, но не значимых для прикладных задач) с помощью набора правил. Авторы также уточняют частеречевую разметку для особых форм глаголов, что позволяет принять во внимание больший спектр потенциальных терминов. Результаты применения такого

метода могут использоваться для решения таких задач, как создание и пополнение онтологий и терминологий.

Методы, основанные на внешних источниках знаний, используют существующие терминологические ресурсы (тезаурусы, словари, онтологии) с целью поиска вхождений, как правило, предметных терминов в тексте. Значимость в данном случае исчерпывающим образом определяется внешним источником (те и только те термины являются значимыми, которые в некотором смысле схожи с входящими в источник). Формализация данной категории методов выглядит следующим образом:

*Дано:*

$D$  – множество документов различных тематик;

$T^S = \{t_1^S, \dots, t_M^S\}$  – множество терминов источника знаний  $S$ , определяющего предметную область;

$sim : T \times T \rightarrow \{0,1\}$  – некоторая функция схожести термина документа с термином источника знаний,  $sim(t_1, t_2) = 1$ , если термины  $t_1, t_2$  схожи и  $sim(t_1, t_2) = 0$  в противном случае.

*Найти:*

$T^{S+} = \{t_i^{S+} \mid \forall t_i^{S+}: \exists t^S \in T^S: f(t_i^{S+}, t^S) = 1\}$  – множество терминов, схожих в смысле функции  $f$  с содержащимися во внешнем источнике.

Например, предложенный в работе [4] подход к извлечению биологических терминов основан на приближенном поиске по словарю, который позволяет извлекать из текста не только точные совпадения терминов, но и значимые слова этих терминов. Значимость слова определяется с помощью соответствующей меры значимости, обученной на основе словаря. В системе извлечения информации [5] в качестве исходного источника знаний выступает онтология, которая в процессе преобразуется в словарь, состоящий из категорий и соответствующих им сущностей. Затем используется анализатор, содержащий сформированный на основе данного словаря набор правил, для поиска в тексте совпадений именных групп, представляющих важные медицинские сущности.

Стоит отметить, что задача пополнения онтологии/тезауруса может рассматриваться как частный случай использования методов извлечения терминов на основе внешних источников. При решении данной задачи на входе также имеется внешний ресурс и текст, а также некоторая функция, возвращающая степень схожести термина-кандидата и термина, присутствующего в исходном ресурсе. Функция схожести может быть реализована различными способами: схожесть контекста [6], схожесть векторного представления в рамках некоторой языковой модели [7] и т.д. Частный случай – когда функция схожести

бинарная: возвращает 1 для полностью идентичных терминов и 0 для прочих, что соответствует вышеизложенному определению данной категории методов.

Статистические методы используют анализ большого количества наблюдений для идентификации терминов. Такие методы редко применяются без какого-либо этапа синтаксического или лексического анализа ввиду того, что это может приводить к большому объему нежелательной терминологии [8]. В статистических методах признак значимости термина определяется частотностью термина в наборе предметных документов. Значимость при этом трактуется как «предметность» – соответствие термина интересующей нас предметной области. Формальный вид данной категории методов следующий:

*Дано:*

$D^S = \{d_1^S, \dots, d_N^S\}$  – множество документов предметной области;

$D^G = \{d_1^G, \dots, d_N^G\}$  – множество документов общей тематики;

$m: B(D) \times T$  – статистическая мера, характеризующая частоту встречаемости термина из  $T$  во множестве документов из  $B(D)$ , где  $B(\cdot)$  – булеан множества.

*Найти:*

$T^S: t^S \in T^S \leftrightarrow m(D^S, t^S) > m(D^G, t^S)$  – множество предметных терминов.

Зачастую статистические методы объединяются с другими для достижения лучших результатов. Например, в [9] используются коллокации (частный случай N-грамм), дополненные частеречной и статистической фильтрацией для извлечения терминов-кандидатов из предметных текстов на немецком языке. Данный метод включает четыре этапа. На первом этапе выполняется предварительная обработка текста, включающая разбиение текста на предложения и другие структурные составляющие, разбиение текста на токены и определение их частей речи. На втором этапе извлекаются однословные термины с помощью существующих процедур, основанных на частотных характеристиках, морфологии и списках морфем. На третьем этапе на основе грамматических свойств извлекаются коллокационные кандидаты-словосочетания. И четвертым этапом производится фильтрация словосочетаний-кандидатов, удаляются тривиальные словосочетания (содержащие определенные общие глаголы), а также общеупотребительные словосочетания. Термины, извлеченные таким методом, могут использоваться для создания и пополнения предметных терминологий.

Кроме того, для извлечения терминов могут использоваться инструменты извлечения именованных сущностей (Named Entity Recognition, NER), нацеленные на извлечение и классификацию именованных сущностей, которые могут рассматриваться как особые типы терминов. В классическом NER выделяют такие типы, как названия организаций, локаций и имена людей. Но NER находит применение и в конкретных предметных областях. Например, в биомедицине выделяют названия генов, белков и связанных с ними биологические или генетические термины, а также наименования болезней, лекарств и др. NER может быть реализован с помощью правил, словарей или машинного обучения. Например, в работе [10] предлагается метод извлечения и классификации биомедицинских терминов на основе NER без учителя. Для извлечения терминов-кандидатов авторы используют выделение именных групп с фильтрацией по TF-IDF (Term Frequency – Inverse Document Frequency), а для классификации этих кандидатов применяется мера схожести с шаблонами классов, разработанными авторами на основе внешних источников.

Помимо рассмотренных выше методов, распространенной практикой является использование гибридных методов, комбинирующих различные подходы. Наиболее популярной является комбинация лингвистических и статистических методов в том или ином виде. К примеру, в [11] используют решающие деревья для отбора и ранжирования терминов-кандидатов на основе трех различных наборов признаков. Первый набор включает в себя структурные признаки, полученные с помощью системы GROBID [12] (библиотека машинного обучения для извлечения, анализа и реструктуризации необработанных документов). Второй набор состоит из фразеологических и информативных (насколько термин отражает смысл документа в рамках набора документов) признаков, полученных различными статистическими методами. И третий набор содержит лексические и семантические признаки, полученные с помощью таких внешних источников знаний, как GRISP [13] и Википедия.

**2.2. Комбинированный метод извлечения терминов для задач анализа тематических обсуждений.** Предлагаемый нами комбинированный метод извлечения терминов использует комбинацию вышеупомянутых категорий методов. В нашем случае множество значимых терминов не ограничивается предметными, но также должно включать некоторое подмножество общеупотребительных терминов, ассоциированных с предметными.

Соответственно, значимость термина определяется источником знаний, определяющим множество предметных терминов, и некоторой метрикой ассоциированности общеупотребительных терминов с элементами этого множества. Для расчета последней в рамках метода предлагается использовать количественные показатели пользовательской реакции и структуру коммуникации в рамках тематического обсуждения. Формально предлагаемый метод можно представить следующим образом:

*Дано:*

$D$  – множество документов; документ представляет собой пост (сообщение) или комментарий;

$l(d)$  – количество лайков документа  $d$ ;

$r(d)$  – количество репостов документа  $d$ ;

$c(d)$  – количество комментариев документа  $d$ ;

$v(d)$  – количество просмотров документа  $d$ ;

$T^S$  – множество терминов базового источника знаний;

$P$  – множество правил извлечения терминов;

$M$  – языковая модель для извлечения именованных сущностей;

$ner: M \times D \rightarrow B(T)$  – функция извлечения именованных сущностей из документов с помощью модели  $M$ .

$L \subseteq D \times D$  – асимметричное транзитивное отношение «является откликом», определяющее на множестве документов структуру коммуникации.  $d_1 L d_2$  означает, что документ (пост или комментарий)  $d_2$  является откликом на документ  $d_1$ . Структура коммуникации состоит из веток обсуждения, представляющих собой последовательности  $Br$  документов из  $D$  ( $Br \subseteq D$ ), удовлетворяющие условию:

$$Br = \{d_1, \dots, d_N\}: \forall i < j, d_i L d_j.$$

Метод извлечения терминов включает следующие основные шаги:

1) Извлечение из коллекции документов  $D$  множества предметных терминов  $T^{S+}$  на основе внешнего источника знаний  $T^S$  и функции схожести  $sim$ .

2) Извлечение из коллекции документов  $D$  множества терминов  $T^P$ , удовлетворяющих заданному набору правил  $P$ .

3) Извлечение из коллекции документов  $D$  множества именованных сущностей  $T^M$  с помощью функции  $ner$ .

4) Выделение из множества  $T^P \cup T^M$  множества терминов  $T^A$ , ассоциированных с терминами из  $T^S \cup T^{S+}$ .

5) Результат работы метода есть объединение множеств извлеченных предметных терминов  $T^{S+}$  и значимых общеупотребительных терминов  $T^A$ .

Отбор терминов для включения в множество  $T^A$  осуществляется на основании метрики ассоциированности, которая зависит от контекста, частоты употребления термина  $s$  и интенсивности пользовательских реакций на документы-источники  $s$ . Метрика может быть представлена функционалом вида:

$$assc(t, s) = F(ct(t, s), q(s), fd(s)): t \in T^S \cup T^{S+}, s \in T,$$

где

$ct: T \times T \rightarrow \mathbb{R}$  – функция оценки контекста, удовлетворяющая следующим условиям:

$$ct(t_1, t_2) > 0, t_1 \in d_i, t_2 \in d_j \leftrightarrow \exists Br: d_i, d_j \in Br;$$

$$ct(t_1, t_2) > ct(t_1, t_3), t_1 \in d_i, t_2 \in d_j, t_3 \in d_k \leftrightarrow \\ \leftrightarrow \exists Br: d_i, d_j, d_k \in Br \wedge |i - j| < |i - k|;$$

$$ct(t_1, t_2) = 0, t_1 \in d_i, t_2 \in d_j \leftrightarrow \nexists Br: d_i, d_j \in Br.$$

$q(s)$  – частота встречаемости термина  $s$ ;

$fd(s) = \sum_{i=1}^n fd(d_i): s \in d_i$  – суммарная оценка реакции на термин  $s$ ;

$fd(d_i) = f(l(d_i), r(d_i), c(d_i), v(d_i))$  – обобщенная оценка интенсивности реакции на документ  $d_i$ .

Следует отметить, что каждый документ может быть как источником некоторого инфополюса, так и откликом на него. При этом отклики могут вызывать свои собственные отклики, и в таком случае первые могут рассматриваться как источники. Вместе с тем адресат/инициатор отклика не всегда выражен в явном виде (например, комментарий или репост), т.к. вполне возможно появление отклика на какую-то популярную публикацию или событие без явной привязки к оному.

Сравнительная характеристика рассмотренных методов извлечения терминов представлена в таблице 1.

Таблица 1. Сравнение методов извлечения терминов

Метод	Используемые исходные данные	Ожидаемый результат и ключевая особенность
Метод на основе правил	$D = \{d_1, \dots, d_N\}$ – множество (коллекция) документов различных тематик; $P = \{p_1, \dots, p_K\}$ – множество правил извлечения терминов, где $p_i: D \rightarrow B(T)$ , $T$ – множество всевозможных терминов.	$T(D) = \bigcup_i T_i$ – множество соответствующих правилам терминов из коллекции документов $D$ , где $T_i = \bigcup_{j=1}^N p_i(d_j)$ – множество терминов, соответствующих правилу $p_i$ . Термины не различаются по признаку «предметности».
Метод на основе внешних источников	$D$ – множество документов различных тематик; $T^S = \{t_1^S, \dots, t_M^S\}$ – множество терминов источника знаний $S$ , определяющего предметную область; $sim: T \times T \rightarrow \{0,1\}$ – функция схожести терминов.	$T^{S+} = \{t_i^{S+}   \forall t_i^{S+}: \exists t^S \in T^S: f(t_i^{S+}, t^S) = 1\}$ – множество терминов, схожих в смысле функции $f$ с содержащимися во внешнем источнике. «Предметность» термина определяется внешним источником.
Статистический метод	$D^S$ – множество документов предметной области; $D^G$ – множество документов общей тематики; $m: B(D) \times T$ – статистическая мера, характеризующая частоту встречаемости термина в документах.	$T^S: t^S \in T^S \leftrightarrow m(D^S, t^S) > m(D^G, t^S)$ – множество предметных терминов. «Предметность» термина определяется соотношением предметной и обще-тематической коллекций документов.
Предлагаемый комбинированный метод	$D$ – множество документов; $l(d), r(d), c(d), v(d)$ – количественные индикаторы пользовательской реакции на документ $d$ ; $T^S$ – множество терминов базового источника знаний; $P$ – множество правил извлечения терминов; $M$ – языковая модель для извлечения именованных сущностей; $per: M \times D \rightarrow B(T)$ – функция извлечения именованных сущностей из документов с помощью модели $M$ . $L \subseteq D \times D$ – асимметричное транзитивное отношение «является откликом», определяющее на множестве документов структуру коммуникации.	$T^{S+}, T^A$ – множества предметных и значимых общепотребительных терминов. Предметные термины определяются внешним источником, значимые общепотребительные – ассоциированностью с предметными и интенсивностью пользовательских реакций.

Стоит отметить, что в результате применения комбинированного метода предполагается пополнение источника знаний извлеченными терминами. Это связано с тем, что значимость, в том числе, определяется ассоциированностью, которая, в свою очередь, зависит от структуры откликов, а отклики могут содержать новые термины. Под новыми терминами в данном случае можно понимать как термины, существовавшие до этого и ставшие значимыми, так и неологизмы, появившиеся с течением времени и ставшие значимыми в рамках решаемой прикладной задачи.

**2.3. Обзор ресурсов и инструментов для извлечения терминов на русском языке.** Рассмотренные выше методы извлечения терминов, применительно к русскоязычным текстам, могут быть реализованы с помощью ряда ресурсов и инструментов.

**2.3.1. Внешние источники знаний.** В качестве внешних источников знаний могут выступать тезаурусы, онтологии, базы знаний, графы знаний и другие необходимые для решения рассматриваемой задачи ресурсы. К сожалению, подобных русскоязычных ресурсов в открытом доступе довольно мало, поэтому, зачастую, исследовательскими группами предпринимаются попытки создания их с нуля для решения той или иной задачи.

Среди существующих ресурсов общего назначения можно выделить RuWordNet [14] – наиболее распространенный русскоязычный тезаурус, представляющий собой результат трансформации тезауруса RuThes [15] в формат WordNet. По состоянию на 2016 год в нем содержатся: 29297 синсетов (наборов синонимов) существительных, 12865 синсетов прилагательных и 7636 синсетов глаголов. Всего RuWordNet содержит 111,5 тысяч слов и выражений русского языка. Также стоит отметить AGROVOC [16] – многоязычный тезаурус Продовольственной и сельскохозяйственной организации ООН, охватывающий такие области, как продовольствие, сельское хозяйство, окружающую среду и др. Он насчитывает более 41 000 понятий, включающих свыше 994 000 терминов на 42 различных языках, в том числе и на русском. Еще один известный ресурс с поддержкой русского языка – DBpedia [17] – открытая многоязычная база знаний, охватывающая множество различных предметных областей, поскольку является результатом извлечения структурированных данных из статей Википедии. DBpedia может быть полезна при решении таких задач, как интеграция данных, распознавание именованных сущностей, выявление тем и ранжирование документов.

**2.3.2. Инструменты NER.** Для решения задачи NER для русского языка можно выделить следующие библиотеки, реализованные на языке Python, – SpaCy [18], Natasha [19] и DeepPavlov [20].

SpaCy использует методы машинного обучения для определения в исходном тексте интервалов (непрерывных последовательностей слов), которые представляют собой сущности. В SpaCy для извлечения именованных сущностей по умолчанию применяется обученная на новостных текстах языковая модель, которая позволяет идентифицировать в тексте такие типы сущностей, как локации, организации и персоны. Также, в зависимости от специфики задачи, имеется возможность обучения собственной модели для выявления особых типов сущностей.

Библиотека Natasha, в отличие от SpaCy, ориентирована на обработку только русского языка и, соответственно, демонстрирует более точные результаты. В Natasha используется набор предопределенных шаблонов, основанных на правилах, для сопоставления и извлечения упоминаний сущностей из текста. В ней, как и в SpaCy, также используется предварительно обученная модель для извлечения стандартных типов сущностей из текста.

Библиотека DeepPavlov содержит набор предварительно обученных современных NLP моделей для анализа, в том числе, русскоязычных текстов. В реализации NER для маркировки последовательностей используется архитектура BiLSTM-CRF (Bidirectional LSTM with Conditional Random Field). Как и в рассмотренных выше библиотеках, DeepPavlov позволяет распознавать стандартные типы сущностей. Также DeepPavlov позволяет настраивать или обучать свои собственные модели NER на пользовательских размеченных наборах данных, что обеспечивает гибкость при работе с конкретными типами сущностей или предметными областями.

**2.3.3. Инструменты для формирования лингвистических правил.** Лингвистические правила используются для задания точных шаблонов, на основе которых из текста будут извлекаться необходимые языковые единицы. Для построения таких правил могут применяться различные парсеры, например, Томига-парсер [21] или Yargy-парсер [22], а также морфологические анализаторы, например, rymorphu2 [23].

Томига-парсер от Яндекс позволяет извлекать структурированные данные из текста на естественном языке с помощью контекстно-свободных грамматик и словарей ключевых слов. Контекстно-свободная грамматика состоит из набора правил-продукций, которые определяют, как символы и слова могут

комбинироваться для формирования действительных предложений. В данном случае правила грамматики применяются без учета контекста или окружающих слов в предложении. В основе Томита-парсера лежит алгоритм GLR-парсинга (Generalized left-to-right algorithm). Парсер также позволяет описать свои грамматики и добавлять словари для нужного языка. В Yargy-парсере – аналоге Томита-парсера для Python – правила для извлечения сущностей также описываются с помощью контекстно-свободных грамматик и словарей. Данная библиотека является частью проекта Natasha, рассмотренного выше, и, соответственно, ее основой. В Yargy-парсере реализован алгоритм Эрли – алгоритм синтаксического анализа предложения по контекстно-свободной грамматике.

Библиотека `ru morphology2`, используемая внутри Yargy-парсера и `SpaCy` для нормализации слов, может выступать в качестве самостоятельного инструмента для формирования частеречевых правил, в соответствии с которыми будут извлекаться термины конкретных частей речи и в нужной форме. Библиотека использует словари и граммы `OpenCorpora` [24] для приведения слов к нормальной форме, а также получения нужной формы слов и их грамматической информации (число, род, падеж, часть речи и т.д.).

**2.4. Программно-алгоритмическая реализация комбинированного метода извлечения терминов.** В данной работе для совместного извлечения не только предметных, но и значимых общепотребительных терминов использовалась комбинация последовательно применяемых методов на основе словарей, инструментов NER и правил, дополненных статистической фильтрацией. Схема комбинированного метода представлена на рисунке 2.

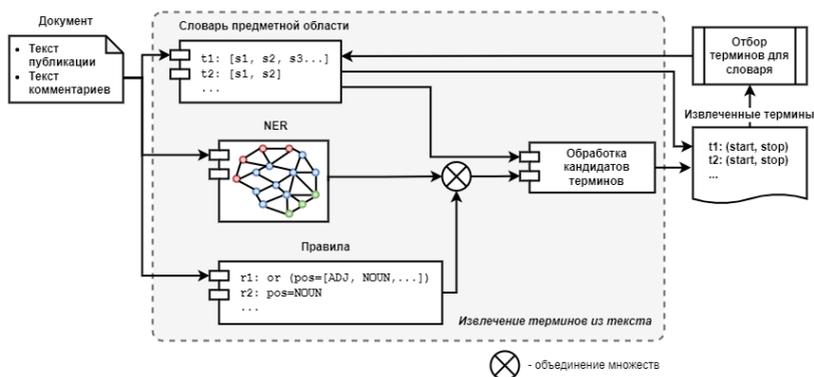


Рис. 2. Схема метода извлечения терминов

На первом этапе используется словарь для выделения предметных терминов, далее применяется NER для выделения терминов, которые могут быть как предметными (в рамках нашей задачи), так и общеупотребительными, и затем применяется набор правил для выделения общеупотребительных терминов (уни- би- и триграмм). Каждый отдельный этап («Словарь», NER, «Правила») представлен отдельным модулем в рамках разрабатываемого метода в целом. Общий алгоритм извлечения терминов описан в алгоритме 1.

Алгоритм 1. Извлечение терминов из текста предлагаемым методом

<b>Ввод:</b> Текстовые документы $D$ , базовый словарь $T^S$ , множество правил $P$ , языковая модель $M$ .
<b>Вывод:</b> Списки извлеченных терминов $T = T^{S+} \cup T^A$ , где $t$ имеет вид $(text(t), (start(t), stop(t)))$ , где $text(t)$ – текстовое представление термина, полученное из документа $d$ , $start(t)$ – позиция начала термина в документе $d$ , $stop(t)$ – позиция окончания.
Для каждого $d_i \in D$ выполнить
Извлечь из $d_i$ предметные термины $T^{S+}$ с помощью словаря
Извлечь из $d_i$ именованные сущности $T^M$ с помощью $M$
Извлечь из $d_i$ словосочетания $T^{P1}$ с помощью Правила для выделения биграмм и триграмм (P1)
Извлечь из $d_i$ униграммы $T^{P2}$ с помощью Правила для выделения униграмм (P2)
Выделить из множества $T^P \cup T^M$ множество терминов $T^A$ , ассоциированных с терминами из $T^S \cup T^{S+}$
<b>Конец цикла</b>

Словарь выступает инструментом однозначной идентификации терминов. Он представляет собой набор терминов предметной области, для каждого из которых хранится связанный с ним синсет, состоящий из синонимов или псевдонимов к данному термину и включающий в себя текстовое представление псевдонима и регулярное выражение для его поиска. Регулярные выражения могут быть заданы в формате точного или частичного совпадения, а также в свободной форме. Для частичного совпадения используется выделение основы текстового представления, дополняемое возможным окончанием слова в виде регулярного выражения. Начальный словарь формируется экспертом на основе анализа предметных текстов с помощью инструментов NER. В дальнейшем, при поддержке эксперта, словарь должен пополняться терминами, извлеченными в результате использования метода. Этап извлечения терминов с использованием словаря описан в алгоритме 2.

## Алгоритм 2. Извлечение терминов с помощью словаря

<b>Ввод:</b> Документ $d$ , базовый словарь $T^S$
<b>Вывод:</b> Список предметных терминов $T^{S+}$
Для <b>каждого</b> термина $t_i^S \in T^S$ выполнить
<b>Если</b> $t_i^S$ в $d$
Определить границы $t_i^S.start$ и $t_i^S.stop$ в $d$
Записать термин $t_i^S$ и его границы в $T^{S+}$
<b>Конец цикла</b>

Инструменты NER позволяют выделять именованные сущности, такие как имена людей, названия организаций и локаций, рассматриваемые нами как наиболее релевантные типы терминов в рамках решаемой задачи. В ходе разработки модуля был рассмотрен ряд инструментов NER: библиотека SpaCy [18], Natasha [19] и предварительно обученная модель BERT (ner\_collection3\_bert) [25]. Для сравнения инструментов был проведен эксперимент, в рамках которого с помощью них были извлечены именованные сущности из имеющегося датасета ВКонтакте и набора нормативно-правовых документов. Наборы извлеченных именованных сущностей были оценены экспертно, и было принято решение использовать модель BERT, т.к. она выдала наименьшее количество ложных кандидатов и ошибок. Таким образом, на вход модулю NER подается документ, где каждое его предложение анализируется моделью BERT и извлекаются все возможные термины-кандидаты и их границы в рамках текущего документа (алгоритм 3).

## Алгоритм 3. Извлечение именованных сущностей с помощью языковой модели

<b>Ввод:</b> Документ $d$ , языковая модель $M$
<b>Вывод:</b> Список именованных сущностей $T^M$
Для <b>каждого</b> предложения $s_i \in d$ выполнить
Извлечь из $s_i$ именованные сущности $t_j^M$ с помощью $M$
Определить границы $t_j^M.start$ и $t_j^M.stop$ в $d$
Записать извлеченные сущности и их границы в $T^M$
<b>Конец цикла</b>

Набор используемых правил для извлечения кандидатов терминов был создан эмпирическим путем на основе имеющихся данных и исходя из решаемой задачи. Первое правило (алгоритм 4) позволяет извлекать термины произвольной длины с минимумом предположений об их структуре (термин может состоять только из

существительных, прилагательных, причастий или числительных). Это правило основано на контекстно-свободных грамматиках, реализованных с помощью Yargy-парсера [22]. Отобранные данным правилом кандидаты фильтруются с помощью модели коллокации (реализация коллокаций из библиотеки NLTK [26]), обученной на имеющемся датасете, и дополнительно по TF для кандидата (среднее значение TF входящих в него токенов). При подсчете TF учитываются стоп-слова, список которых был получен из открытых источников и дополнен сетевым сленгом в результате анализа имеющихся данных. Максимальная длина N-грамм была ограничена до 3 токенов, что сузило круг интересующих нас словосочетаний до би- и триграмм, так как словосочетания большей длины крайне редко были полезны. Такой вывод был сделан на основе анализа результатов тестирования модуля.

#### Алгоритм 4. Извлечение би- и триграмм с помощью P1

<b>Ввод:</b> Документ $d$ , модель коллокации $C$ , список значений $TF$ для документа $d$ , Парсер
<b>Вывод:</b> Список словосочетаний $T^{P1}$
Выделить из $d$ словосочетания $N$ с помощью Парсера
Для <b>каждого</b> словосочетания $n_i \in N$ выполнить
$TF(n_i) = \text{sum}(TF(tk, n_i)) / \text{len}(n_i)$ , где $tk$ – токен в $n_i$
<b>Если</b> $n_i$ не содержит стоп-слов <b>и</b> $TF(n_i) \geq \text{mean}(TF(d))$ <b>и</b> $C(n_i) \geq \text{mean}(C)$
Определить границы $n_i.start$ и $n_i.stop$ в $d$
Записать словосочетания $n_i$ и его границы в $T^{P1}$
<b>Конец цикла</b>

Для выбора униграмм используется частеречевое правило (часть речи может быть только существительным), дополненное фильтрацией по TF с порогом в виде среднего TF по текущему документу (алгоритм 5). Минимальная длина токена была установлена на 3 символа.

#### Алгоритм 5. Извлечение униграмм с помощью P2

<b>Ввод:</b> Документ $d$ , список значений $TF$ для документа $d$
<b>Вывод:</b> Список униграмм $T^{P2}$
Для <b>каждого</b> токена $tk_i \in d$ выполнить
<b>Если</b> $tk_i$ не является стоп-словом <b>и</b> $tk_i$ является существительным <b>и</b> $\text{len}(tk_i) \geq 3$ <b>и</b> $TF(tk_i) \geq \text{mean}(TF(d))$
Определить границы $tk_i.start$ и $tk_i.stop$ в $d$
Записать униграмму $tk_i$ и ее границы в $T^{P2}$
<b>Конец цикла</b>

Все извлеченные кандидаты дополнительно проходят проверку на стоп-слова: для биграмм и триграмм – это вхождение стоп-слова в кандидат, для униграмм – полное совпадение.

Набор правил может быть расширен для извлечения другой интересующей информации, например, названий улиц, номеров домов и т.д. Но на данном этапе, в рамках рассматриваемой задачи, мы ограничились описанными выше правилами для извлечения произвольных уни- би- и триграмм.

На этапе обработки терминов-кандидатов в первую очередь выполняется проверка на ассоциированность кандидата с предметным термином из словаря. Дополнительно исключаются полные совпадения извлеченных терминов на разных этапах метода, с приоритетом для вышестоящего этапа обработки. Предполагается, что, чем раньше был извлечен термин, тем больше информации о нем есть (например, термин, извлеченный словарем, однозначно идентифицирован, в отличие от униграммы, извлеченной последним этапом). Пример работы метода представлен на рисунке 8.

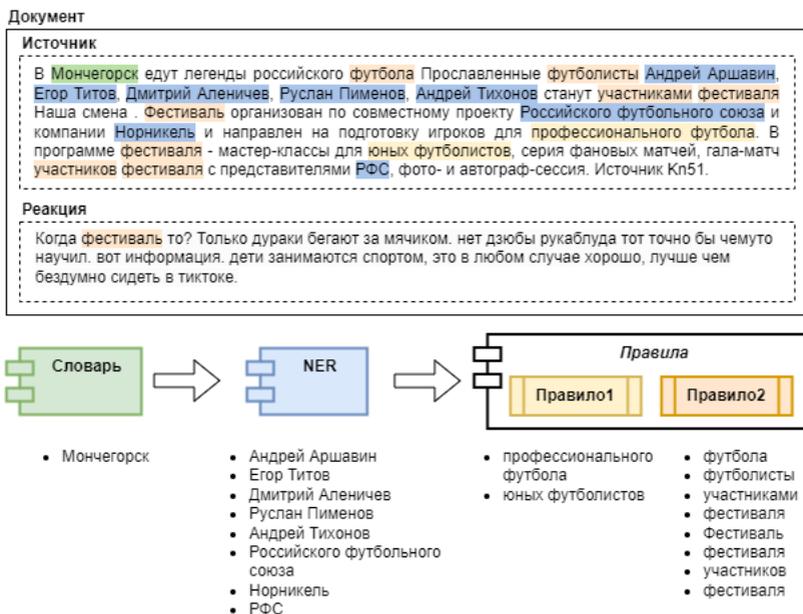


Рис. 8. Пример работы отдельных этапов метода на документе

**3. Результаты.** Проверка метода осуществлялась на примере прикладной задачи мониторинга социально-политической обстановки в регионе. Первоначальный словарь предметных терминов был сформирован экспертами в результате анализа набора из 18 нормативно-правовых документов, а именно уставов крупных организаций и предприятий Мурманской области.

Формирование словаря предметных терминов происходило в три этапа. Первым этапом экспертно из нормативно-правовых документов был получен исходный список терминов. Вторым этапом данный список был расширен с помощью внешнего источника знаний – Википедии. Для этого специализированным программным инструментом получения данных с веб-ресурсов были собраны связанные с исходными терминами в рамках страницы Википедии новые термины. Последним этапом экспертами, в результате анализа датасета, для каждого из терминов были сформированы связанные с ними синсеты. Таким образом, в исходный словарь было добавлено 146 предметных терминов.

Датасет, на котором проводились эксперименты, представляет собой выборку публикаций с комментариями из 8 групп социальной сети ВКонтакте по Мурманской области за приблизительно 2 года. Статистика по данному датасету приведена в таблице 2.

Таблица 2. Статистика по датасету из ВКонтакте

Показатель	Значение
Количество публикаций в датасете после очистки	50683
Количество публикаций, содержащих текст	49990
Средняя длина текста публикации (в символах)	311
Количество публикаций с комментариями	22084
Количество публикаций, содержащих текст и комментарии	21401
Количество комментариев в датасете после очистки	236444
Средняя длина одного комментария	77
Средняя длина всех комментариев к одной публикации	361

В связи с особенностями текста в датасете, необходимо было очистить его от нежелательных символов, таких как эмодзи или бессмысленные повторения различных знаков и конструкций, например, от автоматически добавляемых ссылок на пользователей в ответах на комментарии, от телефонных номеров или ссылок. На первом этапе использовались регулярные выражения для очистки от нежелательных конструкций, а на втором этапе – белый список символов, включающий в себя кириллицу, латиницу, цифры и ряд специальных символов (`[^А-Яа-яЁёА-Za-z0-9,.\!?\#\''\—\s]`). Всего

полученный датасет содержит 50683 преобработанных документа, в которых есть текст публикации и/или комментариев.

На данном этапе исследования ассоциированными считались все термины-кандидаты, выделенные в документе, в котором есть хотя бы один предметный термин, при этом интенсивность пользовательских реакций на документ-источник не учитывалась. Иначе говоря, контекст совместного употребления был задан объединением публикации с ее комментариями без учета их ветвления. Таким образом, для учета структуры коммуникации имеющийся датасет был разбит на документы, содержащие публикации и комментарии к ним.

Для оценки эффективности предлагаемого метода из датасета были выбраны и экспертно размечены 320 документов (пример на рисунке 9). Длина текста публикации и комментариев была ограничена в диапазоне от 100 до 600 символов. Такой размер документов был выбран с учетом средней длины предложения в русском языке. Исходя из решаемой прикладной задачи, эксперты размечали только те документы, в которых содержался хотя бы один термин, имеющий прямое отношение к Мурманской области. В таблице 3 представлены результаты оценки метода на размеченном датасете. В качестве оценочных метрик использовались точность, полнота и мера F1.

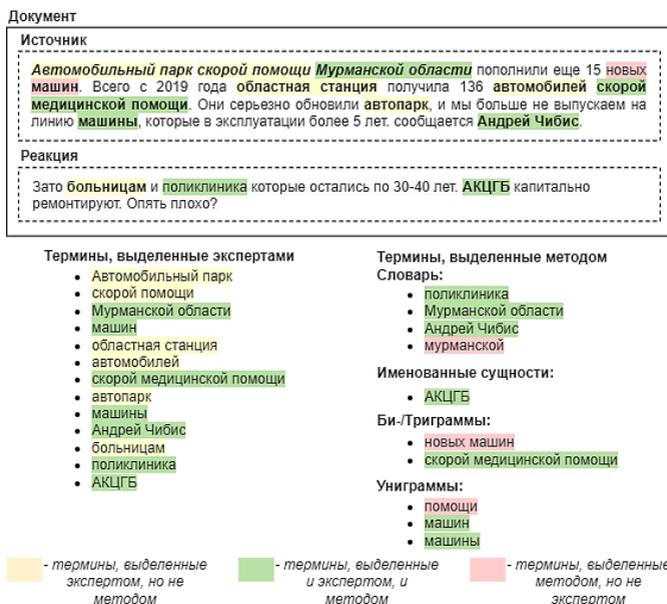


Рис. 9. Пример публикации, размеченной экспертами и методом

Таблица 3. Результаты проверки работы метода на тестовом наборе

Показатель	Значение
Количество терминов, размеченных экспертами	2169
Количество терминов, выделенных методом	2278
Количество терминов, выделенных и экспертами, и методом	1488
Количество терминов, выделенных экспертами, но не выделенных методом	681
Количество терминов, выделенных методом, но не выделенных экспертами	790
Precision	0,69
Recall	0,65
F1 Score	0,67

Было осуществлено сравнение эффективности разработанной реализации предложенного метода с готовыми инструментами извлечения терминов RuATE [27] и ruterextract [28] на имеющемся размеченном датасете. RuATE представляет собой Python-реализацию ряда алгоритмов извлечения терминов, среди которых наиболее точным на тестовом наборе показал себя ComboBasic [29]. Ruterextract – Python-библиотека для извлечения ключевых слов из текстов на русском языке. Результатом ее применения является упорядоченный по частоте употребления список ключевых слов. Результаты сравнения представлены в таблице 3.

Таблица 4. Результаты сравнения метода с инструментами RuATE и ruterextract

Реализация	Precision	Recall	F1 Score
RuATE	0,16	0,08	0,11
ruterextract	0,38	0,14	0,2
Предложенный метод	0,69	0,65	0,67

Точность предложенного метода на тестовом наборе составила 69%, что значительно превышает результативность рассмотренных в качестве аналогов готовых инструментов извлечения терминов. Принимая во внимание трудноформализуемость решаемой задачи, такая точность предложенного метода, на наш взгляд, делает его применимым для практического использования.

**4. Обсуждение.** Следует отметить, что рассматриваемая в данной работе задача извлечения терминов при условии множественности и динамичности рассматриваемых предметных областей является трудноформализуемой, поскольку отсутствуют объективные критерии отнесения некоторой лексической конструкции

к числу терминов (предметных или общеупотребительных). В частности, существующие публикации по данной тематике демонстрируют расхождения во мнениях экспертов, поскольку на вопрос «что является предметным термином?» не всегда имеется однозначный ответ. Возможно, единственным объективным критерием «термичности» лексической конструкции является относительная частота ее употребления в предметных текстах и текстах общей тематики [30]. Однако такой подход к идентификации терминов неприменим для рассматриваемой в данной работе категории прикладных задач. Таким образом, в нашем случае невозможно обойтись без явных или опосредованных экспертных оценок. С другой стороны, прикладной мониторинг социальных медиа предполагает обработку больших объемов данных за ограниченное время, что повышает требования к уровню автоматизации решения соответствующих задач обработки информации.

Решение задач на основе анализа данных социальных медиа встречает схожие проблемы с соотношением предметных и общеупотребительных терминов. Например, в [31] предлагается система CiCui для извлечения терминов на основе машинного обучения с целью пополнения онтологий в рамках задач управления чрезвычайными ситуациями, в которые вовлечено большое количество людей и экспертов различных предметных областей с общими целями. В качестве источника текстов для анализа авторы используют социальные медиа. Сначала предлагаемая авторами система извлекает термины-кандидаты из предварительно обработанного текста при помощи разработанных ими лингвистических шаблонов и затем очищает полученный набор кандидатов с помощью статистических классификаторов, обученных на ранее размеченных данных. В работе [32] авторами предлагается модель для извлечения терминов и классификации текстов в рамках интеллектуального анализа социальных медиа для различных задач маркетинга, анализа конкурентной информации или рынка. Применение предлагаемой авторами модели рассматривается на примере задачи выявления и классификации компонентов автомобилей на основе форумных обсуждений, хотя она может быть использована и для других типов социальных медиа, предметных областей и задач. В таком контексте задачи структура форумов, в некотором роде, дает людям небольшую долю экспертного знания, направляя их в соответствии с их нуждами на специализированный раздел форума, но в дальнейшем люди общаются, используя ту терминологию, которой владеют, и необязательно экспертную. Сначала форумные обсуждения

классифицировались на основе структуры форумов, далее из обсуждений извлекались термины-кандидаты в виде униграмм. Затем для каждой выделенной на основе структуры форумов категории выбирались наиболее значимые термины с помощью таких методов, как критерий прироста информации, критерий хи-квадрат, корреляция релевантности документа и значение выбора Робертсона. Отобранные значимые термины, категории и сами обсуждения сохраняются в базу данных обсуждений, которая используется для обучения набора бинарных классификаторов, которые, в свою очередь, используются для классификации новых публикаций.

В сфере медицины также встречаются задачи, решаемые с помощью анализа данных социальных медиа. Например, [33] предлагают три метода извлечения концептов побочных реакций на лекарства с помощью LLDA (Labeled LDA), байесовского классификатора и условных случайных полей. В этой задаче важно связать интерпретацию побочных эффектов, выраженную общеупотребительными терминами, с медицинским термином, для чего и используются предлагаемые авторами методы. В [34] авторы предлагают алгоритм MSMTC для классификации текстов социальных медиа, использующий общеупотребительные термины, относящиеся к тематике здоровья. На первом этапе тексты преобразуются и размечаются с помощью имеющегося словаря медицинской терминологии. Далее обе задачи, классификации текстов и извлечение терминов, решаются одновременно с использованием двухканальной структуры и обучения генеративно-состязательной сети. Первый канал обрабатывает предложения, как они есть, а второй – со скрытой медицинской терминологией, в результате получая векторные представления предложений. Результат работы второго канала используется для извлечения важных слов, которые, как предполагается, считаются общеупотребительными терминами тематики здоровья, и эти извлеченные термины добавляются в начальный словарь. Результаты работы обоих каналов используются для классификации текстов с учетом извлеченной общеупотребительной терминологии.

Предложенный в данной работе метод извлечения терминов, основанный на комбинировании различных подходов, показал достаточно высокую эффективность на рассмотренной тестовой задаче. При этом стоит заметить, что, с одной стороны, совмещение различных по своей природе методов может обеспечить синергетический эффект, но, с другой стороны, возрастают затраты на реализацию, несмотря на то, что сильные стороны одних методов

могут компенсировать слабые стороны других. Так, необходимость в постоянном пополнении словаря новыми терминами может быть компенсирована отбором кандидатов с помощью NER или специальных правил. Однако правила для отбора кандидатов придется разрабатывать на основе анализа большого количества данных и исходя из решаемой задачи, следовательно, они не всегда могут быть использованы для решения других задач этим же методом. То есть, полностью не исключается необходимость использования экспертных знаний. Также стоит отметить, что часто методы анализа текстов, используемые в качестве компонентов составной технологии, ограничены рамками языка, для которого существуют необходимые инструменты (частеречевая разметка, средства реализации правил, NER решения) и ресурсы (размеченные корпуса, тезаурусы и т.п.).

Метод применим как для социальных медиа, так и для других источников, где преобладают текстовые данные и прослеживается структура «документ-отклик». При этом наличие в рамках социальной медиа или другого источника индикаторов пользовательской реакции не является принципиальным, но потенциально снижает качество извлечения терминов, поскольку значение обобщенной оценки интенсивности реакции на документ для всех документов будет идентичным (нулевым).

В качестве перспективы развития предложенного метода можно рассматривать оперирование не только «плоскими», но и вложенными сущностями. В частности, на этапе обработки кандидатов существует возможность решения задачи выделения вложенных терминов или сущностей (nested NER). Пример такой возможности можно наблюдать на рисунке 9, где можно было дополнительно выделить, в том числе, термин «Автомобильный парк скорой помощи Мурманской области».

**5. Заключение.** В данной работе рассмотрена проблема автоматизированного извлечения терминов из естественно-языковых текстов для последующего их использования при построении формализованных систем знаний (онтологий, тезаурусов, графов знаний) в рамках задачи мониторинга тематических обсуждений в социальных медиа.

Извлечение терминов является ключевой проблемой автоматизированного построения систем знаний на основе текстов на естественном языке, поскольку обеспечивает формирование базовой системы понятий, используемой затем в прикладных задачах интеллектуальной обработки информации. В то же время, данная задача относится к категории трудноформализуемых, поскольку

критерий качества ее решения зависит от прикладной задачи, в которой предполагается использовать результат.

Рассматриваемая в данной работе прикладная задача характеризуется тем, что в формируемую систему знаний необходимо включать как понятия из нескольких различных предметных областей, так и значимые общеупотребительные понятия, используемые аудиторией социальных медиа в рамках тематических обсуждений. Кроме того, формируемая система знаний является динамичной как с точки зрения состава охватываемых ею предметных областей, так и состава релевантных понятий, подлежащих включению в систему.

Для автоматизированного извлечения терминов из текстов в настоящее время создано большое количество методов и технологий, которые можно разделить на три основные категории – методы на основе правил, на основе словарей или внешних источников знаний, а также статистические. Однако прямое использование упомянутых методов для решения рассматриваемой здесь категории задач затруднительно, поскольку, как правило, все они предполагают наличие некоторой единственной предметной области, в рамках которой формируется множество терминов, обозначающих соответствующие понятия. Помимо этого, такие методы не принимают во внимание особенности, присущие текстам социальных медиа. Таким образом, применение только классических методов не покрывает все интересующие нас термины и, следовательно, не обеспечивает достаточной полноты в рамках решаемой задачи.

Для решения задачи извлечения терминов в указанных выше условиях, характерных для рассматриваемой категории прикладных задач, в данной работе предложен комбинированный метод, совмещающий в себе подходы на основе внешних источников знаний, инструментов NER и правил. Особенностью метода является учет структуры коммуникации и пользовательских реакций при принятии решения о включении некоторого термина в число значимых. Результаты экспериментов на тестовом наборе подтверждают эффективность предложенной комбинации подходов к извлечению терминов, позволяющей извлечь как предметные, так и значимые общеупотребительные термины, важные для задачи мониторинга и анализа тематических обсуждений в социальных медиа. Разработанный метод значительно превосходит по точности существующие инструменты извлечения терминов, реализованные в виде программных библиотек на языке Python.

**Литература**

1. Матусевич А.А. Общение в социальных сетях: прагматический, коммуникативный, лингвостилистический аспекты характеристики: дис. ... канд. филол. наук: 10.02.01. Киров. 2016. 190 с.
2. Mulkiewicz A., Marciniak M., Kupsc A. Rule-based information extraction from patients' clinical data // *Journal of Biomedical Informatics*. 2009. vol. 42. no 5. pp. 923–936.
3. Golik W., Bossy R., Ratkovic Z., Nedellec C. Improving term extraction with linguistic analysis in the biomedical domain // *RCS*. 2013. vol. 70. no. 1. pp. 157–172.
4. Zhou X., Zhang X., Hu X. MaxMatcher: Biological Concept Extraction Using Approximate Dictionary Lookup // *PRICAI 2006: Trends in Artificial Intelligence Lecture Notes in Computer Science*. 2006. pp. 1145–1149.
5. Yehia E., Boshnak H., AbdelGaber S., Abdo A., Elzanfaly. D.S. Ontology-based clinical information extraction from physician's free-text notes // *Journal of Biomedical Informatics*. 2019. vol. 98. no. 103276.
6. Lomov P., Malozemova M., Shishaev M. Training and application of neural-network language model for ontology population // *Software engineering perspectives in intelligent systems: Proceedings of 4th Computational Methods in Systems and Software*. 2020. vol. 1295. pp. 919–926.
7. Пимешков В.К., Диковицкий В.В., Шишаев М.Г. Формирование тренировочных наборов данных для нейросетевого классификатора в задаче извлечения понятий и отношений из естественно-языковых текстов // *Сборник Региональной научно-практической конференции-студенческой научной школы филиала МАГУ в г. Апатиты*. 2021. С. 158–170.
8. Pazienza M.T., Pennacchiotti M., Zanzotto F.M. Terminology Extraction: An Analysis of Linguistic and Statistical Approaches // *Knowledge Mining. Studies in Fuzziness and Soft Computing*. 2005. pp. 255–279.
9. Heid U. Extracting terminologically relevant collocations from German technical texts // *Terminology and Knowledge Engineering Proceedings*. 1999. vol. 99. pp. 242–255.
10. Zhang S., Elhadad N. Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts // *Journal of Biomedical Informatics*. 2013. vol. 46. no 6. pp. 1088–1098.
11. Lopez P., Romary L. HUMB: Automatic Key Term Extraction from Scientific Articles in GROBID // *Proceedings of the 5th International Workshop on Semantic Evaluation*. 2010. pp. 248–251.
12. GROBID. URL: <https://github.com/kermitt2/grobid> (дата обращения: 13.11.2023).
13. Lopez P., Romary L. GRISP: A Massive Multilingual Terminological Database for Scientific and Technical Domains // *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. 2010. pp. 2269–2276.
14. RuWordNet. URL: <https://ruwordnet.ru/ru/> (дата обращения: 13.11.2023).
15. RuThes. URL: <http://www.labinform.ru/pub/ruthes/> (дата обращения: 13.11.2023).
16. AGROVOC. URL: <https://www.fao.org/agrovoc/> (дата обращения: 13.11.2023).
17. DBpedia. URL: <https://www.dbpedia.org/> (дата обращения: 13.11.2023).
18. SpaCy. URL: <https://spacy.io/> (дата обращения: 07.12.2023).
19. Natasha. URL: <https://github.com/natasha/natasha> (дата обращения: 24.10.2023).
20. DeepPavlov. URL: <https://docs.deepavlov.ai/en/master/> (дата обращения: 07.12.2023).
21. Томита-парсер. URL: <https://yandex.ru/dev/tomita/> (дата обращения: 13.11.2023).
22. Yargy parser. URL: <https://github.com/natasha/yargy> (дата обращения: 24.10.2023).
23. Korobov M. Morphological Analyzer and Generator for Russian and Ukrainian Languages // *Analysis of Images, Social Networks and Texts. Communications in Computer and Information Science*. 2015. pp. 320–332.
24. OpenCorpora. URL: <https://opencorpora.org/> (дата обращения: 13.11.2023).



V. PIMESHKOV, M. NIKONOROVA, M. SHISHAEV  
**A COMBINED TERM EXTRACTION METHOD FOR THE  
PROBLEM OF MONITORING THEMATIC DISCUSSIONS IN  
SOCIAL MEDIA**

*Pimeshkov V., Nikonorova M., Shishaev M. A Combined Term Extraction Method for the Problem of Monitoring Thematic Discussions in Social Media.*

**Abstract.** Term extraction is an important stage in the automated construction of knowledge systems based on natural language texts, since it provides the formation of a basic concept system, which is then used in applied problems of intellectual information processing. The article discusses the problem of automated extraction of terms from natural language texts for their further use in the construction of formalized knowledge systems (ontologies, thesauruses, knowledge graphs) within the problem of monitoring thematic discussions in social media. This problem is characterized by the need to include in the formed knowledge system both concepts from several different domains, and some general concepts used by the audience of social media within thematic discussions. In addition, the generated knowledge system is dynamic both in terms of the composition of the domains it covers and the composition of relevant concepts to be included in the system. The use of existing classical methods for term extraction in this case is difficult, since they are focused on extracting terms within one domain. Based on this, to solve the problem under consideration, a combined method is proposed, combining approaches based on dictionaries, NER tools and rules. The results of the experiments demonstrate the effectiveness of the proposed combination of approaches to term extraction, which makes it possible to extract terms for the problem of monitoring and analyzing thematic discussions in social media. The developed method significantly exceeds the precision of the considered term extraction tools. As a further direction of research, the possibility of developing a method for solving the problem of identifying nested terms or entities is considered.

