

ISSN 2713-3192
DOI 10.15622/ia.2021.20.3
<http://ia.spcras.ru>

ТОМ 20 № 3

**ИНФОРМАТИКА
И АВТОМАТИЗАЦИЯ**

**INFORMATICS
AND AUTOMATION**



**Санкт-Петербург
2021**

INFORMATICS AND AUTOMATION

Volume 20 № 3, 2021

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

A. A. Ashimov	Prof., Dr. Sci., Academician of the National Academy of Sciences of the Republic of Kazakhstan, Almaty, Kazakhstan
N. P. Veselkin	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
I. A. Kalyaev	Prof., Dr. Sci., Academician of RAS, Taganrog, Russia
Yu. A. Merkur'yev	Prof., Dr. Sci., Academician of the Latvian Academy of Sciences, Riga, Latvia
A. I. Rudskoi	Prof., Dr. Sci., Academician of RAS, St. Petersburg, Russia
V. Sgurev	Prof., Dr. Sci., Academician of the Bulgarian Academy of Sciences, Sofia, Bulgaria
B. Ya. Sovetov	Prof., Dr. Sci., Academician of RAE, St. Petersburg, Russia
V. A. Soyfer	Prof., Dr. Sci., Academician of RAS, Samara, Russia

Editorial Board

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

Editor: I. O. Novikova

Technical editor: A. K. Molotilova

Interpreter: E. N. Mesheryakova

Art editor: N. A. Dormidontova

Editorial office address

14-th line V.O., 39, SPIIRAS, St. Petersburg, 199178, Russia,

e-mail: 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 Nanotechnology and Information Technologies of the Russian Academy of Sciences © St. Petersburg Federal Research Center of the Russian Academy of Sciences, 2021

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

Том 20 № 3, 2021

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

А. А. Ашимов	академик Национальной академии наук Республики Казахстан, д-р техн. наук, проф., Алматы, Казахстан
Н. П. Веселкин	академик РАН, д-р мед. наук, проф., Санкт-Петербург, РФ
И. А. Каляев	академик РАН, д-р техн. наук, проф., Таганрог, РФ
Ю. А. Меркурьев	академик Латвийской академии наук, д-р, проф., Рига, Латвия
А. И. Рудской	академик РАН, д-р техн. наук, проф., Санкт-Петербург, РФ
В. Сгурев	академик Болгарской академии наук, д-р техн. наук, проф., София, Болгария
Б. Я. Советов	академик РАН, д-р техн. наук, проф., Санкт-Петербург, РФ
В. А. Сойфер	академик РАН, д-р техн. наук, проф., Самара, РФ

Редакционная коллегия

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

Выпускающий редактор: И. О. Новикова
Технический редактор: А. К. Молотилова

Переводчик: Е. Н. Мещерякова
Художественный редактор: Н. А. Дормидонтова

Адрес редакции

199178, г. Санкт-Петербург, 14-я линия В.О., д. 39
e-mail: 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», 2021
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

Artificial Intelligence, Knowledge and Data Engineering

A. Velichko, A. Karpov ANALYTICAL REVIEW OF AUTOMATIC SYSTEMS FOR DEPRESSION DETECTION BY SPEECH	497
A. Vorobev, G. Vorobeva MODEL OF INFORMATION INTERACTION BETWEEN ELEMENTS OF MULTILEVEL SYSTEM OF DIGITAL TWINS	530
M. Voronov, V. Pimenov FORMALIZATION OF REGULATORY TEXTS	562
A. Korsakov, A. Bakhshiev, L. Astapova, L. Stankevich BEHAVIORAL FUNCTIONS IMPLEMENTATION ON SPIKING NEURAL NETWORKS S.	591
Altaf, S. Iqbal, M. Soomro EFFICIENT ALGORITHM FOR NATURAL LANGUAGE CLASSIFICATION TO DETECT DUPLICATE UNSUPERVISED FEATURES	623