**Keywords:** text mining, term extraction, social media, knowledge extraction.

## References

1. Matusевич А.А. Обshнение в social'nyh setjah: pragmaticheskij, kommunikativnyj, lingvostilisticheskiy aspekty harakteristiki [Communication in social networks: pragmatic, communicative, linguostylistic aspects of characteristics]. Kirov. 2016. 190 p. (In Russ.).
2. Mykowiecka A., Marciniak M., Kupsc A. Rule-based information extraction from patients' clinical data. *Journal of Biomedical Informatics*. 2009. vol. 42. no 5. pp. 923–936.
3. Golik W., Bossy R., Ratkovic Z., Nedellec C. Improving term extraction with linguistic analysis in the biomedical domain. *RCS*. 2013. vol. 70. no. 1. pp. 157–172.
4. Zhou X., Zhang X., Hu X. MaxMatcher: Biological Concept Extraction Using Approximate Dictionary Lookup. *PRICAI 2006: Trends in Artificial Intelligence Lecture Notes in Computer Science*. 2006. pp. 1145–1149.
5. Yehia E., Boshnak H., AbdelGaber S., Abdo A., Elzanfaly. D.S. Ontology-based clinical information extraction from physician's free-text notes. *Journal of Biomedical Informatics*. 2019. vol. 98. no. 103276.
6. Lomov P., Malozemova M., Shishaev M. Training and application of neural-network language model for ontology population. *Software engineering perspectives*

- in intelligent systems: Proceedings of 4th Computational Methods in Systems and Software. 2020. vol. 1295. pp. 919–926.
7. Pimeshkov V.K. Dikovitsky V.V. Shishaev M.G. [Formation of training data sets for a neural network classifier in the problem of extracting concepts and relations from natural language texts] Sbornik Regional'noj nauchno-prakticheskoy konferencii-studencheskoj nauchnoj shkoly filiala MAGU v g. Apatity [Collection of the Regional scientific and practical conference-student scientific school of Apatity branch of MASU]. 2021. pp. 158-170. (In Russ.).
  8. Paziienza M.T., Pennacchiotti M., Zanzotto F.M. Terminology Extraction: An Analysis of Linguistic and Statistical Approaches. Knowledge Mining. Studies in Fuzziness and Soft Computing. 2005. pp. 255–279.
  9. Heid U. Extracting terminologically relevant collocations from German technical texts. Terminology and Knowledge Engineering Proceedings. 1999. vol. 99. pp. 242–255.
  10. Zhang S., Elhadad N. Unsupervised biomedical named entity recognition: Experiments with clinical and biological texts. Journal of Biomedical Informatics. 2013. vol. 46. no 6. pp. 1088–1098.
  11. Lopez P., Romary L. HUMB: Automatic Key Term Extraction from Scientific Articles in GROBID. Proceedings of the 5th International Workshop on Semantic Evaluation. 2010. pp. 248–251.
  12. GROBID. Available at: <https://github.com/kermitt2/grobid> (accessed 13.11.2023).
  13. Lopez P., Romary L. GRISP: A Massive Multilingual Terminological Database for Scientific and Technical Domains. Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). 2010. pp. 2269–2276.
  14. RuWordNet. Available at: <https://ruwordnet.ru/ru/> (accessed 13.11.2023).
  15. RuThes. Available at: <http://www.labinform.ru/pub/ruthes/> (accessed 13.11.2023).
  16. AGROVOC. Available at: <https://www.fao.org/agrovoc/> (accessed 13.11.2023).
  17. DBpedia. Available at: <https://www.dbpedia.org/> (accessed 13.11.2023).
  18. SpaCy. Available at: <https://spacy.io/> (accessed 07.12.2023).
  19. Natasha. Available at: <https://github.com/natasha/natasha> (accessed 24.10.2023).
  20. DeepPavlov. Available at: <https://docs.deeppavlov.ai/en/master/> (accessed 07.12.2023).
  21. Томита-парсер. Available at: <https://yandex.ru/dev/tomita/> (accessed 13.11.2023).
  22. Yargy parser. Available at: <https://github.com/natasha/yargy> (accessed 24.10.2023).
  23. Korobov M. Morphological Analyzer and Generator for Russian and Ukrainian Languages. Analysis of Images, Social Networks and Texts. Communications in Computer and Information Science. 2015. pp. 320–332.
  24. OpenCorpora. Available at: <https://opencorpora.org/> (accessed 13.11.2023).
  25. BERT NER-models. Available at: <https://docs.deeppavlov.ai/en/master/features/models/NER.html#6.-Models-list> (accessed 24.10.2023).
  26. Bird S., Klein E., Loper E. Natural Language Processing with Python. O'Reilly Media Inc. 2009. 482 p.
  27. Lu K. kevinlu1248/pyate: Python Automated Term Extraction (Version v0.5.3). Zenodo. 2021. DOI: 10.5281/zenodo.5039289.
  28. Ruterextract. Available at: <https://pypi.org/project/ruterextract/> (accessed 07.12.2023).
  29. Astrakhantsev N. ATR4S: toolkit with state-of-the-art automatic terms recognition methods in Scala. Language Resources and Evaluation. 2018. vol. 52. no. 3. pp. 853–872.

30. Hatty A., Schlechtweg D., Dorna M., im Walde S.S. Predicting Degrees of Technicality in Automatic Terminology Extraction. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. pp. 2883–2889.
31. Zhang X., Panizzon R., Musacchio M., Ahmad K. Terminology Extraction for and from Communications in Multi-disciplinary Domains. Proceedings of the LREC 2016 Workshop “EMOT: Emotions, Metaphors, Ontology and Terminology during Disasters”. 2016. pp. 34–39.
32. Abrahams A.S., Jiao J., Fan W., Wang G.A., Zhang Z. What’s buzzing in the blizzard of buzz? Automotive component isolation in social media postings. Decision Support Systems. 2013. vol. 55. no. 4. pp. 871–882.
33. Yates A., Goharian N., Frieder O. Extracting Adverse Drug Reactions from Social Media. Proceedings of the AAAI Conference on Artificial Intelligence. 2015. vol. 29. no. 1. DOI: 10.1609/aaai.v29i1.9527.
34. Liu K., Chen L. Medical Social Media Text Classification Integrating Consumer Health Terminology. IEEE Access. 2019. vol. 7. pp. 78185–78193.

**Pimeshkov Vadim** — Postgraduate student, research intern, Laboratory of information technologies for regional development management, IIMM KSC RAS. Research interests: natural language processing, knowledge extraction from natural language texts. The number of publications — 7. pimeshkov@iimm.ru; 24A, Fersman St., 184209, Apatity, Russia; office phone: +7(815)557-9689.

**Nikonorova Marina** — Postgraduate student, research engineer, Laboratory of information technologies for regional development management, IIMM KSC RAS. Research interests: natural language analysis, knowledge extraction, ontologies. The number of publications — 11. nikonorova@iimm.ru; 24A, Fersman St., 184209, Apatity, Russia; office phone: +7(815)557-9689.

**Shishaev Maxim** — Ph.D., Dr.Sci., Professor, Chief researcher, Laboratory of information technologies for regional development management, IIMM KSC RAS; Professor of the department, Informatics and computing engineering department, Apatity branch of MAU. Research interests: information systems, regional development, knowledge engineering, artificial intelligence, machine learning, natural language processing. The number of publications — 167. shishaev@iimm.ru; 24A, Fersman St., 184209, Apatity, Russia; office phone: +7(815)557-9248.

**Acknowledgements.** This work was supported by the Ministry of Science and Higher Education of the Russian Federation (No.122022800551-0, FMEZ-2022-0007). The authors express their gratitude to Ivan Vishnyakov, System administrator of IIMM KSC RAS, for providing the experimental dataset.

А.В. СМИРНОВ, А.В. ПОНОМАРЕВ, Н.Г. ШИЛОВ, Т.В. ЛЕВАШОВА,  
Н.Н. ТЕСЛЯ

## КОНЦЕПЦИЯ ПОСТРОЕНИЯ КОЛЛАБОРАТИВНЫХ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ: ПОДХОД И АРХИТЕКТУРА ПЛАТФОРМЫ

*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.**

**Аннотация.** В статье описывается общая концепция построения коллаборативных систем поддержки принятия решений, в которых коллективы, осуществляющие поддержку принятия решений, а) формируются гибко в соответствии с задачей и б) состоят как из людей-экспертов, так и из интеллектуальных агентов, реализующих те или иные методы искусственного интеллекта. Проводится анализ ключевых проблем создания коллаборативных систем поддержки принятия решений, основанных на взаимодействии человека и искусственного интеллекта. В частности, выделены следующие проблемы: обеспечение интероперабельности (взаимопонимания) между разнородными участниками коллектива, согласование различающихся позиций участников, обеспечение доверия между участниками, обеспечение эффективности планирования совместных действий и соблюдение баланса между предопределенными потоками работ и самоорганизацией. Сформированы принципы построения подобных систем, предлагающие решения выделенных проблем. В частности, предлагается онтолого-ориентированное представление информации о проблеме (применение мультиаспектных онтологий), набор методов для мониторинга деятельности команды, схема репутации, элементы объяснимого искусственного интеллекта, а также применение механизма ограниченной самоорганизации. Предложенная концепция положена в основу программной платформы для создания коллаборативных систем поддержки принятия решений, основные архитектурные положения которой также представлены в статье. Применение платформы иллюстрируется на примере из области рационального управления дорожной инфраструктурой и создания коллаборативной системы поддержки принятия решений для разработки мероприятий по снижению аварийности.

**Ключевые слова:** поддержка принятия решений, коллаборативные системы, онтологии, ограниченная самоорганизация, человеко-машинное взаимодействие.

**1. Введение.** Современный этап развития технологий интеллектуальной поддержки принятия решений характеризуется тремя относительно независимыми тенденциями. Первая тенденция, это активное применение искусственного интеллекта (ИИ) (в особенности, глубоких нейронных сетей, но не только) – во многих задачах нейросетевые агенты позволяют принимать более точные решения, чем это сделали бы люди (особенно, люди, обладающие средней квалификацией), при этом производительность подобных систем и их доступность оказываются крайне высокими.

Вторая тенденция – это коллаборативность, или необходимость совместной работы. Проблемы, возникающие при управлении сложными динамическими системами (большими коммерческими организациями и государственными структурами), требуют комплексного подхода, использования компетенций и знаний, не всегда присутствующих у одного эксперта или одной модели.

Наконец, третья тенденция – это динамический характер взаимодействий. Эта тенденция проявляется, в частности, через появление и развитие краудсорсинга и основанных на нем моделей сотрудничества. Особенность здесь заключается в том, что при быстром изменении внешних условий у современных субъектов хозяйствования может не быть организационных структур и ресурсов, готовых к принятию решений, и может потребоваться их оперативное динамическое формирование.

Соответственно, разработки, находящиеся на пересечении трех этих тенденций, представляют собой методы, технологии и инструментальные средства коллаборативной поддержки принятия решений, где коллаборации а) состоят как из людей-экспертов, так и из интеллектуальных агентов, реализующих те или иные методы ИИ, б) формируются гибко в соответствии с задачей.

В статье описывается разработанная авторами общая концепция построения подобных систем, включая подход к решению ключевых проблем обеспечения взаимодействия между разнородными участниками (людьми и ИИ-агентами). Предложенная концепция легла в основу программной платформы для создания коллаборативных систем поддержки принятия решений, основные архитектурные положения которой также представлены в статье. Применение платформы иллюстрируется на примере из области рационального управления дорожной инфраструктурой и создания коллаборативной СППР (КСППР) для разработки мероприятий по снижению аварийности.

**2. Обзор существующих исследований.** Предлагаемые в статье концепция построения КСППР и реализующая ее архитектура базируются на современных разработках в нескольких областях исследований, интегрируя их в уникальную комбинацию, обладающую качественно новыми свойствами. Охарактеризуем основные направления исследований и соответствующие результаты.

**2.1. Поддержка динамических гетерогенных коллективов.** Идея динамического формирования коллектива не связанных между собой организационными связями (или слабо связанных) людей, взаимодействующих через Интернет, и комплекс мер, нацеленных на

обеспечение качественного решения задачи таким коллективом, обычно обсуждается в рамках краудсорсинга или крауд-вычислений. В большинстве систем краудсорсинга участники индивидуально выполняют относительно простые задачи, играя роль своеобразного «вычислительного устройства» [1], следуя заранее определенной процедуре. Показательным примером здесь является концепция «человеко-машинной облачной среды», предлагающей абстрагирование человеческих ресурсов аналогично вычислительным [2]. Однако в рамках адаптации краудсорсинга к более сложным задачам (чем, например, не требующая особой квалификации разметка изображений) показано, что для сложных задач заранее запрограммированная процедура (рабочий процесс) может существенно ограничивать эффективность группы (например, [3]), поэтому потенциал краудсорсинга применительно к сложным задачам тесно связан с механизмами динамического формирования команд и адаптации процедур их совместной работы «на лету». Применение крауд-вычислений для решения сложных задач требует гибкости процедур и способности участников корректировать их в определенных рамках. В этом направлении имеется несколько исследовательских работ (например, [4 – 6]), здесь одной из значимых разработок является идея «флэш-организации» (flash organization) [4]. Однако такие «флэш-организации» состоят только из людей, а процессы обработки информации и принятия решений включают как людей, так и ИИ-агентов.

**2.2. Поддержка коллаборативных процессов.** Совместная работа, в том числе, направленная на принятие решений или на поддержку принятия решений, является многоуровневой активностью, поскольку включает в себя согласование позиций различных участников, как на синтаксическом, так и на семантическом (а порой и на телеологическом или ценностном уровнях) [7, 8]. При рассмотрении вопроса синтаксического и семантического согласования в рамках гетерогенных человеко-машинных коллективов особого внимания заслуживают результаты проекта Dicode, реализованного в рамках европейской программы FP7-ICT [9]. Важность этих результатов заключается в том, что для обмена информацией о задаче между экспертами и ИИ-агентами эта задача должна иметь структурированное представление, одним из вариантов которого является граф связи аргументов. Однако явное кодирование всех аргументов может оказаться достаточно сложным.

Проблеме поддержки процессов сотрудничества и повышения их эффективности уделяется большое внимание как в области теории

принятия решений и управления в целом, так и в области систем коллективной поддержки принятия решений.

Одной из распространенных практик для повышения эффективности процессов сотрудничества является выделение в команде специальной роли – фасилитатора [10]. Исследования групповой работы привели к появлению ряда методологий, которые должны использовать фасилитаторы (например, дизайн-мышление [11]). Однако нехватка опытных фасилитаторов является ограничивающим фактором, особенно в случае крупномасштабных систем коллективного интеллекта, основанных на свободном участии (например, массовом) [12].

Поиску путей преодоления этого ограничения посвящены, например, следующие отрасли исследований: проектирование сотрудничества [13] и автоматизированная фасилитация [11, 12, 14].

Подход к поддержке коллаборации, реализованный в платформе, представленной в этой статье, предполагает структурированное представление обсуждения (например, [12]), однако конкретное представление значительно отличается от тех, что рассматриваются в существующих исследованиях. В предлагаемой платформе информация о проблеме представлена с помощью нескольких онтологий (основанных на методологиях поддержки принятия решений) [15], в то время как существующие исследовательские работы рассматривают структурированное представление, типичное для систем обсуждения [12, 14].

### **2.3. Взаимодействие между человеком и ИИ-агентом.**

Существует множество статей, в которых исследуются сценарии сотрудничества между ИИ-агентами и людьми-экспертами в ходе поддержки принятия решений [16 – 25].

В рамках такого сотрудничества ИИ-агенты обычно решают задачи анализа больших объемов информации, прогнозирования, планирования, генерации рекомендуемых решений, предотвращения опасных ситуаций, представления текущей ситуации в удобочитаемой форме и другие. Человек проверяет корректность работы ИИ-агентов, обучает их и принимает окончательные решения. Эксперты и ИИ-агенты могут совместно решать задачи сбора информации о текущей ситуации и построения контекста.

В результате анализа исследований, посвященных взаимодействию ИИ-агентов с человеком, было выделено пять типов взаимодействий [17, 26]: информирование (инициированное ИИ-агентом), запрос к агенту ИИ (инициируемый экспертом), ответ (не сопровождаемый объяснениями ответ ИИ-агента эксперту),

объяснение (ответ ИИ-агента на запрос эксперта, сопровождаемый объяснениями) и обучение ИИ-агента человеком.

В результате анализа были выделены четыре группы КСППР: 1) человек принимает решения, ИИ-агенты помогают человеку; 2) человек принимает решения и обучает ИИ-агентов; 3) человек является лицом, принимающим решения (ЛПР), и вместе с ИИ-агентами участвует в процессах взаимного обучения; 4) любой участник команды, человек или ИИ-агент, может принимать решения. Платформа, представленная в этой статье, охватывает первые три сценария – во всех сценариях решения принимаются человеком.

В сценариях с высокой ценой ошибки сотрудничество человека и ИИ предполагает взаимодополняющую деятельность, в которой индивидуальные сильные стороны человека и ИИ-агентов объединяются для оптимизации результата совместной деятельности, в отличие от принципа разделения труда, где сильные стороны людей и ИИ-агентов используются для принятия окончательного решения в соответствии с выбранной стратегией принятия решений. В сценариях, где задачи, которые будут решать ИИ-агенты, известны, сотрудничество отводит роль ЛПР человеку и обеспечивает получение обратной связи от человека, способного обучать ИИ-агентов. Сценарии, основанные на принципах равенства между людьми и ИИ-агентами, реализуют принципы делегирования задач (решений) и могут использоваться, если участники способны адекватно оценивать свои возможности и возможности друг друга в решении задач.

**3. Подход к построению КСППР.** Ключевые решения по организации КСППР, основанной на взаимодействии людей-экспертов и ИИ-агентов, диктуются проблемами (вызовами), характерными как для коллаборативных систем в целом, так и для взаимодействия между человеком и ИИ в широком смысле. В данном разделе рассмотрены основные такие проблемы и их решения, предлагаемые в рамках подхода к организации КСППР.

**3.1. Проблемы организации КСППР, основанных на взаимодействии человека и ИИ.** Первой фундаментальной проблемой организации КСППР является проблема обеспечения интероперабельности (или взаимопонимания) между разнородными участниками команды, вовлеченной в коллаборативную поддержку принятия решений. Интероперабельность должна обеспечиваться на разных уровнях – начиная с синтаксического, представляющего знаковую основу, одинаково интерпретируемую разнородными участниками команды, заканчивая семантическим, заключающимся

в одинаковом понимании текущей ситуации, состава команды и ее цели. Для решения этой проблемы к настоящему времени разработан широкий спектр подходов, начиная от специализированных протоколов и языков представления знаний, заканчивая подходами, основанными на анализе естественного языка.

Второй проблемой организации КСППР является согласование различающихся позиций. Даже если первая проблема успешно решена, и каждый участник команды (независимо от своей природы) одинаково понимает задачу и цель принятия решения, могут возникнуть существенные различия в предлагаемых разными участниками наборах альтернатив и в их оценке. С одной стороны, эта ситуация абсолютно нормальна для КСППР, потому что сама коллаборативность привносится в нее именно в силу различий в имеющейся у участников информации и в компетенциях участников. Вместе с тем, в результате работы команды должно быть предложено некоторое согласованное решение, что требует применения специальных процедур обсуждения и, возможно, корректировки начальных позиций участников.

Третья проблема – доверие. В определенном смысле доверие также оказывается связанным с согласованием, поскольку при различии позиций восприятие позиции другого участника может очень сильно варьироваться в зависимости от имеющегося образа этого участника. Однако это является и самостоятельной проблемой, поскольку доверие важно не только для согласования различающихся позиций, но и на всех других этапах функционирования команды. В свою очередь, проблема доверия в КСППР, основанных на участии человека и ИИ, естественным образом декомпозируется на доверие между людьми-экспертами и доверие между экспертами и ИИ-агентами. Данная проблема хорошо известна и для нее предложен ряд как общих решений, включающих репутационные схемы [27], так и более частных, к которым можно отнести механизмы отслеживания источников информации (provenance) и объяснения (что оказывается особенно актуально для формирования доверия к ИИ-агентам) [28].

Наконец, четвертая проблема – это поиск эффективной степени гибкости планирования действий группы. Как уже указывалось в разделе 2, полностью предопределенные процессы оказываются наиболее эффективны применительно к простым, рутинным задачам, но не очень подходят для сложных задач принятия решений, когда последовательность необходимых действий не может быть определена заранее.

Следует заметить, что данные проблемы можно считать базовыми (основополагающими), их также можно конкретизировать и уточнять, как это сделано, например, в работе [29]. Однако уточненный перечень, в конечном итоге, определяется перечисленными базовыми проблемами.

**3.2. Принципы построения КСППР.** Предлагаемые принципы построения КСППР, с одной стороны, определяют контекст применения рассматриваемого класса КСППР, с другой – предлагают ответ на перечисленные выше проблемы организации КСППР.

Процесс поддержки принятия решений осуществляет команда, состоящая из людей-экспертов и интеллектуальных агентов (программных объектов, действующих на основе ИИ, ИИ-агентов). Команда создается для решения конкретной проблемы поддержки принятия решений, после решения которой она расформируется. Пул потенциальных участников обеспечивает доступ к совокупности возможных участников, которыми могут быть сотрудники предприятия, члены профессионального сообщества и т.д., а также ИИ-агенты. В задачах поддержки принятия решений члены команды часто не взаимозаменяемы и обладают уникальными способностями, поэтому важно отобрать участников с достаточными и взаимодополняющими навыками и знаниями, чтобы они эффективно работали над проблемой.

За окончательное решение отвечает конечный пользователь (ЛПР), который формулирует проблему и инициирует организацию команды. Команда обеспечивает поддержку принятия решения посредством набора действий, направленных на выявление критериев и альтернатив, оценку альтернатив и определение важности критериев. Реализация этих действий может привести к задачам сбора и обработки информации о сформулированной проблеме, а также к необходимости оценки гипотетических сценариев поддержки принятия решений.

Для решения проблемы интероперабельности применяется структурированное представление информации о проблеме и совместной деятельности (командной работе) в форме онтологии. В ИТ-индустрии с момента появления концепции Семантической паутины онтологии зарекомендовали себя как достаточно эффективный инструмент решения проблем интероперабельности. Для смешанных человеко-машинных команд онтологии могут обеспечивать представление, приемлемое как для людей-экспертов, так и для ИИ-агентов, и гарантировать функциональную совместимость разнородных участников команды. Поскольку прямое

использование языка онтологий может потребовать определенных знаний в области логики и инженерии знаний, использование онтологий может дополняться некоторым интуитивно понятным интерфейсом конечного пользователя. Очевидно, что построение новой онтологии для каждого процесса поддержки принятия решений нецелесообразно, но основной онтологией, необходимой для представления требуемых знаний, является онтология проблемной области, определяющая понятия проблемной области и связи между ними. Наличие онтологии проблемной области является обязательным условием для предлагаемого подхода. За последние десятилетия для многих областей было разработано множество высококачественных онтологий и КСППР могут использовать их. Следует отметить, что различные процессы поддержки принятия решений в одной и той же области могут использовать одну и ту же онтологию проблемной области.

Поскольку в ходе работы над проблемой команда оперирует несколькими видами взаимосвязанной информации, отражающей различные взгляды на проблему, предлагается применение аппарата мультиаспектных онтологий. Информационный фрагмент (факт, утверждение) может играть несколько ролей в командной работе. Это, в частности, вытекает из сложной структуры команды и обуславливает необходимость использования мультиаспектных онтологий, обеспечивающих непротиворечивое представление информации с позиций различных аспектов проблемы и взглядов участников команды [30]. Мультиаспектная онтология включает три уровня: локальный, аспектный и глобальный. Каждый аспект может быть представлен определенным формализмом. Локальный уровень представляет концепции и отношения, наблюдаемые только с одной точки зрения. Уровень аспекта представляет концепции и отношения локального уровня, которые являются общими для двух или более аспектов. Он определяет формализм мультиаспектной онтологии. Глобальный уровень является общей частью для аспектов мультиаспектной онтологии, представленной с помощью формализма мультиаспектной онтологии. Понятия, представленные на этом уровне, связаны с понятиями аспектного уровня. В данной работе используется аспект поддержки принятия решений, предложенный в составе многоаспектной онтологии в [30].

Для решения проблемы согласования различающихся позиций применяются компоненты, осуществляющие мониторинг деятельности команды. Конфликтные ситуации, которые могут возникнуть в ходе совместной деятельности, распознаются и разрешаются. Это касается

различных трудностей и ошибок в понимании текущей ситуации, намерений и целей участников и т.д.

Для решения проблемы взаимного доверия применяется комплекс мер, включающий как схему репутации, используемую в ходе формирования команды, так и элементы объяснимого ИИ. В частности, поскольку взаимодействие основано во многом на использовании онтологий, то речь идет, в первую очередь, о применении онтолого-ориентированного объяснимого ИИ. Так, важными аспектами взаимодействия между участниками являются объяснение результатов, полученных ИИ-агентами, человеку и обучение, т.е. передача знаний от человека-эксперта ИИ-агенту. Оба эти процесса основаны на онтологическом представлении информации и предполагают существование онтолого-ориентированных методов объяснения и использования онтологий как априорных знаний. Если задача (некоторая часть всей проблемы поддержки принятия решений), решаемая агентом ИИ, может быть выражена как онтологический вывод, то этот вывод (возможно, переформулированный на естественном языке) является действительным объяснением. Это достаточно мощный метод, и недавние публикации показывают, что в некоторых случаях его можно использовать даже с нейросетевыми агентами (например, [31]). Однако есть и другие методы. Например, концепты онтологии могут быть связаны с исходными признаками модели или их комбинациями, что делает традиционные методы объяснения моделей машинного обучения более понятными (например, [32]).

Наконец, для решения проблемы определения эффективной гибкости процесса планирования предлагается использование ограниченной самоорганизации [33]. Особенностью этого подхода является то, что, с одной стороны, команда может гибко планировать свою деятельность, с другой, присутствует набор механизмов, «мягко» направляющих деятельность команды посредством социальных норм и рекомендаций.

**3.3. Концептуальная модель КСППР.** Разработанные принципы нашли отражение в обобщенной концептуальной модели КСППР (рисунок 1).



Рис. 1. Обобщенная концептуальная модель КСППР

Исходной точкой для функционирования такой СППР является взаимодействие между ЛПР и командой, состоящей из людей-экспертов и ИИ-агентов. ЛПР формулирует проблемную ситуацию для разрешения которой – поиска возможных решений и их оценки – формируется команда, состоящая из экспертов-людей и из ИИ-агентов (программные компоненты, реализующие какие-либо интеллектуальные функции обработки информации и способные принимать участие в коллективной работе над задачей). ЛПР может управлять процессом формирования команды и координировать ее работу. Команда, в свою очередь, формируется динамически под конкретную проблемную ситуацию. При этом, основными характеристиками команды являются [34]: модель организации команды (схема подчинения), способы распределения задач и ответственности.

Работа над проблемной ситуацией декомпозируется в решение частных задач и интеграцию их результатов. Основные типы таких задач выделены, например, в работе [35]: распознавание, предсказание, решение и действие. Эти типы могут быть использованы, в частности, в описании ИИ-агентов, для облегчения автоматизированного распределения задач.

Команда для поддержки принятия решения ЛПР при работе с заданной проблемной ситуацией формируется в рамках выбранной модели организации команды из экспертов и ИИ-агентов, каждый из которых характеризуется набором возможностей по решению тех или иных задач. Характеристики команды в целом (модель организации, распределения задач и ответственности) определяют ее состав и правила взаимодействия участников, они используются как при формировании (например, подборе участников, определении ролей), так и транслируются участникам в виде норм и правил взаимодействия, которые могут носить как неформальный характер, так и подкрепляться функциями платформы, обеспечивающей совместную работу. Участие в команде, как правило, характеризуется ролью, определяемой как совокупность обязанностей по решению тех или иных задач и ответственности за результат.

Основными процессами, обеспечивающими работу над решением проблемы, является обмен информацией и знаниями (в результате которого снижается общая неопределенность, связанная с проблемной ситуацией), обучение, понимаемое как передача норм и навыков, а также достижение согласия между участниками команды. Конкретные механизмы реализации этих процессов определяют итоговую конфигурацию КСППР.

В основе предлагаемого подхода к построению КСППР лежит представление проблемной ситуации (и прочей информации, имеющей отношение к принятию решения) с помощью нескольких онтологий, раскрывающих различные аспекты проблемы: терминология и закономерности проблемной области, структура задачи принятия решения как таковой, логика аргументации и прочие. Данные аспекты могут проблемно-ориентированным образом объединяться в единую онтологию с помощью аппарата мультиаспектных онтологий [30, 36].

Представленная подобным образом информация о проблеме содержится в онтолого-ориентированном интеллектуальном пространстве [37, 38] (доступная участникам системы «классная доска», хранящая множество высказываний в терминах онтологии и обеспечивающая удобные механизмы доступа и оповещения), к которому имеют доступ как ИИ-агенты, так и эксперты.

Достоинством онтолого-ориентированного представления является то, что информация оказывается доступна как людям (напрямую или через специализированный интерфейс), так и ИИ-агентам, которые могут использовать возможности структурированных языков запросов (например, SPARQL) для извлечения релевантных высказываний из описания текущего состояния проблемы.

Взаимодействие между участниками команды, осуществляющей поддержку принятия решений, происходит опосредованно – через изменение содержимого интеллектуального пространства.

ИИ-агенты в предлагаемой схеме могут реализовывать широкий спектр алгоритмов по поиску и обработке информации. По характеру исходной и результирующей информации можно выделить две группы таких агентов: агенты первого вида осуществляют преобразования символической информации на основе знаний, то есть, и на входе, и на выходе у этих агентов символическая информация. Агенты же второго вида осуществляют преобразование «сырых» данных в символическую информацию. На вход таким агентам попадают изображения, видео-, аудиоданные или просто численные наблюдения, на выходе же у них высказывания, основанные на онтологии проблемной области. Для организации подобных агентов оказываются особенно востребованными технологии нейро-символического интеллекта [39], поскольку эта парадигма позволяет получать символическую интерпретацию производимых преобразований, что оказывается важным для организации коммуникации (в том числе, с учетом объяснения).

Одним из существенных аспектов коллаборативной поддержки принятия решений является отслеживаемость (provenance), понимаемая как возможность отследить источник для каждого утверждения (касающегося перечня альтернатив или их оценок), которые сформулированы в ходе совместной работы над задачей. Подобная возможность играет ключевую роль, например, в формировании и использовании механизмов доверия, существенно снижающих накладные расходы на верификацию. В предлагаемой концептуальной модели отслеживаемость обеспечивается двумя механизмами. Во-первых, все изменения в представлении проблемы снабжаются метаданными в соответствии с онтологией PROV-O [40], позволяющей описать источники информации (участника системы, предоставившего информацию, основания); во-вторых, агенты помимо самого результата работы добавляют и объяснение – почему именно такой результат был получен (для агентов, осуществляющих преобразование символической информации, это цепочка логического

вывода, а для агентов, осуществляющих преобразование «сырых» данных в символичные – это те или иные виды онтолого-ориентированного объяснения).

#### **4. Архитектура платформы для создания КСППР.**

Предложенная концептуальная модель нашла воплощение в архитектуре платформы для создания КСППР, основные положения которой представлены в данном разделе статьи. Платформа предоставляет как набор инструментов, обеспечивающих создание КСППР, ориентированных на принятие решений в конкретной проблемной области, так и набор базовых сервисов, поддерживающих функционирование таких КСППР.

**4.1. Пользователи и сценарии.** Выделяется три категории пользователей с соответствующими ролями: *Конечный пользователь* (ЛПР), *Участник (эксперт)*, и *Разработчик агентов*. *Конечный пользователь* обращается к КСППР (функционирующей на платформе) с некоторой задачей, решение которой требует совместных усилий команды людей-экспертов и ИИ-агентов. *Участник (эксперт)* предпринимает действия, способствующие решению задачи. *Разработчик агентов* регистрирует ИИ-агентов, которые, используя механизмы платформы, также могут принимать участие в решении задачи.

Значительное количество сценариев, обеспечивающих собственно взаимодействие в рамках команды (просмотр информации о проблеме, ввод сведений, просмотр плана и его уточнение и некоторые другие), являются общими для двух ролей – *Конечный пользователь (ЛПР)* и *Участник (эксперт)*. Специфическим сценарием, характерным для *Конечного пользователя*, является формулировка проблемы, включающая создание исходного текстового описания проблемы, для которой требуется использование КСППР, а также задание дополнительных параметров (временные ограничения, политика компенсации (если применимо) и других). В результате выполнения этого сценария 1) в системе регистрируется новая проблема, и для нее создается рабочее пространство, 2) участникам, привлечение которых к решению проблемы может оказаться перспективным, рассылаются приглашения. Под рабочим пространством понимается комплекс элементов интерфейса и связанных с ними функций, направленных на отображение состояния решаемой проблемы и дополнение его новой информацией. Специфическими вариантами использования, характерными для *Участника (эксперта)*, является заполнение профиля участника и работа с поступившими приглашениями присоединиться к решению

задач. Профиль участника объединяет в себе информацию как сообщенную пользователем, так и обобщающую опыт его участия в решении задач, и используется для подбора участников в ходе рассылки приглашений. Основным параметром профиля, который *Участник* может корректировать, является описание имеющихся компетенций (как на естественном языке, так и в привязке к онтологии компетенций).

Важным строительным блоком организации взаимодействия, реализующим идеи ограниченной самоорганизации, является понятие *Действие* (или action point). Это действие, которое должен совершить тот или иной участник в ходе работы над задачей. Оно служит как для реализации элементов планирования работ (обязательства предоставить тот или иной артефакт), так и при реализации координационных протоколов (например, голосования). *Действия* создаются в результате интерпретации высказываний участников системой или явным образом. Прецедент *Выполнение действия* представляет собой взаимодействие с элементом пользовательского интерфейса, представляющим соответствующее *Действие*. Характер взаимодействия может варьироваться в зависимости от типа *Действия* (например, для голосования – это выбор одного из представленных вариантов, для загрузки артефакта – выбор файла и нажатие кнопки загрузки).

Основными прецедентами, доступными для роли *Разработчик агентов*, являются создание ИИ-агента, загрузка очередной версии, просмотр информации об ИИ-агентах и управление их состоянием.

Сценарии, доступные ИИ-агентам, связаны с получением доступа к онтолого-ориентированному представлению содержимого рабочего пространства. Так, агенты могут оформлять подписку на определенные онтологические конструкции (подграфы) и активироваться в тех случаях, когда вследствие изменения содержимого рабочего пространства множество подграфов, соответствующих подписке, изменяется. Кроме того, сервисы могут осуществлять запросы к онтолого-ориентированному представлению рабочего пространства и вносить в него изменения.

**4.2. Компоненты.** Исполнение компонентов системы происходит на трех вычислительных узлах: веб-сервере, сервере баз данных и пользовательском устройстве (рисунок 2).

На веб-сервере размещается не только компонент, обеспечивающий отображение страниц веб-приложения, но и все компоненты, связанные с управлением интеллектуальным пространством, исполнением ИИ-агентов (менеджер ИИ-агентов,

среда выполнения ИИ-агентов), формированием команды, рекомендацией паттернов взаимодействия, интерпретацией естественного языка, а также инфраструктура для отложенного выполнения задач.

На сервере баз данных размещается СУБД, под управлением которой работает сама база данных веб-приложения.

Наконец, взаимодействие с клиентом происходит через веб-браузер, запущенный на устройстве пользователя и предоставляющий вычислительную среду для выполнения кода, связанного с внешним видом и логикой поведения некоторых элементов интерфейса.

Компонент формирования команды предоставляет набор алгоритмов для создания новых команд и расширения существующих. Он поддерживает набор примитивов, таких как компетентность, роль, история участия, и конкретная КСППР, созданная с помощью платформы, может настроить политику формирования команды, используя эти примитивы и заранее определенные политики, или даже реализовать новую политику, адаптированную к проблемной области.

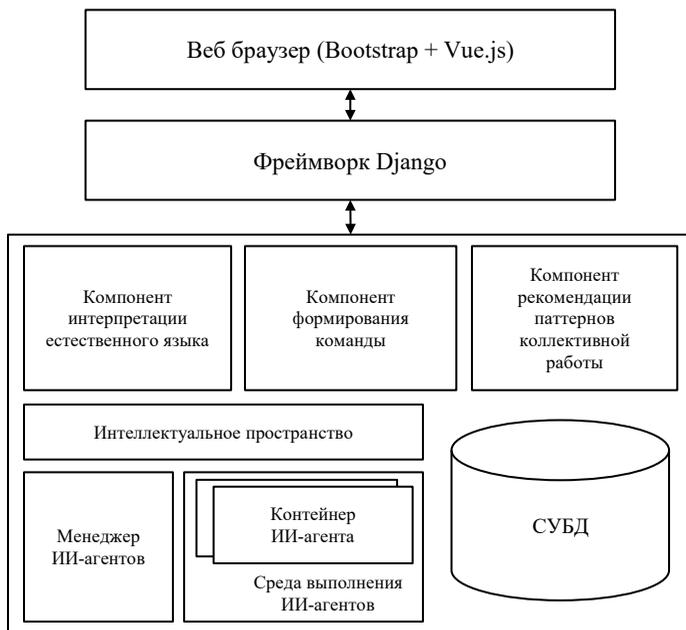


Рис. 2. Архитектура платформы

Компонент интерпретации естественного языка играет важную роль в анализе и преобразовании текстовых обсуждений в структурированное представление. В частности, он определяет роли отдельных сообщений и фрагментов информации (если это альтернативное описание, оценка) и их взаимозависимости (подтверждающие утверждения, противоречащие утверждения). Результирующее структурное представление согласуется с участником команды в интерактивном режиме и затем может использоваться ИИ-агентами, что устраняет необходимость каждому агенту анализировать текстовый поток.

Компонент рекомендации паттернов коллективной работы предназначен для идентификации непродуктивных ситуаций в коллективной работе (поляризации мнений, преждевременной фокусировки на одной из возможных альтернатив) и рекомендации коллективных действий по выходу из подобных ситуаций. Для решения этой задачи компонент осуществляет периодический анализ текущей ситуации обсуждения (ее структурированного представления) и, используя знания в области совместной работы и модель машинного обучения, выявляет непродуктивные ситуации и предлагает способы их преодоления. Сами рекомендуемые паттерны (способы преодоления непродуктивных ситуаций), извлеченные из литературы, сохранены в виде базы знаний. Для идентификации паттернов, соответствующих ситуации, компонент рекомендации использует два метода: 1) метод, основанный на сопоставлении с образцом, 2) метод, основанный на машинном обучении. Первый метод опирается на совокупность правил оценки ситуаций, которая также была установлена на основе анализа профильной литературы, он не требует фактических накопленных данных о деятельности группы. Второй метод (основанный на машинном обучении) требует наличия определенной истории решения задач. Оба метода опираются на формальное представление текущей ситуации обсуждения – граф, вершины в котором соответствуют альтернативам, критериям, оценкам альтернатив. Так, в первом методе применяется поиск в этом графе подграфа, соответствующего непродуктивной ситуации, а во втором – граф дополняется признаками (например, количество альтернатив, количество голосов в пользу каждой и пр.) и используется для обучения модели, определяющей целесообразность применения того или иного паттерна.

Интеллектуальное пространство содержит онтолого-ориентированное представление текущего состояния и обеспечивает доступ к нему со стороны ИИ-агентов. Интеллектуальное пространство реализовано на языке Java с применением библиотеки Jena.

Выполнение ИИ-агентов происходит с использованием технологии контейнеризации (в частности, Docker) – для каждого ИИ-агента выделяется свой Docker-контейнер, доступ к состоянию проблемы производится через среду выполнения ИИ-агентов. Развертывание, выполнение и мониторинг ИИ-агентов контролируется менеджером агентов.

Все компоненты платформы (кроме интеллектуального пространства) реализованы на языке Python 3, оперативная база данных работает под управлением СУБД PostgreSQL.

В таблице 1 показана связь между основными проблемами, возникающими при построении КСППР, принципами, лежащими в основе предлагаемого подхода, а также компонентами и архитектурными решениями платформы создания КСППР.

Таблица 1. Реализация основных принципов в архитектуре

<b>Проблема</b>	<b>Принцип</b>	<b>Компонент/ Механизм информационного взаимодействия</b>
Интероперабельность	Структурированное (онтолого-ориентированное) представление информации	Интеллектуальное пространство, Компонент интерпретации естественного языка
Согласование различающихся позиций	Мониторинг работы команды, идентификация непродуктивных ситуаций и рекомендация паттернов коллективных действий	Компонент рекомендации паттернов взаимодействия
Доверие	Схема репутации, история участия, Объяснимый ИИ	Компонент формирования команды, Интерфейс взаимодействия с ИИ-агентом, поддерживающий предоставление объяснений
Баланс между гибкостью и контролем	Ограниченная самоорганизация	Совокупность средств, обеспечивающих возможность гибко планировать работы (схема репутации, рекомендация паттернов коллективной работы)

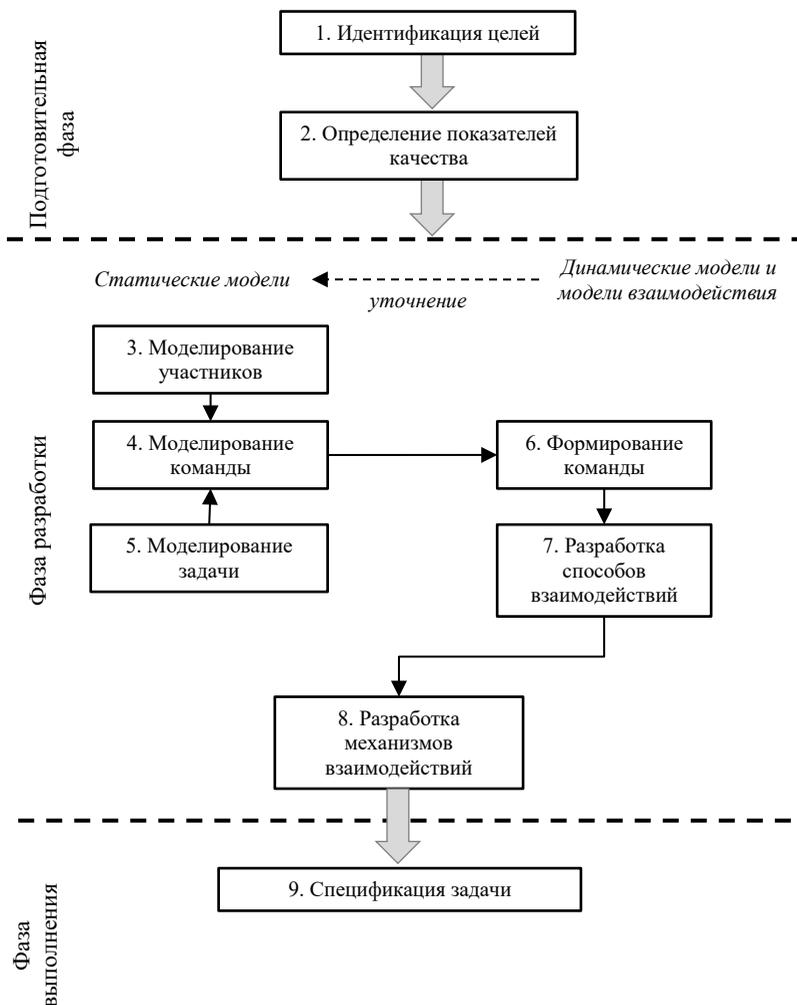


Рис. 3. Методология разработки КСППР

**4.3. Создание КСППР на основе платформы.** Процесс разработки коллаборативной СППР, в которой сотрудничают эксперты и ИИ-агенты, состоит из трех фаз (рисунок 3).

1) Подготовительная фаза. Основными задачами данной фазы являются идентификация целей КСППР, выбор участников и определение критериев для последующей оценки системы.

2) Фаза разработки. Это основная фаза. На ней создаются и определяются основные модели и механизмы. Эти модели и механизмы сильно зависят от возможностей коллаборативной человеко-машинной среды.

3) Фаза выполнения. На данной фазе реализуются процессы поддержки принятия решений, то есть осуществляется совместная деятельность участников по разрешению проблемной ситуации, сформулированной ЛПП на основе разработанных (на фазе разработки) моделей; совместная деятельность поддерживается выбранными или разработанными (на фазе разработки) механизмами. Далее по тексту проблемная ситуация рассматривается как задача ЛПП.

Распределение различных видов деятельности участников по фазам выглядит следующим образом.

1) Идентификация целей СППР. Первоочередная цель любой СППР – обеспечение качества поддержки принятия решений. Тем не менее, СППР может преследовать и дополнительные цели, например, развитие компетенций участников (применимо для внутрикорпоративной СППР).

2) Определение измеримых показателей качества для оценки идентифицированных целей. Показатели качества предлагаются различными методами и инструментами, которые также оказывают поддержку в их измерении. Например, модель оценки качества решений [41] может быть использована соответственно для оценки качества решений, модель для измерения эффективности СППР [42] предлагает методы и измеряемые факторы для оценки эффективности СППР, общая диагностическая система [43] предоставляет оценочные концепты для измерения и оценки компетенций и т.д.

3) Моделирование участников. Деятельность заключается в построении моделей участников (профилей). Здесь определяются основные характеристики экспертов и ИИ-агентов, которые влияют на эффективность команды, участниками которой они являются. Например, компетенции участников, их доступность и другие.

4) Моделирование команды. Данный вид деятельности связан с созданием профиля команды и определением ее основных характеристик. В зависимости от решаемой командой задачи, механизма координации, процедуры формирования команды и ее размера [44] возможен выбор между различными моделями команд. Например, модель Катценбаха и Смита [45] подходит для команд с небольшим количеством участников, которые объединились на определенное время, чтобы решить конкретную задачу; модель ЛаФасто и Ларсона [46] может быть использована, когда индивидуальные характеристики

участников (личные черты, навыки, способности и т.п.) являются основой для формирования команды; модель Роббинса и Джуди [47] полезна для согласования четырех аспектов производительности команды (контекст, состав, проект работы, процесс).

5) Моделирование задачи. Деятельность направлена на построение модели задачи, которая будет решаться командой как задача поддержки принятия решений. Она связана с разработкой или адаптацией онтолого-ориентированных моделей для а) спецификации задачи ЛПП как онтолого-ориентированного представления знаний проблемной области; б) спецификации задачи ЛПП как онтолого-ориентированного представления задачи поддержки принятия решений; в) онтолого-ориентированного представления структуры дискуссии как последовательности взаимодействий участников.

6) Формирование команды. Деятельность связана с разработкой процедуры формирования команды участников, которая учитывает цели системы, показатели качества и основные характеристики участников.

7) Разработка способов взаимодействий. Деятельность направлена на разработку или выбор способов взаимодействий участников, например, взаимодействия, основанные на ролях участников, непосредственные взаимодействия и другие.

8) Разработка механизмов взаимодействий. Данный вид деятельности связан с разработкой или выбором механизмов, которые поддерживают выбранные или разработанные в процессе предыдущей деятельности виды взаимодействий. Такие механизмы определяются моделью взаимодействия.

9) Спецификация задачи. Деятельность нацелена на использование механизмов мультиаспектной онтологии для получения обще-интерпретируемой модели задачи ЛПП.

Как сказано выше, фаза разработки является основной фазой процесса разработки коллаборативной СППР. Деятельность на этой фазе включает в себя два вида деятельности: 1) разработка статических (информационных) моделей для основных объектов КСППР, которыми являются участники, команда и задача, и 2) разработка динамических моделей (процессов), которые обеспечивают достижение целей КСППР. Данная фаза считается завершенной, когда получено множество непротиворечивых моделей и, при необходимости, механизмов, поддерживающих взаимодействие участников.

Зависимости между видами деятельности по обеспечению всего процесса разработки КСППР определяют последовательность их выполнения. Цели КСППР и задача определяют требования к команде.

В свою очередь эти требования определяют требования к участникам, из которых формируется команда. Следовательно, первоначально строится модель задачи, а затем модель команды. Модели участников могут создаваться независимо от указанных двух видов деятельности. Модель задачи, модель команды, модели участников и требования к участникам являются исходными данными для процедуры формирования команды. Спецификация динамических моделей часто требует уточнения статических (например, для реализации некоторого механизма требуется наличие у объекта определенных характеристик). Этот факт представлен на рисунке 3 в виде пунктирной стрелки между этими двумя видами моделей.

Фаза выполнения связана с присвоением значений переменным модели задачи ЛПР как задачи поддержки принятия решений.

**5. Сценарий применения КСППР при принятии решений о выработке рекомендаций по снижению аварийности.** Возможности платформы могут быть продемонстрированы на примере создания КСППР для разработки рекомендаций, направленных на устранение и профилактику возникновения участков концентрации дорожно-транспортных происшествий (ДТП). Участком концентрации ДТП является участок автомобильной дороги протяженностью не более 1000 метров за пределами населенного пункта, 200 метров в населенном пункте или на перекрестке дорог, на котором произошло три и более ДТП одного вида или пять и более ДТП независимо от их вида в течение 12 месяцев, в результате чего погибли или получили ранения люди [48].

Сотрудничество, лежащее в основе этого сценария, предполагает, что партнерство между людьми и программными агентами предпочтительнее полной автоматизации [18]. ЛПР инициирует сотрудничество, предоставляя необходимую информацию команде ИИ-агентов и экспертов-людей. Агенты начинают предоставлять информацию и рекомендации команде экспертов с объяснениями. Эксперты могут согласиться со списком рекомендаций, согласиться с некоторыми рекомендациями или отклонить рекомендации агентов. Если эксперты достигают какого-то соглашения, они передают согласованные рекомендации ЛПР. Если агенты не могут предоставить какие-либо рекомендации, то эксперты и ИИ-агенты взаимодействуют для совместной разработки таких рекомендаций. Эксперты пересматривают совместно разработанные рекомендации и, если они их устраивают, передают рекомендации ЛПР. На основании рекомендаций, полученных от экспертов, ЛПР готовит заключение о рекомендуемых

мерах. Если команда не может предложить какие-либо рекомендации, то она информирует об этом ЛПР и предоставляет объяснение неудачи.

В этом сценарии в совместную команду входят ЛПР, три ИИ-агента и три человека-эксперта.

В группу ИИ-агентов входят:

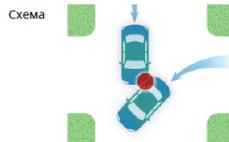
- Аналитик происшествий – агент, предоставляющий информацию об участках концентрации ДТП, произошедших на конкретной административной территории, и анализирующий карточки учета происшествий (рисунок 4);

- Аналитик знаков – агент, предоставляющий информацию о дорожных знаках, установленных на участках концентрации дорожно-транспортных происшествий;

- Рекомендующий агент – ИИ-агент, предоставляющий рекомендации мероприятий по снижению и профилактике аварийности. Этот агент основан на знаниях, база знаний для него составлена с учетом Методических рекомендаций по выбору эффективных некапиталоемких мероприятий по снижению аварийности в местах концентрации дорожно-транспортных происшествий на автомобильных дорогах общего пользования [48] и Руководства по устранению и профилактике возникновения участков концентрации ДТП при эксплуатации автомобильных дорог [49] (пример исходных данных для построения базы знаний приведен на рисунке 5).