Mathematical Modeling, Numerical Methods

K. Krotov, A. Skatkov CONSTRUCTION OF COMPLEX SCHEDULES FOR EXECUTING TASK PACKAGES AT FORMING SETS IN SPECIFIED DIRECTIVE TERMS	654
V. Kosjukov, M. Medvedev, V. Pshikhopov METHOD FOR OPTIMIZING OF MOBILE ROBOT TRAJECTORY IN REPELLER SOURCES FIELD	690
A. Gvozdarev, P. Patralov PROBABILISTIC ANALYSIS OF GENERALISED STATISTIC MODEL FOR MULTIPATH CHANNEL OF SIMO SYSTEM WITH FADING AND CORRELATED SHADOWING	727

СОДЕРЖАНИЕ

Искусственный интеллект, инженерия данных и знаний

А.Н. Величко, А.А. Карпов АНАЛИТИЧЕСКИЙ ОБЗОР СИСТЕМ АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ ДЕПРЕССИИ ПО РЕЧИ	497
А.В. Воробьев, Г.Р. Воробьева МОДЕЛЬ ИНФОРМАЦИОННОГО ВЗАИМОДЕЙСТВИЯ ЭЛЕМЕНТОВ МНОГОУРОВНЕВОЙ СИСТЕМЫ ЦИФРОВЫХ ДВОЙНИКОВ	530
М.В. Воронов, В.М. Пименов ФОРМАЛИЗАЦИЯ РЕГУЛЯТИВНЫХ ТЕКСТОВ	562
А.М. Корсаков, А.В. Бахшиев, Л.А. Астапова, Л. А. Станкевич РЕАЛИЗАЦИЯ ПОВЕДЕНЧЕСКИХ ФУНКЦИЙ НА СПАЙКОВЫХ НЕЙРОННЫХ СЕТЯХ	591
С. Альтаф, С. Икбал, М. Соомро ЭФФЕКТИВНЫЙ АЛГОРИТМ КЛАССИФИКАЦИИ ЕСТЕСТВЕННОГО ЯЗЫКА ДЛЯ ОБНАРУЖЕНИЯ ПОВТОРЯЮЩИХСЯ НЕКОНТРОЛИРУЕМЫХ ПРИЗНАКОВ	623

Математическое моделирование и прикладная математика

К.В. Кротов, А.В. Скатков ПОСТРОЕНИЕ КОМПЛЕКСНЫХ РАСПИСАНИЙ ВЫПОЛНЕНИЯ ПАКЕТОВ ЗАДАНИЙ ПРИ ФОРМИРОВАНИИ КОМПЛЕКТОВ В ЗАДАННЫЕ ДИРЕКТИВНЫЕ СРОКИ	654
В.А. Костюков, М.Ю. Медведев, В.Х. Пшихопов МЕТОД ОПТИМИЗАЦИИ ТРАЕКТОРИИ МОБИЛЬНОГО РОБОТА В ПОЛЕ ИСТОЧНИКОВ-РЕПЕЛЛЕРОВ	690
А.С. Гвоздарев, П.Е. Патралов ВЕРОЯТНОСТНЫЙ АНАЛИЗ ОБОБЩЕННОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ МНОГОЛУЧЕВОГО КАНАЛА SIMO СИСТЕМЫ С ЗАМИРАНИЯМИ И КОРРЕЛИРОВАННЫМИ ЗАТЕНЕНИЯМИ	727

А.Н. ВЕЛИЧКО, А.А. КАРПОВ
**АНАЛИТИЧЕСКИЙ ОБЗОР СИСТЕМ АВТОМАТИЧЕСКОГО
ОПРЕДЕЛЕНИЯ ДЕПРЕССИИ ПО РЕЧИ**

Величко А.Н., Карпов А.А. Аналитический обзор систем автоматического определения депрессии по речи.

Аннотация. В последние годы в медицинской и научно-технической среде возрос интерес к задаче автоматического определения наличия депрессивного состояния у людей. Депрессия является одним из самых распространенных психических заболеваний, непосредственно влияющих на жизнь человека. В данном обзоре представлены и проанализированы последние работы за два года на тему определения депрессивного состояния у людей. Приведены основные понятия, относящиеся к определению депрессии, описаны как одномодальные, так и многомодальные корпусы, содержащие записи информантов с установленным диагнозом депрессии, а также записи контрольных групп, состоящих из людей без депрессии.

Рассмотрены как теоретические исследования, так и работы, в которых представлены автоматические системы для определения депрессивного состояния. Последние включают в себя различные варианты систем определения депрессии — от одномодальных до многомодальных. Часть рассмотренных систем решает задачу регрессивной классификации, предсказывая степень тяжести депрессии (отсутствие, слабая, умеренная, тяжелая), а другая часть — задачу бинарной классификации, предсказывая наличие заболевания у человека или его отсутствие. Представлена оригинальная классификация методов, использующихся для вычисления информативных признаков по трем коммуникативным модальностям (аудио, видео и текстовая информация). Описаны современные методы, используемые для определения депрессии в каждой из модальностей и в совокупности. Наиболее популярными методами моделирования и распознавания депрессии в рассмотренных работах являются нейронные сети. В ходе аналитического обзора выявлено, что основными признаками депрессии считаются психомоторная заторможенность, которая влияет на все коммуникативные модальности, и сильная корреляция с аффективными величинами валентности, активации и доминанции, при этом наблюдается обратная корреляция между депрессией и агрессией. Выявленные корреляции подтверждают взаимосвязь аффективных расстройств с эмоциональными состояниями человека. В множестве рассмотренных работ наблюдается тенденция объединения модальностей для улучшения качества определения депрессии.

Ключевые слова: автоматическое определение депрессии, компьютерная паралингвистика, речевые технологии, машинное обучение

1. Введение. Согласно данным ВОЗ [1], депрессия является распространенным психическим расстройством и одной из основных болезней, которые приводят к ухудшению жизнедеятельности человека, и может стать причиной инвалидности. На 2018 год во всем мире около 264 млн. человек во всех возрастных группах страдали от депрессии [2].

В последние 10 лет возрос интерес к системам автоматического определения депрессии. На это повлияли многие причины — тяжесть заболевания и повсеместная распространенность, отсутствие лаборатор-

ных тестов или процедур для диагностики депрессии и т.д. На данный момент наличие заболевания определяется путем беседы со специалистом-психотерапевтом и заполнения различного рода опросников: состояния здоровья (Patient Health Questionnaire, PHQ) [3], шкала депрессии Бека (Beck Depression Inventory) [4], самооценки депрессивных симптомов (Quick Inventory of Depressive Symptoms - Self Report, QIDS-SR) [5], шкала Гамильтона для оценки депрессии (Hamilton Rating Scale for Depression, HRSD [6] и др. Кроме того, профессиональная оценка может варьироваться в зависимости от компетентности специалиста и методов диагностики, которые он использует. Таким образом, на данный момент не существует объективного метода диагностики депрессии.

Многие работы представляют автоматические системы для определения состояния депрессии — существуют как одномодальные, так и многомодальные системы. Кроме того, часть систем решает задачу регрессии (определяя степень тяжести заболевания), а часть — задачу бинарной классификации (для определения наличия заболевания или его отсутствия). Задача определения депрессии была неоднократно представлена на соревнованиях AVEC (Audio-Visual Emotion Challenge) в 2013 [7], 2014 [8], 2016 [9], 2017 [10] и 2019 годах [11].

Отмеченные в данной работе системы опираются на гипотезу о том, что эмоциональное состояние диктора оказывает существенное влияние на акустические характеристики (спектральные и просодические) его речи. Лингвистические и нелингвистические факторы оказывают влияние на фонетические характеристики речи. Среди таких факторов можно отметить: физическое и психическое состояния говорящего, различные патологии мышления и психические болезни, ряд болезней, влияющих непосредственно на возможность речеобразования и др. [12].

Данная работа проведена с целью показать значимость существующих проблем, которые вызваны депрессией, описать применяемые на данный момент подходы, которые используются специалистами, и выполнить аналитический обзор существующих автоматических систем определения депрессии.