**ОБЩИЕ СВЕДЕНИЯ**

Дата	19.10.2023	Время	11:50
Широта	60.0237	Долгота	30.2505
Номер ДТП	400086496	Вид ДТП	Столкновение



Адрес	Приморский район, г Санкт-Петербург пр-кт Королёва 48 к.1		
Дорога			
Значение дороги	Местного значения (дорога местного значения, включая относящиеся к собственности поселений, муниципальных районов, городских округов)		
Категория дороги			
Категория улицы	Улицы и дороги местного значения научно-производственных промышленных и коммунально-складских районов		

**ДОРОЖНЫЕ УСЛОВИЯ**

Объекты УДС на месте ДТП	Нерегулируемый перекрёсток равнозначных улиц (дорог)
Объекты УДС вблизи места ДТП	Многоквартирные жилые дома
Недостатки транспортно-эксплуатационного содержания улично-дорожной сети	Не установлены
Факторы, оказывающие влияние на режим движения	Сведения отсутствуют
Состояние погоды	Пасмурно
Состояние проезжей части	Мокрое
Освещение	Светлое время суток
Изменения в режиме движения	Режим движения не изменялся

Рис. 4. Пример карточки учета ДТП

Дорожные условия на участках концентрации ДТП	Основные мероприятия по повышению безопасности дорожного движения	Виды ДТП, на которые влияет мероприятие
1	2	3
Кривые в плане малого радиуса	Реконструкция участка с увеличением радиуса кривой в плане до нормативных значений или до значений, обеспечивающих плавность трассы; устройство виража (при его отсутствии), уширение проезжей части в пределах кривой; устройство разделительных островков; увеличение расстояния видимости (устройство зеркал, обеспечивающих заблаговременную видимость встречных автомобилей и устройство срезок видимости)	Опрокидывания и встречные столкновения транспортных средств
Крутые подъемы и спуски	Реконструкция продольного профиля с уменьшением продольных уклонов и увеличением радиусов вертикальных кривых; уширение проезжей части в пределах вертикальных кривых; устройство дополнительной полосы движения на подъеме; устройство аварийных объездов на спусках (для дорог в горной местности)	Встречные и попутные столкновения транспортных средств

Рис. 5. Пример методических рекомендаций, используемых для наполнения базы знаний (из ОДМ 218.4.004-2009)

В состав группы экспертов входят:

- ЛПР;
- ФДА – представитель Федерального дорожного агентства;
- Муниципальная администрация – представитель администрации;
- Инспектор полиции – представитель отдела ГИБДД.

Сценарий выполняется в рамках регулярного обследования, направленного на устранение и предотвращение аварий на участках скопления ДТП.

Реализация представленного сценария осуществляется с использованием предлагаемой платформы КСППР. Данная платформа обеспечивает формирование задач, подбор исполнителей для решения задачи согласно требуемым компетенциям и отслеживание хода решения задач. Исполнителями могут быть как люди, так и ИИ-агенты, взаимодействующие с платформой через REST API.

Для работы с задачами платформа предоставляет веб-интерфейс, через который пользователи могут просматривать общий список задач, отслеживать уведомления о приглашениях на новую задачу или о новых событиях в задачах, исполнителем которых уже является пользователь. Для каждой задачи доступно подробное описание, предполагаемые сроки ее выполнения, требуемые компетенции, возможное вознаграждение исполнителей и разбивка на

подзадачи. При наличии подзадач отображается график их зависимости и временная шкала со сроками выполнения подзадач. Обсуждение задач и подзадач осуществляется через встроенный мессенджер, в котором пользователи могут обмениваться своими решениями и комментировать их, вырабатывая совместное решение задачи. В обсуждении также участвуют агенты, выбранные на основе выполняемых ими функций (аналогично компетенциям для экспертов).

В соответствии с рассматриваемым сценарием при обнаружении места концентрации дорожно-транспортных происшествий ЛПР создает задачу, описывая координаты места и перечень дорожно-транспортных происшествий, а также указывая перечень компетенций, связанных с получением детальной информации о дорожно-транспортных происшествиях и соответствующем участке дорожной сети, формированием рекомендаций и оценкой эффекта предлагаемых рекомендаций. С использованием механизмов платформы осуществляется подбор программного обеспечения и экспертов для сбора данных и формирования рекомендаций, а также экспертов для обсуждения рекомендаций; формируются подзадачи; осуществляется взаимодействие между командой и ЛПР; предоставляется список возможных решений с обоснованием.

**6. Заключение.** В статье рассматривается концепция построения нового поколения систем поддержки принятия решений, основанных на следующих ключевых характеристиках. Во-первых, это активное применение ИИ на различных стадиях подготовки решения – начиная от анализа ситуации и формирования альтернатив, до их оценки и сопоставления. Во-вторых, это необходимость совместной работы, возникающая тогда, когда компетенций одного специалиста оказывается недостаточно для комплексного охвата проблемы и выработки возможных решений. Наконец, в-третьих, это динамический характер коллективов (команд), формируемых для работы над проблемой, что характерно для работы над нетипичными проблемами, принятия решений в изменяющихся условиях, когда не существует устоявшихся структур и их необходимо оперативно формировать. Каждая из этих характеристик как открывает новые перспективы с точки зрения оперативности и качества принимаемых решений, так и привносит ряд трудностей в структуру системы поддержки принятия решений и способ ее организации. Например, активное использование ИИ сопровождается возможными сомнениями в надежности решений, получаемых с его помощью, что вызвано недостаточной объяснимостью, привлечение нескольких участников

чревато возникновением противоречий, а динамический характер коллективов снижает взаимное доверие участников.

В статье выделены ключевые проблемы, присущие этому новому поколению систем поддержки принятия решений, основанных на взаимодействии между людьми-экспертами и ИИ-агентами, и предлагается подход к построению таких систем, обеспечивающий решение выделенных проблем. Отличительными особенностями предлагаемого подхода являются онтологическое представление информации о проблеме, модели взаимодействия и интерфейсы, поддерживающие объяснимый ИИ, поддержка совместной работы в виде настраиваемых алгоритмов формирования команды и мониторинга взаимодействия.

В рамках подхода делается несколько допущений, которые могут ограничить его применимость. Так, ключевую роль в предложенном подходе играет онтолого-ориентированное представление информации о проблеме, что позволяет существенно упростить взаимодействие внутри разнородного коллектива (людей-экспертов и ИИ-агентов). Таким образом, предлагаемый подход предназначен, в первую очередь, для построения КСППР в проблемных областях, где либо уже есть качественные онтологии, либо они могут быть относительно легко созданы. Кроме того, применение подхода целесообразно для решения достаточно сложных задач, не только требующих знаний и опыта нескольких экспертов, но выходящих за пределы жестко определенных сценариев взаимодействия.

Основные идеи подхода реализованы в программной платформе, которая предоставляет набор механизмов и интерфейсов, упрощающих разработку КСППР: поддержку формирования команд и совместной работы, интерфейсы для определения, развертывания ИИ-агентов и управления ими, а также набор структурированных представлений, облегчающих взаимодействие между экспертами-людьми и ИИ-агентами.

В статье также показан пример сценария сотрудничества, который можно реализовать с использованием предлагаемой платформы – принятие решения о выборе мероприятий по снижению аварийности.

### **Литература**

1. Schall D. Service-Oriented Crowdsourcing: Architecture, Protocols and Algorithms. New York: Springer New York, 2012. 94 p.
2. Smirnov A., Shilov N., Ponomarev A., Schekotov M. Human-computer cloud: Application platform and dynamic decision support // CLOSER 2019 – Proceedings

- of the 9th International Conference on Cloud Computing and Services Science. 2019. vol. 1. pp. 120–131.
3. Retelny D., Bernstein M.S., Valentine M.A. No Workflow Can Ever Be Enough: How Crowdsourcing Workflows Constrain Complex Work // Proc. ACM Human-Computer Interact. 2017. vol. 1. no. 2.
  4. Valentine M.A., Retelny D., To A., Rahmati N., Doshi T., Bernstein M.S. Flash Organizations: Crowdsourcing Complex Work By Structuring Crowds As Organizations // Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems – CHI '17. 2017. pp. 3523–3537.
  5. Salehi N. et al. Huddler: Convening stable and familiar crowd teams despite unpredictable availability // Proc. ACM Conf. Comput. Support. Coop. Work (CSCW). 2017. pp. 1700–1713.
  6. Lykourantzou I., Vinella F., Ahmed F., Papastathis C., Papangelis K., Khan V.-J., Masthoff J. Self-organization in online collaborative work settings // Collect. Intell. 2022. vol. 1(1). no. 263391372210780.
  7. Terveen L.G. Overview of human-computer collaboration // Knowledge-Based Syst. 1995. vol. 8. no. 2–3. pp. 67–81.
  8. Elmarzouqi N., Garcia E., Lapayre J.-C. CSCW from Coordination to Collaboration. 11th International Conference CSCWD. 2008. pp. 87–98.
  9. Karacapilidis N., Tampakas V. On the Exploitation of Collaborative Argumentation Structures for Inducing Reasoning Behavior // Proceedings of the 18th International Conference on WWW/Internet 2019. IADIS Press, 2019. pp. 78–84.
  10. Adla A., Zarate P., Soubie J.-L. A Proposal of Toolkit for GDSS Facilitators // Gr. Decis. Negot. 2011. vol. 20. no. 1. pp. 57–77.
  11. Bittner E., Shoury O. Designing Automated Facilitation for Design Thinking: A Chatbot for Supporting Teams in the Empathy Map Method // Proc. 52nd Hawaii Int. Conf. Syst. Sci. 2019. pp. 227–236.
  12. Gu W., Moustafa A.A., Ito T., Zhang M., Yang C. A case-based reasoning approach for automated facilitation in online discussion systems // KICSS 2018 – 13th Int. Conf. Knowledge, Inf. Creat. Support Syst. Proc. 2018. vol. 30. pp. 719–742. DOI: 10.1007/s10726-021-09731-4.
  13. Kolfshoten G.L., De Vreede G.J. The collaboration engineering approach for designing collaboration processes // Proceedings of the 13th International Workshop, CRIWG 2007. 2007. vol. 4715 LNCS. pp. 95–110.
  14. Ito T. et al. D-Agree: Crowd Discussion Support System Based on Automated Facilitation Agent // Proc. AAAI Conf. Artif. Intell. 2020. vol. 34. no. 09. pp. 13614–13615.
  15. Смирнов А.В., Левашова Т.В., Пономарев А.В. Онтологическая модель поддержки принятия решений на основе человеко-машинного коллективного интеллекта // Искусственный интеллект и принятие решений. 2020. № 3. pp. 48–60.
  16. Lee M.H., Siewiorek D.P., Smailagic A., Bernardino A., Bermúdez i Badia S.B. A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment // Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021. pp. 1–14.
  17. Bosch K., Bronkhorst A. Human-AI cooperation to benefit military decision making // Proceedings of Specialist Meeting Big Data & Artificial Intelligence for Military Decision Making. 2018.
  18. Pohl J. Collaborative Decision-Support and the Human-Machine Relationship // A Decision-Making Tools Workshop. San Luis: Collaborative Agent Design Research Center, 2019. pp. 21–46.

19. Chen J., Lim C.P., Tan K.H., Govindan K., Kumar A. Artificial intelligence-based human-centric decision support framework: an application to predictive maintenance in asset management under pandemic environments // *Ann. Oper. Res.* 2021. DOI: 10.1007/s10479-021-04373-w.
20. Bouabdallaoui Y., Lafhaj Z., Yim P., Ducoulombier L., Bennadji B. Predictive Maintenance in Building Facilities: A Machine Learning-Based Approach // *Sensors*. 2021. vol. 21(4), no. 1044.
21. Kase S.E. et al. The Future of Collaborative Human-Artificial Intelligence Decision-Making for Mission Planning // *Front. Psychol.* 2022. vol. 13. DOI: 10.3389/fpsyg.2022.850628.
22. Puranam P. Human-AI collaborative decision-making as an organization design problem // *Journal of Organization Design*. 2021. vol. 10. no. 2. pp. 75–80.
23. Lai V., Carton S., Bhatnagar R., Liao V., Zhang Y., Tan C. Human-AI Collaboration via Conditional Delegation: A Case Study of Content Moderation // *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022. pp. 1–18.
24. Cortes C., DeSalvo G., Mohri M. Learning with Rejection // *Proceedings of the 27th International Conference on Algorithmic Learning Theory, ALT 2016*. Cham: Springer, 2016. vol. 9925. pp. 67–82.
25. Fugener A., Grahl J., Gupta A., Ketter W. Cognitive Challenges in Human – Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation // *Inf. Syst. Res.* 2022. vol. 33. no. 2. pp. 678–696.
26. Dellermann D., Calma A., Lipusch N., Weber T., Weigel S., Ebel P. The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems // *Proceedings of the 52nd Annual Hawaii International Conference on System Sciences*. 2019. pp. 274–283.
27. Peleteiro A. et al. Using reputation and adaptive coalitions to support collaboration in competitive environments // *Eng. Appl. Artif. Intell.* 2015. vol. 45. pp. 325–338.
28. Burkart N., Huber M.F. A survey on the explainability of supervised machine learning // *Journal of Artificial Intelligence Research*. 2021. vol. 70. pp. 245–317.
29. Smirnov A., Ponomarev A., Levashova T. Towards a Methodology for Developing Human-AI Collaborative Decision Support Systems // *International Conference on Computer-Human Interaction Research and Applications*. Springer, Cham, 2023. pp. 69–88.
30. 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.
31. de Sousa Ribeiro M., Leite J. Aligning Artificial Neural Networks and Ontologies towards Explainable AI // *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021. vol. 35. no. 6. pp. 4932–4940.
32. Seeliger A., Pfaff M., Krcmar H. Semantic web technologies for explainable machine learning models: A literature review // *Joint Proceedings of the 6th International Workshop on Dataset PROFILING and the 1st Workshop on Semantic Explainability with the 18th International Semantic Web Conference (ISWC)*. 2019. vol. 2465. pp. 30–45.
33. Smirnov A., Ponomarev A. Stimulating Self-Organization in Human-Machine Collective Intelligence Environment // *2021 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*. IEEE, 2021. pp. 94–102.
34. Xiong W., Fan H., Ma L., Wang C. Challenges of human—machine collaboration in risky decision-making // *Frontiers of Engineering Management*. 2022. vol. 9. no. 1. pp. 89–103.

35. Dellermann D., Calma A., Lipusch N., Weber T., Weigel S., Ebel P. The future of Human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems // Proc. Annu. Hawaii Int. Conf. Syst. Sci. 2019. pp. 274–283.
36. Sandkuhl K., Shilov N., Smirnov A. Facilitating Digital Transformation by Multi-Aspect Ontologies: Approach and Application Steps // IFAC-PapersOnLine. 2019. vol. 52. no. 13. p. 1609–1614.
37. Korzun D.G., Balandin S.I., Gurtov A.V. Deployment of Smart Spaces in Internet of Things: Overview of the Design Challenges. Conference on Internet of Things and Smart Spaces. 2013. pp. 48–59.
38. Roffia L., Morandi F., Kiljander J., D’Elia A., Vergari F., Viola F., Bononi L., Cinotti T.S. A Semantic Publish-Subscribe Architecture for the Internet of Things // IEEE Internet Things J. 2016. vol. 3. no. 6. pp. 1274–1296.
39. Shilov N., Ponomarev A., Smirnov A. The Analysis of Ontology-Based Neuro-Symbolic Intelligence Methods for Collaborative Decision Support // Informatics Autom. 2023. vol. 22. no. 3. pp. 576–615.
40. The PROV Ontology. URL: <https://www.w3.org/TR/prov-o/> (accessed: 28.01.2020).
41. Spetzler C., Winter H., Meyer J. Decision Quality: Value Creation from Better Business Decisions. Wiley, 2016. 256 p.
42. Fayoumi A.G. Evaluating the Effectiveness of Decision Support System: Findings and Comparison // International Journal of Advanced Computer Science and Applications. 2018. vol. 9. no. 10. pp. 195–200.
43. Straka G.A. Measurement and evaluation of competence. Luxembourg, 2004. 263–311 p.
44. Is Your Team Too Big? Too Small? What’s the Right Number? // Knowledge at Wharton Podcast. Knowledge at Wharton. 2006. vol. 14.
45. Katzenbach J.R., Smith D.K. The Wisdom of Teams: Creating the High-Performance Organization. Reprint ed. Harvard: Harvard Business Review Press, 2015. 304 p.
46. LaFasto F., Larson C. When Teams Work Best: 6,000 Team Members and Leaders Tell What it Takes to Succeed. 1st ed. SAGE Publications, Inc, 2001. 256 p.
47. Robbins S.P., Judge T.A. Organizational Behavior. 12th ed. Upper Saddle River: Prentice Hall, 2006. 792 p.
48. Методические рекомендации по выбору эффективных некапиталоемких мероприятий по снижению аварийности в местах концентрации дорожно-транспортных происшествий на автомобильных дорогах общего пользования: ОДМ 218.6.025–2017. Москва, 2020. 46 p.
49. Руководство по устранению и профилактике возникновения участков концентрации ДТП при эксплуатации автомобильных дорог: ОДМ 218.4.004–2009. Москва, 2009. 94 с.

**Смирнов Александр Викторович** — д-р техн. наук, профессор, главный научный сотрудник, руководитель лаборатории, лаборатория интегрированных систем автоматизации, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: системы поддержки принятия решений, интеллектуальные системы, интеллектуальное управление конфигурациями виртуальных и сетевых организаций, логистика знаний. Число научных публикаций — 400. [smir@iias.spb.su](mailto:smir@iias.spb.su); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; т.р.: +7(812)328-8071.

**Пономарев Андрей Васильевич** — канд. техн. наук, доцент, старший научный сотрудник, лаборатория интегрированных систем автоматизации, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: коллективный интеллект, крауд-вычисления,

рекомендательные системы, машинное обучение. Число научных публикаций — 70. [ronomarev@iias.spb.su](mailto:ronomarev@iias.spb.su); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-8071.

**Шилов Николай Германович** — канд. техн. наук, доцент, старший научный сотрудник, лаборатория интегрированных систем автоматизации, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: искусственный интеллект, управление знаниями, управление онтологиями, моделирование и конфигурирование сложных систем, машинное обучение. Число научных публикаций — 300. [nick@iias.spb.su](mailto:nick@iias.spb.su); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-8071.

**Левашова Татьяна Викторовна** — старший научный сотрудник, лаборатория интегрированных систем автоматизации, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: управление знаниями, онтологии, контекстно-управляемые системы, человеко-машинное сотрудничество, поддержка принятия решений. Число научных публикаций — 150. [tatiana.levashova@iias.spb.su](mailto:tatiana.levashova@iias.spb.su); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-8071.

**Тесля Николай Николаевич** — канд. техн. наук, старший научный сотрудник, лаборатория интегрированных систем автоматизации, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: управление знаниями, человеко-машинное взаимодействие, онтологии, интеллектуальные пространства, геоинформационные системы, обработка текста на естественном языке. Число научных публикаций — 80. [teslya@iias.spb.su](mailto:teslya@iias.spb.su); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-8071.

**Поддержка исследований.** Работа выполнена при финансовой поддержке РФФИ (проект № 22-11-00214), отдельные компоненты (в частности, компонент рекомендации паттернов взаимодействия) разработаны в рамках бюджетной темы FFZF-2022-0005 СПб ФИЦ РАН.

A. SMIRNOV, A. PONOMAREV, N. SHILOV, T. LEVASHOVA, N. TESLYA  
**A CONCEPTION OF COLLABORATIVE DECISION SUPPORT  
SYSTEMS: APPROACH AND PLATFORM ARCHITECTURE**

*Smirnov A., Ponomarev A., Shilov N., Levashova T., Teslya N.* **A Conception of Collaborative Decision Support Systems: Approach and Platform Architecture.**

**Abstract.** The paper describes a general conception of collaborative decision support systems, in which teams providing decision support a) are formed flexibly in accordance with the problem and b) consist of both human experts and intelligent agents implementing AI methods and techniques. An analysis of the key problems of creating collaborative decision support systems based on the collaboration of humans and AI is carried out, the following problems are highlighted: ensuring interoperability (mutual understanding) between heterogeneous team members, reconciling differing positions of participants, ensuring trust between participants, ensuring the effectiveness of joint actions planning and maintaining a balance between predefined workflows and self-organization. Principles for constructing such systems have been formed, offering solutions to the identified problems. In particular, it is proposed to employ an ontology-oriented representation of information about the problem (in the form of multi-aspect ontology), a set of methods for monitoring team activities, reputation scheme, elements of explainable AI, as well as mechanisms of limited self-organization. The proposed concept forms the basis of a software platform for the development of collaborative decision support systems, the main architectural provisions of which are also presented in the paper. The use of the platform is illustrated by an example from the field of rational management of road infrastructure and the creation of a collaborative DSS for the development of measures to reduce road accidents.

**Keywords:** decision support system, collaborative systems, ontologies, guided self-organization, human-computer interaction.

## References

1. Schall D. *Service-Oriented Crowdsourcing: Architecture, Protocols and Algorithms.* New York: Springer New York, 2012. 94 p.
2. Smirnov A., Shilov N., Ponomarev A., Schekotov M. Human-computer cloud: Application platform and dynamic decision support. CLOSER 2019 – Proceedings of the 9th International Conference on Cloud Computing and Services Science. 2019. vol. 1. pp. 120–131.
3. Retelny D., Bernstein M.S., Valentine M.A. No Workflow Can Ever Be Enough: How Crowdsourcing Workflows Constrain Complex Work. *Proc. ACM Human-Computer Interact.* 2017. vol. 1. no. 2.
4. Valentine M.A., Retelny D., To A., Rahmati N., Doshi T., Bernstein M.S. Flash Organizations: Crowdsourcing Complex Work By Structuring Crowds As Organizations. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems – CHI '17.* 2017. pp. 3523–3537.
5. Salehi N. et al. Huddler: Convening stable and familiar crowd teams despite unpredictable availability. *Proc. ACM Conf. Comput. Support. Coop. Work (CSCW).* 2017. pp. 1700–1713.
6. Lykourantzou I., Vinella F., Ahmed F., Papastathis C., Papangelis K., Khan V.-J., Masthoff J. Self-organization in online collaborative work settings. *Collect. Intell.* 2022. vol. 1(1). no. 263391372210780.

7. Terveen L.G. Overview of human-computer collaboration. *Knowledge-Based Syst.* 1995. vol. 8. no. 2–3. pp. 67–81.
8. Elmarzouqi N., Garcia E., Lapayre J.-C. CSCW from Coordination to Collaboration. 11th International Conference CSCWD. 2008. pp. 87–98.
9. Karacapılıdīs N., Tampakas V. On the Exploitation of Collaborative Argumentation Structures for Inducing Reasoning Behavior. *Proceedings of the 18th International Conference on WWW/Internet 2019.* IADIS Press, 2019. pp. 78–84.
10. Adla A., Zarate P., Soubie J.-L. A Proposal of Toolkit for GDSS Facilitators. *Gr. Decis. Negot.* 2011. vol. 20. no. 1. pp. 57–77.
11. Bittner E., Shoury O. Designing Automated Facilitation for Design Thinking: A Chatbot for Supporting Teams in the Empathy Map Method. *Proc. 52nd Hawaii Int. Conf. Syst. Sci.* 2019. pp. 227–236.
12. Gu W., Moustafa A.A., Ito T., Zhang M., Yang C. A case-based reasoning approach for automated facilitation in online discussion systems. *KICSS 2018 – 13th Int. Conf. Knowledge, Inf. Creat. Support Syst. Proc.* 2018. vol. 30. pp. 719–742. DOI: 10.1007/s10726-021-09731-4.
13. Kolfšchoten G.L., De Vreede G.J. The collaboration engineering approach for designing collaboration processes. *Proceedings of the 13th International Workshop, CRIWG 2007.* vol. 4715 LNCS. pp. 95–110.
14. Ito T. et al. D-Agree: Crowd Discussion Support System Based on Automated Facilitation Agent. *Proc. AAAI Conf. Artif. Intell.* 2020. vol. 34. no. 09. pp. 13614–13615.
15. Smirnov A., Levashova T., Ponomarev A. [Ontological model of decision support based on human-machine collective intelligence] // *Artificial Intelligence and Decision Making – Iskusstvennyj intellekt i prinyatie reshenij*. 2020. no. 3. pp. 48–60. (In Russ.).
16. Lee M.H., Siewiorek D.P., Smailagic A., Bernardino A., Bermúdez i Badia S.B. A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 2021. pp. 1–14.
17. Bosch K., Bronkhorst A. Human-AI cooperation to benefit military decision making. *Proceedings of Specialist Meeting Big Data & Artificial Intelligence for Military Decision Making.* 2018.
18. Pohl J. Collaborative Decision-Support and the Human-Machine Relationship. *A Decision-Making Tools Workshop.* San Luis: Collaborative Agent Design Research Center, 2019. pp. 21–46.
19. Chen J., Lim C.P., Tan K.H., Govindan K., Kumar A. Artificial intelligence-based human-centric decision support framework: an application to predictive maintenance in asset management under pandemic environments. *Ann. Oper. Res.* 2021. DOI: 10.1007/s10479-021-04373-w.
20. Bouabdallaoui Y., Lafhaj Z., Yim P., Ducoulombier L., Bennadji B. Predictive Maintenance in Building Facilities: A Machine Learning-Based Approach. *Sensors.* 2021. vol. 21(4). no. 1044.
21. Kase S.E. et al. The Future of Collaborative Human-Artificial Intelligence Decision-Making for Mission Planning. *Front. Psychol.* 2022. vol. 13. DOI: 10.3389/fpsyg.2022.850628.
22. Puranam P. Human-AI collaborative decision-making as an organization design problem. *Journal of Organization Design.* 2021. vol. 10. no. 2. pp. 75–80.
23. Lai V., Carton S., Bhatnagar R., Liao V., Zhang Y., Tan C. Human-AI Collaboration via Conditional Delegation: A Case Study of Content Moderation. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 2022. pp. 1–18.

24. Cortes C., DeSalvo G., Mohri M. Learning with Rejection. Proceedings of the 27th International Conference on Algorithmic Learning Theory, ALT 2016. Cham: Springer, 2016. vol. 9925. pp. 67–82.
25. Fugener A., Grahl J., Gupta A., Ketter W. Cognitive Challenges in Human – Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation. Inf. Syst. Res. 2022. vol. 33. no. 2. pp. 678–696.
26. Dellermann D., Calma A., Lipusch N., Weber T., Weigel S., Ebel P. The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems. Proceedings of the 52nd Annual Hawaii International Conference on System Sciences. 2019. pp. 274–283.
27. Peleteiro A. et al. Using reputation and adaptive coalitions to support collaboration in competitive environments. Eng. Appl. Artif. Intell. 2015. vol. 45. pp. 325–338.
28. Burkart N., Huber M.F. A survey on the explainability of supervised machine learning. Journal of Artificial Intelligence Research. 2021. vol. 70. pp. 245–317.
29. Smirnov A., Ponomarev A., Levashova T. Towards a Methodology for Developing Human-AI Collaborative Decision Support Systems. International Conference on Computer-Human Interaction Research and Applications. Springer, Cham, 2023. pp. 69–88.
30. 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.
31. de Sousa Ribeiro M., Leite J. Aligning Artificial Neural Networks and Ontologies towards Explainable AI. Proceedings of the AAAI Conference on Artificial Intelligence. 2021. vol. 35. no. 6. pp. 4932–4940.
32. Seeliger A., Pfaff M., Krcmar H. Semantic web technologies for explainable machine learning models: A literature review. Joint Proceedings of the 6th International Workshop on Dataset PROFILING and the 1st Workshop on Semantic Explainability with the 18th International Semantic Web Conference (ISWC). 2019. vol. 2465. pp. 30–45.
33. Smirnov A., Ponomarev A. Stimulating Self-Organization in Human-Machine Collective Intelligence Environment. 2021 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA). IEEE, 2021. pp. 94–102.
34. Xiong W., Fan H., Ma L., Wang C. Challenges of human—machine collaboration in risky decision-making. Frontiers of Engineering Management. 2022. vol. 9. no. 1. pp. 89–103.
35. Dellermann D., Calma A., Lipusch N., Weber T., Weigel S., Ebel P. The future of Human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems. Proc. Annu. Hawaii Int. Conf. Syst. Sci. 2019. pp. 274–283.
36. Sandkuhl K., Shilov N., Smirnov A. Facilitating Digital Transformation by Multi-Aspect Ontologies: Approach and Application Steps. IFAC-PapersOnLine. 2019. vol. 52. no. 13. p. 1609–1614.
37. Korzun D.G., Balandin S.I., Gurtov A.V. Deployment of Smart Spaces in Internet of Things: Overview of the Design Challenges. Conference on Internet of Things and Smart Spaces. 2013. pp. 48–59.
38. Roffia L., Morandi F., Kiljander J., D’Elia A., Vergari F., Viola F., Bononi L., Cinotti T.S. A Semantic Publish-Subscribe Architecture for the Internet of Things. IEEE Internet Things J. 2016. vol. 3. no. 6. pp. 1274–1296.
39. Shilov N., Ponomarev A., Smirnov A. The Analysis of Ontology-Based Neuro-Symbolic Intelligence Methods for Collaborative Decision Support. Informatics Autom. 2023. vol. 22. no. 3. pp. 576–615.

40. The PROV Ontology. Available at: <https://www.w3.org/TR/prov-o/> (accessed: 28.01.2020).
41. Spetzler C., Winter H., Meyer J. Decision Quality: Value Creation from Better Business Decisions. Wiley, 2016. 256 p.
42. Fayoumi A.G. Evaluating the Effectiveness of Decision Support System: Findings and Comparison. International Journal of Advanced Computer Science and Applications. 2018. vol. 9. no. 10. pp. 195–200.
43. Straka G.A. Measurement and evaluation of competence. Luxembourg, 2004. 263–311 p.
44. Is Your Team Too Big? Too Small? What's the Right Number? Knowledge at Wharton Podcast. Knowledge at Wharton. 2006, vol. 14.
45. Katzenbach J.R., Smith D.K. The Wisdom of Teams: Creating the High-Performance Organization. Reprint ed. Harvard: Harvard Business Review Press, 2015. 304 p.
46. LaFasto F., Larson C. When Teams Work Best: 6,000 Team Members and Leaders Tell What it Takes to Succeed. 1st ed. SAGE Publications, Inc, 2001. 256 p.
47. Robbins S.P., Judge T.A. Organizational Behavior. 12th ed. Upper Saddle River: Prentice Hall, 2006. 792 p.
48. Metodicheskie rekomendacii po vyboru effektivnyh nekapitaloemkih meropriyatij po snizheniyu avarijnosti v mestah koncentracii dorozhno-transportnyh proisshestvij na avtomobil'nyh dorogah obshchego pol'zovaniya [Methodological recommendations for the selection of effective non-capital-intensive measures to reduce accident rates in places where traffic accidents are concentrated on public roads]: ODM 218.6.025–2017. Moscow, 2020. 46 p. (In Russ.).
49. Rukovodstvo po ustraneniyu i profilaktike vozniknoveniya uchastkov koncentracii DTP pri ekspluatatsii avtomobil'nyh dorog [Guidelines for eliminating and preventing the occurrence of accident concentration areas during the operation of highways]: ODM 218.4.004-2009. Moscow, 2009. 94 p. (In Russ.).

**Smirnov Alexander** — Ph.D., Dr.Sci., Professor, Chief researcher, head of the laboratory, Computer-aided integrated systems laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: decision support systems, intelligent systems, intelligent configuration management in virtual and network organizations, knowledge logistics. The number of publications — 400. [smir@iias.spb.su](mailto:smir@iias.spb.su); 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-8071.

**Ponomarev Andrew** — Ph.D., Associate Professor, Senior researcher, Computer-aided integrated systems laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: collective intelligence, crowd computing, recommender systems, applied machine learning. The number of publications — 70. [ponomarev@iias.spb.su](mailto:ponomarev@iias.spb.su); 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-8071.

**Shilov Nikolay** — Ph.D., Associate Professor, Senior researcher, Computer-aided integrated systems laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: artificial intelligence, knowledge management, ontology management, complex system modelling and configuration, machine learning. The number of publications — 300. [nick@iias.spb.su](mailto:nick@iias.spb.su); 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-8071.

**Levashova Tatiana** — Senior researcher, Computer-aided integrated systems laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: knowledge management, ontologies, context-based systems, human-

machine collaboration, decision support. The number of publications — 150. [tatiana.levashova@iias.spb.su](mailto:tatiana.levashova@iias.spb.su); 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-8071.

**Teslya Nikolay** — Ph.D., Senior researcher, Computer-aided integrated systems laboratory, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: knowledge management, human-machine interaction, ontologies, smart spaces, geographic information systems, natural language processing. The number of publications — 80. [teslya@iias.spb.su](mailto:teslya@iias.spb.su); 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-8071.

**Acknowledgements.** This research is funded by the Russian Science Foundation (grant 22-11-00214), some platform components (in particular, the collaboration patterns recommendation component) are developed in the scope of state research topic FFZF-2022-0005 of SPC RAS.

А.А. ДВОЙНИКОВА, И.А. КАГИРОВ, А.А. КАРПОВ  
**МЕТОД РАСПОЗНАВАНИЯ СЕНТИМЕНТА И ЭМОЦИЙ  
В ТРАНСКРИПЦИЯХ РУССКОЯЗЫЧНОЙ РЕЧИ  
С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ПЕРЕВОДА**

*Двойникова А.А., Кагиров И.А., Карпов А.А. Метод распознавания сентимента и эмоций в транскрипциях русскоязычной речи с использованием машинного перевода.*

**Аннотация.** В статье рассматривается проблема распознавания сентимента и эмоций пользователей в русскоязычных текстовых транскрипциях речи с использованием словарных методов и машинного перевода. Количество имеющихся информационных ресурсов для анализа сентимента текстовых сообщений на русском языке очень ограничено, что существенно затрудняет применение базовых методов анализа сентимента, а именно, предобработки текстов, векторизации с помощью тональных словарей, традиционных классификаторов. Для решения этой проблемы в статье вводится новый метод на основе автоматического машинного перевода русскоязычных текстов на английский язык. Частичный перевод предполагает перевод отдельных лексем, не включенных в русскоязычные тональные словари, тогда как полный перевод подразумевает перевод всего текста целиком. Переведенный текст анализируется с использованием различных англоязычных тональных словарей. Экспериментальные исследования для решения задачи распознавания сентимента и эмоций были проведены на текстовых транскрипциях многомодального русскоязычного корпуса RAMAS, извлеченных из аудиоданных экспертным путем и автоматически с использованием системы распознавания речи. В результате применения методов машинного перевода достигается значение взвешенной F-меры распознавания семи классов эмоций 31,12 % и 23,74 %, и трех классов сентимента 75,37 % и 71,60 % для экспертных и автоматических транскрипций русскоязычной речи корпуса RAMAS, соответственно. Также в ходе экспериментов было выявлено, что использование статистических векторов в качестве метода преобразования текстовых данных позволяет достичь значение показателя взвешенной F-меры на 1-5 % выше по сравнению с использованием конкатенированного (статистического и тонального) вектора. Таким образом, эксперименты показывают, что объединение всех англоязычных тональных словарей позволяет повысить точность распознавания сентимента и эмоций в текстовых данных. В статье также исследуется корреляция между длиной вектора текстовых данных и его репрезентативностью. По результатам экспериментов можно сделать вывод, что использование лемматизации для нормализации слов текстовых транскрипций речи позволяет достичь большей точности распознавания сентимента по сравнению со стеммингом. Использование предложенных методов с полным и частичным машинным переводом позволяет повысить точность распознавания сентимента и эмоций на 0,65–9,76 % по показателю взвешенной F-меры по сравнению с базовым методом распознавания сентимента и эмоций.

**Ключевые слова:** машинный перевод, тональные словари, распознавание эмоций, сентимент-анализ, тональные вектора.

**1. Введение.** Анализ тональности текста (сентимент-анализ, англ. sentiment analysis) – это область компьютерной лингвистики, связанная с методами определения эмоциональной полярности текста

на естественном языке. Анализ тональности является частным случаем извлечения информации (англ. data mining), однако он не подразумевает извлечения имен сущностей, ограничиваясь только эмоциональной окраской текстов [1].

Анализ тональности текстовых сообщений актуален во многих сферах человеческой деятельности: оценка качества товаров и услуг, мониторинг общественного мнения, прогнозы на основе новостных подборок в Интернете и т.п. [2 – 3]. Другой важной областью является межчеловеческое взаимодействие в виртуальном пространстве, подразумевающее коммуникацию на естественном языке [4]. Благодаря важности перечисленных областей, анализ сентимента является динамичной и быстро развивающейся отраслью компьютерной лингвистики и методов анализа естественного языка в целом.

В общем случае анализ сентимента сводится к отнесению конкретного текста или группы текстов к определенному классу в зависимости от эмоциональной валентности текста. В существующих исследованиях используют различные классификации сентимента: бинарная (негативный, позитивный), тернарная (негативный, нейтральный, позитивный) и многоуровневую (от сильно негативного до сильно позитивного) [5].

Стоит отметить смежную область – автоматический анализ эмоций в тексте. В классической работе [6], теоретические результаты из которой фактически заложили основы современных исследований в области распознавания эмоций, было выделено шесть базовых эмоций, и для их нахождения используется две группы методов: основанные на словарях и основанные на корпусах. Первый метод достаточно прямолинеен и напрямую зависит от доступных словарей эмоций для конкретного языка. Второй метод подразумевает построение математической модели на основе текстов, предварительно размеченных экспертами. Несмотря на наличие определенной корреляции между определением сентиментом и эмоций, следует иметь в виду, что эмоции – это выражение психофизиологических состояний индивида, а сентимент – отношение говорящего к определенной теме [7].

Настоящая статья посвящена улучшению методов автоматического анализа сентимента и эмоций русскоязычных транскрипций речи за счет использования машинного перевода текстовых данных. Количество информационных ресурсов (корпусов данных, тональных словарей) для анализа текстов на русском языке на сегодня остается достаточно ограниченным

[8 – 9], в связи с чем предлагается метод, основанный на применении автоматического машинного перевода русскоязычных текстов на английский язык.

Автоматический машинный перевод текстовых данных используется в существующих исследованиях для распознавания сентимента [10 – 11] и эмоций [12]. Такой метод актуален для малоресурсных языков, потому что из текстов на таких языках достаточно сложно извлечь лингвистические признаки. Другой причиной может являться маленький объем данных. Поэтому машинный перевод используется в прямом виде (для извлечения лингвистических признаков на другом языке или увеличении объема тональных словарей) [11, 13] или как двойной обратный перевод (для аугментации обучающего набора данных).

Часто производится двойной обратный перевод: с исходного языка на другой выбранный, затем с выбранного обратно на исходный. В большинстве случаев в качестве промежуточного языка выбирается английский язык [10]. Метод машинного перевода для улучшения точности классификации сентимента или эмоций используется для менее ресурсных языков, например для турецкого [10], словацкого [11], иврита [14], польского [12], испанского [14], русского [15] и других.

**2. Исследовательский корпус речевых данных.** Для экспериментальных исследований использован русскоязычный многомодальный корпус данных RAMAS [16]. Он содержит 581 аудио- и видеозапись с участием десяти актеров общей продолжительностью 395 минут. Особенностью корпуса является то, что актеры попарно разыгрывали диалоги по диадическим сценариям. Сценарии были составлены таким образом, чтобы каждый диктор проявил одну из шести основных эмоций: радость, страх, удивление, гнев, грусть, отвращение. Также в корпусе присутствуют монологи каждого диктора с эмоционально нейтральной речью. Для распознавания сентимента и эмоций в качестве меток текста использовались значения эмоций, указанные в сценариях.

Корпус RAMAS размечен только по классам эмоций, разметка по сентиментам высказываний не проводилась. Для построения системы распознавания сентимента необходимо сгруппировать классы эмоций по принципу валентности эмоций. Обоснованностью такого преобразования с точки зрения психологии являются работы американского психолога Дж. Рассела. Диаграмма эмоций Рассела [17] – модель, созданная для описания и классификации эмоций на основе двух основных измерений: валентности и активации. Валентность

относится к тому, насколько положительно или отрицательно оценивается эмоция. Активация в диаграмме Рассела относится к уровню возбуждения психики человека, связанному с эмоцией. Основываясь на диаграмме Рассела в настоящем исследовании для выделения групп сентимента такие эмоции как радость, удивление группируются в положительный класс сентимента, страх, грусть, гнев и отвращение – негативный класс, а нейтральное состояние относится к нейтральному классу. Таким образом, все высказывания корпуса RAMAS группируются в три класса сентимента.

Следует отметить, что материал, содержащийся в корпусе данных, содержит шумы, речь одновременно нескольких дикторов или организаторов. Кроме того, авторы RAMAS не предоставили расшифровок (орфографических транскрипций) речи дикторов. Поэтому для экспериментальных исследований из аудиозаписей диалогов были извлечены транскрипции реплик дикторов экспертным (человеком-аудитором) и автоматическим методом. Для автоматического распознавания речи (APP) использовались сервисы SpeechKit<sup>1</sup> от компании Яндекс и SpeechRecognition<sup>2</sup> от компании Google. Итоговый текстовый корпус составил 535 транскрипций произнесенных текстов, извлеченных экспертным методом, и 263 текста с использованием систем APP. Разница в количестве извлеченных текстов (535 и 263) обусловлена особенностями данных, содержащихся в корпусе RAMAS, а именно, большим количеством шумов и наложений речи нескольких дикторов. Поэтому от большого числа транскрипций, полученных автоматическим методом, пришлось отказаться. Также стоит отметить, что автоматические транскрипции содержат в себе грамматические и лексические ошибки. В ходе эксперимента оба набора транскрипций – как экспертные, так и автоматические – были отдельно использованы для проверки предложенных гипотез.

На рисунке 1 представлено распределение текстов в каждом классе сентимента и эмоций для экспертных и автоматических транскрипций.

---

<sup>1</sup><https://cloud.yandex.ru/services/speechkit>

<sup>2</sup><https://cloud.google.com/speech-to-text/>

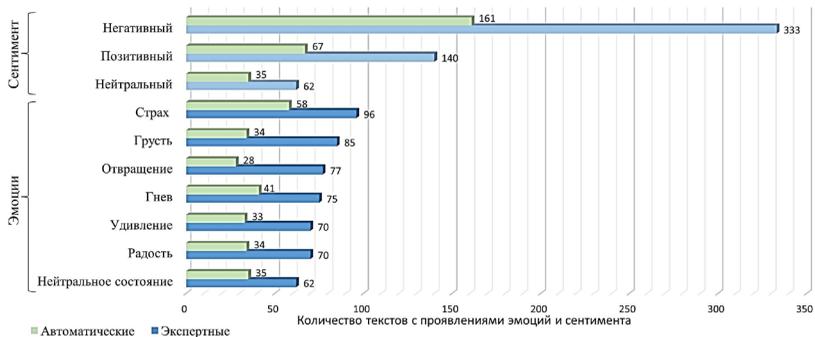


Рис. 1. Гистограмма количества текстов с проявлениями сентимента и эмоций в экспертных (синий) и автоматических (зеленый) транскрипциях корпуса RAMAS

В результате анализа диалогов, содержащихся в базе данных RAMAS, было выявлено количественное соотношение трех видов сентимента – позитивной, негативной и нейтральной, с преобладанием негативного сентимента как для экспертных транскрипций, так и для автоматических. Как можно заметить из рисунка 1 распределение сентиментов в данных сильно несбалансировано, при этом количество текстов в классах эмоций почти одинаковое.

Для анализа экспертных и автоматических транскрипций корпуса RAMAS были построены диаграммы размаха длин текстовых данных, они представлены на рисунке 2.

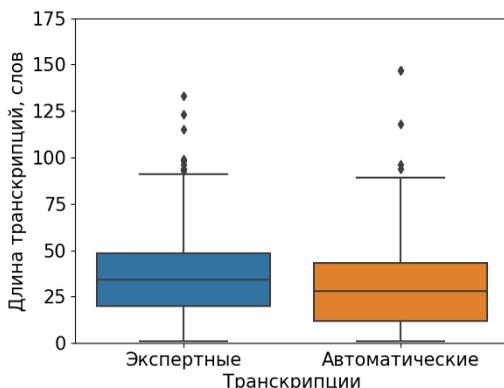


Рис. 2. Диаграммы размаха длин транскрипций (в словах) из корпуса RAMAS

Из рисунка 2 видно, что максимальное значение количества слов в высказываниях (за исключением выбросов) достигает 90. При этом медианное значение длин всех текстов находится в значении около 30-35 слов.

**3. Метод распознавания сентимента и эмоций.** Для распознавания сентимента и эмоций пользователей предложен метод, представленный в виде блок-схемы на рисунке 3.

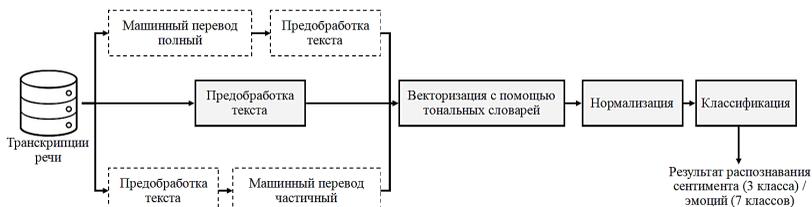


Рис. 3. Блок-схема метода распознавания сентимента и/или эмоций в русскоязычных текстовых транскрипциях речи

Все серые блоки со сплошным контуром образуют базовый метод. Блоки с пунктирным контуром выполняются опционально. Также в статье исследуются методы машинного перевода для увеличения точности распознавания сентимента и эмоций в русскоязычных текстовых данных, рассматривается возможность использования полного и частичного перевода текстовых данных.

Блок предобработки текстовых данных включает в себя токенизацию (разделение текста на токены – слова), понижение регистра, удаление пунктуации и стоп-слов, а также лемматизацию или стемминг. При лемматизации слово приводится к его начальной форме (лемме) с использованием морфологических правил и словарей, позволяющих получить слово в одной из его форм в контексте и привести к словарной форме. Стемминг, наоборот, сводится к удалению всех аффиксов словоформы, не входящих в основу слова. Для стемминга используются эвристические правила и алгоритмы, которые могут приводить к ошибкам, если правила не учитываются в полной мере.

Для блока векторизации необходимо использование тональных словарей. Под векторизацией понимается преобразование текстовых данных в числовой формат. Векторизация с использованием тональных словарей выполняется по алгоритму, представленному на рисунке 4.

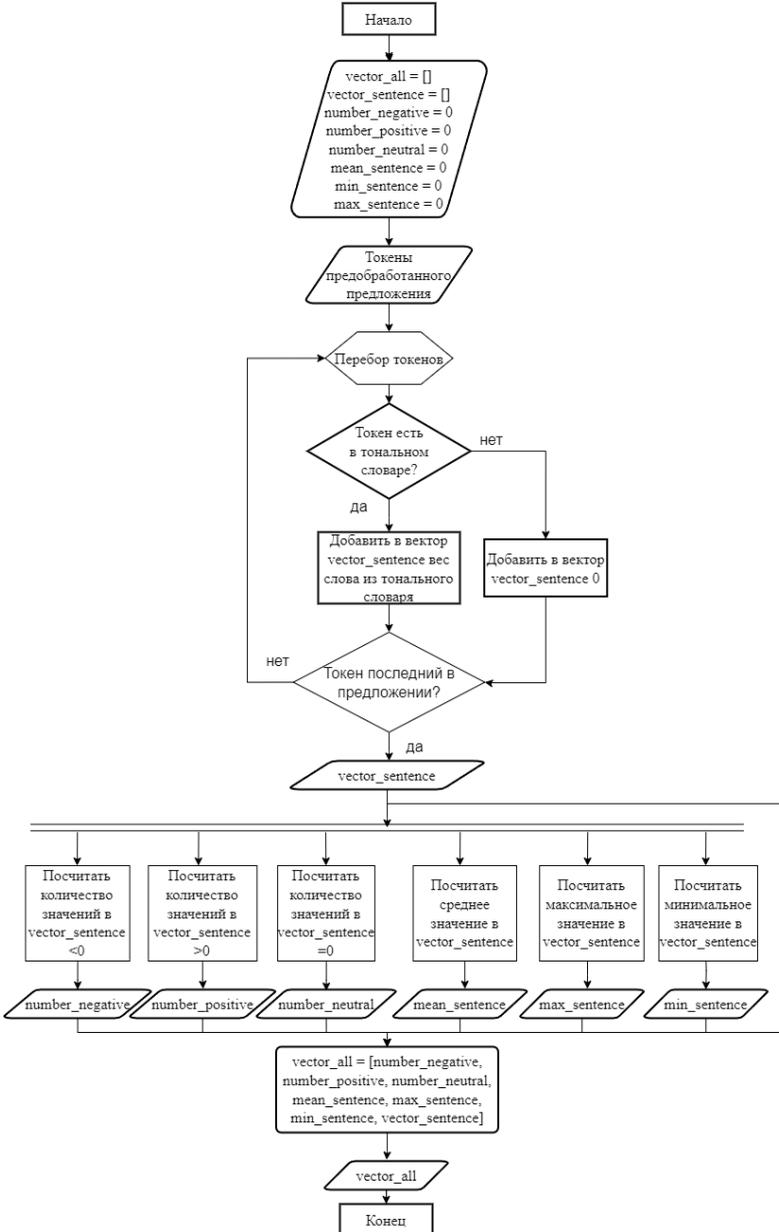


Рис. 4. Блок-схема алгоритма векторизации текстовых предложений с использованием тональных словарей

На рисунке 4 переменная `vector_sentence` является списком, в котором хранятся значения весов из тонального словаря для каждого токена предложения. Переменные `number_negative`, `number_positive`, `number_neutral`, `mean_sentence`, `max_sentence`, `min_sentence`, `sum_pos_sentence` и `sum_neg_sentence` означают количество отрицательных, положительных и нейтральных весов в `vector_sentence`, а также среднее, максимальное и минимальное значения `vector_sentence`, сумма весов положительных и отрицательных значений в `vector_sentence`. Все восемь переменных образуют статистический вектор проявлений сентимента и/или эмоций в тексте. Для распознавания сентимента используются тональные словари, веса слов в которых обозначают степень положительного или отрицательного значения слова, для распознавания эмоций – веса обозначают принадлежность к различным эмоциям.