В разделе 2 приводится определение депрессии, влияние заболевания на различные аспекты жизни человека, описание используемых на данный момент специалистами методов выявления депрессии. В разделе 3 описаны существующие базы данных, на основе которых возможно обучение автоматической системы определения депрессии. Одно- и многомодальные системы, представленные за последние два года, описаны в разделах 4 и 5. Раздел 6 содержит выводы, сделанные в результате аналитического обзора, также представлена классификация наиболее

эффективных методов для построения таких систем. На основе проведенного анализа в главе 6 также описаны требования, которые могут быть выдвинуты к автоматическим системам определения депрессии. Выводы по проделанной работе описываются в разделе 7.

2. Основные определения и медицинские характеристики депрессии. Построение автоматической системы включает в себя в том числе и понимание рассматриваемой задачи. В случае с определением депрессии возможна векторизация и использование некоторых признаков депрессии для обучения моделей, а именно тех признаков, которые проявляются вербально и невербально, и на которые обращает внимание специалист при личной беседе с пациентом.

Аффективные состояния — это психические состояния, которые характеризуются заметной эмоциональной окрашенностью: эмоциональные состояния, состояние аффекта, настроение и т.п. Изменения аффективного состояния являются естественной характеристикой поведения людей. Однако, когда эти изменения становятся интенсивными, длятся продолжительное время и при этом ухудшается жизнедеятельность человека, есть вероятность, что может проявиться аффективное расстройство. В отличие от кратковременных эмоций, настроение — длительное по времени аффективное состояние и, следовательно, клиническая депрессия — это расстройство настроения, которое может длиться неделями, месяцами, и даже годами, изменяясь в тяжести заболевания, если не получено соответствующее лечение.

Расстройства настроения, несомненно, касаются естественных эмоциональных состояний. В частности, схема поведения людей, страдающих от таких расстройств настроения, как униполярная депрессия, показывает сильную временную корреляцию с аффективными величинами валентности, активации и доминанции [8]. Данные аффективные величины используются для определения эмоционального состояния человека. Специалисты выделяют два противоположных аффективных расстройства (расстройств настроения) — депрессию (или большое депрессивное расстройство, БДР) и манию [1]. В психиатрии они обозначаются при помощи терминов униполярная депрессия (пациентов беспокоит депрессивное состояние) и биполярное аффективное расстройство (БАР, пациенты переживают как депрессию, так и манию). При этом депрессия и маниа могут проявляться одновременно, что приводит к смешанному аффективному эпизоду. Кроме того, маниа и депрессия могут проявляться в менее тяжелой форме (гипомания и дистимия соответственно), или маниа и депрессия могут быстро сменяться, что называют быстрой циркуляцией фаз. На рисунке 1 показаны течения описанных заболеваний, приведенные

в работе [13] (для удобства восприятия два графика течения заболеваний были объединены в один, проведен перевод надписей на графиках, а также график был представлен в монохромном виде). Верхняя граница рисунка обозначает состояние мании, нижняя – состояние депрессии, а средняя – нормальное состояние. Пунктиром показаны менее тяжелые состояния гипомании и дистимии, которые также являются отклонением от нормального состояния, по горизонтальной оси указано течение времени, а по вертикальной – валентность заболевания.

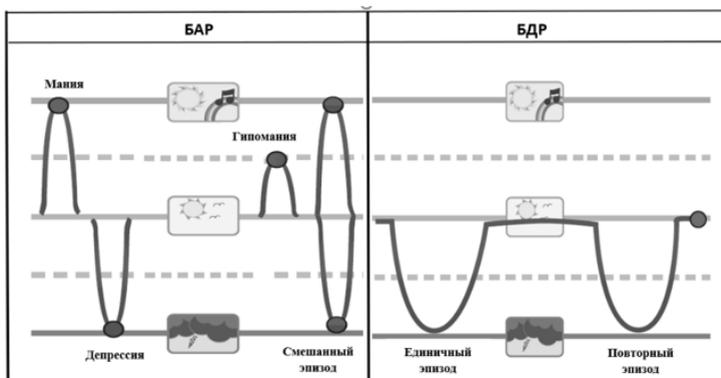


Рис. 1. Фазы течения заболевания при БАР и при БДР [13]