Предложенный метод векторизации текста имеет ряд недостатков: 1) вектора разных предложений имеют различную длину, 2) значения векторов находятся в разных диапазонах. Для решения первого недостатка все вектора обрезаются или дополняются нулями до длины 98. Выбор такого значения обусловлен максимальным значением длин всех текстов (рисунок 2) в исследуемом корпусе. Затем все вектора нормализуются по следующей формуле (Min-Max нормализация):

$$x_{norm}(i) = \frac{x(i) - \min(x)}{\max(x) - \min(x)}, \quad (1)$$

где  $x_{norm}(i)$  и  $x(i)$  – нормированное и исходное  $i$  значения вектора,  $\max(x)$  и  $\min(x)$  – максимальное и минимальное значения данного вектора.

Для определения наиболее подходящего классификатора проведено уменьшение размерности данных при помощи линейного дискриминантного анализа (англ. Linear Discriminant Analysis, LDA<sup>3</sup>). Преимущества метода LDA заключается в том, что при преобразовании исходных векторов в пространство необходимой размерности учитываются метки классов. Суть алгоритма LDA заключается в максимизации линейных дискриминант (представленных в виде осей на рисунке 5) каждого класса. Значения линейных дискриминант на осях не имеют информативного характера,

---

<sup>3</sup>[https://scikit-learn.org/stable/modules/generated/sklearn.discriminant\\_analysis.LinearDiscriminantAnalysis.html](https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html)

они необходимы для расчета расстояния Махаланобиса между группами (классами). Значение расстояния Махаланобиса отображает степень разделимости классов. На рисунке 5 изображены вектора исследуемого корпуса данных в двумерном признаковом пространстве.

Как показывает анализ результатов LDA (рисунок 5), достаточно плохо разделяются классы данных как для сентимента, так и для эмоций. В связи с этим в настоящем исследовании для классификации данных использован ядерный метод опорных векторов (англ. Kernel Support Vector Machine, Kernel SVM<sup>4</sup>), являющийся быстрым и точным методом решения задач, связанных с классификацией текстовых данных [5].

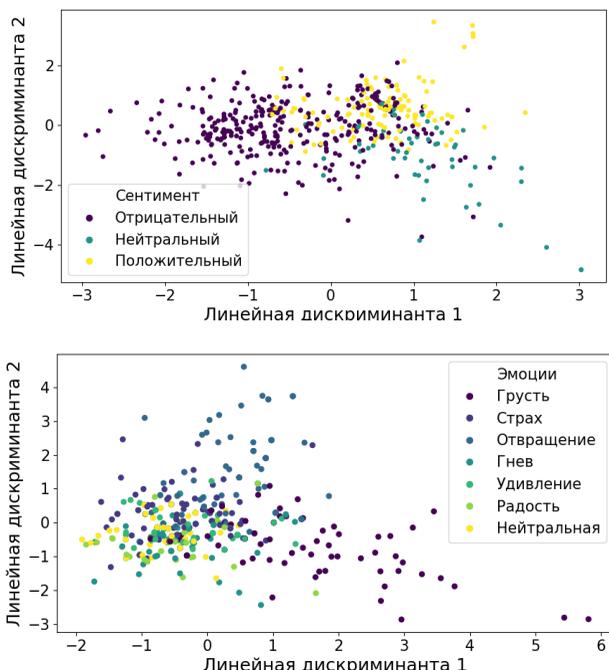


Рис. 5. Распределение векторов в двухмерном пространстве по классам сентимента (сверху) и эмоций (снизу)

**Методы машинного перевода.** Эмоциональная окраска конкретных языковых единиц, с точки зрения лингвистики,

<sup>4</sup><https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC>

представляет собой проявление модальности – структурно сложной и многоплановой семантической зоны языка. Согласно [18], модальные значения концентрируются вокруг двух ядерных значений – отношения говорящего к ситуации и статуса самой ситуации по отношению к реальному миру. В предложении, содержащем модальный компонент, всегда заключено отношение говорящего к тому, что он сообщает, иными словами, оценка ситуации.

Эмоциональная окраска текста коррелирует с одним из важнейших типов модальности – оценочной модальностью («хорошо – плохо»), или этической оценкой. Характерно, что значения оценочной модальности имеют тенденцию выражаться лексическими средствами [19], а не грамматическими. Таким образом, вероятность того, что одни и те же значения из области оценочной модальности, связанные с этической оценкой ситуации, будут выражаться сходными средствами для большинства языковых пар, достаточно высока, и прямой перевод лексики, маркированной в отношении оценочной модальности («эмоционально окрашенные слова») с высокой долей вероятности позволит расширить существующие тональные словари для русского языка.

Для русского языка существует ограниченное количество тональных словарей, относительно английского языка. Поэтому применение автоматического машинного перевода позволяет решить проблему недостатка информационных ресурсов русского языка и позволит извлекать большее количество информации о проявлениях сентимента и эмоций в русскоязычных текстах. Для машинного перевода использовалась встроенная модель машинного перевода модуля Translation библиотеки googletrans<sup>5</sup> языка Python. В работе проводились исследования по использованию двух видов машинного перевода: частичного и полного.

Частичный перевод подразумевает под собой перевод на английский язык только тех слов в тексте, которых нет в русскоязычном тональном словаре. Такой перевод выполняется после блока предобработки текстов на русском языке для того, чтобы перевод выполнялся корректно – переводились только леммы слов, а не использовались различные морфологические формы слова. Полный перевод выполняется с исходным русскоязычным текстом для того, чтобы сохранить синтаксическую структуру предложений. После перевода текст векторизуется с использованием англоязычных тональных словарей.

---

<sup>5</sup> <https://pypi.org/project/googletrans/>

**4. Тональные словари.** Одним из методов распознавания сентимента и эмоций в текстовых данных является использование словарей оценочной лексики (тональных словарей, словарей эмоциональных слов). Словари оценочной лексики содержат лексические единицы (слова и словосочетания), каждой из которых присвоена некоторая «эмоциональная оценка». Таким образом, каждой единице словаря задается вес принадлежности к эмоциональному классу [20]. Разметка может быть бинарная, тернарная и многоклассовая (содержащая больше трех классов).

Одним из основных недостатков этого метода является сильная зависимость от наличия тональных словарей. В том случае, если информационное обеспечение для конкретного языка (в данном случае – для русского) недостаточно, созданная система автоматического анализа текста предсказуемо даст неудовлетворительные результаты. Составление новых словарей является трудоемкой задачей, требующей длительного времени и дополнительного привлечения специалистов. Ниже в таблице 1 представлено описание тональных словарей для русского языка, которые использовались в данном исследовании.

Как видно, анализ тональности имеет определенные информационные ресурсы для русского языка. Однако, в отличие от других основных мировых языков (в первую очередь, английского), системы анализа русскоязычных текстов достигают меньшей точности распознавания сентимента по сравнению с англоязычными, что может быть обусловлено меньшим объемом тональных словарей и обучающих баз данных. Так, в работе [21] демонстрируется, что точность анализа сентимента для английского языка на сегодняшний день выше, чем для другого славянского языка – чешского. В таблице 1 также представлено описание существующих англоязычных тональных словарей, которые использовались в настоящем исследовании. Значения сентимента  $[-1, 1]$  и  $[-4, 4]$  означают регрессивную аннотацию данных по сентименту. Чем больше диапазон значений сентимента, тем больше вариативность меток аннотаций.

Как можно заметить из таблицы 1, количество и объем тональных словарей в свободном доступе для английского языка выше относительно русскоязычных словарей. При этом количество словарей для английского языка, имеющих оценочную лексику по различным классам эмоций, значительно больше, чем для русского языка. Более представительные информационные ресурсы для английского языка могут помочь извлекать большее количество репрезентативных

признаков для анализа сентимента и эмоций в текстовых данных. Регрессивные оценки сентимента присутствуют только в англоязычных словарях.

## 5. Экспериментальные исследования

**5.1. Нормализация слов транскрипций.** Для решения задачи распознавания сентимента и эмоций необходимо выполнить предобработку текстовых данных. В зависимости от различных типов данных, языка и пр. стоит подбирать эффективную комбинацию методов предобработки, с помощью которой можно достичь наиболее высокой точности распознавания сентимента и эмоций. Одним из главных методов предобработки является вид нормализации слов в текстах, который может значительно влиять на точность распознавания сентимента и эмоций. Нормализация слов в текстовых данных может выполняться двумя методами: 1) лемматизация и 2) стемминг. Стоит учесть, что при использовании стемминга слов в тексте в качестве предобработки данных и векторизации текста с использованием тональных словарей, необходимо также обрабатывать слова из тональных словарей стеммингом.

Таблица 1. Описание тональных словарей

Название словаря	Число слов	Количество и содержание классов
Русскоязычные тональные словари		
RuSentiLex <sup>6</sup> [9]	16057	3 сентимента, смешанная оценка
LinisCrowd <sup>7</sup> [22]	7545	Сильно отрицательные, отрицательные, нейтральные, положительные, сильно положительные
WordNetAffect <sup>8</sup> [23]	2401	6 эмоций (радость, страх, гнев, печаль, отвращение, удивление)
Англоязычные тональные словари		
SentiWordNet <sup>9</sup>	206942	Сентимент [-1, 1]
NRC <sup>10</sup> [24]	1515	Сентимент [-1, 1], 8 эмоций
Bing Liu's Opinion Lexicon <sup>11</sup> [25]	6787	7 эмоций, 3 сентимента
Vader <sup>12</sup> [26]	7520	Сентимент [-4, 4]

<sup>6</sup> <https://www.labinform.ru/pub/rusentilex/index.htm>

<sup>7</sup> <http://linis-crowd.org/>

<sup>8</sup> [http://lilu.fcim.utm.md/resourcesRoRuWNA\\_ru.html](http://lilu.fcim.utm.md/resourcesRoRuWNA_ru.html)

<sup>9</sup> <https://github.com/aesuli/SentiWordNet>

<sup>10</sup> <https://www.saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

<sup>11</sup> <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>

<sup>12</sup> <https://github.com/cjhutto/vaderSentiment>

Экспериментальные исследования по определению наиболее эффективного (позволяющего достичь наибольшей точности распознавания сентимента и эмоций) метода нормализации слов в русскоязычных транскрипциях корпуса RAMAS представлены в таблице 2. Лемматизация слов выполнялась с помощью инструмента *rumorphu2*, а стемминг с помощью *nltk*. Для проведения данных исследований экспертные транскрипции (ЭТ) и автоматические (АТ) транскрипции векторизовались с использованием объединенного тонального словаря, включающего *RuSentiLex*, *LinisCroes* и *WordNetAffect*. Объединение тональных словарей позволяет извлечь из текстов большее количество эмоционально репрезентативной информации [15]. В качестве машинного классификатора использовался *Kernel SVM* с подбором гиперпараметров для каждого эксперимента. В данном исследовании использовались следующие ядра: линейное, полиномиальное, сигмовидное, ядро радиальной базисной функции. Обучение классификатора происходило с помощью трехблочной кросс-валидации. Разделение данных на три блока происходило стратифицированным способом. Эксперименты проводились для различных задач классификации: распознавание семи классов эмоций – радость, удивление, грусть, страх, гнев, отвращение, нейтральное состояние и трех классов сентимента – негативный, нейтральный, позитивный. В качестве сравнительной метрики оценки эффективности предложенных методов использовалась взвешенная F-мера. Для многоклассовой классификации она вычисляется по формуле (2).

$$wF = \frac{\sum_{i=0}^m (F_i * N_i)}{m}, \quad (2)$$

где  $wF$  – взвешенная F-мера для всех классов  $m$ ,  $F_i$  – взвешенная F-мера для  $i$  класса,  $N_i$  – количество экземпляров в  $i$  классе.

Таблица 2. Результаты распознавания сентимента и эмоций в зависимости от метода нормализации слов, взвешенная F-мера (%)

Метод нормализации	Эмоции		Сентимент	
	ЭТ	АТ	ЭТ	АТ
<b>Лемматизация</b>	<b>27,42</b>	<b>23,09</b>	<b>65,61</b>	<b>67,79</b>
<b>Стемминг</b>	26,67	21,87	61,47	61,20

Как видно из таблицы 2, лемматизация в качестве метода нормализации слов в текстовых транскрипциях превосходит стемминг на 1–6% по показателю взвешенной F-меры в зависимости от типа транскрипций (экспертные или автоматические) и задачи классификации (сентимент или эмоции). Более высокое качество классификации сентимента и эмоций при использовании лемматизации может быть связано с возникающими проблемами омонимии в тональных словарях при использовании стемминга.

**5.2. Машинный перевод.** В данном исследовании проверяется гипотеза об эффективности использования машинного перевода для распознавания сентимента и эмоций в текстовых транскрипциях речевых высказываний. Обзор существующих исследований показал, что английский язык имеет большое количество информационно-лингвистических ресурсов, и качество распознавания сентимента в англоязычных текстах выше, по сравнению с русскоязычными. Поэтому для экспериментальных исследований использовался машинный перевод на английский язык. Для того, чтобы определить наиболее репрезентативный англоязычный тональный словарь для распознавания сентимента, был произведен полный автоматический перевод всех транскрипций корпуса RAMAS на английский язык. Векторизация англоязычных текстовых данных была выполнена как при помощи каждого тонального словаря английский тональных слов, описанных в разделе 5, так и объединенном тональном словаре. Результаты экспериментов представлены в таблице 3.

Таблица 3. Результаты распознавания сентимента при использовании полного машинного перевода на английский язык транскрипций корпуса RAMAS, взвешенная F-мера (%)

Тональный словарь	Экспертные транскрипции	Автоматические транскрипции
SentiWordNet	63,64	61,51
NRC [24]	65,33	57,36
Bing Liu's Opinion Lexicon [25]	67,91	<b>66,62</b>
Vader [26]	71,80	66,52
Объединенный	<b>75,37</b>	64,77

Как можно заметить, объединенный англоязычный тональный словарь для распознавания сентимента в переведенных на английский язык транскрипциях показывает лучшую точность только при анализе экспертных транскрипций корпуса RAMAS. Наиболее высокая точность распознавания сентимента в переведенных автоматических транскрипциях достигается при использовании во время векторизации

англоязычного тонального словаря Bing Liu's Opinion Lexicon [25]. Это может быть связано с тем, что автоматические транскрипции содержат в себе много зашумленных данных, не относящихся к речи анализируемого диктора. Слишком большое количество распознанных тональных слов различных дикторов может привести к ошибкам в распознавании сентимента в высказываниях анализируемого диктора. Поэтому в дальнейших экспериментальных исследованиях для анализа экспертных транскрипций будет использоваться объединенный англоязычный тональный словарь, а для автоматических транскрипций – Bing Liu's Opinion Lexicon [25].

В ходе экспериментов были исследованы и сравнены два типа машинного перевода: полный и частичный. При полном переводе переводится весь текст, а при частичном – только отдельные лексические единицы. При этом для перевода отбирается только такая лексика, которая не вошла в русскоязычные тональные словари. Результаты экспериментальных исследований по выявлению наиболее эффективного метода (показывающего наибольшую точность) к машинному переводу русскоязычных текстовых данных на английский язык для распознавания сентимента и эмоций представлены в таблице 4.

Из результатов экспериментов видно, что использование полного перевода на английский язык является более эффективным методом для распознавания сентимента и эмоций только в экспертных транскрипциях. Такой метод позволяет достичь точности классификации семи классов эмоций и трех классов сентимента по показателю взвешенной F-меры 31,12% и 75,37%, соответственно. Полученные результаты превосходят другие методы анализа текстовых транскрипций (использование оригинального текста с русскими тональными словарями и частичный перевод с английскими словарями) на 5-10% по показателю взвешенной F-меры.

Таблица 4. Результаты распознавания сентимента и эмоций при использовании различных методов машинного перевода, взвешенная F-мера (%)

Транскрипции	Тип перевода	Эмоции	Сентимент
Экспертные	Русскоязычный текст	27,42	65,61
	Полный перевод	<b>31,12</b>	<b>75,37</b>
	Частичный перевод	26,21	71,79
Автоматические	Русскоязычный текст	23,09	67,79
	Полный перевод	22,78	66,62
	Частичный перевод	<b>23,74</b>	<b>71,60</b>

Для автоматических транскрипций корпуса RAMAS наиболее эффективным методом достичь наибольшей точности распознавания сентимента и эмоций является использование частичного перевода на английский язык. Такие результаты могут быть обоснованы низким качеством автоматических транскрипций, которые в свою очередь зависят от качества записи корпуса данных. Можно предположить, что наличие фраз сторонних дикторов в анализируемом высказывании сильно зашумляет эмоциональную окраску всего текста.

**5.3. Репрезентативные признаки.** В настоящем исследовании предлагается извлекать текстовые признаки при помощи тональных словарей. В предложенном методе вектор транскрипций состоит из статистического вектора и тонального вектора. Статистический вектор представляет собой набор статистических показателей всего анализируемого высказывания и состоит из следующих параметров: количество положительных, отрицательных, нейтральных слов, среднее значение сентимента всех слов, сумма весов положительных и отрицательных слов, максимальное и минимальное значение весов (всего восемь значений). Тональный вектор содержит информацию о последовательности эмоциональных слов в высказывании.

В экспериментах выше использовался конкатенированный вектор из статистического и тонального векторов. Для приведения всех векторов к единой длине все вектора либо обрезались, либо дополнялись нулями до значения 98 (8 – статистический вектор и 90 – тональный вектор). В данном экспериментальном исследовании исследуется гипотеза о том, является ли статистический вектор полностью репрезентативным для анализа сентимента. Под репрезентативностью понимается содержание характеристик в малой выборке, отражающей характеристики генеральной совокупности. Суть эксперимента заключалась в следующем: использовалось различное количество показателей статистического вектора (от двух до восьми), использовались различные комбинации статистических показателей, с каждой комбинацией статистических показателей векторов строился классификатор распознавания сентимента для всех типов данных (оригинальные русскоязычные тексты, полный и частичный переводы на английский язык), затем результаты усреднялись по комбинациям в зависимости от количества использованных статистических показателей. Полученные результаты представлены на рисунке 6 в виде гистограммы, по оси абсцисс которой располагаются количество параметров (в комбинации) в статистическом векторе для различных типов данных, по оси ординат – точность распознавания трех классов сентимента по показателю взвешенной F-меры.

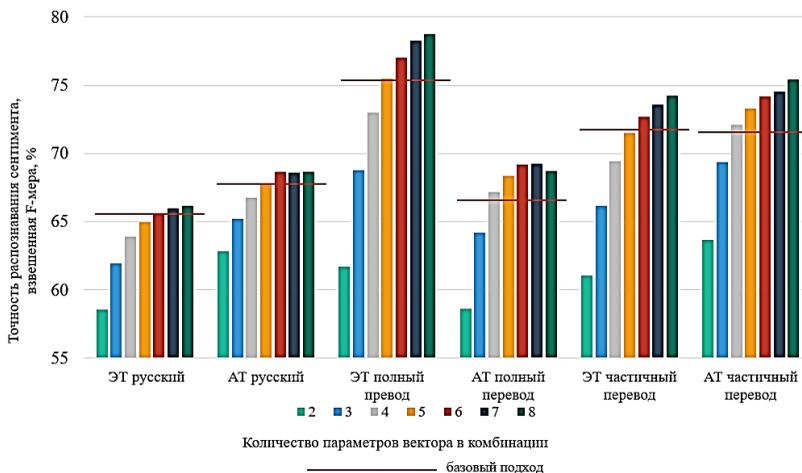


Рис. 6. Гистограмма влияния количества параметров статистического вектора на точность распознавания сентимента

На рисунке 6 красной линией отмечен базовый результат, полученный в ходе предыдущих экспериментальных исследований с использованием конкатенированного вектора (статистического и тонального) текстовых транскрипций, представленный в таблице 4. Как можно заметить из результатов экспериментов, использование только статистических показателей при анализе сентимента в текстовых данных позволяет достичь более высокой точности относительно использования конкатенированного вектора. В большинстве случаев, чем больше количество статистических показателей, тем выше точность распознавания сентимента. Однако, даже использование минимум шести статистических показателей позволяют превзойти базовый метод (в котором используется конкатенированный вектор) примерно на 1-5% по показателю взвешенной F-меры. Это может быть обосновано тем, что классификатор при анализе только статистических показателей сосредотачивается на более репрезентативной информации о сентименте. Под репрезентативностью понимается сконцентрированная информация о сентименте, находящаяся в статистическом векторе, когда тональный вектор содержит в себе разряженную информацию. Также можно сделать вывод о том, что последовательность эмоциональных слов (тональный вектор) для автоматического распознавания сентимента не важна. Более того, использование тонального вектора может только зашумлять информацию о тональности в высказываниях, что делает данный вектор

не репрезентативным для классификатора. Еще одним преимуществом использования только статистического вектора является сокращение времени обучения классификаторов распознавания сентимента. В предыдущих исследованиях авторов данной статьи [15] для распознавания сентимента использовался только тональный вектор на автоматических транскрипциях корпуса RAMAS, где достигнуто значение 43,31% взвешенной F-меры, что на 22% ниже использования только статистического вектора для аналогичной задачи.

Также в ходе экспериментов было установлено какие именно статистические показатели вносят больший вклад в распознавание сентимента в текстовых транскрипциях. Для выделения такой информации была использована оригинальная методика, описанная ниже. На рисунке 7 представлена гистограмма, по оси абсцисс которой представлены статистические показатели, а по оси ординат усредненная точность распознавания сентимента по показателю взвешенной F-меры в %. Усреднение точности происходило на основе предыдущего эксперимента по принципу усреднения полученной точности распознавания сентимента для каждого вектора, в котором встречался анализируемый показатель. Например, параметр количество негативных слов в предыдущем эксперименте встречался в различных  $M$  векторах. Для получения усредненной точности рассматриваемого параметра вычислялись точности распознавания сентимента с использованием каждого из векторов  $M$  и находилось среднее арифметическое этих значений точности.

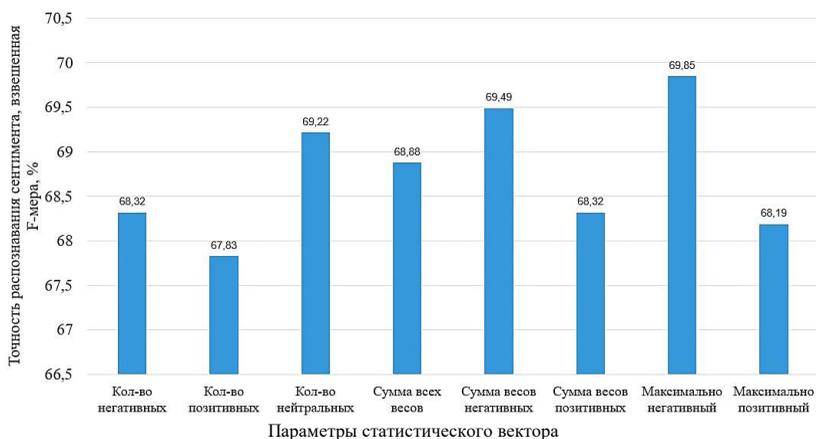


Рис. 7. Гистограмма вклада параметров статистического вектора в точность распознавания сентимента

Гистограмма на рисунке 7 свидетельствует о том, что каждый статистический показатель достаточно важен для распознавания сентимента. Так, разброс между крайними значениями показателей составляет менее 2% (в диапазоне от 67,75 до 69,75%). При этом такие показатели как максимальный вес негативного слова, сумма весов негативных слов и количество нейтральных слов вносят больший вклад в распознавание сентимента в текстовых транскрипциях речевых высказываний.

**6. Обсуждение результатов.** Вследствие проведенных экспериментов получены несколько теоретических и эмпирических результатов, которые имеют значение в контексте автоматического анализа текстов на естественных языках. Во-первых, лемматизация оказалась более предпочтительным методом, повышая точность распознавания как для экспертных транскрипций, так и для автоматических. По-видимому, лемматизация позволяет обеспечить лучший результат распознавания нормализованного текста за счет того, что учитывает контекст и частеречную принадлежность слова. Кроме того, лемматизация позволяет снизить количество ошибок, которые могут возникнуть из-за неправильного обрезания слов при стемминге. С другой стороны, лемматизация, в отличие от стемминга, требует больших временных и вычислительных ресурсов.

Другим важным результатом является возможность объединения тональных словарей. Тональные словари являются одним из определяющих факторов для определения сентимента текстов, однако могут быть неполными и/или не учитывать различные факторы. Объединение данных из различных источников предсказуемо повышает количество обучающих данных. В то же время, привлечение языкового материала на иностранных языках повышает качество распознавания по той причине, что основным способом выражения оценочной модальности (эмоциональной окраски текста) в языках мира является именно лексика, а не синтаксис или морфология. Лексические единицы, маркированные в отношении оценочной модальности, проявляют устойчивость, совпадая по тональности в различных языках. Таким образом, машинный перевод оказывается эффективным инструментом, позволяющим пополнять обучающие данные напрямую из других языков. Следует иметь в виду, что однозначный машинный перевод с одного естественного языка на другой пока невозможен, поэтому машинный перевод должен, в идеале, сопровождаться экспертной проверкой. Так или иначе, даже прямолинейное применение машинного перевода в представленных в рамках настоящей статьи экспериментах позволило поднять значение

F-меры распознавания на 1-4% для эмоций, и на 5-10% для сентимента.

Наконец, последним важным выводом является возможность применения статистического вектора на этапе векторизации текста вместо конкатенированного тонального и статистического. В самом деле, объединение тональных и статистических векторов может увеличить размерность векторного пространства, что может привести к проблемам с памятью и производительностью модели определения тональности. Как показали эксперименты, использование только статистических векторов может быть более эффективным для достижения большей точности распознавания сентимента, а также уменьшения времени автоматического анализа текстовых данных.

**7. Заключение.** Статья посвящена разработке метода распознавания сентимента и эмоций в орографических транскрипциях речи дикторов. Предложенный метод основан на использовании тональных словарей для векторизации текстов, а также машинного перевода русскоязычных транскрипций на английский язык. В статье описываются экспериментальные исследования с методами предобработки текстовых данных, машинного перевода (полного и частичного), выделения репрезентативных признаков из тональных векторов.

По результатам проведенного исследования можно сделать следующие выводы:

1) Лемматизация в качестве метода нормализации слов в текстовых транскрипциях речи позволяет повысить точность распознавания сентимента и эмоций.

2) Полный перевод на английский язык русскоязычных экспертных транскрипций речи увеличивает точность распознавания эмоций (семь классов) и сентимента (три класса).

3) Для автоматических транскрипций речи частичный перевод на английский язык помогает увеличить точность распознавания сентимента и эмоций.

4) Использование статистического вектора в качестве векторизации текстовых транскрипций речи позволяет получить точность выше относительно использование конкатенированного (статистического и тонального) вектора.

Для достижения наилучших результатов в распознавании эмоций в русскоязычных текстах необходимо комбинировать различные методы и подходы, включая машинный перевод и использование объединенных тональных словарей. Важно также учитывать специфику текста и задачу, которую необходимо решить,

при выборе подходящего метода. В целом, исследования в области анализа сентимента и эмоций в русскоязычных текстах необходимо продолжать с целью поиска эффективных методов, с помощью которых можно достичь наибольшей точности.

### Литература

1. Николаев И.С., Митренина О.В., Ландо Т.М. Прикладная и компьютерная лингвистика // М.:ЛЕНАНД. 2017. 320 с.
2. Carosia A.E.O., Coelho G.P., Silva A.E.A. Analyzing the Brazilian financial market through portuguese sentiment analysis in social media // Applied Artificial Intelligence. 2020. vol. 34. no. 1. pp. 1–19.
3. Smetanin S. The applications of sentiment analysis for Russian language texts: Current challenges and future perspectives // IEEE Access. 2020. vol. 8. pp. 110693–110719. DOI: 10.1109/ACCESS.2020.3002215.
4. Карпов А.А., Юсупов Р.М. Многомодальные интерфейсы человеко-машинного взаимодействия // Вестник Российской академии наук. 2018. Т. 88. № 2. С. 146–155.
5. Dvoynikova A., Verkholiyak O., Karpov A. Analytical review of methods for identifying emotions in text data // CEUR-WS. 2020. vol. 2552. pp. 8–21.
6. Ekman P. An Argument for Basic Emotions // Cognition and Emotion. 1992. vol. 6(3-4). pp. 169–200.
7. Dvoynikova A., Karpov A. Bimodal sentiment and emotion classification with multi-head attention fusion of acoustic and linguistic information // Computational Linguistics and Intellectual Technologies. 2023. vol. 22. pp. 51–61.
8. Viksna R., Jekabsons G. Sentiment analysis in Latvian and Russian: A survey // Applied Computer Systems. 2018. vol. 23. no. 1. pp. 45–51.
9. Loukachevitch N., Levchik A. Creating a general Russian sentiment lexicon // Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). 2016. pp. 1171–1176.
10. Demirtas E., Pechenizkiy M. Cross-lingual polarity detection with machine translation // Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining. 2013. pp. 1–8.
11. Reichel J., Benko L. The Influence of a Machine Translation System on Sentiment Levels // RASLAN 2022 Recent Advances in Slavonic Natural Language Processing. 2022. pp. 201–208.
12. Zygadlo A., Kozlowski M., Janicki A. Text-Based emotion recognition in English and Polish for therapeutic chatbot // Applied Sciences. 2021. vol. 11(21). no. 10146.
13. Nandwani P., Verma R. A review on sentiment analysis and emotion detection from text // Social Network Analysis and Mining. 2021. vol. 11(1). no. 81.
14. Hartung K., Herygers A., Kurlekar S.V., Zakaria K., Volkan T., Gröttrup S., Georges M. Measuring Sentiment Bias in Machine Translation // International Conference on Text, Speech, and Dialogue. 2023. pp. 82–93.
15. Двойникова А.А. Сентимент-анализ транскрипции разговорной речи при помощи автоматического машинного перевода // Сборник трудов IX Конгресса молодых ученых. 2021. С. 199–203.
16. Perepelkina O., Kazimirova E., Konstantinova M. RAMAS: Russian Multimodal Corpus of Dyadic Interaction for studying emotion recognition // PeerJ Preprints. 2018. vol. 6. no. e26688v1.
17. Russell J.A. A circumplex model of affect // Journal of personality and social psychology. 1980. vol. 39. no. 6. pp. 1161–1178.

18. Плунгян В.А. Введение в грамматическую семантику: Грамматические значения и грамматические системы языков мира // М.: РГГУ. 2011. 672 с.
19. Goddard C., Wierzbicka A. Semantic and Lexical Universals // *Studies in Second Language Acquisition*, 1996. vol. 18(4). 520 p.
20. Котельников Е.В., Разова Е.В., Котельникова А.В., Вычерганин С.В. Современные словари оценочной лексики для анализа мнений на русском и английском языках (аналитический обзор) // *Научно-техническая информация. Серия. 2020. Т. 2. С. 16–33.*
21. Hercig T., Brychcín T., Svoboda L., Konkol M. Uwb at semeval-2016 task 5: Aspect based sentiment analysis // *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, 2016. pp. 342–349.
22. Koltsova O.Y., Alexeeva S., Kolcov S. An opinion word lexicon and a training dataset for Russian sentiment analysis of social media // *Computational Linguistics and Intellectual Technologies*, 2016. vol. 15. pp. 277–287.
23. Strapparava C., Valitutti A. Wordnet affect: an affective extension of wordnet // *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, 2004. pp. 1083–1086.
24. Mohammad S.M., Turney D.P. Crowdsourcing a word-emotion association lexicon // *Computational Intelligence*, 2013. vol. 29(3). pp. 436–465.
25. Hu M., Liu B. Mining and summarizing customer reviews // *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004. pp. 168–177.
26. Hutto C., Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text // *Proceedings of the international AAAI conference on web and social media*, 2014. vol. 8, no. 1. pp. 216–225.

**Двойникова Анастасия Александровна** — младший научный сотрудник, лаборатория речевых и многомодальных интерфейсов, Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук". Область научных интересов: искусственный интеллект, машинное обучение, нейронные сети, сентимент-анализ, анализ аффективных состояний человека. Число научных публикаций — 20. dvoynikova.a@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-0421.

**Кагиров Ильдар Амирович** — научный сотрудник, лаборатория речевых и многомодальных интерфейсов, Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук". Область научных интересов: корпусная лингвистика, малоресурсные языки. Число научных публикаций — 40. kagirov@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-0421.

**Карпов Алексей Анатольевич** — д-р техн. наук, профессор, руководитель лаборатории, лаборатория речевых и многомодальных интерфейсов, Федеральное государственное бюджетное учреждение науки "Санкт-Петербургский Федеральный исследовательский центр Российской академии наук". Область научных интересов: речевые технологии, автоматическое распознавание речи, обработка аудиовизуальной речи, многомодальные человеко-машинные интерфейсы, компьютерная паралингвистика и другие. Число научных публикаций — 350. karpov@iias.spb.su; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-0421.

**Поддержка исследований.** Раздел 4 выполнен в рамках бюджетной темы СПб ФИЦ РАН (№ FFZF-2022-0005), остальные исследования выполнены при финансовой поддержке Российского научного фонда, проект № 22-11-00321.

A. DVOYNIKOVA, I. KAGIROV, A. KARPOV  
**A METHOD FOR RECOGNITION OF SENTIMENT  
AND EMOTIONS IN RUSSIAN SPEECH TRANSCRIPTS USING  
MACHINE TRANSLATION**

*Dvoynikova A., Kagirov I., Karpov A. A Method for Recognition of Sentiment and Emotions in Russian Speech Transcripts Using Machine Translation.*

**Abstract.** This paper addresses the issue of user emotions and sentiment recognition in transcripts of Russian speech samples using lexical methods and machine translation. The availability of data for sentiment analysis in Russian texts is quite limited, thus this paper proposes a new approach which is based on automatic machine translation of Russian texts into English. Additionally, the paper presents the results of experimental research regarding the impact of partial and full machine translation on emotion and sentiment recognition. Partial translation means translating single lexemes not included in Russian sentiment dictionaries, while full translation implies translating the entire text. A translated text is further analyzed using different English sentiment dictionaries. Experiments have demonstrated that the combination of all English sentiment dictionaries enhances the accuracy of emotion and sentiment recognition in text data. Furthermore, this paper explores the correlation between the length of the text data vector and its representativity. Experimental research for emotion and sentiment recognition tasks was conducted with the use of expert and automatic transcripts of the multimodal Russian corpus RAMAS. Based on the experimental results, one can conclude that the use of word lemmatization is a more effective approach for normalizing words in speech transcripts compared to stemming. The use of the proposed methods involving full and partial machine translation allows for an improvement in sentiment and emotion recognition accuracy by 0.65-9.76% in terms of F-score compared to the baseline approach. As a result of the application of machine translation methods to expert and automatic transcriptions of the Russian speech corpus RAMAS, an accuracy in recognition of 7 emotion classes was achieved at 31.12% and 23.74%, and 3 sentiment classes at 75.37% and 71.60%, respectively. Additionally, the experiments revealed that the use of statistical vectors as a text data vectorization method results in an a 1-5% increase in F-score value compared to concatenated (statistical and sentiment) vectors.

**Keywords:** machine translation, sentiment dictionaries, emotion recognition, sentiment analysis, sentiment vectors.

## References

1. Nikolaev I.S., Mitrenina O.V., Lando T.M. *Prikladnaya i komp'yuternaya lingvistika [Applied and computational linguistics]*. M.: LENAND, 2017. 320 p (In Russ.).
2. Carosia A.E.O., Coelho G.P., Silva A.E.A. Analyzing the Brazilian financial market through portuguese sentiment analysis in social media. *Applied Artificial Intelligence*. 2020. vol. 34. no. 1. pp. 1–19.
3. Smetanin S. The applications of sentiment analysis for Russian language texts: Current challenges and future perspectives. *IEEE Access*. 2020. vol. 8. pp. 110693–110719. DOI: 10.1109/ACCESS.2020.3002215.
4. Karpov A.A., Yusupov R.M. Multimodal interfaces of human-computer interaction, *Herald of the Russian Academy of Sciences*. 2018. vol. 88. no. 2. pp. 146–155.
5. Dvoynikova A., Verkholyak O., Karpov A. Analytical review of methods for identifying emotions in text data. *CEUR-WS*. 2020. vol. 2552. pp. 8–21.

6. Ekman P. An Argument for Basic Emotions. *Cognition and Emotion*. 1992. vol. 6(3-4). pp. 169–200.
7. Dvoynikova A., Karpov A. Bimodal sentiment and emotion classification with multi-head attention fusion of acoustic and linguistic information. *Computational Linguistics and Intellectual Technologies*. 2023. vol. 22. pp. 51–61.
8. Viksna R., Jekabsons G. Sentiment analysis in Latvian and Russian: A survey. *Applied Computer Systems*. 2018. vol. 23. no. 1. pp. 45–51.
9. Loukachevitch N., Levchik A. Creating a general Russian sentiment lexicon. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 2016. pp. 1171–1176.
10. Demirtas E., Pechenizkiy M. Cross-lingual polarity detection with machine translation. *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*. 2013. pp. 1–8.
11. Reichel J., Benko E. The Influence of a Machine Translation System on Sentiment Levels. *RASLAN 2022 Recent Advances in Slavonic Natural Language Processing*. 2022. pp. 201–208.
12. Zygadlo A., Kozlowski M., Janicki A. Text-Based emotion recognition in English and Polish for therapeutic chatbot. *Applied Sciences*. 2021. vol. 11(21). no. 10146.
13. Nandwani P., Verma R. A review on sentiment analysis and emotion detection from text // *Social Network Analysis and Mining*. 2021. vol. 11(1). no. 81.
14. Hartung K., Herygers A., Kurlekar S.V., Zakaria K., Volkan T., Gröttrup S., Georges M. Measuring Sentiment Bias in Machine Translation. *International Conference on Text, Speech, and Dialogue*. 2023. pp. 82–93.
15. Dvoynikova A.A. [Sentiment analysis of transcription of spoken speech using automatic machine translation] *Sbornik trudov IX Kongressa molodyh uchenykh [Proceedings of the IX Congress of Young Scientists]*. 2021. pp. 199–203. (In Russ.).
16. Perepelkina O., Kazimirova E., Konstantinova M. RAMAS: Russian Multimodal Corpus of Dyadic Interaction for studying emotion recognition. *PeerJ Preprints*. 2018. vol. 6. no. e26688v1.
17. Russell J.A. A circumplex model of affect. *Journal of personality and social psychology*. 1980. vol. 39. no. 6. pp. 1161–1178.
18. Plungjan V.A. Vvedenie v grammaticheskuyu semantiku: Grammaticheskie znachenija i grammaticheskie sistemy jazykov mira [Introduction to Grammatical Semantics: Grammatical Meanings and Grammatical systems of the languages of the world]. M.: RSUH. 2011. 672 p (In Russ.).
19. Goddard C., Wierzbicka A. Semantic and Lexical Universals. *Studies in Second Language Acquisition*, 1996. vol. 18(4). 520 p.
20. Kotel'nikov E.V., Razova E.V., Kotel'nikova A.V., Vyhegzhani S.V. [Modern dictionaries of evaluation vocabulary for the analysis of opinions in Russian and English (analytical review)]. *Nauchno-tehnicheskaja informacija. Serija – Scientific and technical information. Series*. 2020. vol. 2. pp. 16–33 (In Russ.).
21. Hercig T., Brychcin T., Svoboda L., Konkol M. Uwb at semeval-2016 task 5: Aspect based sentiment analysis. *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*. 2016. pp. 342–349.
22. Koltsova O.Y., Alexeeva S., Kolcov S. An opinion word lexicon and a training dataset for Russian sentiment analysis of social media. *Computational Linguistics and Intellectual Technologies*. 2016. vol. 15. pp. 277–287.
23. Strapparava C., Valitutti A. Wordnet affect: an affective extension of wordnet. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*. 2004. pp. 1083–1086.
24. Mohammad S.M., Turney D.P. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*. 2013. vol. 29(3). pp. 436–465.

25. Hu M., Liu B. Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004. pp. 168–177.
26. Hutto C., Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the international AAAI conference on web and social media. 2014. vol. 8. no. 1. pp. 216–225.

**Dvoynikova Anastasia** — Junior researcher, Laboratory of speech and multimodal interfaces, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: artificial intelligence, machine learning, neural networks, sentiment analysis, human affective states analysis. The number of publications — 20. dvoynikova.a@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-0421.

**Kagirov Ildar** — Researcher, Laboratory of speech and multimodal interfaces, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: corpus linguistics, low-resource languages. The number of publications — 40. kagirov@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-0421.

**Karpov Alexey** — Ph.D., Dr.Sci., Professor, Head of the laboratory, Laboratory of speech and multimodal interfaces, St. Petersburg Federal Research Center of the Russian Academy of Sciences. Research interests: speech technology, automatic speech recognition, audio-visual speech processing, multimodal human-computer interfaces, and computational paralinguistics. The number of publications — 350. karpov@iias.spb.su; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-0421.

**Acknowledgements.** Research presented in Section 4 was financially supported by a state research grant for SPC RAS (Topic No. FFZF-2022-0005), the remainder of the study was funded by RSF (Project 22-11-00321).

Е.В. ВЛАСОВ, Н.П. КРАСНЕНКО  
**КАСКАДНЫЙ КЛАССИФИКАТОР ДЛЯ ОБНАРУЖЕНИЯ  
И ИДЕНТИФИКАЦИИ ПТИЦ В ВИДЕОПОТОКЕ**

*Власов Е.В., Красненко Н.П. Каскадный классификатор для обнаружения и идентификации птиц в видеопотоке.*

**Аннотация.** В статье представлен разработанный метод и прототип программы для определения наличия птиц в видеопотоке данных в режиме реального времени. Этот метод основан на использовании каскадного классификатора, который был применен для решения задачи обнаружения и идентификации птиц в биоакустической установке отпугивания птиц в аэропорту Томска. В рамках исследования был использован каскадный классификатор Виолы-Джонса, который является одной из реализаций алгоритма каскад Хаара. Этот алгоритм позволяет с высокой точностью и скоростью обнаруживать объекты на изображениях и видео. В данном случае классификатор был обучен на наборе данных, содержащем изображения птиц, что позволило достичь высокой точности обнаружения и идентификации птиц на видео. Также приведены результаты оценки возможностей созданного классификатора и продемонстрирована его высокая результативность. В ходе исследования были использованы различные методы машинного обучения и анализа видеоданных, что позволило получить точные и надежные результаты. В целом, данная работа представляет собой инновационный подход к решению актуальной задачи защиты аэропортов от птиц. Применение разработанного метода позволило повысить эффективность работы биоакустической установки отпугивания птиц и обеспечить безопасность полетов в аэропорту Томска, снизив вероятность столкновения самолетов с птицами. Новизна работы заключается в применении метода Виолы-Джонса к задаче обнаружения и идентификации птиц с оценкой его результативности. Таким образом, представленная в статье работа является важным вкладом в развитие методов обнаружения и идентификации объектов на видео и может быть использована в других областях, где требуется автоматическое обнаружение и классификация объектов в видеопотоке данных.

**Ключевые слова:** авиационная орнитология и безопасность, видеонаблюдение, обнаружение и идентификация птиц.

**1. Введение.** Птицы являются объектом исследований как любителей, так и профессиональных орнитологов различных направлений. Одним из таких направлений является авиационная орнитология [1 – 5], где занимаются птицами, представляющими опасность для полетов воздушных судов.

Согласно статистическим данным [6 – 8], столкновения самолётов с птицами на территории России и в мире происходят регулярно на протяжении всего года и являются основным повреждающим фактором самолетов биологического происхождения. Ежегодно прослеживается тенденция к увеличению случаев столкновений, что обусловлено ростом интенсивности авиаперевозок, снижением шумности двигателей самолетов, привыканием птиц к окружающей обстановке. Наибольшее количество столкновений

происходит на территории аэродрома и вблизи от него на высотах до 300 м. Поэтому в основе эффективного обеспечения орнитологической безопасности полётов воздушных судов лежат знания орнитологической обстановки на территории лётного поля и его окрестностей. Практика в области орнитологического обеспечения безопасности полётов показывает, что при всём богатстве орнитофауны в столкновениях с воздушными судами участвует определённый спектр видов птиц, более-менее специфичный для каждого конкретного аэропорта, и их необходимо вовремя обнаруживать, идентифицировать и отпугивать.

Правила орнитологического обеспечения безопасности полетов на аэродроме [9] включают в себя мероприятия по визуальному и радиолокационному контролю за орнитологической обстановкой, по оперативному оповещению экипажей воздушного судна при возникновении опасности столкновений с птицами, отпугиванию скоплений птиц, предотвращению условий, способствующих концентрации птиц и т.д.

Для отпугивания птиц от взлетно-посадочной полосы аэродромов используются различные методы и средства [1, 4, 5, 10], но наибольшее применение получили биоакустические отпугиватели [5, 11 – 19], автоматически воспроизводящие звуковые сигналы криков хищных птиц, а также криков «бедствия» и «тревоги» для различных видов птиц. Их общим недостатком является привыкание птиц к постоянным звуковым сигналам при отсутствии реальной опасности.

Этот недостаток пытаются компенсировать путем использования систем видеонаблюдения за окружающей обстановкой путем установки видеокамер совместно с акустическими излучателями [14]. Видеоинформация по двухсторонней оптоволоконной линии связи передается на удаленный пункт дистанционного мониторинга (автоматизированное рабочее место диспетчера) для принятия решения по работе отпугивателей. Такое решение по отпугиванию птиц с использованием модуля (блока), осуществляющего функции наблюдения, обнаружения, распознавания птиц, предложено в работах [15, 16]. Функции обнаружения реализуются на основе данных камер видеонаблюдения (тепловизионных и дневных стереокамер), которые постоянно сканируют окружающую среду над охраняемой территорией на предмет входящих в неё летающих объектов. Эти данные через блок управления передаются в центр дистанционного мониторинга, где обрабатываются, анализируются и принимается решение. Как только летящий объект обнаружен, и система распознает его как птицу, она

ориентирует посредством электропривода в это направление акустический модуль отпугивания, который излучает соответствующий звуковой сигнал, вызывающий эффект испуга птицы, для изменения курса её полета и покидания контролируемой зоны.

Таким образом, для повышения эффективности работы биоакустических отпугивателей птиц целесообразно использовать, в дополнение к ним, систему компьютерного (технического) зрения, определяющую наличие птиц в зоне действия устройства и работающую только в этом случае с выборочным сигналом звукового воздействия на конкретный фиксируемый вид птицы. Желательно чтобы такая установка работала автономно. Для автоматизации её работы прежде всего необходимо решить задачу по обнаружению и распознаванию птиц в видеопотоке поступающих данных.

В последнее время научные направления по цифровой обработке изображений, компьютерному зрению, робототехнике и искусственному интеллекту, связанные с поставленной задачей, получили активное развитие и реализацию [20 – 35]. В них описаны алгоритмы для обнаружения и распознавания различных объектов из изображений и видеоданных. Так в работе [26] представлен обзор по объектному обнаружению за последние 20 лет по 296 источникам цитирования. Для обнаружения и распознавания птиц используются различные методы и алгоритмы [27 – 35]. Они, конечно, характеризуются различными возможностями в реализации, в том числе за счет применения различных цифровых видеокамер и обрабатывающих компьютеров. В последнее время для решения данной задачи стали использоваться нейросетевые технологии.

В данной работе при разработке алгоритма обнаружения и идентификации птиц использовался метод каскадного классификатора Хаара [23]. Это эффективный метод обнаружения объектов, предложенный Полом Виолой и Майклом Джонсом [33]. Данный подход основан на машинном обучении, где Каскадная функция обучается на большом количестве положительных и отрицательных изображений. Алгоритм получается относительно нетребовательным к вычислительной мощности компьютера и позволяет использовать его в автоматических полевых установках для обнаружения птиц. Он может также использоваться для мониторинга популяций птиц: для отслеживания и анализа поведения, численности и миграции птиц в определенных районах.

## **2. Применение методов технического зрения для формирования обучающей выборки и обучения классификатора.**

Как показывает обзор рынка разработанных и используемых биоакустических средств отпугивания, то проблему привыкания птиц, за исключением отдельных случаев [15, 16, 18, 19], пытаются решить применением режима случайной трансляции наборов сигналов, а также режима работы по командам диспетчера, не учитывая наличие птиц и видоспецифичность звуковых отпугивающих птиц сигналов. В то же время, используя методы компьютерного зрения и каскадный классификатор для распознавания образов в видеопотоке получаемых данных, можно запускать отпугивание в автоматическом режиме и только при наличии птиц в поле зрения биоакустической установки. Ниже рассматривается построение технологии такой работы и программы [36, 37], реализующей задачу обнаружения и идентификации птиц в аэропорту г. Томска.

**2.1. Среда разработки программы.** В качестве среды разработки программы выбран Qt Creator – кроссплатформенная свободная интегрированная среда [38], уже знакомая авторам, для разработки на языках C, C++ и QML. Разработана Trolltech (Digia) для работы с фреймворком Qt.

Для реализации обработки изображений в поступающем видеопотоке данных использована OpenCV – библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения с открытым кодом [39–41]. Она реализована на языке C/C++, также разрабатывается для Python, Java, Ruby, Matlab, Lua и других языков.

**2.2. Формирование обучающей выборки.** В качестве обучающих данных выбраны изображения птиц полученные в их естественной среде обитания, а также контрпримеры в той же среде. Для их формирования реализован метод обнаружения движущихся объектов. Реализацию данного метода рассмотрим на примере рисунка 1.



Рис. 1. Пример исходного изображения с птицами

Для упрощения работы цветное изображение преобразуется в чёрно-белое. Для этого используем функцию библиотеки OpenCV:

```
void cv::cvtColor(InputArray src, OutputArray dst, int code),
```

где `InputArray src` – входное изображение,

`OutputArray dst` – выходное изображение,

`int code` – код преобразования.

На рисунке 2 показан результат преобразования исходного цветного изображения в чёрно-белое с кодом преобразования `COLOR_BGR2GRAY`.



Рис. 2. Фрагмент преобразованного изображения

Теперь сравним два последовательно снятых кадра. Для наглядного сравнения они представлены на рисунке 3 с использованием полупрозрачного наложения друг на друга.