Согласно диагностическому и статистическому руководству по психическим расстройствам V издания (Diagnostic and Statistical Manual of mental disorders V, DSM-V) [14] для постановки диагноза депрессии необходимо, чтобы на протяжении как минимум двух недель присутствовали 5 или более симптомов (включая как минимум один из основных: подавленное настроение и/или потеря интереса и утрата способности получать удовольствие от приятной ранее деятельности):

- подавленное настроение;
- потеря интереса и утрата способности получать удовольствие от приятной ранее деятельности;
- расстройства сна и аппетита;
- психомоторное возбуждение/заторможенность;
- повышенная утомляемость и снижение энергичности;
- сниженная самооценка, чувство никчемности или неадекватное чувство вины;
- снижение способности к концентрации внимания или заторможенное мышление;
- суицидальные тенденции.

Крайней формой выражения депрессии является самоубийство, а риск самоубийства у пациентов с депрессией в течение жизни составляет 15%. В работе [15] был проведен сравнительный мета-анализ, который показал 3428 факторов риска суицида среди 365 лонгитюдных (длительных) исследований. Авторы сделали вывод, что все обозначенные факторы риска недостаточно точно могут предсказать суицид, что частично может быть вызвано методологическими ограничениями существующей литературы. По мнению авторов работы [15], исследования показателей оценивания рисков суицида с использованием клинического инструментария, предсказывающего суицид, также недоработаны, а последние систематические обзоры показывают, что на данный момент нет инструментария для определения риска суицида, который показывал бы высокую точность. В работе [16] был проведен статистический анализ на основе англоязычных рецензированных статей. В рассмотренных авторами исследованиях предикативные модели должны были предсказать смерть путем самоубийства или попытку самоубийства. Была проведена серия симуляций на гипотетической популяции индивидов с целью выявления преимуществ в статистическом моделировании, которые могли бы повысить способность предсказания попыток суицида и смертей. Было отмечено, что системы предсказания суицидального поведения должны проходить клинические тесты до внедрения в систему здравоохранения, чтобы показывать высокую точность именно в клинических случаях, а также предполагать более общие результаты, включая суицидальные тенденции, поскольку ошибочные гипотезы системы (как ложноположительные, так и ложноотрицательные) могут навредить пациентам.

Задача определения депрессии относится к задаче определения деструктивных паралингвистических явлений, которые также включают в себя такие явления, как агрессия, ложь и пр. Связь между агрессией и депрессией в известных теориях выражена неявно. Редкие случаи депрессии предположительно были обнаружены в обществах и группах, где агрессия могла быть сразу выражена, например, в военных сражениях и там, где процент случаев насильственной смерти велик. Тем не менее, немногие межкультурные работы, существующие на момент написания статьи [17], не поддерживали гипотезу об обратной корреляции между агрессией и депрессией. Большинство рассмотренных автором статьи [17] работ подтверждали, что обязательным компонентом депрессии является чувство вины. Несмотря на то, что автор утверждает, что рассмотренные им работы имели методологические недоработки, сочетание признаков депрессии (депрессивное настроение, суточные циркадные изменения, усталость, бессонница, потеря интереса, потеря веса, периодичность и

двухфазовая природа заболевания) все-таки обнаруживается во всех культурах.

Перед автоматическими системами определения депрессии, в том числе, стоит задача проверки данной гипотезы, а также использование выявленных корреляций с другими деструктивными паралингвистическими явлениями и аффективными величинами для получения более точного предсказания автоматической системы о наличии или отсутствии заболевания.

3. Многомодальные базы данных для исследования депрессии. Для создания автоматической системы определения депрессии необходимо иметь данные для обучения моделей, которые содержат как речь информантов с установленной депрессией, так и информантов, у которых не было депрессии на момент записи. Среди работ, которые будут рассматриваться в данном аналитическом обзоре, встречаются одномодальные и многомодальные корпуса. В таблице 1 представлены систематизированные данные об описанных корпусах, а именно: язык, количество дикторов / авторов текстов, методы оценивания заболевания, количество данных, доступность. Большинство корпусов является многомодальным, однако описаны также и одномодальные корпуса: Mundt и SH2-FS (Free Speech) (аудио), RusNeuroPsych и Текстовый корпус эссе (текст), корпус информации из профилей социальной сети «Вконтакте» (Текст, изображения). Кроме того, некоторые многомодальные корпуса содержат лингвистическую информацию: Pitt (ручная транскрипция), DAIC (автоматическая транскрипция), General Psychotherapy Corpus (транскрипция). В корпусе DAIC также содержатся записи сенсора глубины и физиологические данные (гальваническая проводимость кожи, ЭКГ, дыхание). В столбце Язык сокращения "Австрал." от "Австралийский" и "Амер." от "Американский"

AViD-Corpus [7] содержит 340 аудио- и видеозаписей, где участники (292 участника) взаимодействуют с компьютером, в то время как их действия записываются камерой и микрофоном. Количество записей каждого участника варьируется от 1 до 4. Длина видео составляет от 20 до 50 минут, в среднем — 25 минут. Общая длительность видео составляет 240 часов. Возраст участников — от 18 до 63 лет, средний возраст — 31,5 лет. Поведение участников в ходе записей было определено заданием: произнесение букв, произнесение букв громким голосом, произнесение букв с улыбкой, повышение громкости голоса при выполнении задания.

Участники читали отрывки из новелл и преданий, пели, рассказывали истории из своего прошлого, рассказывали вымышленные истории для тематического апперцепционного теста (апперцепция – восприятие, узнавание на основе прежних переживаний). Аудиоданные были записаны

Таблица 1. Сравнение данных в корпусах, содержащих данные людей с депрессией

Корпус	Язык	Количество человек	Оценка заболеваний	Количество данных	Доступность
AViD-Corpus [7]	Немецкий	292	Шкала Бека-2	240 часов	По запросу
BlackDog [19]	Австрал. английский региолект	30	QIDS-SR	509 минут	Не доступен
Pitt [20]	Амер. английский региолект	19	HRSD, QIDS-SR	355 минут	Не доступен
Distress Analysis Interview Corpus (DAIC) [21]	Амер. английский региолект	Не известно	5 опросников	621 сессия	По запросу
Mundt [23]	Амер. английский региолект	35	Текущее лечение, шкала Гамильтона	Не известно	Не доступен
General Psychotherapy Corpus [24]	Амер. английский региолект	Не известно	Оценка специалиста	1300 сессий	Платный доступ (30 дней бесплатного пользования)
SH2-FS (Free Speech) [25]	Амер. английский региолект	887	PHQ-9	16 часов, 5937 аудио файлов	Не доступен
RusNeuroPsych [22]	Русский	447 (246 детей до 18 лет и 209 взрослых)	3 опросника	643 (252 текста детей и 392 текста взрослых)	Открыт
Текстовый корпус эссе [46]	Русский	164	10 опросников	164 текста	Не доступен
Корпус информации из профилей социальной сети «Вконтакте» [47]	Русский	1330	Шкала депрессии Бека	Не известно	Не доступен

с использованием микрофонной гарнитуры с частотой 41 кГц, 16 бит. Видеоданные были записаны с использованием различных кодеков и кадровых частот, а затем сконвертированы в формат до 30 кадров в секунду с разрешением 640x480 пикселей, 24 бита на пиксель. Аннотация проводилась согласно опроснику депрессии, шкале Бека-2 [18]. Данный опросник является вторым пересмотром опросника Бека, принятым в 1996 году. Он содержит 21 вопрос, где каждый вопрос оценивается по шкале от 0 до 3

в соответствии с тяжестью симптома. Суммарный балл составляет от 0 до 63, при том, чем выше значение, тем серьезнее симптомы депрессии. Средний уровень депрессии по шкале Бека у информантов данного корпуса был 15,1 для обучающего набора и 14,8 — для отладочного набора, стандартное отклонение 12,3 и 11,8 соответственно. Каждой записи было присвоено единственное значение метки наличия депрессии.