Рис. 3. Фрагмент изображения двух наложенных последовательных кадров

Преобразуем оба исходных кадра, изображённых на рисунке 3, также в чёрно-белые с кодом преобразования `COLOR_BGR2GRAY` и сравним их, используя функцию библиотеки `OpenCV`:

```
void absdiff(InputArray src1, InputArray src2, OutputArray dst),
```

где `InputArray src1` – первое входное изображение,

`InputArray src2` – второе входное изображение,

`OutputArray dst` – выходное изображение (результат сравнения).

Результат сравнения двух кадров представлен на рисунке 4.

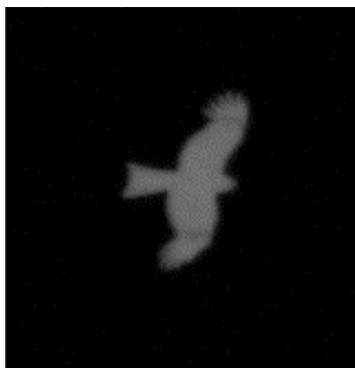


Рис. 4. Фрагмент результата сравнения двух кадров

В результате сравнения двух кадров на рисунке 4 видны 3 подсвеченные области по форме повторяющие движущиеся объекты (птицы).

Далее преобразуем изображение, полученное путём сравнения двух кадров (рис. 4), используя порог. Эта функция определена в библиотеке OpenCV:

```
double threshold( InputArray src, OutputArray dst,  
double thresh, double maxval, int type);
```

где InputArray src – входное изображение,  
OutputArray dst – выходное изображение,  
double thresh – порог минимального значения

яркости пикселя,

double maxval – максимальное значение яркости пикселя,

int type – тип порога.

Зададим порог минимального значения яркости пикселя равный 15, максимальное значение яркости пикселя равное 255, тип порога THRESH\_BINARY. Результат применения порога (бинаризации) к предыдущему изображению (рисунок 4) представлен на рисунке 5.

Теперь все изменения пикселей выделены в отдельные области, которые можно идентифицировать как отдельные объекты.



Рис. 5. Фрагмент изображения (рисунок 4) с применением порога (бинаризации)

Из полученного изображения, представленного на рисунке 5, алгоритм пронумеровывает каждый объект для дальнейшего обращения к нему. Результат представлен на рисунке 6.



Рис. 6. Фрагмент изображения с выделенным на нём объектом

Слежение за каждым объектом осуществляется путём наложения двух последовательно снятых кадров друг на друга. Пример наглядного наложения двух кадров, с выделенными на них объектами, представлен на рисунке 7.

Из рисунка 7 видно, что, сравнивая два кадра с предварительно выделенными на них объектами, можно отследить траекторию перемещения этих объектов. Результат работы алгоритма по идентификации и слежению за объектами представлен на рисунке 8.

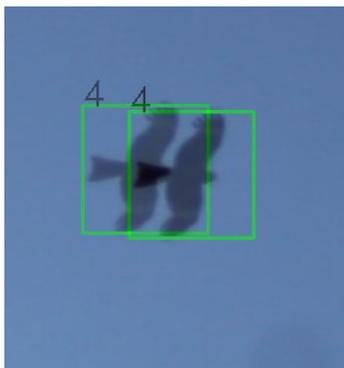


Рис. 7. Фрагмент изображения двух наложенных последовательных кадров, с выделенными на них объектами



Рис. 8. Демонстрация работы алгоритма по идентификации и отслеживанию траекторий объектов

Таким образом, используя алгоритм по идентификации и отслеживанию траекторий объектов и видеоряд с изображением птиц, достаточно добавить функцию сохранения каждого объекта в виде отдельных изображений для формирования обучающей выборки.

Результат работы алгоритма формирования обучающих данных виден на рисунке 9.

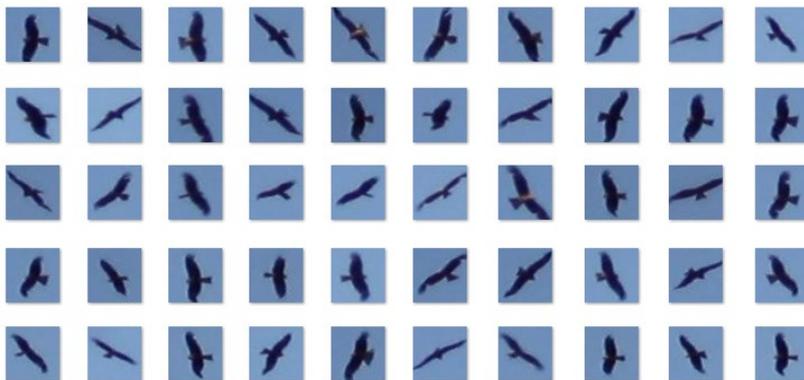


Рис. 9. Изображения объектов распознавания

В качестве контрпримеров для обучения берутся различные изображения каких либо других объектов, которые могут находиться

в среде обитания искомого объекта (птицы). Они приводят к ложным срабатываниям классификатора и используются для повторного переобучения. На рисунке 10 приведены такие примеры.



Рис. 10. Изображения контрпримеров

Сформированные таким образом обучающие данные применяются для обучения классификатора.

**2.4. Каскадный классификатор.** Здесь Каскадная функция обучается на большом количестве положительных (птиц) и отрицательных изображений. Ключевую роль в методе Виолы-Джонса играют признаки. Они используются для описания объектов на изображении и помогают определить, является ли объект птицей или нет. Признаки могут быть различными, например, форма объекта, его размер, цвет, текстура и т.д. Признаки Хаара являются более сложными функциями, которые описывают форму объектов на изображении, и они строятся на основе примитивов Хаара. Примитивы Хаара представляют собой прямоугольные области, которые делятся на левую и правую части. Каждая часть имеет свою яркость, которая сравнивается с яркостью другой части. Если яркость левой части больше, то объект считается обнаруженным (рисунок 11).

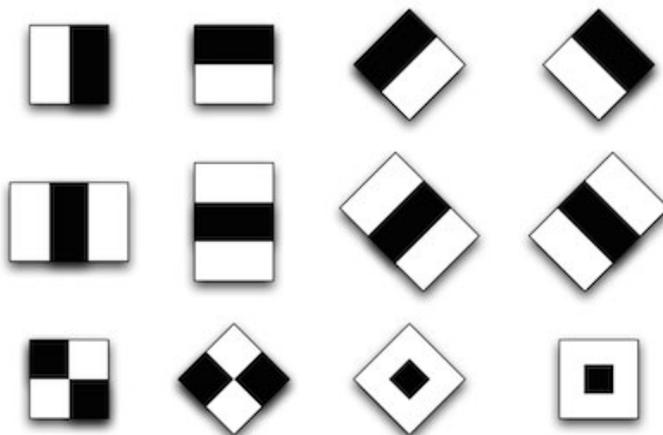


Рис. 11. Прimitives Хаара

Значение примитива вычисляется по формуле:

$$F = X - Y,$$

где  $X$  – сумма яркостей точек в светлой части примитива, а  $Y$  – сумма яркостей точек в темной части примитива.

Для работы с большим массивом данных используется интегральное представление изображения – матрица, в каждом элементе которой находится сумма яркости пикселей, размещенных выше и левее данного элемента. Рассчитывается каждый элемент по формуле:

$$L(x, y) = I(x, y) - L(x - 1, y - 1) + L(x, y - 1) + L(x - 1, y),$$

где  $I(x, y)$  – яркость пикселя исходного изображения,  $L(x, y)$  – полученный элемент матрицы.

Применение интегрального представления позволяет сократить количество вычислений для расчета признаков Хаара.

Изначально алгоритму требуется много положительных изображений (изображений искомого объекта – птицы) и отрицательных изображений (изображений, не содержащих искомого объекта) для обучения классификатора. Затем нужно извлечь из него значения примитивов, показанных на рисунке 11.

Перебирая все возможные размеры и местоположения каждого примитива, можно рассчитать множество признаков, характерных

изображению. Но даже окно 24 на 24 пикселя дает более 160000 функций. Среди всех этих характеристик, большинство не имеет значения. Поэтому каждый признак применяют ко всем обучающим изображениям находя лучший порог, который классифицирует объекты на положительные и отрицательные. Очевидно, что будут ошибки или неверные классификации. Выбираются признаки с минимальным количеством ошибок. Это означает, что именно они наиболее точно классифицируют изображения. Вначале каждому изображению присваивается одинаковый вес. После каждой классификации веса неправильно классифицированных изображений увеличиваются. Затем выполняется тот же процесс. Рассчитываются новые коэффициенты ошибок. Процесс продолжается до тех пор, пока не будет достигнута требуемая точность или частота ошибок, или не будет найдено необходимое количество признаков для удовлетворения заданных параметров обучения.

**2.5. Обучение классификатора.** Для обучения использовались программы `opencv_createsamples.exe` и `opencv_traincascade.exe`, входящие в состав библиотеки OpenCV и не имеющие графического интерфейса.

В качестве обучающих данных выбирались изображения птиц в их естественной среде обитания, а также контрпримеры в той же среде. Примеры изображения этих данных представлены на рисунках 9 и 10.

В процессе обучения использовались 500 изображений с птицами и 500 изображений контрпримеров. Сначала загружались файлы примеров и контрпримеров изображений в формате `jpg` или `png` в программу `opencv_createsamples.exe`, а затем в программу `opencv_traincascade.exe` с необходимыми задаваемыми параметрами (`command line arguments`) [42].

Обучение классификатора проводилось на мобильном процессоре Intel Core i7-4900 и заняло 1 день 15 часов 49 минут 52 секунды. Работа программы завершилась на 13-м уровне вычислений, когда требуемые параметры по вероятности правильного обнаружения (0.999) и ложной тревоги (0.4) были достигнуты. В результате работы этих двух программ сформировался `xml`-файл, который в дальнейшем использовался для обнаружения птиц на изображениях.

**2.6. Применение классификатора.** Для проверки работы обученного классификатора применялась небольшая программа на языке C++ с использованием функции `detectMultiScale`, входящей в состав библиотеки OpenCV, к которой подключался обученный `xml`-файл. На выходе функции `detectMultiScale` получался массив

прямоугольников с обнаруженными объектами, которые в дальнейшем наносились на исходное изображение для наглядного представления работы программы. Результат работы программы приведён на рисунке 12.

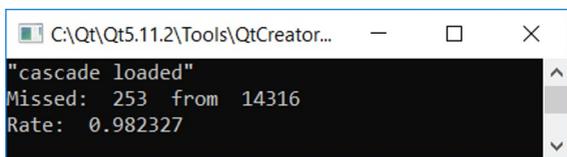


Рис. 12. Результат работы программы

Из рисунка можно наглядно оценить работу классификатора. Видно, что есть пропуски птиц. Это связано с мелким размером этих объектов (большим удалением от видеокамеры), так как обучающий шаблон имеет размер 25 на 25 пикселей, что фактически является минимальным размером окна поиска. Также есть и ложноположительный результат: в левом нижнем углу алгоритм неверно определил листу дерева как птицу.

Время обработки одного кадра размером 1920x1080 пикселей составляет 135 миллисекунд на процессоре Intel Core i7-1165G7. Таким образом получается обеспечить потоковую обработку видеозображений в режиме реального времени с частотой 7.4 кадра в секунду.

Для проверки процентного соотношения пропуска целей на классификатор подавали изображения из обучающей выборки с положительными примерами (рисунке 9) и подсчитывали соотношение отвергнутых изображений к общему их количеству. Результаты представлены на рисунке 13.



```
C:\Qt\Qt5.11.2\Tools\QtCreator...  
"cascade loaded"  
Missed: 253 from 14316  
Rate: 0.982327
```

Рис. 13. Результат подсчёта соотношения неверно отвергнутых изображений птиц к общему количеству изображений

Как видно из результатов подсчёта (рисунок 13), точность обнаружения по обучающей выборке составляет 98.2%. Высокий процент точности обусловлен тем, что объекты на проверочных изображениях сделаны в схожей с обучением среде.

Результаты проверки количества неверно отвергнутых изображений других видов птиц и сделанных в другой среде (примеры изображений на рисунке 14) приведены на рисунке 15.

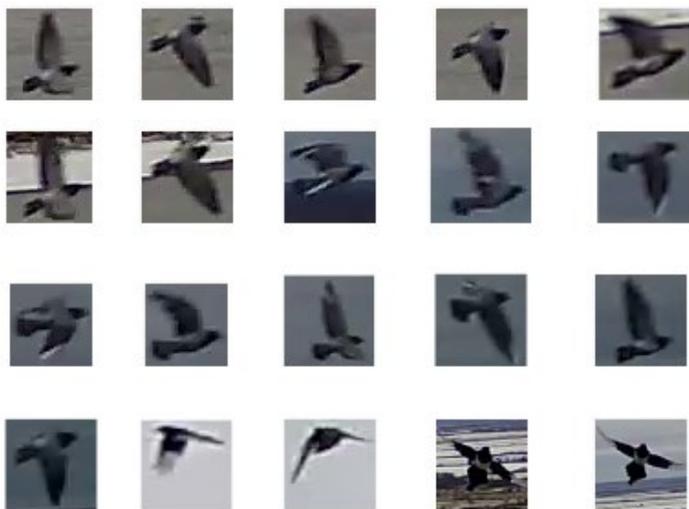
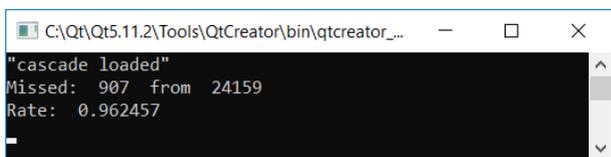


Рис. 14. Изображения птиц для проверки классификатора



```
C:\Qt\Qt5.11.2\Tools\QtCreator\bin\qtcreator...  
"cascade loaded"  
Missed: 907 from 24159  
Rate: 0.962457
```

Рис. 15. Результат подсчёта соотношения неверно отвергнутых изображений птиц к общему количеству изображений

Из результатов подсчета, представленных на рисунках 13 и 15 видно, что при подаче на классификатор изображений, снятых в той же или близкой среде, количество неверно отвергнутых изображений птиц составляет 1.7673%, а при подаче изображений, полученных в другой среде и с другими видами птиц 3.7543%. Во втором случае, для уменьшения пропусков по обнаружению птиц, необходимо повторно дообучить классификатор к новой незнакомой ему среде, а также к новым видам птиц.

**3. Заключение.** Таким образом в данной статье рассмотрен метод построения и применения каскадного классификатора для определения наличия птиц в видеопотоке данных. Данный метод позволяет автоматически распознавать птиц в реальном времени. Его положительной особенностью является низкая требовательность к вычислительной мощности компьютера, что дает возможность его реализации в автономных установках по обнаружению и отпугиванию птиц.

Также следует отметить, что данный метод более стабильно работает в малоизменчивой и предсказуемой окружающей среде. Сильные отличия реальной среды от среды, в которой обучался классификатор могут дать неприемлемый конечный результат. Решить данную проблему можно путём выбора однообразного фона, например неба. Но в таком случае отсекается часть обзриваемого сектора пространства и, в частности, теряется возможность обнаружения птиц, находящихся на земле, в траве, кустах. Эту проблему на качественном уровне могут решить нейросети, но они весьма требовательны к вычислительной мощности в процессе обучения.

Классификатор реализован в биоакустической установке [18, 19], проходящей испытания в аэропорту г. Томска. Все исходные материалы могут быть получены при обращении к авторам.

### Литература

1. Рогачев А.И., Лебедев А.М. Орнитологическое обеспечение безопасности полетов // М.: изд-во «Транспорт». 1984. 126 с.
2. Силаева О.Л., Ильичёв В.Д., Золотарев С.С. Основные направления авиационной орнитологии // Вестник Российского университета дружбы народов. Серия: Экология и безопасность жизнедеятельности. 2010. № 5. С. 10–14.
3. Рыжов С.К. Столкновения с птицами. Актуальные аспекты // Труды общества независимых исследователей авиационных происшествий. Москва. 2013. № 25. С. 175–179.
4. Desoky A.A.S. A review of bird control methods at airports // Global journal of science frontier research (E). 2014. vol. 14(2). pp. 40–50.
5. Кухта А.Е., Большакова Н.П., Мацюра А.В. Концептуальные подходы к орнитологическому обеспечению безопасности полётов воздушных судов //

- Вестник Тувинского государственного университета. Естественные и сельскохозяйственные науки. 2017. № 2. С. 96–105.
6. Официальный сайт Федерального агентства воздушного транспорта, Росавиация. Статистика столкновений с птицами и другими животными. URL: <https://favt.gov.ru/dejatelnost-bezopasnost-poletov-stolknoveniya-ptici-stat/> (дата обращения: 21.05.2022).
  7. Официальный сайт Международной организации гражданской авиации, электронный бюллетень. Анализ столкновений с дикими животными (IBIS) за 2008–2015 годы. URL: [https://www.icao.int/safety/IBIS/2008%20-%202015%20Wildlife%20Strike%20Analyses%20\(IBIS\)%20-%20RU.pdf](https://www.icao.int/safety/IBIS/2008%20-%202015%20Wildlife%20Strike%20Analyses%20(IBIS)%20-%20RU.pdf) (дата обращения: 19.04.2022).
  8. Официальный сайт Министерства транспорта Российской Федерации. Динамика статистических показателей воздушного транспорта Российской Федерации в области столкновений с птицами. URL: <https://favt.gov.ru/dejatelnost-bezopasnost-poletov-stolknoveniya-ptici/> (дата обращения: 20.12.2021).
  9. Рогачев А.И., Ростовский В.А., Шергалин Е.Э. Руководство по орнитологическому обеспечению полётов в гражданской авиации (РООП ГА – 89) // Министерство гражданской авиации СССР. Москва: Воздушный транспорт, 1989. 32 с.
  10. Ильичёв В.Д., Силаева О.Л., Золотарёв С.С., Бирюков В.А., Нечваль Н.А., Якоби В.Э., Титков А.С. Защита самолётов и других объектов от птиц // М.: Товарищество науч. изд. КМК, 2007. 320 с.
  11. Мащора А.В., Яковлев Р.В., Уланов П.Н. Обзор акустических средств для отпугивания птиц // *Acta Biologica Sibirica*. 2016. Т. 2. № 4. С. 141–148. DOI: 10.14258/abs.v2i4.1724.
  12. Отраслевая группа авиационной орнитологии. Отпугивание птиц биоакустическим методом. Проект «Универсал-Акустик». URL: <http://www.otpugivanie.narod.ru/means-control/Universal-Acoustic.html> (дата обращения 21.02.2020).
  13. Биоакустическое оборудование для отпугивания птиц (БАСОП). URL: (<https://aviasvet.ru/push-birds/>) (дата обращения 21.02.2020).
  14. BirdGard. URL: <https://www.birdgard.com/product-page-for-international-customers/> (дата обращения 21.02.2020).
  15. Bird collision avoidance system. URL: <https://www.volacom.com/bird-collision-avoidance-system> (дата обращения 21.02.2020).
  16. Vassilev V.M., Vassileva L.I., Karsch M.A., Petkov K.P., Petkov P.K., Larre J.C. Animal collision avoidance system // Patent US 8,598,998 B2. 2013.
  17. Vasilyev A.F., Neginsky I.V., Protopopov A.G., Yakimets A.L. System of ornithological protection of airfields // Institute of Scientific Communications Conference. Cham: Springer International Publishing. 2020. pp. 307–314.
  18. Власов Е.В., Кузьмин А.А., Раков А.С. Биоакустический комплекс обнаружения и отпугивания птиц в аэропортах // Сборник избранных статей научной сессии ТУСУРа по материалам Международной научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР. 2019. С. 22–24.
  19. Красненко Н.П., Кухта А.Е., Раков А.С. Радиофизические методы в обеспечении орнитологической безопасности объектов и территорий // Шарыгинские чтения. Четвертая международная конференция ведущих научных школ в области радиолокации, радионавигации и радиоэлектронных систем передачи информации. Материалы конференции. Томск: Изд-во Томск. гос. ун-та систем упр. и радиоэлектроники, 2022. С. 5–14.

20. Фисенко В.Т., Фисенко Т.Ю. Компьютерная обработка и распознавание изображений: учебное пособие // СПб: СПбГУ ИТМО. 2008. 192 с.
21. Гонсалес Р., Вудс Р. Цифровая обработка изображений: пер. с англ. // М.: Техносфера. 2002. 1104 с.
22. Berger W. Deep Learning Haar Cascade Explained // <http://www.willberger.org>. 2017. Available at: <http://www.willberger.org/cascade-haar-explained/> (accessed 22.12.2019).
23. Прохоренок Н.А. OpenCV и Java. Обработка изображений и компьютерное зрение // СПб.: БХВ-Петербург. 2018. 320 с.
24. Форсайт Д., Понс Ж. Компьютерное зрение. Современный подход: пер с англ. / под ред. А.В. Назаренко // М.: Вильямс. 2004. 926 с.
25. Шапиро Л., Стокман Д. Компьютерное зрение: учебное пособие для вузов: пер. с англ. // М.: БИНОМ. Лаборатория знаний. 2009. 752 с.
26. Zou Z., Shi Z., Guo Y., Ye J. Object detection in 20 years: A survey // 2019. arXiv:1905.05055v2 [cs.CV].
27. Verstraeten W.W., Vermeulen B., Struckens J., Lhermitte S., Van der Zande D., Van Ranst M., Coppin P. Webcams for bird detection and monitoring: A demonstration study // *Sensors*. 2010. vol. 10. no. 4. pp. 3480–3503. DOI: 10.3390/s100403480.
28. Yoshihashi R., Kawakami R, Iida M., Naemuva T. Bird detection and species classification with time-lapse images around a wind farm: dataset construction and evaluation // *Wind Energy*. 2017. vol. 20. no. 12. pp. 1983–1995. DOI: 10.1002/we.2135.
29. Reyes E. A comparison of image processing techniques for bird detection // A Thesis for degree of master of science in electrical engineering faculty. USA, San Luis Obispo: California Polytechnic State University, 2014. 105 p.
30. Jampens R.T., Hernandez F., Vandecasteele F., Verstockt S. Automatic detection, tracking and counting of birds in marine video content // *Proceedings of sixth International conference on image processing theory, tools and applications (IPTA)*. 2016. pp. 1–6. DOI: 10.1109/IPTA.2016.7821031.
31. Niemi J., Tantt J.T. Deep learning case study for automatic bird identification // *Applied sciences*. 2018. vol. 8(11). no. 2089. DOI: 10.3390/app8112089.
32. Mirudwe A., Nyirenda J., Dufouv E Automating bird detection based on webcam captured images using deep learning // *EPIC Series in Computing. Proceedings of the 43rd conference of the South African institute of computer scientists and information technologists*. 2022. vol. 85. pp. 62–76.
33. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features // *Proceeding of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR*. 2001. vol. 1. 9 p. DOI: 10.1109/CVPR.2001.990517.
34. Hong S.-J., Han Yu., Kim S.-Y., Lee A.-Y., Kim G. Application of Deep-Learning Methods to Bird Detection Using Unmanned Aerial Vehicle Imagery // *Sensors*. 2019. vol. 19(7). no. 1651. DOI: 10.3390/s19071651.
35. Weinstein B.G., Ganner L., Saccomanno V.R., Steinkraus A., Ortega A., Brush K., et. al. A general deep learning model for bird detection in high-resolution airborne imagery // *Ecological Application*. 2022. vol. 32. no. 8.
36. Кузьмин А.А., Власов Е.В., Красненко Н.П. Программа идентификации птиц в видеопотоке и воспроизведения звуковых сигналов (программа для ЭВМ) // Свидетельство о государственной регистрации программ для ЭВМ № 2021612566 от 19.02.2021.
37. Власов Е.В., Красненко Н.П. Программа идентификации птиц в видеопотоке и набора статистики (программа для ЭВМ) // Свидетельство о государственной регистрации программ для ЭВМ № 2022683820 от 08.12.2022.

38. Qt Creator – кроссплатформенная IDE для разработки приложений [Электронный ресурс]. URL: <https://www.qt.io/product/development-tools/> (дата обращения: 01.10.2018).
39. Справочник по OpenCV [Электронный ресурс]. URL: <https://opencv-tutorial.ru> (дата обращения: 01.10.2018).
40. Библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения с открытым кодом [Электронный ресурс]. URL: <https://opencv.org/> (дата обращения: 01.10.2018).
41. Работа каскада Хаара в OpenCV [Электронный ресурс]. URL: <https://habr.com/ru/company/recognitor/blog/228195/> (дата обращения: 01.10.2018).
42. Обучение каскадного классификатора [Электронный ресурс]. URL: [https://docs.opencv.org/3.4/dc/d88/tutorial\\_traincascade.html](https://docs.opencv.org/3.4/dc/d88/tutorial_traincascade.html) (дата обращения: 01.10.2018).

**Власов Евгений Викторович** — инженер I категории, лаборатория акустических исследований, Институт мониторинга климатических и экологических систем Сибирского отделения РАН (ИМКЭС СО РАН). Область научных интересов: обработка информации, математическая статистика, программирование, системы видеонаблюдения. Число научных публикаций — 6. [evvicv1@gmail.com](mailto:evvicv1@gmail.com); проспект Академический, 10/3, 634055, Томск, Россия; р.т.: +7(3822)492-418.

**Красненко Николай Петрович** — д-р физ.-мат. наук, профессор, главный научный сотрудник, заведующий лабораторией, лаборатория акустических исследований, Институт мониторинга климатических и экологических систем Сибирского отделения РАН (ИМКЭС СО РАН); профессор кафедры, кафедра радиотехнических систем радиотехнического факультета, Томский государственный университет систем управления и радиоэлектроники (ТУСУР). Область научных интересов: исследования и разработки в области атмосферной акустики, системы дистанционного зондирования атмосферы. Число научных публикаций — 500. [krasnenko@imces.ru](mailto:krasnenko@imces.ru); проспект Академический, 10/3, 634055, Томск, Россия; р.т.: +7(3822)492-418.

**Поддержка исследований.** Работа выполнена за счет гранта Российского научного фонда № 22-29-00750 «Исследование аспектов орнитологической безопасности аэропортов и разработка научно-методических основ построения оптико-акустического аппаратно-программного комплекса отпугивания птиц», <https://rscf.ru/project/22-29-00750/>, на ИМКЭС СО РАН.

E. VLASOV, N. KRASNENKO  
**CASCADE CLASSIFIER FOR THE DETECTION AND  
IDENTIFICATION OF BIRDS IN A VIDEOSTREAM**

*Vlasov E., Krasnenko N. Cascade Classifier for the Detection and Identification of Birds in a Videostream.*

**Abstract.** A method and a prototype of the program for detecting the presence of birds in the video data flow in real time are presented in the paper. The method is based on the cascade classifier solving the problem of bird detection and identification with the use of a bioacoustic bird scaring system deployed at the Tomsk airport. In our research, the Viola-Jones cascade classifier representing one of the implementations of the Haar cascade algorithm has been used. This algorithm allows objects to be detected in images and videos with high accuracy and rate. In this case, the classifier was leaned on the data set containing images of birds that allowed us to reach high accuracy of bird detection and identification in the videos. The possibilities of the developed classifier are also estimated, and its high productivity is shown. In this study, various methods of machine learning and video data analysis are used to obtain exact and reliable results. As a whole, the present work is an innovative approach to a solution to the urgent problem of airport protection from birds. The application of the developed method has allowed the operating efficiency of the bioacoustic bird scaring system to be increased together with the safety of flights at the Tomsk airport, thereby decreasing the probability of airplane collisions with birds. The novelty of the work consists of the application of the Viola–Jones method for solving the problem of bird detection and identification and estimating its efficiency. Thus, this work is an important contribution to the development of methods for detecting and identifying objects in videos and can also be used in other fields of science and technology in which the automatic detection and classification of objects in the video data flow is required.

**Keywords:** aviation ornithology and safety, video observation, bird detection and identification.

### References

1. Rogatchyov A.I., Lebedev A.M. Ornitologicheskoe obespechenie bezopasnosti poletov [Ornithological Safety of Flights]. Moscow: Publishing House: Transport, 1984. 126 p. (In Russ.).
2. Silaeva O.L., Ilyichev V.D., Zolotarev S.S. [Main direction of avian ornithology]. Vestnik Rossijskogo universiteta druzhby narodov. Serija: Jekologija i bezopasnost' zhiznedejatel'nosti – Bulletin of People' Friendship University of Russia, Series Ecology and Safety. 2010. no. 5. pp. 10–14. (In Russ.).
3. Ryzhov S.K. [Strikes with Birds. Topical Aspects]. Trudy Obshchestva Nezavisimykh Rassledovatelei Aviatsionnykh Proisshestvii – Proceedings of the Society of Independent Investigators of Aviation Accidents. Moscow. 2013. no. 25. pp. 175–179. (In Russ.).
4. Desoky A.A.S. A review of bird control methods at airports. Global journal of science frontier research (E). 2014. vol. 14(2). pp. 40–50.
5. Kukhta A.E., Bolshakova N.P., Matsyura A.V. [Conceptual approaches to ornithological safety of flights of aircraft]. Vestnik Tuvinskogo gosudarstvennogo universiteta. Estestvennye i sel'skohozjajstvennye nauki – Vestnik of Tuvan State University. Natural and Agricultural sciences. 2017. no. 2. pp. 96–105. (In Russ.).

6. Oficial'nyj sajt Federal'nogo agentstva vozdušnogo transporta, Rosaviacija. Statistika stolknovenij s pticami i drugimi zhivotnymi [Official web site of Federal Agency of Air Transport, Rosaviation, Statistics of collisions with birds and other animals]. Available at: <https://favt.gov.ru/dejatelnost-bezopasnost-poletov-stolknoveniya-pticistat/> (accessed 21.05.2022). (In Russ.).
7. Oficial'nyj sajt Mezhduнародной organizacii grazhdanskoj aviacii, jelektronnyj bjulleten'. Analiz stolknovenij s dikimi zhivotnymi (IBIS) za 2008–2015 gody [Official web site of International Civil Aviation Organization, Analysis of collisions with wild animals (IBIS) for 2008–2015, Electronic Bulletin]. Available at: [https://www.icao.int/safety/IBIS/2008%20-%202015%20Wildlife%20Strike%20Analyses%20 \(IBIS\) %20-%20RU.pdf](https://www.icao.int/safety/IBIS/2008%20-%202015%20Wildlife%20Strike%20Analyses%20(IBIS)%20-%20RU.pdf) (accessed 19.04.2022). (In Russ.).
8. Oficial'nyj sajt Ministerstva transporta Rossijskoj Federacii. Dinamika statističeskikh pokazatelej vozdušnogo transporta Rossijskoj Federacii v oblasti stolknovenij s pticami [Official web site of Ministry of Transport of the Russian Federation, Dynamics of statistics of air transport of the Russian Federation in the field of collisions with birds]. Available at: <https://favt.gov.ru/dejatelnost-bezopasnost-poletov-stolknoveniya-ptici/> (accessed 20.12.2021). (In Russ.).
9. Rogachev A.I., Rostovskii V.A., Shergalin E.E. Rukovodstvo po ornitologičeskomu obespečeniju poljotov v grazhdanskoj aviacii (ROOP GA – 89) [Ornithological Flight Safety Guide for Civil Aviation (OFSGCA– 89)]. Ministry of Civil Aviation of the USSR. Moscow: Air Transport, 1989. 32 p. (In Russ.).
10. Il'ichev V.D., Silaeva O.L., Zolotarev S.S., Biryukov V.A., Nechval' N.A., Yakobi V.E., Titkov A.S. Zashhita samoljotov i drugih ob'ektov ot ptic [Protection of Aircrafts and Other Objects from Birds]. Moscow: KMK Sci. Publ. Assoc., 2007. 320 p. (In Russ.).
11. Matsyura A.V., Yakovlev R.V., Ulanov P.N. Acoustic bird repellents – a brief review. *Acta Biologica Sibirica*. 2016. vol. 2. no. 4. pp. 141–148. DOI: 10.14258/abs.v2i4.1724. (In Russ.).
12. Oficial'nyj sajt otraslevoj gruppy aviacionnoj ornitologii. Otpugivanie ptic bioakustičeskimi metodami. Proekt: Universal-Akustik [Official web site of Sectoral Aviation Ornithology Group. Bioacoustic Method for Scaring of Birds. Project: Universal Acoustic]. Available at: <http://www.otpugivanie.narod.ru/means-control/Universal-Acoustic.html>. (accessed 21.02.2020). (In Russ.).
13. Oficial'nyj sajt kompanii Aviasvetotekhnika. Bioakustičeskoe oborudovanie dlja otpugivanija ptic (BASOP) [Official web site of the company Airlight Engineering. Bioacoustic Equipment for Scaring Birds (BASOP)]. Available at: <https://aviasvet.ru/push-birds/>. (accessed 21.02.2020). (In Russ.).
14. BirdGard. Available at: <https://www.birdgard.com/product-page-for-international-customers/> (accessed 21.02.2020).
15. Bird collision avoidance system. Available at: <https://www.volacom.com/bird-collision-avoidance-system> (accessed 21.02.2020).
16. Vassilev V.M., Vassileva L.I., Karsch M.A., Petkov K.P., Petkov P.K., Larre J.C. Animal collision avoidance system. Patent US 8,598,998 B2. 2013.
17. Vasilyev A.F., Neginsky I.V., Protopopov A.G., Yakimets A.L. System of ornithological protection of airfields. Institute of Scientific Communications Conference. Cham: Springer International Publishing, 2020. pp. 307–314.
18. Vlasov E.V., Kuz'min A.A., Rakov A.S. Bioacoustic system for detecting and scaring bird at airports. Materials of the International Scientific-Technical Conference for Students, Post Graduates, and Young Scientists, Scientific Session «TUSUR-2019». 2019. pp. 22–24. (In Russ.).

19. Krasnenko N.P., Kukhta A.E., Pakov A.S. Radio physical methods for ensuring ornithological safety of objects and territories. Sharygin readings. The Fourth International Conference of Leading Schools of Thought in the Field of Radar Detection and Ranging and Navigation and Radio-Electronic Systems of Information Transfer. Conference Materials. Tomsk: Publishing House of Tomsk State University of Control Systems and Radioelectronics. 2022. pp. 5–14. (In Russ.).
20. Fisenko V.T., Fisenko T.Yu. Computer processing and recognition of images: A Textbook. Saint Petersburg: SPbSU ITMO. 2008. 192 p.
21. Gonzalez R.C., Woods R.E. Digital imageprocessing (3rd Ed.). Pearson Education, Inc. 2008.
22. Berger W. Deep Learning Haar Cascade Explained. <http://www.willberger.org>. 2017. Available at: <http://www.willberger.org/cascade-haar-explained/> (accessed 22.12.2019).
23. Prokhorenok N.A. OpenCV and Java. Images Processing and Computer Vision. Saint Petersburg: BHV-Petersburg. 2018. 320 p. (In Russ.).
24. Forsyth D.A., Ponce J. Computer Vision. A Modern Approach. (2nd Ed.). Pearson. 2012. 828 p.
25. Shapiro L., Stockman G. Computer vision. Pearson Education, Inc. 2001. 752 p.
26. Zou Z., Shi Z., Guo Y., Ye J. Object detection in 20 years: A survey. 2019. arXiv:1905.05055v2 [cs.CV].
27. Verstraeten W.W., Vermeulen B., Struckens J., Lhermitte S., Van der Zande D., Van Ranst M., Coppin P. Webcams for bird detection and monitoring: A demonstration study. *Sensors*. 2010. vol. 10. no. 4. pp. 3480–3503. DOI: 10.3390/s100403480. DOI: 10.3390/s100403480.
28. Yoshihashi R., Kawakami R, Iida M., Naemuva T. Bird detection and species classification with time-lapse images around a wind farm: dataset construction and evaluation. *Wind Energy*. 2017. vol. 20. no. 12. pp. 1983–1995. DOI: 10.1002/we.2135.
29. Reyes E. A comparison of image processing techniques for bird detection. A Thesis for degree of master of science in electrical engineering faculty. USA, San Luis Obispo: California Polytechnic State University, 2014. 105 p.
30. Jampens R.T., Hernandez F., Vandecasteele F., Verstockt S. Automatic detection, tracking and counting of birds in marine video content. Proceedings of sixth International conference on image processing theory, tools and applications (IPTA). 2016. pp. 1–6. DOI: 10.1109/IPTA.2016.7821031.
31. Niemi J., Tantt J.T. Deep learning case study for automatic bird identification. *Applied sciences*. 2018. vol. 8(11). no. 2089. DOI: 10.3390/app8112089.
32. Mirudwe A., Nyirenda J., Dufouvg E Automating bird detection based on webcam captured images using deep learning. EPIC Series in Computing. Proceedings of the 43rd conference of the South African institute of computer scientists and information technologists. 2022. vol. 85. pp. 62–76.
33. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features. Proceeding of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR. 2001. vol. 1. 9 p. DOI: 10.1109/CVPR.2001.990517.
34. Hong S.-J., Han Yu., Kim S.-Y., Lee A.-Y., Kim G. Application of Deep-Learning Methods to Bird Detection Using Unmanned Aerial Vehicle Imagery. *Sensors*. 2019. vol. 19(7). no. 1651. DOI: 10.3390/s19071651.
35. Weinstein B.G., Ganner L., Saccomanno V.R., Steinkraus A., Ortega A., Brush K., et. al. A general deep learning model for bird detection in high-resolution airborne imagery. *Ecological Application*. 2022. vol. 32. no. 8.

36. Kuz'min A.A., Vlasov E.V., Krasnenko N.P. Program for identification of birds in a videostream and reproduction of sound signals, Certificate of State Registration of Computer Program No. 2021612566. 2021. (In Russ.).
37. Vlasov E.V., Krasnenko N.P. Program for identification of birds in a videostream and a set of statistics. Certificate of State Registration of Computer Program No. 2022683820. 2022 (In Russ.).
38. Qt Creator – cross-platform IDE for development of application. Available at: <https://www.qt.io/product/development-tools/> (accessed 01.10.2018).
39. Directory on OpenCV. Available at: [https://opencv-tutorial.readthedocs.io/\\_/downloads/en/latest/pdf/](https://opencv-tutorial.readthedocs.io/_/downloads/en/latest/pdf/) (accessed 01.06.2020).
40. OpenCV – Open Source Computer Vision Library. Available at: <https://opencv.org/> (accessed 01.10.2018).
41. Work Haar cascade in OpenCV. Available at: <https://habr.com/ru/company/recognitor/blog/228195/> (accessed 01.10.2018). (In Russ.).
42. Cascade Classifier Training. Available at: [https://docs.opencv.org/3.4/dc/d88/tutorial\\_traincascade.html](https://docs.opencv.org/3.4/dc/d88/tutorial_traincascade.html) (accessed 01.10.2018).

**Vlasov Evgeny** — Engineer, Laboratory of acoustic researches, Institute of Monitoring of Climatic and Ecological Systems SB RAS. Research interests: processing of the information, mathematical statistics, programming, systems of video observation. The number of publications — 6. [evvicvl@gmail.com](mailto:evvicvl@gmail.com); 10/3, Akademicheskii Ave., 634055, Tomsk, Russia; office phone: +7(3822)492-418.

**Krasnenko Nikolay** — Ph.D., Dr.Sci., Professor, Chief researcher, head of the laboratory, Laboratory of acoustic researches, Institute of Monitoring of Climatic and Ecological Systems SB RAS; Professor of the department, Department of radio engineering systems, Tomsk State University of Control Systems and Radioelectronics. Research interests: atmospheric acoustics, systems of remote sounding of the atmosphere. The number of publications — 500. [krasnenko@imces.ru](mailto:krasnenko@imces.ru); 10/3, Akademicheskii Ave., 634055, Tomsk, Russia; office phone: +7(3822)492-418.

**Acknowledgements.** This research is supported by RSCF (grant 22-29-00750).

А.Л. ВЕРХОТУРОВ, А.С. СТЕПАНОВ, Л.В. ИЛЛАРИОНОВА  
**ИСПОЛЬЗОВАНИЕ РАДИОЛОКАЦИОННЫХ ДАННЫХ ДЛЯ  
МОНИТОРИНГА СОСТОЯНИЯ ПОСЕВОВ  
СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР НА ЮГЕ ДАЛЬНЕГО  
ВОСТОКА РОССИИ**

*Верхотуров А.Л., Степанов А.С., Илларионова Л.В.* **Использование радиолокационных данных для мониторинга состояния посевов сельскохозяйственных культур на юге Дальнего Востока России.**

**Аннотация.** Использование радиолокационных спутниковых данных в мониторинге сельскохозяйственных культур является перспективным дополнением методов и технологий, базирующихся на анализе мультиспектральных изображений. К основным достоинствам радиолокационных вегетационных индексов относится их чувствительность к поляриметрическим свойствам принимаемого сигнала, а также независимость от облачности. Это особенно важно для территории юга российского Дальнего Востока, муссонный климат которого обеспечивает влажную и облачную погоду в период набора сельскохозяйственными культурами максимальной биомассы. Для оценки возможностей радиолокационных спутниковых данных на примере пахотных земель Хабаровской края и Амурской области были проанализированы 64 снимка космического аппарата Sentinel-1 за период наблюдений с мая по октябрь 2021 года. Для каждого снимка были рассчитаны значения индексов DpRVI, RVI, VH/VV и построены временные ряды для всего периода наблюдений по отдельным полям (всего 342 поля). По мультиспектральным снимкам Sentinel-2 с использованием маски облачности были построены временные ряды NDVI. Были рассчитаны характеристики экстремумов временных рядов для разных типов пахотных земель: сои, овса, и залежи. Показано, что для каждой сельхозкультуры кривые сезонного хода DpRVI, RVI, VH/VV имели характерный вид. Установлено, что индекс DpRVI продемонстрировал наиболее высокую устойчивость – коэффициенты вариации сезонного хода DpRVI были существенно ниже показателей для RVI и VH/VV. Также было выявлено, что сходство между сезонным ходом индексов сохранялось для удаленных друг от друга регионов – Хабаровского края и Амурской области. Были рассчитаны основные характеристики сезонного хода временных рядов радиолокационных индексов в сравнении с NDVI – величина максимума, дата наступления максимума и вариабельность этих показателей. Установлено, во-первых, что значения этих показателей в разных регионах схожи между собой; во-вторых, вариабельность максимума и дня наступления максимума для DpRVI ниже, чем для RVI и VH/VV; в-третьих, вариабельность максимума и дня наступления максимума для DpRVI сопоставима с NDVI. Таким образом, можно сделать вывод о том, что временные ряды радиолокационных индексов DpRVI, RVI, VH/VV для основных типов сельскохозяйственных земель Дальнего Востока имеют отличительные особенности и могут быть использованы в задачах классификации, моделирования урожайности и контроля севооборота.

**Ключевые слова:** данные дистанционного зондирования Земли, Дальний Восток, мониторинг сельскохозяйственных земель, радиолокационные вегетационные индексы, вариабельность, пахотные земли, сельскохозяйственная культура, временной ряд.

**1. Введение.** На сегодняшний день уровень развития спутниковых систем дистанционного зондирования Земли (ДЗЗ)

позволяет проводить эффективный мониторинг и анализ состояния сельскохозяйственных (с/х) культур, как на региональном уровне, так и на уровне всей страны. С увеличением пространственного разрешения и частоты съемки, использование накопленных многолетних архивов спутниковых данных в оптическом диапазоне длин волн и методы их обработки для такого рода задач получили широкое распространение в научном сообществе [1]. Оценка состояния возделываемых культур имеет важное значение: для понимания тенденций развития с/х производства, для обеспечения продовольственной безопасности, а также для борьбы с краткосрочными и долгосрочными угрозами. В связи с этим, остро стоит вопрос разработки специализированных информационных систем и сервисов, которые могут решать как комплексные, так и отдельные научные и научно-технические задачи в этом направлении. Такая работа активно ведется во всем мире. В качестве примеров можно привести систему Министерства сельского хозяйства и продовольствия Канады [2], которая в течение вегетационного сезона выпускает аналитические отчеты и материалы по инвентаризации урожая, основанные на анализе спутниковых данных отдельных территорий и провинций страны. В России можно отметить информационный сервис ВЕГА-PRO [3 – 5], разработанный учеными Института космических исследований РАН. Возможности этой системы позволяют проводить комплексный анализ состояния растительного покрова, лесов и посевов с/х культур, с использованием временных рядов различных вегетационных индексов.

Необходимо отметить, что наряду с данными ДЗЗ оптического диапазона в настоящее время наблюдается возрастающий интерес к развитию методов обработки с радара с синтезированной апертурой (РСА). Электромагнитное излучение микроволнового диапазона частот почти не поглощается и не рассеивается облаками, что позволяет получать изображения земной поверхности и расположенных на ней объектов независимо от времени суток и метеорологических условий. К тому же чувствительность радиоволн к диэлектрическим и геометрическим свойствам радиолокационных целей, сделало данные РСА полезными для мониторинга урожая с/х культур, особенно для регионов, которые ограничены возможностью использования данных оптических спутниковых изображений. С учетом вышесказанного данные РСА могут внести существенный вклад в развитие систем и технологий идентификации и мониторинга состояния посевов.

Известно, что диэлектрические и геометрические свойства растительности с/х угодий влияют на отражательную способность радиолокационного сигнала, который характеризуется коэффициентом обратного рассеяния или удельной эффективной площадью рассеяния (УЭПР). Как показано в [6] по мере набора биомассы пшеницы и подсолнечника, изменялись значения УЭПР. В работе [7] авторы исследовали чувствительность УЭПР двух поляризаций VH/VV (VH – вертикальная поляризация на передачу, горизонтальная на прием; VV – вертикальная поляризация на передачу и на прием) к динамике роста культур. Было отмечено, что отношение VH/VV лучше коррелировало с ростом биомассы, в отличие от отдельных коэффициентов VV и VH. В [8, 9] указывается, что VH/VV применимо и к оценке фенологического состояния культур, характеристике растительности и их классификации. Кроме того, это позволяет разделять культуры по признакам на кукурузу, сою и подсолнух на поздних стадиях их фазы развития.

Особое внимание также стоит обратить на работы, которые используют поляриметрические характеристики для мониторинга и классификации культур. В [9], авторы использовали энтропию Шеннона и общую мощность отраженного радиолокационного сигнала, принятого антенной радиолокатором в двух поляризаций для идентификации фенологических стадий роста пшеницы и канолы. В исследовании [10] сообщается, что при анализе временных рядов по данным РСА изображений таких характеристик как энтропия Шеннона, мера доминирующего механизма рассеяния и анизотропия, были получены более высокие оценки точности классификации культур. В [11] исследуются чувствительность тех же характеристик, но к фенологическим стадиям их роста. В то же самое время можно выделить работу [12], в которой одновременно рассматриваются показатели УЭПР и характеристики поляриметрического разложения радарного сигнала для анализа роста с/х культур.

Стоит отдельно упомянуть о появлении концепции вегетационных индексов по данным РСА изображений. В 2009 г. на основе радиолокационной съемки в четырехполяризационном режиме был разработан индекс RVI (Radar Vegetation Index) [13]. В работах [14 – 16] исследователи указывают на возможность использования RVI для двух поляризаций. Другая исследовательская группа разработала алгоритмы для вычисления индексов Generalized Radar Vegetation Index (GRVI) [17] и Compact-Pol Radar Vegetation Index (CpRVI) [18]. Вместе с тем они вычисляются на основе полно-

поляриметрических PCA изображений, таких спутников, например, как Radarsat-2 и Alos-2.

Общедоступность данных миссии Sentinel-1A/B, параметры её систематической съемки одного и того же участка местности (6 или 12 суток), пространственное разрешение 14м x 14м – побудило исследователей к разработке алгоритмов вычисления индексов с двойной поляризацией. Поэтому, значительный интерес представляют собой результаты следующих работ: Polarimetric Radar Vegetation Index [19] (PRVI), Dual Polarization SAR Vegetation Index (DPSVI) [20] и Dual polarimetric radar vegetation index [21] (DpRVI). В отличие от индексов VH/VV и RVI (с двойной поляризацией), в расчете DpRVI используются данные более низкого уровня обработки, Single Look Complex (SLC). В комплексном формате PCA изображений содержится информация о рассеянии радиолокационного сигнала, которая характеризуется степенью поляризации и мерой доминирующего механизма рассеяния. В свою очередь благодаря этим показателям, как отмечено в [21], индекс DpRVI становится более чувствительным к росту с/х культур и используется как относительно простой и физически интерпретируемый дескриптор растительности.

В последнее время все более актуальными становятся исследования, направленные на использование радиолокационных индексов, как известных ранее, так и вновь предложенных авторами. Такие работы прежде всего обусловлены развитием существующих современных радиолокационных спутниковых группировок (SAOCOM, TerraSAR-X/TanDEM-X, COSMO-SkyMed, Radarsat-2, Alos-2, Кондор-ФКА1 и другие), а также перспективой вывода на орбиту новых российских и зарубежных космических аппаратов с PCA (Кондор-ФКА2, NISAR, Sentinel-1C и другие). Так, например, исследователями был установлен достаточно высокий потенциал использования данных PCA Alos-2 в сравнении с Sentinel-1 на распаханых и покрытых растительностью с/х полях. Авторами была проведена оценка показателей поверхностной влажности почв [22]. Результаты показали, что обратное рассеяние от поверхности почвы, покрытой и не покрытой растительностью, в соответствии с моделью водяного облака WCM (Water Cloud Model, [23]) по данным Alos-2 в L-диапазоне были получены несколько лучшие результаты, чем для Sentinel-1 в C-диапазоне длин волн. Кроме того, вода является уникальным объектом при зондировании её микроволновыми частотами, на которых работают PCA и скаттерометры. Это единственная естественная среда с высокой

диэлектрической проницаемостью, поэтому увеличение доли жидкой воды, содержащейся в почве, снеге и растительности, увеличивает диэлектрические свойства этих сред, тем самым значительно изменяя их характеристики рассеяния и поглощения. УЭПР, измеряемая РСА системами, зависит от диэлектрических свойств поверхностного слоя почвы, шероховатости поверхности и растительности. Поэтому, данные РСА представляют собой особый интерес в исследованиях с/х угодий.

Таким образом, наряду с развитием спутниковых космических аппаратов РСА, недостаточная изученность особенностей сезонного хода радиолокационных вегетационных индексов, в том числе сравнительной устойчивости различных индексов для разных культур в пределах одного или соседних регионов является актуальной проблемой. Например, сравнительная оценка применимости DpRVI и RVI [21] строилась на анализе значений индексов для 4 снимков в период с 13 июня по 31 июля, что оставляло открытым вопрос о поведении временных рядов DpRVI на других этапах вегетационного периода. Также ранее не проводилось сравнение вариабельности временных рядов DpRVI в сравнении с индексом NDVI при оценке одних и тех же с/х полей и предполагалось, что радиолокационные индексы в сельском хозяйстве могут использоваться в основном как вспомогательные. В целом, работы по системному внедрению этих индексов в задачах сельского хозяйства будут иметь не только прикладное, но и фундаментальное значение, особенно для российской науки, в связи отсутствием комплексных исследований в отечественной практике о применении радиолокационных спутниковых данных для мониторинга с/х культур.

Поиск новых, ранее не исследованных, но устойчивых индексов, имеющих наименьшую пространственную вариабельность в течение всего вегетационного сезона, при этом сравнимую с NDVI – важен, в связи с развитием точного и цифрового земледелия. Сравнение пространственной вариабельности сезонного хода разных радарных индексов для отдельных регионов Дальнего Востока еще не проводилось и является новым результатом данного научного направления. Широко используемые индексы в оптическом диапазоне не всегда удовлетворяют потребностям исследователей. В дальнейшем использование устойчивого индекса может помочь в решении разных задач: уточнения контуров полей, контроль севооборота, картографирования, поиск неоднородностей на полях и других. Полученные результаты могут внести весомый вклад

в запланированные исследования временных рядов данных РСА в рамках международного эксперимента JECAM SAR [24].

Исходя из этого, необходимо провести анализ и оценку радарных вегетационных индексов, сравнить их устойчивость во временной серии РСА изображений, установить характерный вид хода роста каждой культуры, исследовать поведение культур для разных регионов. Исследованию этих вопросов посвящена представленная работа, основанная на использовании данных ДЗЗ микроволнового диапазона для анализа усредненных значений радарных вегетационных индексов на с/х полях.

В статье представлены результаты анализа и сравнительной характеристики временных рядов радиолокационных вегетационных индексов по данным съемки РСА. В качестве входных данных использовались РСА изображения спутников Sentinel-1A/B в период с апреля по октябрь 2021 года. Всего было отобрано и оконтурено 342 сельскохозяйственных поля (соя, овес и залежь) в векторном формате на территории Хабаровского района и Амурской области. Для каждого поля с культурой построены временные ряды средних значений вегетационных радиолокационных индексов (RVI, VH/VV и DpRVI). Проведена оценка вариабельности сезонного хода значений этих индексов и характеристик экстремумов в сравнении с индексом NDVI в двух регионах на территории юга Дальнего Востока России.

**2. Материалы и методы.** Территория исследования – Хабаровский край и Амурская область, входящие в состав Дальневосточного федерального округа. Были рассмотрены с/х поля в Хабаровском крае, расположенные в границах Хабаровского муниципального района, на правом берегу реки Амур и в бассейне реки Усури. Этот район характеризуется умеренным муссонным климатом с холодной снежной зимой (средняя температура в январе варьируется от  $-20^{\circ}\text{C}$ ) и тёплым летом, влажным во второй половине (средняя температура в июле превышает  $21^{\circ}\text{C}$ ). Среднегодовое распределение количества осадков находится на уровне 600–700 мм в год. Ближе к северо-западу региона он переходит в резко континентальный климат Восточной Сибири. Благоприятные климатические условия с дерново-подзолистыми почвами позволяют выращивать зерновые и зернобобовые культуры (рисунок 1).

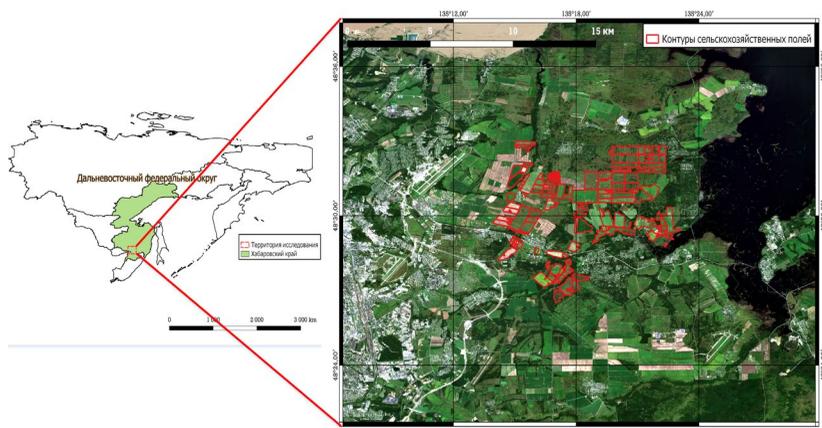


Рис. 1. Территория исследования. Хабаровский край, Хабаровский муниципальный район

Семь районов Амурской области (рисунок 2) (Белогорский, Бурейский, Ивановский, Константиновский, Михайловский, Октябрьский и Тамбовский) расположены на территории Зейско-Бурейнской равнины. Около 75% этой территории являются сельскохозяйственными угодьями. Общая площадь пахотных земель здесь составляет 1,32 млн га. Климат Амурской области в отличие от Хабаровского края более континентальный, так как область отгорожена от океана горами. Характеризуется жарким летом (средняя температура в июле превышает  $21^{\circ}\text{C}$ ), стабильно морозной зимой (средняя температура в январе варьируется от  $-25^{\circ}\text{C}$ ), с количеством осадков 850 мм в год. Причем наибольшее количество осадков выпадает в горной части на востоке области. Сравнительно высокие температуры и продолжительность вегетационного периода позволяют достигать высокие показатели урожая зерновых, технических и других культур.

Для расчета радиолокационных вегетационных индексов использовались спутниковые данные из распределенного архива ASF DAAC (Alaska Satellite Facility Distributed Active Archive Center). Были выбраны четыре сцены Sentinel-1 уровня обработки SLC, полученных с трех подспутниковых трека (таблица 1).

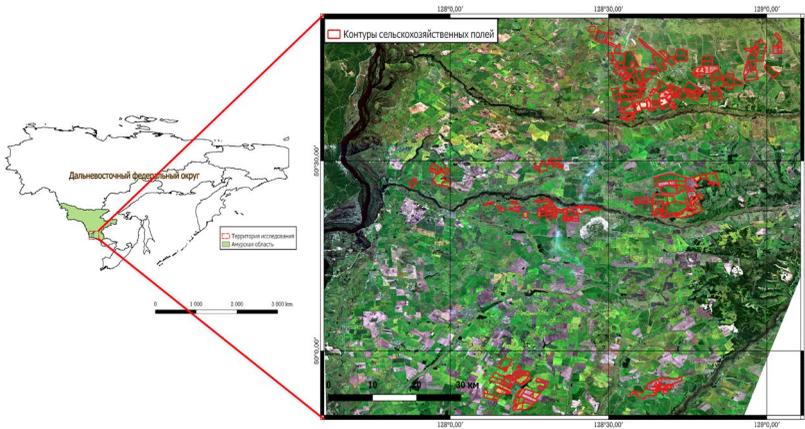


Рис. 2. Территория исследования. Амурская область

С помощью модуля Graph Processing Tool (GPF) программного обеспечения SNAP v.9.0 (<http://step.esa.int/main/>, по состоянию на 30 марта 2024 г.) был выполнен весь процесс обработки 16-ти PCA изображений на каждую сцену за 2021 год. В качестве выходной информации выступал сформированный временной ряд в растровом формате (\*.tif) радиолокационных индексов для каждого поля с культурой. Методика получения таких рядов опирается на преобразование комплексных величин, полученных одновременно в нескольких поляризационных каналах, методом поляриметрической декомпозиции, в которых содержится вся информация о поляриметрических свойствах рассеяния зондируемой поверхности. Данная технология более подробно описана в работах [25, 26]. Основные формулы для расчета индексов VH/VV, RVI и DpRVI:

$$\sigma_{VHV} = \frac{\sigma_{VH}^0}{\sigma_{VV}^0}, \quad (1)$$

$$RVI = \frac{4 \cdot \sigma_{VH}^0}{\sigma_{VV}^0 + \sigma_{VH}^0}, \quad (2)$$

$$DpRVI = 1 - m \cdot \beta = 1 - \sqrt{\frac{4 \cdot |C_2|}{(Tr(C_2))^2}} \cdot \frac{\lambda_1}{\lambda_1 + \lambda_2}, \quad (3)$$

$$C_2 = U_2 \cdot \Sigma \cdot U_2^{-1}, \text{ где } \Sigma = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix},$$

где:

$\sigma_{VH}^0$  – УЭПР (дБ) для VH поляризации;

$\sigma_{VV}^0$  – УЭПР (дБ) для VV поляризации;

$m$  – степень поляризации ( $0 \leq m \leq 1$ ), которая определяется как отношение (средней) интенсивности поляризованной части волны к (средней) общей интенсивности волны;

$\beta$  – мера доминирующего механизма рассеяния, которая определяется из разложения матрицы  $S_2$  на два неотрицательных собственных значения ( $\lambda_1 \geq \lambda_2 \geq 0$ );

$Tr$  – сумма диагональных элементов матрицы;

$| \cdot |$  – определитель матрицы.

Таблица 1. Параметры радиолокационной съемки исследуемой территории

Район исследования	Период наблюдения	Номер подспутникового трека	Пространственное разрешение	Углы съемки	Поляризация
Хабаровский край	С мая по октябрь 2021	90	14м x 14м	38.3°–43.1°	VV, VH
Амурская область	С мая по октябрь 2021	105, 134	14м x 14м	30.22°–32.47°	VV, VH

Для исследуемого района были рассмотрены всего 102 мультиспектральных снимка с разрешением 10м, полученных со спутников Sentinel-2A/B в период с 1 мая по 31 октября 2021 года. В дальнейшем формировались временные ряды NDVI, в которых значения показателей для каждого дня определялись по формулам:

$$NDVI = \frac{NIR - RED}{NIR + RED}, \quad (4)$$

где  $NIR$  – физический коэффициент отражения в ближней инфракрасной области спектра;

$RED$  – физический коэффициент отражения в красной области спектра.

Для оценки пространственной устойчивости радарных индексов для разных с/х полей использовался коэффициент вариации. Оценка степени сходства временных рядов базировалась на применении меры близости (евклидова расстояния). В качестве входной информации выступали данные временной серии индексов VH/VV, RVI, DpRVI и NDVI. Формулы для расчетов оцениваемых параметров следующие:

$$VAR_{VI_i} = \frac{\sigma(VI_i)}{\overline{VI}_i} \cdot 100\%, \quad (5)$$

$$d = \sqrt{\sum_{i=1}^n (\overline{VI}_i^{Khab} - \overline{VI}_i^{Amu})^2}, \quad (6)$$

$$i \in (1..n),$$

где:

$VAR$  – коэффициент вариации;

$VI$  – значение вегетационного индекса (один из VH/VV, RVI, DpRVI или NDVI);

$\sigma(VI_i)$  – стандартное отклонение величины  $VI$  на определенную дату для полей одной с/х культуры;

$\overline{VI}_i$  – среднее значение вегетационного индекса на определенную дату для полей одной с/х культуры;

$n$  – количество спутниковых снимков;

$d$  – евклидово расстояние;

$\overline{VI}_i^{Khab}$  – среднее значение вегетационного индекса на определенную дату для всех полей с одной культурой в Хабаровском крае;

$\overline{VI}_i^{Amu}$  – среднее значение вегетационного индекса на определенную дату для всех полей с одной культурой в Амурской области.

Были рассчитаны средние значения максимума ( $\overline{VI_{max}}$ ) и дня наступления максимума ( $\overline{DOY_{max}}$ ), а также  $\Delta\overline{VI_{max}}$  и  $\Delta\overline{DOY_{max}}$

$$\Delta\overline{VI_{max}} = t(p, f) \cdot \sigma(VI_{max}), \quad (7)$$

$$\Delta\overline{DOY_{max}} = t(p, f) \cdot \sigma(DOY_{max}), \quad (8)$$

где:

$\sigma(VI_{max})$  – стандартное отклонение величины  $VI_{max}$  по всем полям для одной культуры;

$\sigma(DOY_{max})$  – стандартное отклонение величины  $DOY_{max}$  по всем полям для одной культуры;

$t(p, f)$  – значения t-критерия,  $p=005$ .

По данным спутникового мониторинга для каждого из изучаемых отдельных полей Амурской области и Хабаровского края были построены средние временные ряды  $VI$  в период с 1 мая 2021

(121 календарный день) по 31 октября 2021 (304 календарный день) года. На рисунке 3 представлена временная шкала с датами, соответствующими мультиспектральным и радарным изображениям на примере Хабаровского края. Как видно, для этой области исследования было рассмотрено 47 снимков спутника Sentinel-2 (из них 24 были не пригодны из-за высокой облачности) и 16 радиолокационных изображений, полученных со спутника Sentinel-1. Для Амурской области, соответственно, было получено 55 снимков Sentinel-2 и 48 снимков Sentinel-1.

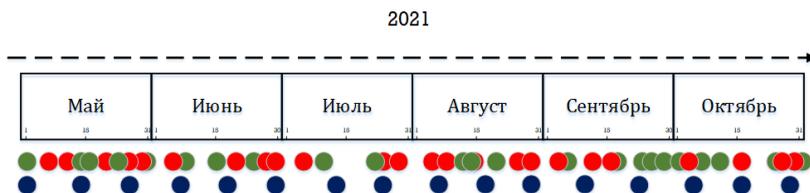


Рис. 3. Спутниковые данные Sentinel-1и Sentinel-2 (Хабаровский край).  
 Зеленым цветом – безоблачные снимки Sentinel-2 (менее 20% облачности),  
 красным цветом – снимки Sentinel-2 с высокой облачностью,  
 синим цветом – радиолокационные снимки Sentinel-1

Для расчета средних значений на дату для отдельных полей предварительно оценивалась однородность выборки. Несмотря на то, что  $DpRVI$ ,  $RVI$ ,  $VH/VV$  практически не зависят от атмосферных явлений, на некоторых участках полей могут наблюдаться аномальные значения этого показателя. Для индекса  $NDVI$  неоднородность выборки напрямую могла быть связана с облачностью. Чтобы свести к минимуму влияние таких шумов, была проведена фильтрация с использованием доверительного интервала: для каждого поля и для каждой даты были рассчитаны среднее значение радиолокационных  $VI$  и  $NDVI$  и стандартное отклонение ( $\sigma$ ). Границы доверительного интервала рассчитывали по следующей формуле

$$VI_d = \bar{VI} \pm t(P, f) * \sigma, \quad (9)$$

где:

$t(P, f)$  – t-критерий Стьюдента, ( $P=0,95$ );

$f$  – число степеней свободы;

$\sigma$  – стандартное отклонение величины  $VI$ .

Значения радиолокационных VI и NDVI для пикселей отдельных полей, не вошедших в доверительный интервал, не участвовали в расчете средних значений по отдельным полям.

По данным Росстата, основной сельскохозяйственной культурой южной части Дальнего Востока, на долю которой приходится более 70% общей площади пахотных земель регионе, являлась соя. На втором месте в последние годы в большинстве муниципальных образований Хабаровского края и Амурской области находился овес. При этом овес часто использовался в многолетнем севообороте с соей, что обуславливает важность сочетанного исследования сезонного хода индексов вегетации этих двух культур. И, наконец, Дальневосточный федеральный округ в 2021-2023 гг. занимал лидирующие места среди федеральных округов РФ по доли залежи: свыше 9% от площади сельскохозяйственных угодий. В настоящей работе авторами были рассмотрены три класса: соя и овес, а также залежные земли. Общая площадь исследуемых полей (342 поля) составила более 13000 га. Площадь рассмотренных отдельных полей в Хабаровском крае составила 3389 га (128 полей), среди них соя – 1147 га (33 поля), овес – 751 га (30 полей), залежь – 1491 га (65 полей). В Амурской области вегетационные индексы рассчитывались для 214 полей общей площадью 9812 га, в том числе 184 поля с соей (8207 га), 16 полей с овсом (623 га), 14 залежных полей (982 га).

**3. Результаты.** На рисунке 4 представлены значения радарных индексов вегетации для полей с соей, овсом, а также залежных земель в Хабаровском крае и Амурской области. Как видно, для каждого рассматриваемого класса, кривые DpRVI, RVI, VH/VV имели схожий вид. Коэффициенты корреляции, рассчитанные между показателями временных рядов для трех видов радарных индексов, находились в диапазоне 0,96-0,99 для разных классов полей. Значения временных рядов сезонного хода радарных VI сои возрастали в период с начала июля до последней декады сентября (рисунок 4(а, б, в)). Это является достаточно логичным, учитывая, что сев сои в южной части Дальнем Востоке обычно проводится в конце мая и в первой половине июня, при этом формирование травяного покрова в 2021 г. на полях в мае могло зависеть от сроков вспашки, обработки гербицидами и прочих мероприятий. Начиная с конца сентября, значения радарных VI для полей с соей достаточно быстро уменьшались, что было связано со стадией увядания растений при низких температурных характеристиках. В это время соя готова к уборке, но точная дата для разных регионов и муниципалитетов различалась в зависимости от погодных условий (вплоть до начала календарной зимы).

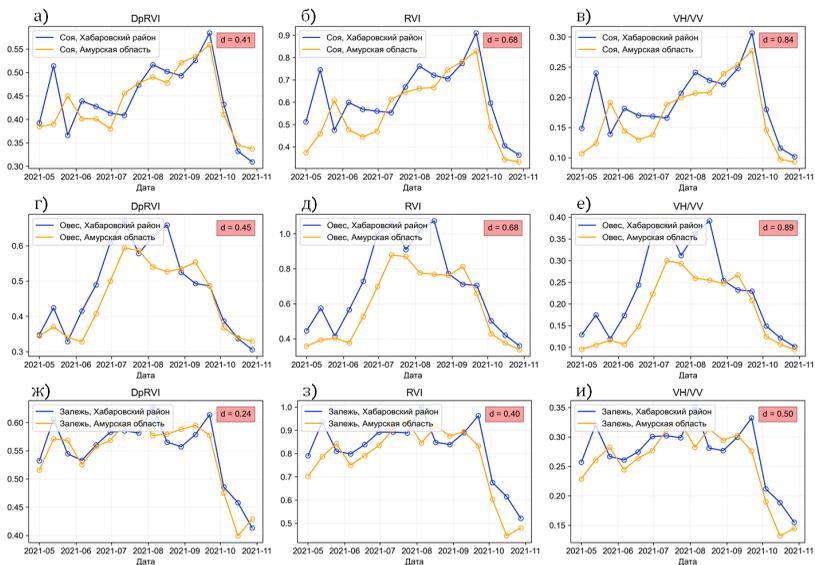


Рис. 4. Средние значения временных рядов евклидова расстояния DpRVI, RVI, VN/VV для сельскохозяйственных земель 2021 года в Хабаровском крае и Амурской области: а), б), в) – соя; г), д), е) – овес; ж), з), и) – залежь

Сев овса проводится в мае; однако можно отметить, что для Хабаровского края наблюдалось раннее повышение значений радарных VI перед посевом культуры (рисунок 4(г, д, е)). Достаточно высокие значения характерны в этот же период также и для полей с соей, что могло быть вызвано как быстрым ростом сорных культур, так и особенностями севооборота – семена овса, оставленные на поле в прошлый год, давали ранние всходы. Максимум радарных VI для овса достигался в июле. Для всей области исследования уборка овса проводилась с последней декады июля по последнюю декаду августа. При этом, на некоторых полях в сентябре мог продолжаться рост значений VI, что обусловлено развитием биомассы многолетних трав после уборки овса, что особенно актуально для Дальнего Востока, где посеы овса зачастую проводятся совместно с многолетними травами. Кривая сезонного хода радарных VI на залежных землях не продемонстрировала выраженных пиков, что связано с естественным ростом травы и отсутствием посевной и уборочной кампании (рисунок 4(ж, з, и)). Тем не менее, различные природные и антропогенные явления могут привести к определенным отклонениям роста травяного покрова на заброшенных пахотных землях.

Проведенный анализ попарного сходства значений временных рядов индексов вегетации для Хабаровского края и Амурской области продемонстрировал, что значения меры близости для индекса DpRVI были минимальны, а для VH/VV – максимальны. Так, для полей с соей  $d=0,41; 0,68; 0,84$ , соответственно для DpRVI, RVI, VH/VV. Для полей с овсом и залежных земель значения  $d$  составили  $0,45; 0,68; 0,89$  и  $0,24; 0,40; 0,50$ . Исследования вариабельности сезонного хода индексов VI для отдельных полей подтвердили наибольшую стабильность и устойчивость индекса DpRVI. Вариабельность трех индексов в период вегетационного цикла для всех исследуемых классов существенно различается (рисунок 5). Так, например, для сои в Хабаровском крае значения VAR в период с мая по первую декаду июля для DpRVI находились на уровне 10-15%, для RVI на уровне 13-18%, и VH/VV – 18-23% (рисунок 5(а)). С середины июля до третьей декады сентября вариабельность DpRVI снизилась до 5-8%, а RVI и VH/VV, соответственно, до 8-12% и 10-15%. В Амурской области соответствующие значения коэффициентов вариации в период максимального набора биомассы – июль-август для DpRVI составили 6-10%, для RVI – 10-18%, для VH/VV – 8%, 12%, 15% (рисунок 5(б)). Более высокие значения коэффициентов вариации для сезонного хода VI для полей с соей в Амурской области могли быть объяснены сортовым разнообразием сои в регионе. На рисунках 5(в) и 5(г) представлены изменения вариабельности значений VI в период с 1 мая по 31 октября для полей с овсом. Как видно, значения коэффициентов вариации DpRVI для отдельных полей на всем периоде наблюдений ниже, чем для индексов RVI и VH/VV соответственно, в среднем на 5-8% и 10-15%. Минимальные значения вариабельности радиолокационных индексов вегетации в Хабаровском крае в период с начала июля до середины августа (от вызревания до уборки) составили 3-10% для DpRVI, 10-13% для RVI, 14-20% для VH/VV (рисунок 5(в)). Дальнейший рост вариабельности связан с особенностями сева овса – для части полей сев проводился с подсевом многолетних трав, что способствовало росту значений индекса, для части полей подсев не использовался. В Амурской области сроки уборки овса в 2021 году в ряде муниципалитетов были смещены в сравнении с другими территориями, что вызвало увеличение значений коэффициентов вариации в августе. Для залежных полей практически для всех точек временных рядов VI и в Хабаровском крае, и в Амурской области наблюдалась закономерность, свойственная индексам для полей с соей и овсом – вариабельность DpRVI ниже, чем вариабельность значений RVI и VH/VV (рисунок 5(д, е)). При этом значения коэффициентов

вариации сезонного хода DpRVI залежи в Хабаровском районе во всем периоде наблюдений, за исключением последней недели октября, не превышали 10% (рисунок 5(д)). Для Амурской области значения коэффициентов вариации сезонного хода DpRVI залежных земель резко возрастают с началом октября, что возможно связано с большой пространственной удаленностью отдельных полей в Амурской области, различными метеорологическими условиями, и соответственно с разной продолжительностью вегетационных циклов (рисунок 5(е)). Пространственная неоднородность в Амурской области естественно, оказывает влияние и на вариабельность значений VI сои, овса и прочих культур.

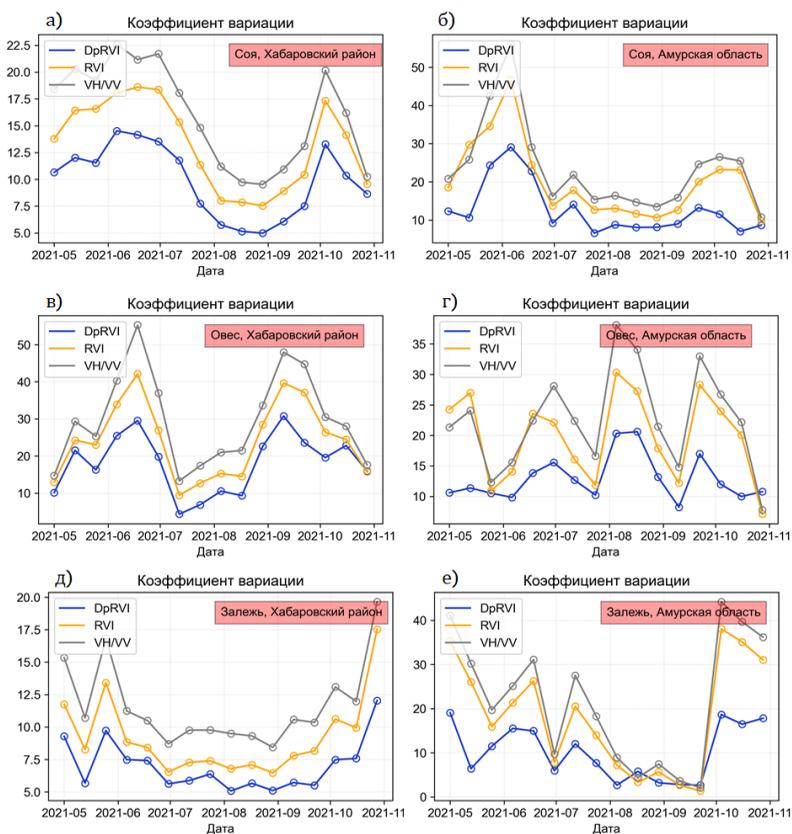


Рис. 5. Средние значения временных рядов коэффициента вариации DpRVI, RVI, Vh/Vv для сельскохозяйственных земель 2021 года в Хабаровском крае и Амурской области: а), б) – соя; в), г) – овес; д), е) – залежь

Основные характеристики временных рядов VI для отдельных полей с посевами сои, овса и залежных земель представлены в таблицах 1-3. Были рассчитаны численные значения среднего максимума VI, среднего значения дня наступления максимума, а также вариабельность этих показателей отдельно для Хабаровского края и Амурской области. Для сравнения, помимо радиолокационных индексов, определялись средние значения максимума и дня наступления максимума для сезонного хода NDVI.

Таблица 2. Характеристики временных рядов VI для отдельных полей с посевами сои в 2021 г. в Хабаровском крае и Амурской области

Регион	Характеристики	VI			
		<i>DpRVI</i>	<i>RVI</i>	<i>VH/VV</i>	<i>NDVI</i>
Хабаровский край	$\overline{VI_{max}} \pm \Delta \overline{VI_{max}}$	0,59±0,08	0,92±0,16	0,31±0,07	0,80±0,11
	$VAR_{VI_{max}}$	6,7	9,1	11,7	6,8
	$\overline{DOY_{max}} \pm \Delta \overline{DOY_{max}}$	263,4±8,2	261,0±21,3	260,7±21,4	238,2±15,0
	$VAR_{DOY_{max}}$	1,5	3,8	4,4	3,4
Амурская область	$\overline{VI_{max}} \pm \Delta \overline{VI_{max}}$	0,60±0,06	0,91±0,17	0,31±0,07	0,86±0,15
	$VAR_{VI_{max}}$	4,4	8,0	10,5	8,6
	$\overline{DOY_{max}} \pm \Delta \overline{DOY_{max}}$	261,5±15,0	257,0±24,0	256,1±27,2	224,5±24,6
	$VAR_{DOY_{max}}$	2,9	4,1	5,1	5,5

Как видно из таблицы 2, средние значения  $\overline{VI_{max}}$  и  $\overline{DOY_{max}}$  сои по Амурской области и Хабаровскому краю не различаются значимо для всех индексов. Коэффициенты вариации среднего максимума для *DpRVI* в Хабаровском крае и Амурской области были равны, соответственно 6,7% и 4,4%, для *RVI* – 9,1% и 8,0%, *VH/VV* – 11,7% и 10,5%. Вариабельность максимума *NDVI*, а также дня наступления максимума *NDVI* была несколько ниже соответствующих показателей для *DpRVI*.

Таблица 3. Характеристики временных рядов VI для отдельных полей с посевами овса в 2021 г. в Хабаровском крае и Амурской области

Регион	Характеристики	VI			
		$DpRVI$	$RVI$	$VH/VV$	$NDVI$
Хабаровский край	$\overline{VI_{max}} \pm \Delta\overline{VI_{max}}$	0,69±0,11	0,96±0,34	0,44±0,20	0,69±0,24
	$VAR_{VI_{max}}$	6,2	11,3	17,8	8,1
	$\overline{DOY_{max}} \pm \Delta\overline{DOY_{max}}$	193,3±33,8	203,0±35,4	200,7±42,4	200,3±34,8
	$VAR_{DOY_{max}}$	6,8	7,8	8,7	5,9
Амурская область	$\overline{VI_{max}} \pm \Delta\overline{VI_{max}}$	0,61±0,16	0,91±0,32	0,31±0,16	0,76±0,19
	$VAR_{VI_{max}}$	11,7	15,6	21,9	13,6
	$\overline{DOY_{max}} \pm \Delta\overline{DOY_{max}}$	199,0±19,2	201,4±22,3	204,1±25,6	195,6±21,3
	$VAR_{DOY_{max}}$	4,3	4,9	5,3	4,9

Для отдельных полей с овсом вариабельность максимума и дня наступления максимума радиолокационных VI выше, чем для полей с соей. В Хабаровском крае для максимума  $VAR_{DpRVI_{max}}$  был равен 6,2%, в Амурской области – 11,7%. Значения показателей для RVI и VH/VV были выше примерно в 1,5-3 раза. Вариабельность максимума NDVI для полей с овсом была на уровне вариабельности DpRVI – 8,1% и 13,6% в зависимости от региона. Значения календарного дня максимума для полей с овсом также отличались большей вариативностью в сравнении с соей, что объясняется особенностями сева овса с подсевом многолетних трав. При этом коэффициенты вариации дня максимума DpRVI для полей с овсом в Хабаровском крае были выше, чем соответствующий показатель для NDVI – 6,8% и 5,9%. В Амурской области, напротив,  $VAR_{DOY_{max}}$  для DpRVI был равен 4,3%, а для NDVI – 4,9%.

Средние значения  $\overline{VI_{max}}$  для радиолокационных индексов и NDVI залежи в Хабаровском крае и Амурской области также соответствуют друг другу (таблица 4). Вариабельность DpRVI для отдельных полей с залежью ниже, чем вариабельность RVI и VH/VV, что соответствует коэффициентам вариации максимума NDVI. Необходимо отметить, что дни наступления максимума для залежи на разных полях обладали большой вариативностью, что вполне

объяснимо, учитывая природу залежи. Низкие значения  $VAR_{DOY_{max}}$  не могут служить показателем устойчивости того или иного индекса.

Таблица 4. Характеристики временных рядов VI для залежных земель в 2021 г. в Хабаровском крае и Амурской области

Регион	Характеристики	VI			
		$DpRVI$	$RVI$	$VH/VV$	$NDVI$
Хабаровский край	$\overline{VI_{max}} \pm \Delta\overline{VI_{max}}$	0,65±0,05	0,97±0,12	0,36±0,05	0,75±0,07
	$VAR_{VI_{max}}$	4,0	5,7	7,5	4,6
	$\overline{DOY_{max}} \pm \Delta\overline{DOY_{max}}$	207,4±97,1	208,1±92,1	208,6±92,8	200,1±64,8
	$VAR_{DOY_{max}}$	23,4	22,2	22,3	13,7
Амурская область	$\overline{VI_{max}} \pm \Delta\overline{VI_{max}}$	0,65±0,10	0,98±0,23	0,36±0,12	0,72±0,09
	$VAR_{VI_{max}}$	6,3	9,3	13,3	5,6
	$\overline{DOY_{max}} \pm \Delta\overline{DOY_{max}}$	196,4±121,9	191,3±111,7	192,4±122,6	195,6±61,3
	$VAR_{DOY_{max}}$	25,4	23,9	25,7	14,1

**4. Выводы.** Проведенное исследование, выполненное в рамках разработки системного подхода к использованию радиолокационных изображений в цифровом земледелии, продемонстрировало, что временные ряды радиолокационных индексов  $DpRVI$ ,  $RVI$ ,  $VH/VV$  для основных типов с/х земель юга Дальнего Востока имеют отличительные особенности и могут применяться в задачах классификации, моделирования урожайности, распознавания культур и прочих задачах цифрового земледелия. Было выявлено, что средние значения временных рядов для полей с соей, овсом и залежью имели характерную форму, при этом сходство между средними VI сохранялось для достаточно удаленных друг от друга регионов – Хабаровского края и Амурской области. Наиболее высокое сходство было отмечено для индекса  $DpRVI$  (d находилось в диапазоне от 0,24 до 0,45) в сравнении с  $RVI$  (d от 0,40 до 0,68) и  $VH/VV$  (d от 0,50 до 0,84). В результате проведенного исследования впервые было установлено, что индекс  $DpRVI$  оказался наиболее устойчивым для основных культур, а также залежных земель Хабаровского края и Амурской области. Значения коэффициентов вариации  $DpRVI$  по отдельным полям с одинаковыми с/х культурами были в 2-3 раза ниже,

чем у радиолокационных индексов RVI и VH/VV. Были рассчитаны основные характеристики сезонного хода временных рядов радиолокационных VI в сравнении с NDVI – величина максимума VI, дата наступления максимума и значения коэффициента вариации для этих показателей. Было установлено, во-первых, что значения этих показателей в разных регионах схожи между собой; во-вторых, вариабельность максимума и дня наступления максимума для DpRVI ниже, чем для RVI и VH/VV; в-третьих, вариабельность максимума и дня наступления максимума для DpRVI сопоставима с NDVI. В целом, использование DpRVI может являться хорошей альтернативой и дополнением к индексам, рассчитанным по мультиспектральным изображениям, в случае отсутствия снимков из-за облачности или иных причин.

Стоит отметить, что индекс DpRVI с двойной поляризацией представляет значительный интерес, так как в запланированных будущих миссиях, таких как Sentinel-1C/D, NISAR и Sentinel-1 Next Generation, не для всех режимов съемки будут использоваться все четыре поляризации, в то время как двойная поляризация – присутствует во многих режимах. Кроме того, необходима дальнейшая оценка и анализ данных поляриметрического режима съемки для HV/HH поляризаций. Коэффициент обратного рассеяния от полога с/х культур может отличаться от VH/VV поляризаций, т.к. передаваемые горизонтально поляризованные электромагнитные волны по-другому будут взаимодействовать в отличие от вертикальной поляризации. Также необходимы исследования зависимости вегетационных индексов от угла падения радиолокационного сигнала радарного изображения. Весьма интересен S-диапазон длин волн для задач сельского хозяйства, поэтому вывод на орбиту российского спутника Кондор-ФКА1 представляет собой многообещающую перспективу.

В связи с прекращением функционирования одного из спутников миссии Sentinel-1, ближайшими работами коллектива авторов предполагаются исследования, направленные на: обработку радарных снимков, расчет и анализ временных рядов DpRVI для некоторых российских и зарубежных спутников, интегрирование и совместное использование оптических и радарных изображений, использование нейросетевых подходов к моделированию урожайности и другие.

### **Литература**

1. Якушев В.П., Захарян Ю.Г., Блохина С.Ю. Состояние и перспективы использования дистанционного зондирования Земли в сельском хозяйстве //

- Современные проблемы дистанционного зондирования земли из космоса. 2022. Т. 19. № 1. С. 287–294.
2. Fisette T., Rollin P., Aly Z., Campbell, L., Daneshfar, B., Filyer, P., Smith A., Davidson A., Shang J., Jarvis I. AAFC annual crop inventory // Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics). 2013. pp. 270–274.
  3. Лупян Е.А., Барталев С.А., Толпин В.А., Жарко В.О., Крашенинникова Ю.С., Оксюкевич А.Ю. Использование спутникового сервиса ВЕГА в региональных системах дистанционного мониторинга // Современные проблемы дистанционного зондирования земли из космоса. 2014. Т. 11. № 3. С. 215–232.
  4. Лупян Е.А., Прошин А.А., Бурцев М.А., Кашницкий А.В., Балашов И.В., Барталев С.А., Бриль А.А., Егоров В.А., Жарко В.О., Константинова А.М., Кобец Д.А., Мазуров А.А., Марченков В.В., Матвеев А.М., Миклашевич Т.С., Плотников Д.Е., Радченко М.В., Стыщенко Ф.В., Сычугов И.Г., Толпин В.А., Уваров И.А., Хвостиков С.А., Ховратович Т.С. Система «Вега-Science»: особенности построения, основные возможности и опыт использования // Современные проблемы дистанционного зондирования земли из космоса. 2021. Т. 18. № 6. С. 9–31.
  5. Денисов П.В., Трошко К.А., Лупян Е.А., Толпин В.А. Возможности и опыт использования информационной системы ВЕГА-PRO для мониторинга сельскохозяйственных земель // Вычислительные технологии. 2022. Т. 27. № 3. С. 66–83.
  6. Macelloni G., Paloscia S., Pampaloni P., Ruisi R., Dechambre M., Valentin R., Chanzy A., Prevot L., Bruguier N. Modelling radar backscatter from crops during the growth cycle // *Agronomie*. 2002. vol. 22. № 6. pp. 575–579.
  7. Blaes X., Defourny P., Wegmuller U., Della Vecchia A., Guerrierio L., Ferrazzoli P. C-band polarimetric indexes for maize monitoring based on a validated radiative transfer model // *IEEE Transactions on Geoscience and Remote Sensing*. 2006. vol. 44. no. 4. pp. 791–800.
  8. McNairn H., Champagne C., Shang J., Holmstrom D., Reichert G. Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2009. vol. 64. no. 5. pp. 434–449.
  9. Mercier A., Betbeder J., Baudry J., Le Roux V., Spicher F., Lacoux J., Roger D., Hubert-Moy L. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2020. vol. 163. pp. 231–256.
  10. Li M., Bijker W. Potential of Multi-Temporal Sentinel-1A Dual Polarization SAR Images for Vegetable Classification in Indonesia // *IEEE International Geoscience and Remote Sensing Symposium*. 2018. pp. 3820–3823.
  11. Harfenmeister K., Itzerott S., Weltzien C., Spengler D. Detecting Phenological Development of Winter Wheat and Winter Barley Using Time Series of Sentinel-1 and Sentinel-2 // *Remote Sensing*. 2021. vol. 13(24). no. 5036.
  12. Bao X., Zhang R., Lv J., Wu R., Zhang H., Chen J., Zhang B., Ouyang X., Liu G.. Vegetation descriptors from Sentinel-1 SAR data for crop growth monitoring // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2023. vol. 203(17). pp. 86–114.
  13. Kim Y., van Zyl J.J. A Time-Series Approach to Estimate Soil Moisture Using Polarimetric Radar Data // *IEEE Transactions on Geoscience and Remote Sensing*. 2009. vol. 47. no. 8. pp. 2519–2527.
  14. Trudel M., Charbonneau F., Leconte R. Using RADARSAT-2 polarimetric and ENVISAT-ASAR dual-polarization data for estimating soil moisture over agricultural fields // *Canadian Journal of Remote Sensing*. 2012. vol. 38. no. 4. pp. 514–527.

15. Kumar S., Rao S., Sharma J. Radar Vegetation Index as an Alternative to NDVI for Monitoring of Soyabean and Cotton // Proceedings of the XXXIII INCA International Congress (Indian Cartographer). 2013. vol. 33. pp. 91–96.
16. Nasirzadehdizaji R., Balik Sanli F., Abdikan S., Cakir Z., Sekertekin A., Ustuner M.. Sensitivity Analysis of Multi-Temporal Sentinel-1 SAR Parameters to Crop Height and Canopy Coverage // Applied Sciences. 2019. vol. 9(4). no. 655.
17. Ratha D., Mandal D., Kumar V., McNairn H., Bhattacharya A., Frery A.C. A Generalized Volume Scattering Model-Based Vegetation Index From Polarimetric SAR Data // IEEE Geoscience and Remote Sensing Letters. 2019. vol. 16. no. 11. pp. 1791–1795.
18. Mandal D., Ratha D., Bhattacharya A., Kumar V., McNairn H., Rao Y.S., Frery A.C. A Radar Vegetation Index for Crop Monitoring Using Compact Polarimetric SAR Data // IEEE Transactions on Geoscience and Remote Sensing. 2020. vol. 58. no. 9. pp. 6321–6335.
19. Chang J.G., Shoshany M., Oh Y. Polarimetric Radar Vegetation Index for Biomass Estimation in Desert Fringe Ecosystems // IEEE Transactions on Geoscience and Remote Sensing. 2018. vol. 56. no 12. pp. 7102–7108.
20. Periasamy S. Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1 // Remote Sensing of Environment. 2018. vol. 217. pp. 537–549.
21. Mandal D., Kumar V., Ratha D., Dey S., Bhattacharya A., Lopez-Sanchez J. M., McNairn H., Rao Y.S. Dual polarimetric radar vegetation index for crop growth monitoring using Sentinel-1 SAR data // Remote Sensing of Environment. 2020. vol. 247. no. 111954.
22. Sekertekin A., Marangoz A.M., Abdikan S. ALOS-2 and Sentinel-1 SAR data sensitivity analysis to surface soil moisture over bare and vegetated agricultural fields // Computers and Electronics in Agriculture. 2020. vol. 171. no. 105303.
23. Ulaby F., Bativala P., Dobson M. Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part I-Bare soil // IEEE Transactions on Geoscience Electronics. 1978. vol. 16. no. 4. pp. 286–295.
24. Дубровин К.Н., Степанов А.С., Верхотуров А.Л., Асеева Т.А. Идентификация сельскохозяйственных культур с использованием радарных изображений // Информатика и автоматизация. 2022. Т. 21. № 2. С. 405–426.
25. Robertson L.D., Davidson A.A., McNairn H., Hosseini M., Mitchell S. Assessment of Multi-Frequency SAR for Crop Type Classification and Mapping // IGARSS-2019 IEEE International Geoscience and Remote Sensing Symposium. 2019. pp. 489–492.
26. Dubrovin K., Stepanov A., Verkhotur A. Cropland Mapping Using Sentinel-1 Data in the Southern Part of the Russian Far East // Sensors. 2023. vol. 23(18). no. 7902.

**Верхотуров Андрей Леонидович** — старший научный сотрудник, лаборатория информационных и вычислительных систем, Вычислительный центр Дальневосточного отделения Российской академии наук (ВЦ ДВО РАН). Область научных интересов: разработка информационных систем обработки спутниковых данных для исследования (мониторинга) состояния природных и сложных технических систем, обработка данных радиолокационной интерферометрии, геоинформационные системы. Число научных публикаций — 51. andrey@ccfebras.ru; улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(4212)227-267.

**Степанов Алексей Сергеевич** — д-р фармацевт. наук, ведущий научный сотрудник, лаборатория селекции зерновых и колосовых культур, Дальневосточный научно-исследовательский институт сельского хозяйства (ДВ НИИСХ). Область научных интересов: сельскохозяйственная экономика прогнозирование урожайности

сельскохозяйственных культур, математическое моделирование в сельском хозяйстве. Число научных публикаций — 98. [stepanfx@mail.ru](mailto:stepanfx@mail.ru); улица Клубная, 13, 680521, Восточное, Россия; р.т.: +7(924)210-9102.

**Илларионова Любовь Викторовна** — канд. физ.-мат. наук, старший научный сотрудник, лаборатория численных методов математической физики, Вычислительный центр Дальневосточного отделения Российской академии наук (ВЦ ДВО РАН). Область научных интересов: машинное обучение, программирование, численные методы. Число научных публикаций — 37. [illarionova\\_l@list.ru](mailto:illarionova_l@list.ru); улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(924)200-494-037.

A. VERKHOTUROV, A. STEPANOV, L. ILLARIONOVA  
**USING SAR DATA FOR MONITORING OF AGRICULTURAL  
CROPS IN THE SOUTH OF THE RUSSIAN FAR EAST**

*Verkhoturov A., Stepanov A., Illarionova L. Using SAR Data for Monitoring of Agricultural Crops in the South of the Russian Far East.*

**Abstract.** The use of SAR data to monitoring agricultural crops is a promising area of research designed to complement existing methods and technologies based on the analysis of multispectral images. The main advantages of vegetation indices calculated from SAR data include their sensitivity to the polarimetric properties of the backscatter intensity, its scattering characteristics, and independence from cloud cover. This is especially important for the territory of the south of the Russian Far East, whose monsoon climate provides humid and cloudy weather during the period when crops gain maximum biomass. For arable lands in the Khabarovsk Territory and the Amur Region, a total of 64 Sentinel-1 SAR images were obtained from May to October 2021. For each date, the values of the DpRVI, RVI, VH/VV indices were calculated and time series were constructed for the entire observation period for individual fields (342 fields in total). NDVI time series were constructed from Sentinel-2 multispectral images using a cloud mask. The characteristics of time series extremes were calculated for different types of arable land: soybeans, oats, and fallows. It was shown that for each crop the seasonal curves DpRVI, RVI, VH/VV had a characteristic appearance. It was found that the DpRVI demonstrated the highest stability – the coefficients of variation of the seasonal variation of DpRVI were significantly lower than those for RVI and VH/VV. It was also revealed that the similarity between the curves of these indices remained for regions quite distant from each other – the Khabarovsk Territory and the Amur Region. The main characteristics of the seasonal variation of time series of radar indices were calculated in comparison with NDVI – the magnitude of the maximum, the date of the maximum and the values of the coefficient of variation for these indicators. It was found, firstly, that the values of these indicators in different regions are similar to each other; secondly, the variability of the maximum and the day of the maximum for DpRVI is lower than for RVI and VH/VV; thirdly, the variability of the maximum and the day of the maximum for DpRVI is comparable to NDVI. Thus, time series of radar indices DpRVI, RVI, VH/VV for the main types of agricultural lands in the south of the Far East have distinctive features and can be used in classification problems, yield modeling and crop rotation control.

**Keywords:** remote sensing data, Far East, agricultural land monitoring, radar vegetation indices, variability, arable land, crops, time series.

## References

1. Yakushev V.P., Zakharyan Yu.G., Blokhina S.Yu. [Current problems and prospects for the use of remote sensing of the Earth in agriculture]. *Sovremennye Problemy Distantionnogo Zondirovaniya Zemli iz Kosmosa – Modern problems of remote sensing of the Earth from space.* 2022. vol. 19. no. 1. pp. 287–294. (In Russ.).
2. Fisette T., Rollin P., Aly Z., Campbell, L., Daneshfar, B., Filyer, P., Smith A., Davidson A., Shang J., Jarvis I. AAFC annual crop inventory. *Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics).* 2013. pp. 270–274.
3. Loupian E.A., Bartalev S.A., Tolpin V.A., Krashennikova Ju.S., Oksjukevich A.Ju. [VEGA satellite service applications in regional remote monitoring systems]. *Sovremennye Problemy Distantionnogo Zondirovaniya Zemli iz Kosmosa – Modern*

- problems of remote sensing of the Earth from space. 2014. vol. 11. no. 3. pp. 215–232. (In Russ.).
4. Loupian E.A., Proshin A.A., Bourtsev M.A., Kashnickij A.V., Balashov I.V., Bartalev S.A., Bril' A.A., Egorov V.A., Zharko V.O., Konstantinova A.M., Kobec D.A., Mazurov A.A., Marchenkov V.V., Matveev A.M., Miklashevich T.S., Plotnikov D.E., Radchenko M.V., Stycenko F.V., Sychugov I.G., Tolpin V.A., Uvarov I.A., Hvostikov S.A., Hovratovich T.S. [Vega-Science system: design features, main capabilities and usage experience]. *Sovremennye Problemy Distantionnogo Zondirovaniya Zemli iz Kosmosa – Modern problems of remote sensing of the Earth from space*. 2021. vol. 18. no. 6. pp. 9–31. (In Russ.).
  5. Denisov P.V., Troshko K.A., Loupian E.A., Tolpin V.A. [Potential and experience of Vega-PRO information system use for monitoring of agricultural lands]. *Vychislitel'nye tehnologii – Computing technologies*. 2022. vol. 27. no. 3. pp. 66–83. (In Russ.).
  6. Macelloni G., Paloscia S., Pampaloni P., Ruisi R., Dechambre M., Valentin R., Chanzy A., Prevot L., Bruguier N. Modelling radar backscatter from crops during the growth cycle. *Agronomie*. 2002. vol. 22. № 6. pp. 575–579.
  7. Blaes X., Defourny P., Wegmuller U., Della Vecchia A., Guerriero L., Ferrazzoli P. C-band polarimetric indexes for maize monitoring based on a validated radiative transfer model. *IEEE Transactions on Geoscience and Remote Sensing*. 2006. vol. 44. no. 4. pp. 791–800.
  8. McNairn H., Champagne C., Shang J., Holmstrom D., Reichert G. Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2009. vol. 64. no. 5. pp. 434–449.
  9. Mercier A., Betbeder J., Baudry J., Le Roux V., Spicher F., Lacoux J., Roger D., Hubert-Moy L. Evaluation of Sentinel-1 & 2 time series for predicting wheat and rapeseed phenological stages. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2020. vol. 163. pp. 231–256.
  10. Li M., Bijker W. Potential of Multi-Temporal Sentinel-1A Dual Polarization SAR Images for Vegetable Classification in Indonesia. *IEEE International Geoscience and Remote Sensing Symposium*. 2018. pp. 3820–3823.
  11. Harfenmeister K., Itzerott S., Weltzien C., Spengler D. Detecting Phenological Development of Winter Wheat and Winter Barley Using Time Series of Sentinel-1 and Sentinel-2. *Remote Sensing*. 2021. vol. 13(24). no. 5036.
  12. Bao X., Zhang R., Lv J., Wu R., Zhang H., Chen J., Zhang B., Ouyang X., Liu G. Vegetation descriptors from Sentinel-1 SAR data for crop growth monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2023. vol. 203(17). pp. 86–114.
  13. Kim Y., van Zyl J.J. A Time-Series Approach to Estimate Soil Moisture Using Polarimetric Radar Data. *IEEE Transactions on Geoscience and Remote Sensing*. 2009. vol. 47. no. 8. pp. 2519–2527.
  14. Trudel M., Charbonneau F., Leconte R. Using RADARSAT-2 polarimetric and ENVISAT-ASAR dual-polarization data for estimating soil moisture over agricultural fields. *Canadian Journal of Remote Sensing*. 2012. vol. 38. no. 4. pp. 514–527.
  15. Kumar S., Rao S., Sharma J. Radar Vegetation Index as an Alternative to NDVI for Monitoring of Soyabean and Cotton. *Proceedings of the XXXIII INCA International Congress (Indian Cartographer)*. 2013. vol. 33. pp. 91–96.
  16. Nasirzadehdizaji R., Balik Sanli F., Abdikan S., Cakir Z., Sekertekin A., Ustuner M. Sensitivity Analysis of Multi-Temporal Sentinel-1 SAR Parameters to Crop Height and Canopy Coverage. *Applied Sciences*. 2019. vol. 9(4). no. 655.
  17. Ratha D., Mandal D., Kumar V., McNairn H., Bhattacharya A., Frery A.C. A Generalized Volume Scattering Model-Based Vegetation Index From Polarimetric

- SAR Data. IEEE Geoscience and Remote Sensing Letters. 2019. vol. 16. no. 11. pp. 1791–1795.
18. Mandal D., Ratha D., Bhattacharya A., Kumar V., McNairn H., Rao Y.S., Frery A.C. A Radar Vegetation Index for Crop Monitoring Using Compact Polarimetric SAR Data. IEEE Transactions on Geoscience and Remote Sensing. 2020. vol. 58. no. 9. pp. 6321–6335.
  19. Chang J.G., Shoshany M., Oh Y. Polarimetric Radar Vegetation Index for Biomass Estimation in Desert Fringe Ecosystems. IEEE Transactions on Geoscience and Remote Sensing. 2018. vol. 56. no. 12. pp. 7102–7108.
  20. Periasamy S. Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1. Remote Sensing of Environment. 2018. vol. 217. pp. 537–549.
  21. Mandal D., Kumar V., Ratha D., Dey S., Bhattacharya A., Lopez-Sanchez J. M., McNairn H., Rao Y.S. Dual polarimetric radar vegetation index for crop growth monitoring using Sentinel-1 SAR data. Remote Sensing of Environment. 2020. vol. 247. no. 111954.
  22. Sekertekin A., Marangoz A.M., Abdikan S. ALOS-2 and Sentinel-1 SAR data sensitivity analysis to surface soil moisture over bare and vegetated agricultural fields. Computers and Electronics in Agriculture. 2020. vol. 171. no. 105303.
  23. Ulaby F., Batlivala P., Dobson M. Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part I-Bare soil. IEEE Transactions on Geoscience Electronics. 1978. vol. 16. no. 4. pp. 286–295.
  24. Robertson L.D., Davidson A.A., McNairn H., Hosseini M., Mitchell S. Assessment of Multi-Frequency SAR for Crop Type Classification and Mapping. IGARSS-2019 IEEE International Geoscience and Remote Sensing Symposium. 2019. pp. 489–492.
  25. Dubrovin K.N., Stepanov A.S., Verkhoturov A.L., Aseeva T.A. [Crop identification using radar images]. Informatika i avtomatizacija –Informatics and Automation. 2022. vol. 21. no. 2. pp. 405–426. (In Russ.).
  26. Dubrovin K., Stepanov A., Verkhoturov A. Cropland Mapping Using Sentinel-1 Data in the Southern Part of the Russian Far East. Sensors. 2023. vol. 23(18). no. 7902.