BlackDog [19] – англоязычный корпус (австралийский английский региолект), собранный в Австралии при помощи 1 камеры и 1 микрофона. Частота дискретизации аудио 44,1 кГц. Метки бинарной классификации были проставлены вручную и содержат категории «тяжелая депрессия» и «здоровые субъекты». Всего было записано 30 информантов для каждого класса, количество мужчин и женщин сбалансированное (30 и 30 соответственно). Интервью содержало вопросы с открытым ответом, а симптомы депрессии измерялись при помощи QIDS-SR. Средний показатель тяжести депрессии у информантов данного корпуса — 19 при разбросе от 14 до 26. Общая длительность записей — 509 минут, из которых 119,3 минуты – чистая речь информантов.

Pitt [20] — англоязычный корпус (американский английский региолект), собранный в Питтсбурге при помощи 4 камер и 2 микрофонов. Частота дискретизации аудио 48 кГц. Метки бинарной классификации содержат категории «тяжелая депрессия» и «легкая депрессия». Кроме того, проводилась ручная транскрипция. Всего было записано 19 информантов для каждого класса, количество мужчин и женщин — 14 и 24 соответственно. Опрос содержал вопросы из клинического интервью HRSD. Средние показатели тяжести депрессии у информантов данного корпуса — 22,4 при диапазоне от 17 до 35 для тяжелой депрессии, и — 2,9 при разбросе от 1 до 7 для легкой депрессии. Эквивалентные показатели по шкале QIDS-SR: средний показатель — 17 при диапазоне от 13 до 26 для тяжелой депрессии, и — 2 при диапазоне от 1 до 5 для легкой депрессии. Общая длительность записей составляет 355 минут, из которых 92 минуты – чистая речь информантов.

Distress Analysis Interview Corpus (DAIC) [21] — многомодальная коллекция клинических интервью. Корпус разработан для симуляции стандартных процессов определения того, имеется ли у человека риск ПТСР (посттравматического стрессового расстройства) и большого депрессивного расстройства. Корпус содержит следующие типы интервью:

- очные интервью (лицом к лицу) между участниками и интервьюером;
- телеконференции, где интервью проводилось с использованием телеконференций;

- «Волшебник Оз» (WoZ) или «Гудвин» (Wizard of Oz, WoZ) – интервью проводилось анимированным виртуальным интервьюером по имени Элли, которую контролировал интервьюер в другой комнате;
- автоматические интервью – интервью проводилось в автоматическом режиме с Элли.

Разработчиками были выбраны две группы жителей Лос-Анджелеса – ветераны военных сил США и гражданские. Все они были проверены опросниками на депрессию, ПТСР и тревожность. Интервью начинались с нейтральных вопросов, затем вопросы становились более специфичными (о симптомах, событиях), а заканчивались фазой спокойствия. Корпус включает в себя аудио-, видеоданные и записи сенсора глубины всех взаимодействий. В коллекции также присутствуют физиологические данные (гальваническая проводимость кожи, ЭКГ, дыхание). Перед и после интервью участники заполняли ряд опросников, включающие базовые вопросы о биографии, измерение психологического стресса, а также измерение текущего настроения. Использовались следующие опросники: The Positive and Negative Affect Scale (PANAS) для оценки настроения, PTSD Checklist – Civilian Version для оценки ПТСР, Patient Health Questionnaire для оценки психического здоровья, Depression module для оценки наличия и уровня депрессии, State-Trait Anxiety Inventory для оценки тревожности.