**Verkhoturov Andrey** — Senior researcher, Laboratory of information and computational systems, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences (CC FEB RAS). Research interests: processing of Earth remote sensing data, processing of radar interferometry data, geoinformation systems. The number of publications — 51. andrey@ccfebras.ru; 65, Kim Yu Chen St., 680000, Khabarovsk, Russia; office phone: +7(4212)227-267.

**Stepanov Aleksey** — Ph.D., Dr.Sci., Leading researcher, Laboratory of breeding cereals and ear crops, Far Eastern Research Institute of Agriculture (FEARI). Research interests: agricultural economics, forecasting the yield of crops, mathematical modeling in agriculture. The number of publications — 98. stepanfx@mail.ru; 13, Clubnaya St., 680521, Vostochnoye, Russia; office phone: +7(924)210-9102.

**Illarionova Lyubov** — Ph.D., Senior researcher, Laboratory of numerical methods of mathematical physics, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences (CC FEB RAS). Research interests: machine learning, numerical methods. The number of publications — 37. illarionova\_l@list.ru; 65, Kim Yu Chen St., 680000, Khabarovsk, Russia; office phone: +7(924)200-494-037.

А.А. Молдовян, Д.Н. Молдовян, Н.А. Молдовян  
**ПОСТКВАНТОВЫЕ ДВУХКЛЮЧЕВЫЕ КРИПТОСХЕМЫ  
НА КОНЕЧНЫХ АЛГЕБРАХ**

*Молдовян А.А., Молдовян Д.Н., Молдовян Н.А. Постквантовые двухключевые криптосхемы на конечных алгебрах.*

**Аннотация.** Одним из направлений разработки практических постквантовых криптографических алгоритмов с открытым ключом является использование конечных алгебр в качестве их алгебраического носителя. Рассматриваются два подхода в этом направлении: 1) построение алгоритмов электронной цифровой подписи со скрытой группой на некоммутативных ассоциативных алгебрах и 2) построение алгоритмов многомерной криптографии с использованием операции экспоненцирования в векторном конечном поле (коммутативной алгебре, являющейся конечным полем) для задания нелинейного отображения с секретной лазейкой. Первый подход включает разработку криптосхем двух типов: основанных на вычислительной трудности а) скрытой задачи дискретного логарифмирования и б) решения большой системы квадратных уравнений. Для второго подхода возникают проблемы обеспечения полной рандомизации цифровой подписи и задания некоммутативных ассоциативных алгебр большой размерности. Обсуждаются способы решения данных проблем. Показана важность исследования строения конечных некоммутативных алгебр с точки зрения декомпозиции на множество коммутативных подалгебр. Другое направление использования конечных алгебр для разработки криптографических алгоритмов с открытым ключом связано с существенным (в 10 и более раз) уменьшением размера открытого ключа в алгоритмах многомерной криптографии. В нем возникает проблема разработки формализованных параметризуемых унифицированных способов задания векторных конечных полей больших размерностей (от 5 до 130) с достаточно большим числом потенциально реализуемых типов и модификаций (до  $2^{500}$  и более), задаваемых различными наборами структурных констант, с помощью которых определяется операция умножения векторов. Предложены варианты указанных способов и топологий нелинейных отображений на векторных конечных полях различных размерностей. Показано, что использование отображений, задающих операцию экспоненцирования в векторных конечных полях, потенциально обеспечивает устранение основного недостатка известных алгоритмов многомерной криптографии, связанного с большим размером открытого ключа.

**Ключевые слова:** постквантовая криптография, многомерная криптография, конечная алгебра, некоммутативная алгебра, векторное конечное поле, нелинейные отображения.

**1. Введение.** Криптографические алгоритмы электронной цифровой подписи (ЭЦП), открытого шифрования и открытого согласования секретного ключа, основанные на вычислительной сложности задачи факторизации (ЗФ) и задачи дискретного логарифмирования (ЗДЛ), имеют достаточно широкое применение в современных информационных системах. В России и других ведущих странах мира действуют стандарты на криптографические алгоритмы с открытым ключом, основанные на ЗДЛ на эллиптической

кривой. Однако, достигнутый в последние годы значительный прогресс в области создания вычислителей нового типа, основанных на принципах квантовой механики, и ожидаемая возможность практической реализации квантовых алгоритмов решения ЗФ и ЗДЛ [1, 2], имеющих полиномиальную по времени вычислительную сложность, обусловили высокую степень актуальности проблемы разработки постквантовых криптографических алгоритмов с открытым ключом [3, 4], стойких к атакам с использованием как обычного, так и квантового компьютера. Откликом на данную проблему является проводимый с 1997 года по настоящее время всемирный конкурс по разработке проектов постквантовых стандартов на криптографические алгоритмы с открытым ключом [5].

Разработка постквантовых криптосхем (алгоритмов и протоколов) связана с использованием вычислительно трудных задач, для которых предположительно не будет найдено полиномиальных алгоритмов решения на гипотетическом квантовом компьютере. Предложены и апробированы различные задачи (отличные от ЗФ и ЗДЛ) в качестве основы для построения постквантовых криптосхем с открытым ключом. В качестве носителей последних, например, используются группы [6, 7], коды, исправляющие ошибки [8, 9], алгебраические решетки [10, 11], трудно обратимые [12] и булевы функции [13].

Одним из наиболее изученных и апробированных направлений разработки постквантовых криптосхем открытым ключом является многомерная криптография, основанная на трудно обратимых нелинейных отображениях с секретной лазейкой [14, 15]. Однако существенным ограничением для практического применения ее алгоритмов является чрезмерно большой размер открытого ключа (от десятков Кбайт при 80-битном уровне стойкости до нескольких Мбайт при 256-битной стойкости) [16, 17]. Исключение составляет алгоритм [18], стойкость которого на настоящий момент мало изучена.

Недавно предложенная парадигма [19, 20] построения нелинейных отображений как операций экспоненцирования в конечных векторных полях потенциально позволяет уменьшить размер открытого ключа алгоритмов многомерной криптографии на один – два десятичных порядка при заданном уровне стойкости. Описанные варианты реализации таких нелинейных отображений используют конкретные векторные конечные поля размерности  $m = 5, 6, 7$ , заданные эвристически над полями нечетной [19] и четной [20] характеристики. При этом предполагается, что эвристический способ также может быть применен и для случаев

$m = 11-97$ . Однако, этот способ затрудняет оптимизацию топологий нелинейных отображений, требующую иметь возможность параметризуемого задания векторных конечных полей с большим числом потенциально реализуемых модификаций при фиксированном значении размерности  $m$ . Это обуславливает важность задачи разработки формализованных унифицированных способов задания векторных конечных полей, представляющих собой частный случай конечных коммутативных ассоциативных алгебр, для развития парадигмы [19, 21].

Сравнительно новым подходом является построение постквантовых алгоритмов электронной цифровой подписи (ЭЦП) со скрытой группой на конечных некоммутативных ассоциативных алгебрах (КНАА) с использованием вычислительной трудности решения больших систем квадратных уравнений [21]. В настоящей статье показывается, что в известных криптосхемах последнего типа имеется проблема ограниченной рандомизации подписи, требующая использования КНАА сравнительно больших размерностей, и рассматривается способ обеспечения полной рандомизации подписи.

**2. Алгебры и векторные поля.** Элементами конечного  $m$ -мерного векторного пространства, заданного над конечным полем  $GF(p^s)$ , где  $p$  – четное или нечетное простое число;  $s$  – натуральное число, являются векторы  $A$ , представляемые в виде упорядоченного набора элементов поля:  $A = (a_0, a_1, \dots, a_{m-1})$ , где  $a_i \in GF(p^s)$  – координаты вектора, или в виде суммы его компонент:

$$A = \sum_{i=0}^{m-1} a_i \mathbf{e}_i, \text{ где } \mathbf{e}_i \text{ – базисные векторы. В векторном пространстве}$$

заданы две стандартные операции: сложение векторов и умножение вектора на скаляр (скалярное умножение). Если дополнительно к этим двум операциям определить операцию векторного умножения, операндами которой являются два произвольных вектора, которая обладает свойствами замкнутости и дистрибутивности слева и справа относительно операции сложения, то получаем конечную  $m$ -мерную алгебру.

Естественным является задание результата операции умножения векторов  $\mathbf{A} = \sum_{i=0}^{m-1} a_i \mathbf{e}_i$  и  $\mathbf{B} = \sum_{j=0}^{m-1} b_j \mathbf{e}_j$  по правилу перемножения каждой компоненты первого вектора с каждой компонентой второго вектора по следующей формуле:

$$\mathbf{AB} = \sum_{i,j=0}^{m-1} a_i b_j (\mathbf{e}_i \mathbf{e}_j)$$

в которой предполагается, что каждое произведение вида  $\mathbf{e}_i \mathbf{e}_j$  должно быть заменено на некоторый базисный вектор или однокомпонентный вектор  $\lambda \mathbf{e}_k$ , где  $\lambda \in GF(p)$  называется структурной константой. Для осуществления такой замены задается так называемая таблица умножения базисных векторов (ТУБВ), в ячейках которой при  $\lambda = 1$  указывается базисный вектор  $\mathbf{e}_k$ . Для определенности будем полагать, что левый множитель в произведении  $\mathbf{e}_i \mathbf{e}_j$  задает строку, а правый – столбец, пересечение которых указывает ячейку, содержащую однокомпонентный вектор  $\lambda \mathbf{e}_k$ . В общем случае можно задать замену  $\mathbf{e}_i \mathbf{e}_j$  на многокомпонентный вектор с сохранением требуемых свойств замкнутости и дистрибутивности, однако в последнем случае проблематично разработать ТУБВ реализующие свойство ассоциативности. В рассматриваемых в данной статье приложениях интерес представляют ассоциативные алгебры. Для задания ассоциативной операции векторного умножения и обеспечения возможности унифицированного параметризуемого задания ТУБВ используется замена произведения  $\mathbf{e}_i \mathbf{e}_j$  на однокомпонентный вектор.

Если ассоциативная операция векторного умножения (или просто операция умножения) является коммутативной (некоммутативной), то имеем конечную коммутативную (некоммутативную) ассоциативную алгебру (КК(Н)АА). Типовой вид ТУБВ [23], по которым могут быть заданы ККАА с глобальной двухсторонней единицей для случая произвольных размерностей  $m \geq 2$  представлен в таблице 1 (с нумерацией базисных векторов от 1 до  $m$ ).

Таблица 1 [23]. Общий вид ТУБВ (где  $\psi = \tau^{-1} \varepsilon \lambda$ ) для задания векторных конечных полей  $GF((p^s)^m)$  в виде  $m$ -мерных конечных алгебр над полем  $GF(p^s)$

$\mathbf{x}$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	...	$\mathbf{e}_{m-1}$	$\mathbf{e}_m$
$:\mathbf{e}_1$	$\tau \mathbf{e}_1$	$\tau \mathbf{e}_2$	$\tau \mathbf{e}_3$	$\tau \mathbf{e}_4$	$\tau \mathbf{e}_5$	$\tau \dots$	$\tau \mathbf{e}_{m-1}$	$\tau \mathbf{e}_m$
$:\mathbf{e}_2$	$:\tau \mathbf{e}_2$	$\varepsilon \mathbf{e}_3$	$\varepsilon \mathbf{e}_4$	$\varepsilon \mathbf{e}_5$	$\varepsilon \dots$	$\varepsilon \mathbf{e}_{m-1}$	$\varepsilon \mathbf{e}_m$	$\psi \mathbf{e}_1$
$:\mathbf{e}_3$	$\tau \mathbf{e}_3$	$\varepsilon \mathbf{e}_4$	$\varepsilon \mathbf{e}_5$	$\varepsilon \dots$	$\varepsilon \mathbf{e}_{m-1}$	$\varepsilon \mathbf{e}_m$	$\psi \mathbf{e}_1$	$\lambda \mathbf{e}_2$
$:\mathbf{e}_4$	$:\tau \mathbf{e}_4$	$\varepsilon \mathbf{e}_5$	$\varepsilon \dots$	$\varepsilon \mathbf{e}_{m-1}$	$\varepsilon \mathbf{e}_m$	$\psi \lambda \mathbf{e}_1$	$\lambda \mathbf{e}_2$	$\lambda \mathbf{e}_3$
$:\mathbf{e}_5$	$\tau \mathbf{e}_5$	$\varepsilon \dots$	$\varepsilon \mathbf{e}_{m-1}$	$\varepsilon \mathbf{e}_m$	$\psi \lambda \mathbf{e}_1$	$\lambda \mathbf{e}_2$	$:\lambda \mathbf{e}_3$	$\lambda \mathbf{e}_4$
...	$:\tau \dots$	$\varepsilon \mathbf{e}_{m-1}$	$\varepsilon \mathbf{e}_m$	$\psi \lambda \mathbf{e}_1$	$\lambda \mathbf{e}_2$	$:\lambda \mathbf{e}_3$	$:\lambda \mathbf{e}_4$	$\lambda \dots$
$\mathbf{e}_{m-1}$	$\tau \mathbf{e}_{m-1}$	$\varepsilon \mathbf{e}_{m-1}$	$\psi \lambda \mathbf{e}_1$	$\lambda \mathbf{e}_2$	$\lambda \mathbf{e}_3$	$:\lambda \mathbf{e}_4$	$\lambda \dots$	$\lambda \mathbf{e}_{m-2}$
$\mathbf{e}_m$	$\tau \mathbf{e}_m$	$\psi \mathbf{e}_1$	$\lambda \mathbf{e}_2$	$:\lambda \mathbf{e}_3$	$:\lambda \mathbf{e}_4$	...	$\lambda \mathbf{e}_{m-2}$	$\lambda \mathbf{e}_{m-1}$

В статье [23] показано, что алгебры с операцией умножения, заданной по ТУБВ данного вида, при выполнении условия делимости числа  $p^s - 1$  на значение  $m$  при многих различных наборах значений независимых структурных констант  $\tau$ ,  $\varepsilon$  и  $\lambda$  являются конечными полями  $GF((p^s)^m)$ , которые будем называть векторными конечными полями. Распределение структурной константы  $\tau$  по ячейкам ТУБВ имеет отличительную специфику, которая состоит в том, что она присутствует в ряде ячеек как скалярный множитель  $\tau^{-1}$ . В качестве критерия формирования векторного конечного поля для заданного набора значений структурных констант в [23] использовано существование вектора, имеющего порядок  $p^{sm} - 1$ . Количество различных наборов значений структурных констант, при которых образуется алгебра, являющаяся полем, определяет количество различных модификаций векторного поля  $GF((p^s)^m)$ , определяемого заданным фиксированным распределением базисных векторов по ячейкам ТУБВ.

Для применения в рамках парадигмы [19] для разработки алгоритмов многомерной криптографии предполагается использование случайно заданной модификации векторного конечного поля, формируемого по ТУБВ с известными распределениями базисных векторов и структурных констант, в качестве элемента секретного ключа. Это определяет интерес к обеспечению возможности задания большого числа различных модификаций поля  $GF((p^s)^m)$ , а значит – к поиску ТУБВ с достаточно большим числом распределений различных независимых структурных констант. Число различных модификаций векторного поля  $GF((p^s)^m)$  можно оценить как  $O((p^s - 1)^k)$ , где  $O(\cdot)$  – обозначение порядка;  $k$  – число независимых структурных констант. Значение  $k$  учитывает как константы, влияющие на формирование векторного поля, так и константы, которые не влияют на формирование поля  $GF((p^s)^m)$  в том смысле, что изменение их значений при фиксированных значениях остальных констант не может привести к формированию векторного поля (если исходный набор значений констант не задавал формирование поля). Примером такого случая является константа  $\tau$  в таблице 1.

Если для заданного значения размерности  $m$  найдено распределение базисных векторов в ТУБВ, обеспечивающее коммутативность и ассоциативность операции умножения, то при малых значениях  $m$  можно эвристически-интуитивным путем найти распределения структурных констант, при которых сохраняются указанные свойства векторного умножения. Однако при больших значениях  $m$  эвристический способ становится неэффективным,

за исключением малого числа типовых случаев, иллюстрируемых, например, распределениями структурных констант  $\varepsilon$ ,  $\lambda$  и  $\tau$  в таблице 1. Далее будут предложены формализованные унифицированные методы построения ТУБВ с параметризуемым заданием распределений базисных векторов и структурных констант. При этом для фиксированного распределения базисных векторов задаются  $2(m - 1)$  различных распределений независимых структурных констант.

Коммутативность операции умножения, задаваемой некоторой ТУБВ со структурными константами, является необходимым, но не достаточным требованием обеспечения возможности формирования векторного конечного поля. Например, ТУБВ, представленная как таблица 2 (задает алгебру с единицей  $(0,0,0,\tau^{-1})$ ), не позволяет найти наборы значений констант  $\delta$ ,  $\lambda$ ,  $\mu$ , и  $\tau$ , при которых формируется поле  $GF((p^s)^m)$ , в том числе и при выполнении условия  $m|p^s - 1$ .

Таблица 2. Задание коммутативной операции умножение без возможности формирования векторного конечного поля

$\times$	$e_0$	$e_1$	$e_2$	$e_3$
$e_0$	$\delta\lambda\tau e_3$	$\delta e_2$	$\lambda e_1$	$\tau^{-1}e_0$
$e_1$	$\delta e_2$	$\delta\mu\tau e_3$	$\mu e_0$	$\tau^{-1}e_1$
$e_2$	$\lambda e_1$	$\mu e_0$	$\lambda\mu\tau e_3$	$\tau^{-1}e_2$
$e_3$	$\tau^{-1}e_0$	$\tau^{-1}e_1$	$\tau^{-1}e_2$	$\tau^{-1}e_3$

На данный момент нет теоретического критерия, по которому можно было бы установить возможность формирования векторных полей по заданной размерности и заданным распределениям базисных векторов и структурных констант в ТУБВ, определяющей коммутативную операцию умножения. В дальнейшем предполагается использование вычислительных экспериментов для ответа на этот вопрос.

Последние выполняются по следующему алгоритму:

1. Задается случайный набор значений структурных констант.
2. Осуществляется проверка наличия вектора, имеющего порядок  $p^{sm} - 1$ , т.е. наличия примитивного элемента поля. Если для  $k$  случайных векторов были получены значения порядка менее  $p^{sm} - 1$ , то перейти к шагу 3, иначе СТОП и вывести сообщение: «При текущем наборе значений структурных констант алгебра является полем  $GF((p^s)^m)$ ».
3. Если  $i < t$ , то сгенерировать новый случайный набор значений структурных констант прирастить счетчик  $i \leftarrow i + 1$  и перейти к шагу 2. В противном случае СТОП и вывести сообщение:

«При данных значениях  $m$ ,  $p$  и  $s$  и данных распределениях базисных векторов и структурных констант векторное конечное поле  $GF((p^s)^m)$  с вероятностью, близкой к 100%, не может быть сформировано».

Выводимое сообщение на шаге 2 (3) имеет нулевую (ненулевую) вероятность ошибки. Достаточно низкую вероятность ошибки вывода на шаге 3 можно обеспечить устанавливая сравнительно большие значения  $k$  и  $t$ , например,  $k = t = 100$ .

**3. Алгоритмы ЭЦП со скрытой группой.** Алгебраическими носителями алгоритмов ЭЦП со скрытой группой являются КНАА. С целью повышения производительности процедур генерации и верификации цифровой подписи могут использоваться прореженные ТУБВ. Примером последних является таблица 3. Для данного применения КНАА важным является знание их строения с точки зрения декомпозиции на множество коммутативных подалгебр. Изучение строения позволяет установить типы содержащихся в КНАА коммутативных мультипликативных групп и значения их порядка, а также получить формулы, описывающие все элементы коммутативной подалгебры по координатам некоторого ее представителя.

Для ряда четырехмерных КНАА, в том числе и алгебры, заданной по таблице 3 [22], строение детально изучено и установлена однотипность их строения. Последнее характеризуется следующими общими моментами:

1. Четырехмерная КНАА разбивается на  $p^2 + p + 1$  коммутативных подалгебр порядка  $p^2$ , относящихся к трем различным типам и попарно пересекающихся строго в множестве скалярных векторов.

2. Подалгебры первого типа являются полями, изоморфными полю  $GF(p^2)$ , и их мультипликативная группа  $\Gamma_1$  имеет циклическое строение и порядок, равный  $\Omega_1 = p^2 - 1$ .

3. Подалгебры второго типа содержат мультипликативную группу  $\Gamma_2$  порядка  $\Omega_2 = (p - 1)^2$ , которая имеет двухмерное циклическое строение (т. е. базис  $\Gamma_2$  включает два вектора одного и того же порядка  $p - 1$ ).

4. Подалгебры третьего типа содержат циклическую мультипликативную группу  $\Gamma_3$  порядка  $\Omega_3 = p(p - 1)$ .

Для числа подалгебр первого  $\eta_1$ , второго  $\eta_2$  и третьего  $\eta_3$  типов получены следующие формулы [22]:

$$\eta_1 = p(p - 1)/2; \quad (1)$$

$$\eta_2 = p(p + 1)/2; \tag{2}$$

$$\eta_3 = p + 1. \tag{3}$$

Из формул (1), (2) и (3) видно, что наибольший интерес для задания скрытой группы, являющейся элементом секретного ключа, представляют мультипликативные группы коммутативных подалгебр первого и второго типов, поскольку их число примерно в  $p$  раз больше числа подалгебр третьего типа. Можно предположить, что указанные моменты строения имеют место для всех четырехмерных КНАА, однако формальное доказательство этого факта не получено. Изучение строения КНАА с размерностями  $m \geq 6$  на данный момент остается открытым вопросом, который имеет существенную значимость для разработки алгоритмов ЭЦП на их основе.

Таблица 3. Задание четырехмерной КНАА по прореженной ТУБВ

$\times$	$\mathbf{e}_0$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$
$\mathbf{e}_0$	$\mu\mathbf{e}_0$	0	0	$\mu\mathbf{e}_3$
$\mathbf{e}_1$	0	$\delta\mathbf{e}_1$	$\delta\mathbf{e}_2$	0
$\mathbf{e}_2$	$\mu\mathbf{e}_2$	0	0	$\lambda\mu\mathbf{e}_1$
$\mathbf{e}_3$	0	$\delta\mathbf{e}_3$	$\delta\lambda\mathbf{e}_0$	0

Для генерации КНАА произвольных четных размерностей может быть применен унифицированный способ задания соответствующих ТУБВ, который характеризуется следующими двумя математическими формулами, описывающими вектор на пересечении  $i$ -й строки и  $j$ -го столбца:

$$\mathbf{e}_i\mathbf{e}_j = \begin{cases} \mathbf{e}_{i+j+d \bmod m} & (i \bmod 2 = 0); \\ \mathbf{e}_{i-j-d \bmod m} & (i \bmod 2 = 1, j \bmod 2 = 0); \\ \lambda\mathbf{e}_{i-j-d \bmod m} & (i \bmod 2 = 1, j \bmod 2 = 1). \end{cases} \tag{4}$$

$$\mathbf{e}_i\mathbf{e}_j = \begin{cases} \mathbf{e}_{i-j-d \bmod m} & (i \bmod 2 = 0, j \bmod 2 = 1); \\ \lambda\mathbf{e}_{i-j-d \bmod m} & (i \bmod 2 = 0, j \bmod 2 = 0); \\ \mathbf{e}_{i+j+d \bmod m} & (i \bmod 2 = 1). \end{cases} \tag{5}$$

В данном способе распределение базисных векторов является параметризуемым, т.е. зависит от выбранного значения параметра  $d$

$(0 \leq d \leq m - 1)$  и его четности (для четных  $d$  используется формула (4), для нечетных  $d$  – формула (5)). Параметризация распределения структурных констант не реализуется, однако это не имеет существенного значения для построения схем ЭЦП со скрытой группой. Существенными представляются возможность формализованного задания КНАА произвольных четных размерностей и знание их строения. Ввиду отсутствия на данный момент результатов изучения строения КНАА для случая  $m \geq 6$  можно применить подход, связанный с экспериментальным определением возможных значений порядка векторов.

Рассмотрим построение алгоритма ЭЦП со скрытой группой, основанного на вычислительной трудности скрытой задачи дискретного логарифмирования (СЗДЛ) в четырехмерной КНАА, заданной по таблице 3 над полем  $GF(p)$ , где простое  $p = 2q + 1$  при 256-битном простом  $q$ . Определим формирование открытого ключа по следующему алгоритму:

1. Сгенерировать случайный базис  $\langle \mathbf{G}, \mathbf{Q} \rangle$  (где  $\mathbf{G}$  и  $\mathbf{Q}$  – векторы порядка  $q$ ) скрытой коммутативной группы  $\Gamma_{\langle \mathbf{G}, \mathbf{Q} \rangle}$  с двухмерной циклическостью ( $\Gamma_{\langle \mathbf{G}, \mathbf{Q} \rangle}$  – подгруппа группы типа  $\Gamma_2$ ).
2. Сгенерировать случайные обратимые векторы  $\mathbf{A}$  и  $\mathbf{B}$  порядка  $p^2 - 1$ , такие, что  $\mathbf{AB} \neq \mathbf{BA}$ , и случайное натуральное число  $x < q$ .
3. Вычислить открытый ключ (размером 96 байт) в виде тройки 32-байтовых векторов  $\mathbf{U}$ ,  $\mathbf{Y}$  и  $\mathbf{Z}$  по секретным значениям  $\mathbf{G}$ ,  $\mathbf{Q}$ ,  $\mathbf{A}$ ,  $\mathbf{B}$  и  $x$ , составляющим 136-байтовый секретный ключ:

$$\mathbf{U} = \mathbf{AG}^x \mathbf{B}^{-1}; \mathbf{Y} = \mathbf{BGB}^{-1} \text{ и } \mathbf{Z} = \mathbf{BQA}^{-1}.$$

Алгоритм генерации ЭЦП к документу  $M$  использует 256-битную хэш-функцию  $f_H$  и включает следующие шаги:

1. Выбрать два случайных натуральных числа  $k < q$  и  $t < q$  и вычислить вектор  $\mathbf{V} = \mathbf{AG}^k \mathbf{Q}^t \mathbf{A}^{-1}$  и первый элемент подписи в виде числа  $e = f_H(M, \mathbf{V})$ .
2. Вычислить второй элемент  $s$  ЭЦП как решение квадратного уравнения  $es^2 - s + xt + t = k \pmod q$  с неизвестным  $s$ . Если это уравнение не имеет решений, то перейти к шагу 1.
3. Вычислить третий элемент  $d$  ЭЦП:  $d = (s^{-1}t - 1) \pmod q$ .
4. Верификация ЭЦП осуществляется с использованием открытого ключа  $(\mathbf{U}, \mathbf{Y}, \mathbf{Z})$  по алгоритму:

1. Вычислить вектор  $\mathbf{V}'$ :

$$\mathbf{V}' = \left( \mathbf{U} \circ \mathbf{Y}^{es} \circ \mathbf{Z} \circ (\mathbf{U} \circ \mathbf{Y} \circ \mathbf{Z})^d \right)^s.$$

2. Если  $e' = f_H(M, \mathbf{V}') = e$ , то ЭЦП верна, иначе подпись ложная.

**4. Второй тип схем ЭЦП со скрытой группой.** Представляет интерес парадигма построения алгоритмов ЭЦП, основанных на трудности решения больших систем степенных уравнений и использующих векторное проверочное уравнение с многократным вхождением подписи. В некоммутативных алгебрах решение уравнений с многократным вхождением неизвестной и со случайными параметрами является вычислительно трудной задачей. Для того, чтобы владелец открытого ключа мог вычислить значение подписи, требуется специальным способом сформировать открытый ключ и соответствующим способом задать механизм рандомизации подписи. Основные моменты реализации алгоритмов указанного типа иллюстрируются следующей схемой ЭЦП, в которой алгебраическим носителем служит четырехмерная КНАА, использованная в алгоритме из предыдущего раздела.

*Личный секретный ключ* генерируется владельцем открытого ключа в виде четырех случайных векторов  $\mathbf{A}, \mathbf{B}, \mathbf{G}, \mathbf{H}$  и натурального числа  $x$  ( $1 < x < q$ ), где векторы  $\mathbf{G}$  и  $\mathbf{H}$  порядка  $q$  составляют базис  $\langle \mathbf{G}, \mathbf{H} \rangle$  скрытой группы;  $\mathbf{A}$  и  $\mathbf{B}$  – обратимые векторы порядка  $p^2 - 1$  или  $(p - 1)^2$ , удовлетворяющие условиям  $\mathbf{AB} \neq \mathbf{BA}$ ,  $\mathbf{AG} \neq \mathbf{GA}$ ,  $\mathbf{BG} \neq \mathbf{GB}$ .

*Открытый ключ* вычисляется в виде набора векторов  $\mathbf{Y}, \mathbf{Z}$  и  $\mathbf{U}$  по следующим формулам:

$$\mathbf{Y} = \mathbf{AGB}, \mathbf{Z} = \mathbf{AG}^x\mathbf{B} \text{ и } \mathbf{U} = \mathbf{AHB}. \quad (6)$$

*Процедура генерации ЭЦП* к электронному документу  $M$  включает следующие шаги:

1. Выбрать случайные натуральные числа  $k$  и  $t$  ( $1 < k < q$ ;  $1 < t < q$ ) и вычислить рандомизирующий вектор-фиксатор:

$$\mathbf{R} = \mathbf{B}^{-1} \mathbf{G}^k \mathbf{H}^t \mathbf{B}. \quad (7)$$

2. Используя 512-битную коллизивно стойкую хэш-функцию  $f_H$ , вычислить 512-битный рандомизирующий элемент подписи в виде хэш-значения  $e = e_1 || e_2 = f_H(M || \mathbf{R})$ , представленного как конкатенация двух 256-битных чисел  $e_1$  и  $e_2$ .

3. Вычислить натуральные значения  $n$  и  $d$ :

$$n = \frac{k - e_1 - e_1^2 - xe_1}{e_1(e_1 + e_2 + 1)} \bmod q;$$

$$d = \frac{t - e_1e_2}{e_1(e_1 + e_2 + 1)} \bmod q.$$

4. Подгоночный элемент подписи вычисляется в виде четырехмерного вектора  $\mathbf{S}$  по формуле:

$$\mathbf{S} = \mathbf{B}^{-1} \mathbf{G}^n \mathbf{H}^d \mathbf{A}^{-1}. \quad (8)$$

*Процедура верификации ЭЦП* выполняется по следующему алгоритму:

1. Вычислить контрольный четырехмерный вектор  $\mathbf{R}'$  по формуле:

$$\mathbf{R}' = \left( (\mathbf{S}\mathbf{Y})^{e_1} \mathbf{S}(\mathbf{U}\mathbf{S})^{e_2} \mathbf{Z}\mathbf{S}\mathbf{Y} \right)^{e_1}. \quad (9)$$

2. Вычислить значение хэш-функции  $e' = f_H(M \parallel \mathbf{R}')$ .

3. Если  $e' = e$ , то ЭЦП признается подлинной, в противном случае подпись отклоняется как ложная.

Постквантовая стойкость этого алгоритма ЭЦП обеспечивается тем, что вычисление секретного ключа по открытому ключу требует нахождения решения системы из следующих пяти квадратных векторных уравнений с неизвестными  $\mathbf{A}$ ,  $\mathbf{B}^{-1}$ ,  $\mathbf{G}$ ,  $\mathbf{G}_x$  ( $\mathbf{G}_x = \mathbf{G}^x$ ) и  $\mathbf{H}$ :

$$\mathbf{Y}\mathbf{B}^{-1} = \mathbf{A}\mathbf{G}, \mathbf{Z}\mathbf{B}^{-1} = \mathbf{A}\mathbf{G}_x, \mathbf{U}\mathbf{B}^{-1} = \mathbf{A}\mathbf{H}, \mathbf{G}\mathbf{G}_x = \mathbf{G}_x\mathbf{G}, \mathbf{G}\mathbf{H} = \mathbf{H}\mathbf{G}. \quad (10)$$

Система уравнений (10) сводится к системе из 20 квадратных уравнений с 20 неизвестными в поле  $GF(p)$ , порядок которого равен 257-битному простому числу. Однако оценка стойкости как вычислительной трудности решения системы (1) является достаточно прямолинейной. При более внимательном рассмотрении описанного алгоритма можно заметить, что в нем используется несколько ограниченный механизм рандомизации подписи, реализуемый формулами (7) и (8). Ограниченность заключается в том, что при

заданном фиксированном секретном ключе векторы  $\mathbf{R}$  и  $\mathbf{S}$  могут принимать только  $\approx p^2$  различных значений из  $\approx p^4$  обратимых векторных значений в КНАА, использованной в качестве алгебраического носителя.

Нетрудно заметить, что при наличии многих различных подписанных документов имеется много различных векторов  $\mathbf{S}_i$ , ( $i = 1, 2, 3, \dots$ ) а также можно вычислить для каждого значения  $\mathbf{S}_i$  соответствующее ему значение  $\mathbf{R}_i$ . Это позволяет построить систему уравнений с существенно меньшим числом векторных уравнений (и неизвестных) по сравнению с системой (10).

Легко видеть, что имея  $k$  подлинных подписей к некоторым документам, из проверочного уравнения (9) можно вычислить  $k$  рандомизирующих векторов  $\mathbf{R}_i$ , каждый из которых по формуле (7) задает уравнение с неизвестным вектором  $\mathbf{V}$  и неизвестным вектором  $\mathbf{Q}_i = \mathbf{G}^{k_i} \mathbf{H}^{t_i}$ , принадлежащим скрытой группе. Последнее означает, что выбирая  $\mathbf{Q}_1$  как неизвестный вектор с четырьмя неизвестными координатами, неизвестные  $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_i$  являются связанными с вектором  $\mathbf{Q}_1$ , а именно, они выбираются из коммутативной подалгебры, представителем которой является вектор  $\mathbf{Q}_1$ . В работе [22] приводится формула, описывающая через координаты вектор  $\mathbf{Q}_1$  и две скалярные переменные все  $p^2$  векторов, принадлежащие указанной подалгебре. Таким образом, уравнение  $\mathbf{R}_1 = \mathbf{V}^{-1} \mathbf{G}^{k_1} \mathbf{H}^{t_1} \mathbf{V}$  задает 8 неизвестных скалярных значений, а каждое из остальных  $k-1$  уравнений вида  $\mathbf{R}_i = \mathbf{V}^{-1} \mathbf{G}^{k_i} \mathbf{H}^{t_i} \mathbf{V}$  добавляет только две независимые скалярные неизвестные. В случае  $k=3$  имеем систему из 3 векторных квадратных уравнений, сводимую к системе из 12 скалярных кубических уравнений с 12 скалярными неизвестными, из которой можно определить координаты неизвестного вектора  $\mathbf{V}$  и координаты одного из векторов, задающих скрытую группу.

С учетом результатов исследования строения использованной в качестве алгебраического носителя четырехмерной КНАА, а именно возможности использования формулы с двумя скалярными переменными, которая описывает все элементы подалгебры, содержащей скрытую группу, в системе (10) последние два векторных уравнения могут быть устранены добавлением условия, что неизвестные  $\mathbf{G}_x$  и  $\mathbf{H}$  принадлежат подалгебре, которая содержит неизвестный вектор  $\mathbf{G}$ . При преобразовании системы векторных уравнений в систему скалярных уравнений указанное условие даст 12 скалярных уравнений с 16 неизвестными скалярными значениями.

Если координаты вектора  $\mathbf{B}$  известны, то число скалярных неизвестных уменьшается до 12.

Таким образом, неполная рандомизация потенциально может быть использована для существенного снижения сложности алгоритма взлома описанной схемы ЭЦП и актуальным является нахождение механизмов обеспечения полной рандомизации подписи, которые дадут возможность векторам  $\mathbf{S}$  и  $\mathbf{R}$  принимать все (или почти все) возможные обратимые значения в используемой КНАА. При этом в рассмотрении механизма такого снижения стойкости существенно использовались результаты исследования строения использованной КНАА. Это показывает, что последние важны не только для разработки алгоритмов ЭЦП со скрытой группой, но и для оценки стойкости.

Для обеспечения полной рандомизации подписи можно задать вычисление подгоночного элемента подписи по формуле:

$$\mathbf{S} = \mathbf{D}\mathbf{G}^n\mathbf{H}^d\mathbf{V}, \quad (11)$$

где  $\mathbf{V}$  – случайный обратимый вектор. Однако это приводит к проблемам с использованием проверочных уравнений с многократным входением подгоночного элемента подписи  $\mathbf{S}$ , которое является важным моментом стойкости к атакам с использованием  $\mathbf{S}$  как подгоночного параметра алгоритма подделки ЭЦП. При применении формулы (11) для вычисления элемента подписи  $\mathbf{S}$  вместо многократного входения  $\mathbf{S}$  в проверочное уравнение можно использовать два различных проверочных уравнения с однократным входением  $\mathbf{S}$ , например, таких:

$$\begin{aligned} \mathbf{R}_1' &= \mathbf{Y}_1 e' \mathbf{T}_1 \mathbf{Z}_1^{es} \mathbf{U}_1 \mathbf{S}, \\ \mathbf{R}_2' &= \mathbf{Y}_2 e' \mathbf{T}_2 \mathbf{Z}_2^{es} \mathbf{U}_2 \mathbf{S}, \end{aligned} \quad (12)$$

где  $e$  и  $e'$  – рандомизирующие элементы подписи;  $\mathbf{Y}_1$ ,  $\mathbf{T}_1$ ,  $\mathbf{Z}_1$ ,  $\mathbf{Y}_2$ ,  $\mathbf{T}_2$ , и  $\mathbf{Z}_2$  – элементы открытого ключа, зависящие от секретного ключа (в виде набора векторов  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{D}$ ,  $\mathbf{N}$ ,  $\mathbf{P}$ ,  $\mathbf{G}$ ,  $\mathbf{H}$  и числа  $x$ ) и вычисляемые по формулам:

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{A}\mathbf{G}^x\mathbf{A}^{-1}; \mathbf{T}_1 = \mathbf{A}\mathbf{G}\mathbf{B}^{-1}; \mathbf{Z}_1 = \mathbf{B}\mathbf{H}\mathbf{B}^{-1}; \mathbf{U}_1 = \mathbf{B}\mathbf{G}\mathbf{D}^{-1}; \\ \mathbf{Y}_2 &= \mathbf{N}\mathbf{H}^x\mathbf{N}^{-1}; \mathbf{T}_2 = \mathbf{N}\mathbf{H}\mathbf{P}^{-1}; \mathbf{Z}_2 = \mathbf{P}\mathbf{G}\mathbf{P}^{-1}; \mathbf{U}_2 = \mathbf{P}\mathbf{H}\mathbf{D}^{-1}, \end{aligned} \quad (13)$$

где  $\mathbf{G}$  и  $\mathbf{H}$  – векторы, образующие базис скрытой группы. При удвоенном проверочном уравнении (12) процедура генерации подписи в виде четверки значений  $(e, e', s, \mathbf{S})$  включает следующие шаги:

1. Сгенерировать случайные натуральные числа  $k_1, k_2, t_1$  и вычислить число  $t_2 = t_1 + k_1 - k_2 \bmod q$ . Затем сгенерировать случайный обратимый вектор  $\mathbf{V}$  и вычислить векторы-фиксаторы  $\mathbf{R}_1 = \mathbf{A}\mathbf{G}^{k_1+2} \mathbf{H}^{t_1} \mathbf{V}$  и  $\mathbf{R}_2 = \mathbf{N}\mathbf{G}^{k_2} \mathbf{H}^{t_2+2} \mathbf{V}$ .

2. Используя 512-битную хэш-функцию  $f_H$ , вычислить 512-битный рандомизирующий элемент подписи в виде хэш-значения  $e \| e' = f_H(M \| \mathbf{R}_1 \| \mathbf{R}_2)$ , представленного как конкатенация двух 256-битных чисел  $e$  и  $e'$ .

3. Вычислить числа  $n = k_1 - e'x \bmod q$ ;  $d = t_2 - e'x \bmod q$   $s = e^{-1}(t_1 - t_2 + e'x) \bmod q$  и вектор  $\mathbf{S} = \mathbf{D}\mathbf{G}^n \mathbf{H}^d \mathbf{V}$ .

Корректность данной схемы ЭЦП легко доказывается, подавая на вход алгоритма верификации значения подписи, сформированной в соответствии с процедурой генерации подписи.

Заметим, что подгоночный элемент подписи  $\mathbf{S}$  и каждый из векторов-фиксаторов  $\mathbf{R}_1$  и  $\mathbf{R}_2$  может принимать любое обратимое значение из КНАА, используемой в качестве алгебраического носителя. Тем не менее, по формулам (11) и (12) и известным подлинным подписям можно составить систему уравнений, по которой можно будет вычислить элементы секретного ключа  $\mathbf{D}$  и  $\mathbf{N}$ , однако это не приводит к снижению стойкости по отношению к атаке, реализующей решение системы квадратных уравнений, составленной из формул (13), связывающих элементы открытого ключа с элементами секретного ключа.

Последнее обусловлено тем, что система, составленная по формулам по формулам (11) и (12), включает значительно больше уравнений за счет того, что каждая тройка уравнений, связанная с одной подписью включает дополнительное векторное неизвестное – уникальное значение вектора  $\mathbf{V}$ . Для получения оценки числа подписей, которые необходимы для составления системы кубических скалярных уравнений, в которой число уравнений равно числу скалярных неизвестных, следует учесть, что каждая подпись задает следующие три векторных уравнения  $\mathbf{S} = \mathbf{D}\mathbf{G}^n \mathbf{H}^d \mathbf{V}$ ,  $\mathbf{R}_1' = \mathbf{A}\mathbf{G}^{k_1+2} \mathbf{H}^{t_1} \mathbf{V}$  и  $\mathbf{R}_2' = \mathbf{N}\mathbf{G}^{k_2} \mathbf{H}^{t_2+2} \mathbf{V}$ , где векторы  $\mathbf{R}_1'$  и  $\mathbf{R}_2'$  вычисляются по проверочным уравнениям (12).

Последние три уравнения представим в виде  $\mathbf{S} = \mathbf{D}\mathbf{G}_1 \mathbf{V}$ ,  $\mathbf{R}_1' = \mathbf{A}\mathbf{G}_2 \mathbf{V}$  и  $\mathbf{R}_2' = \mathbf{N}\mathbf{G}_3 \mathbf{V}$ , где неизвестные векторы  $\mathbf{G}_1 = \mathbf{G}^n \mathbf{H}^d$ ,  $\mathbf{G}_2 = \mathbf{G}^{k_1+2} \mathbf{H}^{t_1}$  и  $\mathbf{G}_3 = \mathbf{G}^{k_2} \mathbf{H}^{t_2+2}$  принадлежат скрытой группе и являются

уникальными неизвестными для каждой подписи, также как и векторная неизвестная  $\mathbf{V} = (v_0, v_1, v_2, v_3)$ . Принадлежность скрытой группе позволяет записать координаты каждого из векторов  $\mathbf{G}_1$ ,  $\mathbf{G}_2$  и  $\mathbf{G}_3$  через координаты фиксированного вектора  $\mathbf{G}_0$ , принадлежащего скрытой группе, и пару скалярных неизвестных (формула (8) в [22]). Последнее определяет 6 уникальных скалярных неизвестных. С учетом уникальных скалярных неизвестных  $v_0, v_1, v_2$  и  $v_3$  устанавливаем, что каждая известная подпись задает 12 скалярных уравнений, 10 уникальных скалярных неизвестных. При этом 16 фиксированных неизвестных (координаты векторных неизвестных  $\mathbf{A}$ ,  $\mathbf{D}$  и  $\mathbf{N}$ ) присутствуют в уравнениях, задаваемых каждой подписью. Для  $k$  известных подлинных подписей имеем систему из  $12k$  скалярных уравнений, включающую  $10k + 12$  скалярных неизвестных.

Решение уравнения  $12k = 10k + 12$  дает значение  $k = 6$  при котором число уравнений и число неизвестных равно 72, тогда как система скалярных кубических уравнений, составленная по формулам (13) включает всего 32 уравнения с 38 скалярными неизвестными (недоопределенность данной системы показывает, что имеются много различных решений, задающих класс эквивалентных секретных ключей).

**5. Трудно обратимые отображения в векторных конечных полях.** В алгоритмах многомерной криптографии открытый ключ  $P$  представляет собой трудно обратимое отображение ( $n$ -мерных векторов в  $u$ -мерное векторное пространство) с секретной лазейкой, заданное в виде набора из  $u$  многочленов второй или более высокой степени с коэффициентами и  $n$  переменными (координатами вектора-прообраза), принимающими значения в поле  $GF(p^s)$ . Значения многочленов задают координаты  $u$ -мерного вектора-образа. Шифрование по ключу  $P$  выполняется путем представления сообщения в виде  $n$ -мерного вектора  $\mathbf{M}$  и вычисления  $u$ -мерного вектора-образа  $\mathbf{C} = P(\mathbf{M})$ . Однозначность расшифровывания шифртекста  $\mathbf{C}$  обеспечивается неравенством  $u \geq n$ .

Прямой атакой на алгоритмы такого типа является решение системы уравнений, задаваемой шифртекстом  $\mathbf{C}$  и набором многочленов  $P$ . В настоящее время разработаны достаточно эффективные способы решения этой задачи [24]. В зависимости от требуемого уровня стойкости выбираются значения  $u$ ,  $n$  и  $p^s$ . При этом вычислительная сложность прямой атаки  $W$  на основе лучших известных способов решения больших систем степенных уравнений экспоненциально зависит от числа уравнений и относительно слабо зависит от порядка поля  $GF(p^s)$  в котором задана система. Таблица 4,

отражающая обобщающие результаты статьи [25], иллюстрирует этот факт для случая  $u = n$ .

Таблица 4. Минимальное число уравнений, обеспечивающее заданный уровень стойкости  $W$  к прямой атаке [25]

Порядок поля $GF(p^s)$	$W = 2^{80}$	$W = 2^{100}$	$W = 2^{128}$	$W = 2^{192}$	$W = 2^{256}$
$p^s = 16$	30	39	51	80	110
$p^s = 31$	28	36	48	75	103
$p^s = 256$	26	33	43	68	93

Для формирования открытого ключа  $P$  обычно составляется некоторый набор многочленов над полем  $GF(p^s)$  малого порядка, задающий нелинейное отображение  $N$ , для которого легко задать процедуру, выполняющую обратное отображение  $N^{-1}$ . Последнее не может служить потайной лазейкой, поэтому для получения отображение, пригодного для использования в качестве  $P$ , выполняется маскирование  $N$  путем преобразования набора многочленов  $N$  в другой набор многочленов, описывающих суперпозицию  $N$  с одним или двумя линейными отображениями. Последние легко представимы в виде набора многочленов первой степени, поэтому результирующий набор многочленов  $P$  легко вычисляется и имеет ту же степень, что и многочлен  $N$ . При этом отображение  $P$  является трудно обратимым, если не знать строение  $P$  как суперпозиции, например,  $P(\mathbf{M}) = L_2(N(L_1(\mathbf{M})))$  для случая двух маскирующих линейных отображений  $L_1$  и  $L_2$ . Для последних легко вычисляются соответствующие обратные отображения  $L_1^{-1}$  и  $L_2^{-1}$ , а секретной лазейкой служит суперпозиция отображений  $P^{-1}(\mathbf{C}) = L_1^{-1}(N^{-1}(L_2^{-1}(\mathbf{C}))) = \mathbf{M}$ .

Атаки, связанные с использованием особенностей формирования открытого ключа  $P$ , например, с поиском эквивалентных представлений  $P$  в виде суперпозиции линейных отображений  $L_1'$  и  $L_2'$  и сравнительно легко обратимого отображения  $N'$  называются структурными атаками. Различные варианты структурных атак на алгоритмы многомерной криптографии представлены в работах [25, 26]. Для обеспечения стойкости к структурным атакам маскирующие отображения  $L_1$  и  $L_2$  задаются наборами многочленов с достаточно большим числом слагаемых, что приводит к большому размеру открытого ключа.

Устранение этого недостатка в рамках парадигмы [19] обеспечивается тем, что нелинейное отображение задается

с использованием операций экспоненцирования в  $m$ -мерных векторных конечных полях  $GF((p^s)^m)$ . Последние являются конечными алгебрами, в которых указанные операции легко задаются наборами многочленов над полем  $GF(p^s)$ , над которым заданы сами алгебры. Обратной по отношению к операции экспоненцирования в поле  $GF((p^s)^m)$  является операция извлечения корня соответствующей степени  $z$  и для многих случаев (которые легко могут быть заданы) выполняется как операция возведения в степень  $Z = z^{-1} \bmod \Omega$ , где  $\Omega = p^{sm} - 1$  – порядок мультипликативной группы поля. Выполнение операции извлечения корня даже в случае малых степеней  $z$  вычислительно непредставима в виде набора многочленов, но может быть легко выполнена при знании всех параметров задания конкретной модификации векторного поля, в котором выполнялась соответствующая операция экспоненцирования.

В парадигме [19] предполагается такое задание набора многочленов открытого ключа  $P$ , из которого вычислительно трудно установить параметры задания (одного или многих) модификаций векторных конечных полей, использованных владельцем открытого ключа для задания нелинейного трудно обратимого отображения  $N$ . При этом обратное отображение  $N^{-1}$  легко реализуется посредством выполнения одного или многих операций извлечения корня в известных модификациях векторных конечных полей, а секретной лазейкой является знание указанных модификаций. В данном подходе структурные атаки связаны с вычислением параметров задания модификаций векторных конечных полей по многочленам открытого ключа. Структурные атаки данного типа оказываются связанными с решением больших систем степенных уравнений. В отличие от систем уравнений, решаемых в случае прямых атак, в случае указанных структурных атак уравнения содержат сравнительно мало слагаемых, однако число неизвестных и значения степеней значительно больше.

Для маскирования параметров задания векторных конечных полей достаточно использование линейных отображений, не приводящих к увеличению размера открытого ключа (например, перестановки координат преобразуемых векторов). При построении нелинейного отображения  $N$  при использовании нескольких операций экспоненцирования возможны различные структуры (топологии) отображения  $N$ , выбираемые в зависимости от параметров разрабатываемого криптографического алгоритма и параметров используемых векторных полей. Частные варианты топологий представлены в статьях [19, 20].

**6. Формализованный способ задания векторных конечных полей с большим числом реализуемых модификаций.** Рассмотрим унифицированный способ генерации ТУБВ, определяющих коммутативную ассоциативную операцию умножения  $m$ -мерных векторов для размерностей  $m = 2^k - 1$ , где  $k$  – натуральное число. В данном способе предполагается, что ТУБВ генерируется по следующей формуле, в которой индексы  $i$  и  $j$  трактуются как двоичные многочлены, представленные битовыми строками, двоичные значения которых равны числам  $i$  и  $j$  (номерам базисных векторов  $1 \leq i, j \leq m$ ):

$$\mathbf{e}_i \mathbf{e}_j = \mathbf{e}_{ijd \bmod f(x)}, \quad (14)$$

где  $d$  – ненулевой двоичный многочлен не выше степени  $k - 1$ ; операция умножения двоичных многочленов  $i, j$  и  $d$  выполняется по модулю неприводимого двоичного многочлена  $f(x)$  степени  $k$ , например,  $f(x) = x^3 + x^2 + 1$  при  $k = 3$ . Параметр  $d$  задает вид распределения базисных векторов. Число различных значений  $d$  равно порядку мультипликативной группы поля  $GF(2^k)$ , а именно, значению  $\Omega = 2^k - 1$ , и формула (14) задает  $\Omega$  различных распределений базисных векторов для фиксированного значения  $m$ .

Параметризация распределений структурных констант в ТУБВ, задаваемых формулой (14), может быть обеспечена внесением условий установки констант, приводящих к следующей унифицированной формуле с трехфакторной параметризацией (параметризации по значениям: размерности  $m$ , параметра  $d$  и параметра  $t$ ):

$$\mathbf{e}_i \mathbf{e}_j = \begin{cases} \delta \mathbf{e}_{ijd \bmod f(x)}, & \text{если } \left( t(\tilde{i} + \tilde{d}) \bmod \Omega \right) + \left( t(\tilde{j} + \tilde{d}) \bmod \Omega \right) < \Omega \\ \lambda \mathbf{e}_{ijd \bmod f(x)}, & \text{если } \left( t(\tilde{i} + \tilde{d}) \bmod \Omega \right) + \left( t(\tilde{j} + \tilde{d}) \bmod \Omega \right) \geq \Omega \end{cases}, \quad (15)$$

где  $t$  – натуральное число ( $1 \leq t < \Omega$ ) задающее распределение независимых структурных констант  $\delta$  и  $\lambda$ ;  $\tilde{i}, \tilde{j}, \tilde{d}$  – индексы (дискретные логарифмы) двоичных многочленов  $i, j, d$  по модулю  $f(x)$  при основании, равном одному из примитивных элементов поля  $GF(2^k)$  с умножением по модулю  $f(x)$ . (Заметим, что при  $t = \Omega$  задается тривиальное распределение структурной константы  $\delta$ , состоящее в ее наличии в каждой клетке ТУБВ.)

**Утверждение 1.** Формула (14) для произвольных значений из области определения параметров  $d$  и  $k$  генерирует ТУБВ с коммутативной и ассоциативной операцией векторного умножения.

*Доказательство.* Свойство коммутативности непосредственно следует из коммутативности операции умножения в поле двоичных многочленов  $GF(2^k)$ . Докажем, что свойство ассоциативности тоже имеет место. Для произвольных трех векторов  $\mathbf{A}$ ,  $\mathbf{B}$  и  $\mathbf{C}$  в соответствии с определением операции векторного умножения имеем:

$$(\mathbf{AB})\mathbf{C} = \sum_{i,j,h=1}^m a_i b_j c_h (\mathbf{e}_i \mathbf{e}_j) \mathbf{e}_h; \quad \mathbf{A}(\mathbf{BC}) = \sum_{i,j,h=1}^m a_i b_j c_h \mathbf{e}_i (\mathbf{e}_j \mathbf{e}_h).$$

Следовательно, если для всевозможных троек базисных векторов  $\mathbf{e}_i$ ,  $\mathbf{e}_j$  и  $\mathbf{e}_h$  выполняется равенство  $(\mathbf{e}_i \mathbf{e}_j) \mathbf{e}_h = \mathbf{e}_i (\mathbf{e}_j \mathbf{e}_h)$ , то векторное умножение ассоциативно, поскольку в этом случае имеем  $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ . В соответствии с формулой (14) получаем:

$$\begin{aligned} (\mathbf{e}_i \mathbf{e}_j) \mathbf{e}_h &= \mathbf{e}_{ijd \bmod f(x)} \mathbf{e}_h = \mathbf{e}_{ijhd^2 \bmod f(x)}; \\ \mathbf{e}_i (\mathbf{e}_j \mathbf{e}_h) &= \mathbf{e}_i \mathbf{e}_{jhd \bmod f(x)} = \mathbf{e}_{ijhd^2 \bmod f(x)} = (\mathbf{e}_i \mathbf{e}_j) \mathbf{e}_h. \end{aligned}$$

**Утверждение 2.** Формула (15) для произвольных значений из области определения параметров  $d$  и  $k$  генерирует ТУБВ, задающую конечную алгебру с глобальной двухсторонней единицей  $\mathbf{U} = (u_1, u_2, \dots, u_\Omega)$ , где все координаты равны нулю, кроме координаты  $u_{d^{-1} \bmod f(x)} = \delta^{-1}$ .

*Доказательство.* Умножая произвольный вектор  $\mathbf{A}$  на  $\mathbf{U}$ , легко показывается выполнимость равенств  $\mathbf{AU} = \mathbf{A}$  и  $\mathbf{UA} = \mathbf{A}$ .

**Теорема.** Формула (15) для произвольных значений из области определения параметров  $d$ ,  $k$  и  $t$  генерирует ТУБВ, задающую коммутативную и ассоциативную операцию векторного умножения.

*Доказательство.* Свойство коммутативности дано в предыдущем доказательстве. С учетом доказанного утверждения 1 и формулы (15) имеем  $(\mathbf{e}_i \mathbf{e}_j) \mathbf{e}_h = \psi_1 \mathbf{e}_u$  и  $\mathbf{e}_i (\mathbf{e}_j \mathbf{e}_h) = \psi_2 \mathbf{e}_u$  при некотором двоичном многочлене  $u$ . Докажем, что во всех случаях имеем  $\psi_1 = \psi_2$ , т.е. свойство ассоциативности векторного умножения имеет место.

Определим переменные  $i'$ ,  $j'$  и  $h'$  следующим образом:

$$i' = t(\tilde{i} + \tilde{d}) \bmod \Omega; \quad j' = t(\tilde{j} + \tilde{d}) \bmod \Omega; \quad h' = t(\tilde{h} + \tilde{d}) \bmod \Omega.$$

Представим формулу (15) в следующем виде:

$$\mathbf{e}_i \mathbf{e}_j = \begin{cases} \delta \mathbf{e}_{ijd \bmod f(x)}, & \text{если } i' + j' < \Omega \\ \lambda \mathbf{e}_{ijd \bmod f(x)}, & \text{если } i' + j' \geq \Omega \end{cases} \quad (16)$$

Можно легко показать, что умножение вектора  $(\mathbf{e}_i \mathbf{e}_j)$  на базисный вектор  $\mathbf{e}_h$  вносит структурную константу  $\lambda$ , если  $(i' + j') \bmod \Omega + h' \geq \Omega$ , или структурную константу  $\delta$ , если  $(i' + j') \bmod \Omega + h' < \Omega$ . Умножение базисного вектора  $\mathbf{e}_i$  на вектор  $(\mathbf{e}_j \mathbf{e}_h)$  вносит структурную константу  $\lambda$ , если  $i' + (j' + h') \bmod \Omega \geq \Omega$ , или структурную константу  $\delta$ , если  $i' + (j' + h') \bmod \Omega < \Omega$ .

Имеем следующие четыре случая.

*Случай 1.* Значения  $i'$ ,  $j'$  и  $h'$  удовлетворяют условию  $i' + j' + h' < \Omega$ . Тогда  $i' + j' < \Omega$  и  $j' + h' < \Omega$  и из (16) следует  $\psi_1 = \delta^2$  и  $\psi_2 = \delta^2 = \psi_1$ .

*Случай 2.* Значения  $i'$ ,  $j'$  и  $h'$  удовлетворяют условиям  $i' + j' < \Omega$  и  $(i' + j') \bmod \Omega + h' \geq \Omega$ . Тогда из (16) следует  $\psi_1 = \delta\lambda$ . Для вычисления  $\psi_2$  рассмотрим подслучаи 2.1 и 2.2:

*Подслучай 2.1.* Имеет место неравенство  $j' + h' < \Omega$  (произведение  $\mathbf{e}_j \mathbf{e}_h$  дает множитель  $\delta$ ). Тогда имеем  $i' + (j' + h') \bmod \Omega \geq \Omega$  (умножение  $\mathbf{e}_i$  на  $(\mathbf{e}_j \mathbf{e}_h)$  дает множитель  $\lambda$ ) и  $\psi_2 = \delta\lambda = \psi_1$ .

*Подслучай 2.2.* Имеет место неравенство  $j' + h' \geq \Omega$  (произведение  $\mathbf{e}_j \mathbf{e}_h$  дает множитель  $\lambda$ ). Тогда имеем  $(j' + h') \bmod \Omega = j' + h' - \Omega$  и из условия  $i' + j' < \Omega$  следует  $i' + (j' + h') \bmod \Omega < \Omega$  (умножение  $\mathbf{e}_i$  на  $(\mathbf{e}_j \mathbf{e}_h)$  дает множитель  $\delta$ ) и  $\psi_2 = \lambda\delta = \psi_1$ .

*Случай 3.* Значения  $i'$ ,  $j'$  и  $h'$  удовлетворяют условиям  $i' + j' \geq \Omega$  (произведение  $\mathbf{e}_i \mathbf{e}_j$  вносит множитель  $\lambda$ ) и  $(i' + j') \bmod \Omega + h' < \Omega$  (умножение  $\mathbf{e}_i \mathbf{e}_j$  на  $\mathbf{e}_h$  дает множитель  $\delta$ ). Тогда  $\psi_1 = \lambda\delta$ . Для вычисления  $\psi_2$  рассмотрим подслучаи 3.1 и 3.2:

*Подслучай 3.1.* Имеет место неравенство  $j' + h' \geq \Omega$  (произведение  $\mathbf{e}_j \mathbf{e}_h$  дает множитель  $\lambda$ ). Тогда  $i' + (j' + h') \bmod \Omega = i' + j' + h' - \Omega \geq i' + j' \geq \Omega$  и умножение  $\mathbf{e}_i$  на  $(\mathbf{e}_j \mathbf{e}_h)$  дает множитель  $\lambda$ , т. е.  $\psi_2 = \lambda\delta = \psi_1$ .

*Подслучай 3.2.* Имеет место неравенство  $j' + h' < \Omega$  (произведение  $\mathbf{e}_j \mathbf{e}_h$  дает множитель  $\delta$ ). Тогда  $i' + (j' + h') \bmod \Omega < \Omega$  (поскольку  $i' + (j' + h') \bmod \Omega = i' + j' + h' - \Omega < j' + h' < \Omega$ ) и умножение  $\mathbf{e}_i$  на  $(\mathbf{e}_j \mathbf{e}_h)$  дает множитель  $\delta$ , т. е.  $\psi_2 = \lambda\delta = \psi_1$ .

Случай 4. Значения  $i', j'$  и  $h'$  удовлетворяют условиям  $i' + j' \geq \Omega$  и  $(i' + j') \bmod \Omega + h' \geq \Omega$ . Из этих двух условий следует  $j' + h' \geq \Omega$ , а с учетом (16) имеем  $\psi_1 = \lambda^2$ . С учетом неравенства  $j' + h' \geq \Omega$  (умножение  $e_j$  на  $e_h$  дает множитель  $\lambda$ ) легко показать  $i' + (j' + h') \bmod \Omega = (i' + j') \bmod \Omega + h' \geq \Omega$ , т. е. умножение  $e_i$  на  $(e_j e_h)$  дает множитель  $\lambda$  и получаем  $\psi_2 = \lambda^2 = \psi_1$ .

Таким образом, для всех случаев и подслучаев выполняется равенство  $\psi_1 = \psi_2$ , т. е. рассматриваемая операция векторного умножения является ассоциативной.

Для фиксированных значений параметров  $m$  и  $d$  доказанная теорема позволяет найти  $\Omega = m - 1$  различных распределений структурной константы  $\lambda$  и столько же различных распределений структурной константы  $\delta$  (с независимыми значениями), задаваемых различными значениями параметра  $t$ . Значения констант, относящихся к различным распределениям, являются независимыми. Легко показать, что любое сочетание этих независимых структурных констант в единой ТУБВ сохраняет свойства коммутативности и ассоциативности операции умножения. С точки зрения возможности задания векторного конечного поля эти две константы имеют различное значение. Если для некоторых двух значений  $\lambda$  и  $\delta$  конечная алгебра не является полем, то путем модифицирования значения  $\lambda$  можно добиться формирования векторного конечного поля. Однако, путем модифицирования значения  $\delta$  при фиксированном исходном значении  $\lambda$  этого добиться нельзя.

Таким образом, доказанная теорема обеспечивает формальное параметризуемое задание  $2m - 2$  (без учета константы с тривиальным распределением для случая  $t = \Omega$ ) независимых структурных констант для значений размерности, представленных в таблице 5.

Таблица 5. Значения размерности, охватываемые предложенным способом параметрического задания распределений структурных констант

Размерность	7	15	31	63	127	255	511
Число констант	12	28	60	124	252	508	1020

**7. Примеры построения ТУБВ.** Рассмотрим варианты практического использования предложенного метода для случая размерности  $m = 7$ . Таблица 6 иллюстрирует случай задания полей  $GF(p^7)$  в виде семимерных конечных алгебр при использовании ненулевых структурных констант  $\eta$  ( $t = 1$ ),  $\gamma$  ( $t = 2$ ),  $\rho$  ( $t = 3$ ),  $\lambda$  ( $t = 4$ ),  $\varepsilon$  ( $t = 5$ ),  $\mu$  ( $t = 6$ ) в унифицированной формуле (15) с умножением

двоичных многочленов  $i, j$  и  $d$  по модулю неприводимого двоичного многочлена  $x^3 + x^2 + 1$  (1101 – в представлении в виде битовой строки) и вычислении индексов номеров-многочленов  $i, j$  и  $d$  при основании индексов, равном многочлену  $x^2 + 1$  (101). Экспериментально подтверждается существование векторных полей  $GF(p^7)$  с глобальной двухсторонней единицей  $U=(0,0,0,0,0,1)$ , заданных по таблице 6 над полем  $GF(p)$  для  $p = 29$  и  $p = 211$  и различных наборов значений структурных констант.

Таблица 6. Задание семимерных векторных конечных полей с помощью базовых констант  $\eta, \gamma, \rho, \lambda, \varepsilon$ , и  $\mu$  (константа  $\Psi = \eta\gamma\rho\lambda\varepsilon\mu$ )

$\times$	$e_1$	$:e_2$	$:e_3$	$:e_4$	$e_5$	$:e_6$	$e_7$
$e_1$	$\varepsilon\lambda\mu e_5$	$\Psi e_7$	$:e\lambda\mu e_2$	$\varepsilon\mu\rho e_3$	$\varepsilon\mu\rho e_6$	$\delta\lambda\mu e_4$	$e_1$
$:e_2$	$\Psi e_7$	$\delta\eta\rho e_3$	$:e\lambda\eta e_4$	$\varepsilon\eta\rho e_6$	$\delta\eta\rho e_1$	$\delta\lambda\eta e_5$	$e_2$
$:e_3$	$:e\lambda\mu e_2$	$\delta\lambda\eta e_4$	$\varepsilon\lambda\eta e_6$	$\varepsilon\lambda\eta e_5$	$\Psi e_7$	$\delta\lambda\eta e_1$	$e_3$
$:e_4$	$\varepsilon\mu\rho e_3$	$\varepsilon\eta\rho e_6$	$\varepsilon\lambda\eta e_5$	$\varepsilon\eta\rho e_1$	$\varepsilon\mu\rho e_2$	$\Psi e_7$	$e_4$
$e_5$	$\varepsilon\mu\rho e_6$	$\delta\eta\rho e_1$	$\Psi e_7$	$\varepsilon\mu\rho e_2$	$\delta\mu\rho e_4$	$\delta\mu\rho e_3$	$e_5$
$:e_6$	$:e\lambda\mu e_4$	$\delta\lambda\eta e_5$	$\delta\lambda\eta e_1$	$\Psi e_7$	$:e\mu\rho e_3$	$\delta\lambda\mu e_2$	$e_6$
$e_7$	$e_1$	$:e_2$	$:e_3$	$:e_4$	$e_5$	$:e_6$	$e_7$

Константы  $\delta$  и  $\lambda$ , присутствующие в формуле (15) при фиксированном значении параметра  $t$  заполняют все ячейки ТУБВ, в каждой из которых присутствует либо  $\delta$ , либо  $\lambda$ . При этом значение  $\delta$  влияет на ненулевую координату глобальной двухсторонней единицы, но не влияет на формирование векторного поля (в том смысле, что, если для текущей пары значений  $\delta$  и  $\lambda$  не сформировалось векторное конечное поле, то подбором значений  $\delta$  этого добиться нельзя). Будем называть константу  $\lambda$  базовой, а  $\delta$  – сопряженной к ней. В следующем примере константы, сопряженные с соответствующими базовыми константами, обозначены добавлением штриха.

Распределения сопряженных (по отношению к соответствующим константам из таблице 6) независимых структурных констант  $\eta'$  ( $t = 1$ ),  $\gamma'(t = 2)$ ,  $\rho'(t = 3)$ ,  $\lambda'(t = 4)$ ,  $\varepsilon'(t = 5)$ ,  $\mu'(t = 6)$  и эвристической константы  $\tau$  представлены в таблице 7. Константа  $\tau$  также не влияет на формирование векторного поля, но влияет на значение ненулевой координаты единичного вектора  $U$ :

$$u_{d^{-1} \bmod f(x)} = \Psi'^{-1} \tau^{-1}.$$

Найденное эвристическим путем распределение константы  $\tau$  показывает, что формально задаваемое формулой (15) множество структурных констант может быть дополнено, расширяя множество потенциально реализуемых модификаций векторного конечного поля для фиксированного распределения базисных векторов. Для многих других значений размерности  $m$  и параметра  $d$  были найдены аналогичные распределения эвристических констант. С точки зрения увеличения числа различных модификаций векторного конечного поля, задаваемого по ТУБВ с фиксированным распределением базисных векторов представляет прикладной интерес эвристический поиск и других распределений дополнительных структурных констант. Однако он может служить только дополнением к разработанному формализованному методу параметризуемого задания различных распределений структурных констант.

Таблица 7. Распределения сопряженных структурных констант  $\gamma', \varepsilon', \lambda', \mu', \eta'$  и  $\rho'$  (константа  $\Psi' = \gamma'\varepsilon'\lambda'\mu'\eta'\rho'$ )

$\times$	$e_1$	$:e_2$	$:e_3$	$:e_4$	$e_5$	$:e_6$	$e_7$
$e_1$	$\delta'\eta'\rho'e_5$	$\tau^{-1}e_7$	$\delta'\eta'\rho'e_2$	$\delta'\lambda'\eta'e_3$	$\delta'\lambda'\eta'e_6$	$\varepsilon'\eta'\rho'e_4$	$\Psi'\tau e_1$
$:e_2$	$\tau^{-1}e_7$	$\varepsilon'\lambda'\mu'e_3$	$\varepsilon'\mu'\rho'e_4$	$\delta'\lambda'\mu'e_6$	$\varepsilon'\lambda'\mu'e_1$	$\varepsilon'\mu'\rho'e_5$	$\Psi'\tau e_2$
$:e_3$	$\delta'\eta'\rho'e_2$	$\varepsilon'\mu'\rho'e_4$	$\delta'\mu'\rho'e_6$	$\delta'\mu'\rho'e_5$	$\tau^{-1}e_7$	$\varepsilon'\mu'\rho'e_1$	$\Psi'\tau e_3$
$:e_4$	$\delta'\lambda'\eta'e_3$	$\delta'\lambda'\mu'e_6$	$\delta'\mu'\rho'e_5$	$\delta'\lambda'\mu'e_1$	$\delta'\lambda'\eta'e_2$	$\tau^{-1}e_7$	$\Psi'\tau e_4$
$e_5$	$\delta'\lambda'\eta'e_6$	$\varepsilon'\lambda'\mu'e_1$	$\tau^{-1}e_7$	$\delta'\lambda'\eta'e_2$	$\varepsilon'\lambda'\eta'e_4$	$\varepsilon'\lambda'\eta'e_3$	$\Psi'\tau e_5$
$:e_6$	$\varepsilon'\eta'\rho'e_4$	$\varepsilon'\mu'\rho'e_5$	$\varepsilon'\mu'\rho'e_1$	$\tau^{-1}e_7$	$\varepsilon'\lambda'\eta'e_3$	$\varepsilon'\eta'\rho'e_2$	$\Psi'\tau e_6$
$e_7$	$\Psi'\tau e_1$	$:\Psi'\tau e_2$	$\Psi'\tau e_3$	$\Psi'\tau e_4$	$\Psi'\tau e_5$	$\Psi'\tau e_6$	$\Psi'\tau e_7$

**8. Топологии нелинейного отображения.** Предложенный метод формального параметризуемого задания многих распределений структурных констант при заданном распределении базисных векторов дает достаточно широкие возможности реализации нелинейного отображения  $N$  с секретной лазейкой (раздел 5). Простейшее строение (топология) представляет собой реализацию  $N$  в виде одной операции экспоненцирования в векторном конечном поле  $GF((p^s)^n)$ , заданной над полем  $GF(p^s)$  набором многочленов, который вычисляется по ТУБВ, использованной для задания поля  $GF((p^s)^n)$ .

Для реализации такой тривиальной топологии представляет интерес использование полей  $GF(p^s)$  четной характеристики, т.е. полей  $GF(2^s)$  при различных степенях  $s$ . Это связано с тем, что в полях такой характеристики операции  $N(X) = X^i$  для значений степени  $i$ , равных

степеням числа 2, записываются набором степенных многочленов, каждый из которых включает всего одно слагаемое, а отображения вида  $N(\mathbf{X}) = \mathbf{X}^i \mathbf{X}^j$ , которые непосредственно могут быть использованы в качестве открытого ключа  $P$ , – набором из  $n$  многочленов, каждый из которых включает не более  $n$  слагаемых [20]. В последнем случае имеем возможность предложить алгоритм шифрования, алгоритм ЭЦП или протокол открытого согласования общего секретного ключа с открытым ключом размера  $\approx sn^2$  бит.

Такая тривиальная топология отображения  $N$ , например, может быть реализована для случая  $n = 31$  (таблица 5),  $s = 15$  и  $N(\mathbf{X}) = \mathbf{X}^{17}$ , в котором реализуются условия делимости числа  $2^s - 1$  на 31 (обеспечивает возможность задания векторного конечного поля) и неделимости  $2^s - 1$  на 17 (обеспечивает однозначность извлечения корня степени 17 в поле  $GF((2^{15})^{31})$ ). В такой реализации обеспечивается 80-битный уровень стойкости (таблица 4) при размере открытого ключа, равном  $\approx 2$  Кбайт.

Аналогичная реализация для случая  $n = 127$  (таблица 5),  $s = 14$  и  $N(\mathbf{X}) = \mathbf{X}^{257}$  обеспечивается 256-битный уровень стойкости (таблица 4) при размере  $P$ , равном  $\approx 32$  Кбайт. По сравнению с известным алгоритмом Rainbow [27] для такого же уровня стойкости имеем примерно 40-кратное уменьшение размера открытого ключа.

В случае реализации тривиальной топологии с заданием векторных конечных полей над полями нечетной характеристики  $p$  отображение  $N$  может быть задано в виде  $N(\mathbf{X}) = \mathbf{X}^2$  и  $N(\mathbf{X}) = \mathbf{X}^3$ . В первом случае процедура расшифровывания шифртекста является неоднозначной и требуется разработка специального алгоритма извлечения квадратных корней, однако обеспечивается существенно меньший размер открытого ключа. Во втором случае всегда можно выбрать значение  $p$ , при котором извлечение корня третьей степени дает однозначный результат. Для этого следует выбрать простое  $p$ , такое, что  $p - 1$  делится на 31, а число  $p^{31} - 1$  не делится на 3 (например,  $p = 311$ ), определяя возможность задания векторного конечного поля  $GF(p^{31})$ , в котором вычисление корня третьей степени осуществляется путем выполнения одной операции возведения в степень  $z = 3^{-1} \bmod (p^{31} - 1)$ :  $\mathbf{X}^{1/3} = \mathbf{X}^z$ . Однако использование отображения  $N(\mathbf{X}) = \mathbf{X}^3$  приводит к тому, что размер открытого ключа становится сравнительно большим, например,  $\approx 30$  Кбайт ( $\approx 2$  Мбайт) при уровне стойкости  $2^{80}$  ( $2^{256}$ ).

Существенное уменьшение размера открытого ключа может быть достигнуто применением каскадной топологии, в которой входной вектор  $\mathbf{X}$  разбивается на несколько подвекторов размерности

$m$  и над каждым из них выполняется операции возведения в куб, перестановка всех  $n$  координат всех подвекторов и умножение новых значений подвекторов на секретные векторы размерности  $m$  (задает маскирующее линейное преобразование). В такой каскадной топологии при значениях  $m = 7$  и  $n = 70$  используются семимерные векторные конечные поля (в общем случае десять различных полей), заданные, например, над  $GF(239)$ , и обеспечивается ожидаемый уровень 192-битной стойкости при размере открытого ключа, равном  $\approx 24$  Кбайт. По сравнению с Rainbow [27] для такого же уровня стойкости имеем примерно 36-кратное уменьшение размера открытого ключа.

При реализации аналогичной каскадной топологии при значениях  $m = 7$  и  $n = 77$  с использованием поля четной характеристики  $GF(64)$  192-битная стойкости обеспечивается при размере открытого ключа, равном  $\approx 3,8$  Кбайт (226-кратное уменьшение размера  $P$  относительно Rainbow [27]).

Другие типы каскадных топологий для задания нелинейного отображения с секретной лазейкой на основе операций экспоненцирования в векторных конечных полях представлены в работах [19, 20].

**9. Заключение.** Конечные некоммутативные и коммутативные алгебры с ассоциативной операцией умножения представляют значительный интерес для разработки на их основе постквантовых двухключевых криптоалгоритмов. Первые служат алгебраическим носителем алгоритмов ЭЦП со скрытой группой, относящихся к двум различным типам, использующим вычислительную трудность двух различных задач: 1) скрытой ЗДЛ и 2) решения большой системы степенных уравнений. В алгоритмах второго типа возникает проблема обеспечения полной рандомизации подписи, для решения которой предложен способ, основанный на использовании приема удвоения уравнения верификации ЭЦП.

Алгебры второго типа представляют существенный интерес для построения криптоалгоритмов с открытым ключом, основанных на трудно обратимых отображениях с потайной лазейкой, предоставляя потенциальную возможность решения проблемы чрезвычайно большого размера открытого ключа в известных алгоритмах данного типа. Предложенный способ параметризуемого унифицированного задания векторных конечных полей решает проблему задания большого числа модификаций таких полей и обеспечивает возможность использования конкретной модификации или конкретного набора модификаций векторного конечного поля

в качестве элемента секретного ключа. Разработанный способ может служить прототипом для разработки других унифицированных способов задания векторных конечных полей с параметризацией распределения структурных констант с целью расширения множества значений размерности, для которых может быть формализован поиск распределений многих структурных констант. Однако это представляет самостоятельную исследовательскую задачу.

### Литература

1. Ekert A., Jozsa R. Quantum computation and Shor's factoring algorithm // *Reviews of Modern Physics*. 1996. vol. 68. no. 3. pp. 733–752.
2. Shor P.W. Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer // *SIAM Journal of Computing*. 1997. vol. 26. pp. 1484–1509.
3. Post-Quantum Cryptography. Proceedings of the 13th International Conference, PQCrypto 2022 // *Lecture Notes in Computer Science*. 2022. vol. 13512.
4. Johansson T., Smith-Tone D. Post-Quantum Cryptography. Proceedings of the 14th International Conference, PQCrypto 2023 // *Lecture Notes in Computer Science*. 2023. vol. 14154.
5. Alagic G, Cooper D., Dang Q., Dang T., Kelsey J., Lichtinger J., Liu Y., Miller C., Moody D., Peralta R., Perlner R., Robinson A., Smith-Tone D., Apon D. Status Report on the Third Round of the NIST Post-Quantum Cryptography Standardization Process // *NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology*. 2022. URL: [https://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=934458](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934458) (дата обращения: 25.02.2024).
6. Han J., Zhuang J. DLP in semigroups: algorithms and lower bounds // *J. Math. Cryptol*. 2022. vol. 16. no. 1. pp. 278–288.
7. Battarbee C., Kahrobaei D., Perret L., Shahandashti S.F. SPDH-Sign: Towards Efficient, Post-quantum Group-Based Signatures // *Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science*. 2023. vol. 14154. pp. 113–138. DOI: 10.1007/978-3-031-40003-2\_5.
8. Vysotskaya V.V., Chizhov I.V. The security of the code-based signature scheme based on the Stern identification protocol // *Applied Discrete Mathematics*. 2022. № 57. С. 67–90. DOI: 10.17223/20710410/57/5.
9. Kosolapov Y.V., Turchenko O.Y. On the construction of a semantically secure modification of the McEliece cryptosystem // *Applied Discrete Mathematics*. 2019. № 45. С. 33–43. DOI: 10.17223/20710410/45/4.
10. Gartner J. NTWE: A Natural Combination of NTRU and LWE // *Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science*, 2023. vol. 14154. pp. 321–353. DOI: 10.1007/978-3-031-40003-2\_12.
11. Lysakov I.V.. Solving some cryptanalytic problems for lattice-based cryptosystems with quantum annealing method // *Mathematical Aspects of Cryptography*. 2023. vol. 14. no. 2. pp. 111–122. DOI: 10.4213/mvk441.
12. Hamlin B., Song F. Quantum Security of Hash Functions and Property-Preservation of Iterated Hashing // *Post-Quantum Cryptography. PQCrypto 2019 / Lecture Notes in Computer Science*. 2019. vol. 11505. pp. 329–349. DOI: 10.1007/978-3-030-25510-7\_18.
13. Agibalov G.P. ElGamal cryptosystems on Boolean functions / *Applied Discrete Mathematics*. 2018. № 42. С. 57–65. DOI: 10.17223/20710410/42/4.

14. Ding J., Petzoldt A., Schmidt D.S. Multivariate Cryptography // *Multivariate Public Key Cryptosystems. Advances in Information Security*. 2020. vol. 80. DOI: 10.1007/978-1-0716-0987-3\_2.
15. Debnath S., Kundu N., Mishra D., Choudhury T. Post-quantum digital signature scheme based on multivariate cubic problem // *Journal of Information Security and Applications*. 2020. vol. 53. DOI: 10.1016/j.jisa.2020.102512.
16. Ding J., Petzoldt A., Schmidt D.S. Oil and Vinegar // *Multivariate Public Key Cryptosystems. Advances in Information Security*. 2020. vol. 80. pp. 89–151. Springer, New York. DOI: 10.1007/978-1-0716-0987-3\_5.
17. Cartor R., Cartor M., Lewis M., Smith-Tone D. IPRainbow // *Post-Quantum Cryptography. PQCrypto 2022. Lecture Notes in Computer Science*. 2022. vol. 13512. pp. 170–184. DOI: 10.1007/978-3-031-17234-2\_9.
18. Beullens W. MAYO: practical post-quantum signatures from oil-and-vinegar maps // *Proceedings of the International Conference on Selected Areas in Cryptography (SAC 2021). Lecture Notes in Computer Science*. 2022. vol. 13203. pp. 355–376.
19. Молдовян А.А., Молдовян Д.Н., Молдовян Н.А. Новый подход к разработке алгоритмов многомерной криптографии // *Вопросы кибербезопасности*. 2023. № 2(54). С. 52–64. DOI: 10.21681/2311-3456-2023-2-52-6.
20. Moldovyan A.A., Moldovyan N.A. Vector finite fields of characteristic two as algebraic support of multivariate cryptography // *Computer Science Journal of Moldova*. 2024. no. 1(94). pp. 46–60. DOI: 10.56415/esjm.v32.04.
21. Duong M.T., Moldovyan D.N., Do B.V., Nguyen M.H. Post-quantum signature algorithms on non-commutative algebras, using difficulty of solving systems of quadratic equations // *Computer Standards and Interfaces*. 2023, vol. 86. no. 103740. DOI: 10.1016/j.csi.2023.103740.
22. Moldovyan D.N. A practical digital signature scheme based on the hidden logarithm problem // *Computer Science Journal of Moldova*. 2021. vol. 29, no. 2(86). pp. 206–226.
23. Moldovyan N.A., Moldovyanu P.A. Vector Form of the Finite Fields GF(pm) // *Bulletin of Academy of Sciences of Moldova. Mathematics*. 2009. no. 3(61). pp. 57–63.
24. Ding J., Petzoldt A., Schmidt D.S. Solving Polynomial Systems // In: *Multivariate Public Key Cryptosystems. Advances in Information Security*. Springer, New York. 2020. vol. 80. pp. 185–248. DOI: 10.1007/978-1-0716-0987-3\_8.
25. Ding J., Petzoldt A. Current State of Multivariate Cryptography // *IEEE Security and Privacy*. 2017. vol. 15. no. 4. pp. 28–36.
26. Qiao S., Han W., Li Y., Jiao L. Construction of Extended Multivariate Public Key Cryptosystems // *International Journal of Network Security*. 2016. vol. 18. no. 1. pp. 60–67.
27. Rainbow Signature. One of three NIST Post-quantum Signature Finalists [on line] 2021. URL: <https://www.pqcrainbow.org/> (дата обращения: 25.02.2024).

**Молдовян Александр Андреевич** — д-р техн. наук, профессор, главный научный сотрудник, лаборатория проблем компьютерной безопасности, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: криптография, постквантовые криптоалгоритмы с открытым ключом, электронная цифровая подпись, криптографические протоколы, компьютерная безопасность. Число научных публикаций — 200. [maa1305@yandex.ru](mailto:maa1305@yandex.ru); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)333-3411.

**Молдовян Дмитрий Николаевич** — канд. техн. наук, доцент кафедры, кафедра информационных систем, Санкт-Петербургский государственный электротехнический

университет «ЛЭТИ». Область научных интересов: криптография, постквантовые двухключевые криптоалгоритмы на алгебрах и нелинейных отображениях, цифровая подпись, информационная безопасность. Число научных публикаций — 100. [mdn.spectr@mail.ru](mailto:mdn.spectr@mail.ru); улица профессора Попова, 5, 199178, Санкт-Петербург, Россия; р.т.: +7(812)234-2772.

**Молдовян Николай Андреевич** — д-р техн. наук, профессор, главный научный сотрудник, лаборатория проблем компьютерной безопасности, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: криптография, постквантовые криптоалгоритмы с открытым ключом, электронная цифровая подпись, блочные шифры, компьютерная безопасность. Число научных публикаций — 230. [nmold@mail.ru](mailto:nmold@mail.ru); 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)333-3411.

**Поддержка исследований.** Работа выполнена при финансовой поддержке РФФИ: проекты № 24-21-00225 (разделы 2, 3 и 4) и № 24-41-04006 (разделы 5, 6, 7 и 8).

A. MOLDOVYAN, D. MOLDOVYAN, N. MOLDOVYAN  
**POST-QUANTUM PUBLIC-KEY CRYPTOSCHEMES ON FINITE ALGEBRAS**

---

*Moldovyan A., Moldovyan D., Moldovyan N. Post-Quantum Public-Key Cryptoschemes on Finite Algebras.*

**Abstract.** One direction in the development of practical post-quantum public-key cryptographic algorithms is the use of finite algebras as their algebraic carrier. Two approaches in this direction are considered: 1) construction of electronic digital signature algorithms with a hidden group on non-commutative associative algebras and 2) construction of multidimensional cryptography algorithms using the exponential operation in a vector finite field (in a commutative algebra, which is a finite field) to specify a nonlinear mapping with a secret trapdoor. The first approach involves the development of two types of cryptoschemes: those based on the computational difficulty of a) the hidden discrete logarithm problem and b) solving a large system of quadratic equations. For the second type, problems arise in ensuring complete randomization of the digital signature and specifying non-commutative associative algebras of large dimension. Ways to solve these problems are discussed. The importance of studying the structure of finite non-commutative algebras from the point of view of decomposition into a set of commutative subalgebras is shown. Another direction is aimed at a significant (10 or more times) reduction in the size of the public key in multivariate-cryptography algorithms and is associated with the problem of developing formalized, parameterizable, unified methods for specifying vector finite fields of large dimensions (from 5 to 130) with a sufficiently large number of potentially implementable types and modifications each type (up to  $2^{500}$  or more). Variants of such methods and topologies of nonlinear mappings on finite vector fields of various dimensions are proposed. It is shown that the use of mappings that specify the exponential operation in vector finite fields potentially eliminates the main drawback of known multivariate-cryptography algorithms, which is associated with the large size of the public key.

**Keywords:** post-quantum cryptography, multivariate cryptography, finite algebra, non-commutative algebra, vector finite field, nonlinear mappings.

---

## References

1. Ekert A., Jozsa R. Quantum computation and Shor's factoring algorithm. *Reviews of Modern Physics*. 1996. vol. 68. no. 3. pp. 733–752.
2. Shor P.W. Polynomial-time algorithms for prime factorization and discrete logarithms on quantum computer. *SIAM Journal of Computing*. 1997. vol. 26. pp. 1484–1509.
3. Post-Quantum Cryptography. *Proceedings of the 13th International Conference, PQCrypto 2022*. Lecture Notes in Computer Science. 2022. vol. 13512.
4. Johansson T., Smith-Tone D. Post-Quantum Cryptography. *Proceedings of the 14th International Conference, PQCrypto 2023*. Lecture Notes in Computer Science. 2023. vol. 14154.
5. Alagic G, Cooper D., Dang Q., Dang T., Kelsey J., Lichtinger J., Liu Y., Miller C., Moody D., Peralta R., Perlner R., Robinson A., Smith-Tone D., Apon D. Status Report on the Third Round of the NIST Post-Quantum Cryptography Standardization Process. NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology. 2022. Available at: [https://tsapps.nist.gov/publication/get\\_pdf.cfm?pub\\_id=934458](https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=934458) (accessed: обращения: 25.02.2024).

6. Han J., Zhuang J. DLP in semigroups: algorithms and lower bounds. *J. Math. Cryptol.* 2022. vol. 16. no. 1. pp. 278–288.
7. Battarbee C., Kahrobaei D., Perret L., Shahandashti S.F. SPDH-Sign: Towards Efficient, Post-quantum Group-Based Signatures. *Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science.* 2023. vol. 14154. pp. 113–138. DOI: 10.1007/978-3-031-40003-2\_5.
8. Vysotskaya V.V., Chizhov I.V. The security of the code-based signature scheme based on the Stern identification protocol. *Applied Discrete Mathematics.* 2022. № 57. C. 67–90. DOI: 10.17223/20710410/57/5.
9. Kosolapov Y.V., Turchenko O.Y. On the construction of a semantically secure modification of the McEliece cryptosystem. *Applied Discrete Mathematics.* 2019. № 45. C. 33–43. DOI: 10.17223/20710410/45/4.
10. Gartner J. NTWE: A Natural Combination of NTRU and LWE. *Post-Quantum Cryptography. PQCrypto 2023. Lecture Notes in Computer Science,* 2023. vol. 14154. pp. 321–353. DOI: 10.1007/978-3-031-40003-2\_12.
11. Lysakov I.V. Solving some cryptanalytic problems for lattice-based cryptosystems with quantum annealing method. *Mathematical Aspects of Cryptography.* 2023. vol. 14. no. 2. pp. 111–122. DOI: 10.4213/mvk441.
12. Hamlin B., Song F. Quantum Security of Hash Functions and Property-Preservation of Iterated Hashing. *Post-Quantum Cryptography. PQCrypto 2019 / Lecture Notes in Computer Science.* 2019. vol. 11505. pp. 329–349. DOI: 10.1007/978-3-030-25510-7\_18.
13. Agibalov G.P. ElGamal cryptosystems on Boolean functions. *Applied Discrete Mathematics.* 2018. № 42. C. 57–65. DOI: 10.17223/20710410/42/4.
14. Ding J., Petzoldt A., Schmidt D.S. Multivariate Cryptography. *Multivariate Public Key Cryptosystems. Advances in Information Security.* 2020. vol. 80. DOI: 10.1007/978-1-0716-0987-3\_2.
15. Debnath S., Kundu N., Mishra D., Choudhury T. Post-quantum digital signature scheme based on multivariate cubic problem. *Journal of Information Security and Applications.* 2020. vol. 53. DOI: 10.1016/j.jisa.2020.102512.
16. Ding J., Petzoldt A., Schmidt D.S. Oil and Vinegar. *Multivariate Public Key Cryptosystems. Advances in Information Security.* 2020. vol. 80. pp. 89–151. Springer, New York. DOI: 10.1007/978-1-0716-0987-3\_5.
17. Cartor R., Cartor M., Lewis M., Smith-Tone D. IPRainbow. *Post-Quantum Cryptography. PQCrypto 2022. Lecture Notes in Computer Science.* 2022. vol. 13512. pp. 170–184. DOI: 10.1007/978-3-031-17234-2\_9.
18. Beullens W. MAYO: practical post-quantum signatures from oil-and-vinegar maps. *Proceedings of the International Conference on Selected Areas in Cryptography (SAC 2021). Lecture Notes in Computer Science.* 2022. vol. 13203. pp. 355–376.
19. Moldovyan A.A., Moldovyan D.N., Moldovyan N.A. A new approach to the development of multivariate cryptography algorithms. *Voprosy kiberbezopasnosti – Cibersecurity questtions.* 2023. no. 2(54). pp. 52–64. DOI: 10.21681/2311-3456-2023-2-52-6.
20. Moldovyan A.A., Moldovyan N.A. Vector finite fields of characteristic two as algebraic support of multivariate cryptography. *Computer Science Journal of Moldova.* 2024. no. 1(94). pp. 46–60. DOI: 10.56415/csjm.v32.04.
21. Duong M.T., Moldovyan D.N., Do B.V., Nguyen M.H. Post-quantum signature algorithms on non-commutative algebras, using difficulty of solving systems of quadratic equations. *Computer Standards and Interfaces.* 2023, vol. 86. no. 103740. DOI: 10.1016/j.csi.2023.103740.

22. Moldovyan D.N. A practical digital signature scheme based on the hidden logarithm problem. *Computer Science Journal of Moldova*. 2021. vol. 29, no. 2(86). pp. 206–226.
23. Moldovyan N.A., Moldovyanu P.A. Vector Form of the Finite Fields GF(pm). *Bulletin of Academy of Sciences of Moldova. Mathematics*. 2009. no. 3(61). pp. 57–63.
24. Ding J., Petzoldt A., Schmidt D.S. Solving Polynomial Systems. In: *Multivariate Public Key Cryptosystems. Advances in Information Security*. Springer. New York. 2020. vol. 80. pp. 185–248. DOI: 10.1007/978-1-0716-0987-3\_8.
25. Ding J., Petzoldt A. Current State of Multivariate Cryptography. *IEEE Security and Privacy*. 2017. vol. 15. no. 4. pp. 28–36.
26. Qiao S., Han W., Li Y., Jiao L. Construction of Extended Multivariate Public Key Cryptosystems. *International Journal of Network Security*. 2016. vol. 18. no. 1. pp. 60–67.
27. Rainbow Signature. One of three NIST Post-quantum Signature Finalists [on line] 2021. Available at: <https://www.pqc rainbow.org/> (accessed: 25.02.2024).

**Moldovyan Alexandr** — Ph.D., Dr.Sci., Professor, Chief researcher, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: cryptography, post-quantum public-key cryptoalgorithms, digital signature, cryptographic protocols, computer security. The number of publications — 200. [maa1305@yandex.ru](mailto:maa1305@yandex.ru); 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)333-3411.

**Moldovyan Dmitriy** — Ph.D., Associate professor of the department, Department of information systems, Saint Petersburg Electrotechnical University "LETI". Research interests: cryptography, post-quantum public-key cryptoalgorithms on algebras and on non-linear mappings, digital signature, information security. The number of publications — 100. [mdn.spectr@mail.ru](mailto:mdn.spectr@mail.ru); 5, Professor Popov St., 199178, St. Petersburg, Russia; office phone: +7(812)234-2772.

**Moldovyan Nikolay** — Ph.D., Dr.Sci., Professor, Chief researcher, Laboratory of computer security problems, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: cryptography, post-quantum public-key cryptoalgorithms, digital signature, block ciphers, computer security. The number of publications — 230. [nmold@mail.ru](mailto:nmold@mail.ru); 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)333-3411.

**Acknowledgements.** This research is supported by RFBR: projects #24-21-00225 (Sections 2, 3, and 4) and #24-41-04006 (Sections 5, 6, 7, and 8).

## 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 материал а выделение подразделов. В основную часть опускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

## Author guidelines

Interaction between each potential author and the Editorial board is realized through the personal account on the website of the journal "Informatics and Automation" <http://ia.spcras.ru/>. At the registration the authors are requested to fill out all data fields in the proposed form. The submissions should be prepared using MS Word 2007, LaTeX. The text of the paper in the main part should not exceed 30 pages. Pages are not numbered; hyphenations are not allowed. Certain figures, tables, listings and formulas are allowed in the main section, and their typography is considered in more detail at the journal web.

---

Signed to print 24.06.2024. Passed for print 01.07.2024.

Printed in Publishing center GUAP.

Address: 67 litera A, B. Morskaya, St. Petersburg, 190000, Russia

---

Founder and Publisher: SPC RAS.

Address: 39 litera A, 14th Line V.O., St. Peterburg, 199178, Russia.

The journal is registered in the Federal Service for Supervision of Communications, Information Technology, and Mass Media,

Registration Certificate (registration number) ПИ № ФС77-79228 dated September 25, 2020

Subscription Index П5513, Russian Post Catalog

---

Подписано к печати 24.06.2024. Дата выхода в свет 01.07.2024.

Формат 60×90 1/16. Усл. печ. л. 19,8. Заказ № 225. Тираж 300 экз., цена свободная.

Отпечатано в Редакционно-издательском центре ГУАП.

Адрес типографии: Б. Морская, д. 67, лит. А, г. Санкт-Петербург, 190000, Россия

---

Учредитель и издатель: СПб ФИЦ РАН.

Адрес учредителя и издателя: 14-я линия В.О., д. 39, лит. А, г. Санкт-Петербург, 199178, Россия

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.

Подписной индекс П5513 по каталогу «Почта России»