При экспериментах по обучению системы определения депрессии и ПТСР по записям речи из данного корпуса авторам корпуса удалось добиться точности распознавания депрессии в 75,0%, а ПТСР – 72,0%. В записях с волшебником Оз в данных также были найдены некоторые признаки стресса: участники, испытывающие стресс, медленнее начинали говорить, и использовали меньше заполненных пауз, чем участники, не испытывающие стресс. Кроме того, от типа стресса зависело какие признаки наиболее предсказуемы. Если стандартное отклонение перед началом речи каждого участника диалога изменялось, это было лучшим признаком для предсказания депрессии. Однако для ПТСР более информативным было определение среднего количества заполненных пауз в сегменте. Время перед началом ответа при личных вопросах и длительность речи при ответах на вопросы для установления контакта указывают на наличие стресса. Участники использовали меньше заполненных пауз при диалоге с агентом, чем при диалоге с человеком. Участники выражали меньше страха или негативных проявлений когда агент представлялся автоматическим, чем когда агент был представлен как управляемый человеком. Кроме того, участники показывали больше эмоций из категории «грусть», когда были уверены, что взаимодействуют с компьютером, а не с человеком.

RusNeuroPsych [22] – текстовый русскоязычный корпус, который содержит 643 текста на русском языке 447 авторов в возрасте от 12 до 35 лет. Корпус разделен на две части – подкорпус «дети» (тексты написаны детьми школьного возраста от 12 до 17) и подкорпус «взрослые» (тексты написаны взрослыми от 18 до 35 лет). Первый подкорпус содержит 252 текста 246 человек, а в метаданных указан их пол, год рождения, исследования их моторного, сенсорного и латерального профиля, а также результаты психологического тестирования, которое использовалось для измерения уровня агрессии, тревожности, ригидности и фрустрации. Второй подкорпус состоит из 392 текстов 209 человек, а в метаданных указан идентификатор участника, пол, год рождения, образование, результаты теста большой пятерки, исследования их моторного, сенсорного и латерального профиля, а также результаты шкалы тревоги и депрессии.

Mundt [23] – корпус содержит записи речи за период в 6 недель 35 информантов (20 женщин и 15 мужчин), чей средний возраст 41,8 лет. Одним из условий записи было начало фармакотерапевтического и/или психотерапевтического лечения депрессии. Участники читали заранее подготовленный текст и описывали свои эмоциональные и физические ощущения. Для оценки наличия или отсутствия депрессии использовалась шкала Гамильтона.

General Psychotherapy Corpus [24] – корпус состоит из 1300 транскрибированных терапевтических сессий, которые покрывают различные клинические подходы. Метаданные представлены на уровне сессий и включают демографические сведения для терапевта и пациента, симптомы, которые испытывает пациент, и сводную информацию, прилагающуюся к каждой сессии, обозначаемую как «название». Каждая сессия состоит из следующих друг за другом коммуникативных ходов, аннотированных как сторона терапевта и сторона клиента. Среди 1262 сессий 881 аннотированы как «без депрессии» и 381 как «с депрессией».

SH2-FS (Free Speech) [25] – содержит записи речи в естественных условиях (дома, на работе, в машине) и оценки по самодиагностическому тесту PHQ-9. Общая длительность аудио в корпусе 16 часов речи 887 участников (436 женщин и 450 мужчин), всего 5937 аудио файлов. На всех записях присутствует фоновый шум.

Текстовый корпус эссе [46] – корпус эссе длиной в одну страницу на тему «Я, другие, мир», записанный с целью определить лингвистические характеристики текстов людей с установленной депрессией и отсутствием депрессии. Также участников просили заполнить 10 опросников. В исследовании участвовали две группы испытуемых: 22 пациента клиники ФГБНУ НЦПЗ с установленной депрессией и 142 студента гуманитарных

и технических вузов Москвы и Кургана, а также взрослые жители этих городов. Корпус информации из профилей социальной сети «Вконтакте» - данные профилей были собраны в период с января 2017 по апрель 2019, также были получены баллы участников по шкале депрессии Бека. Всего была получена информация 1330 профилей, 425 мужчин и 904 женщин, средний возраст 25 лет, а средний показатель по шкале Бека 18,79. Также в корпусе имеется 485121 изображений, собранных из альбомов, аватаров и постов в профилях социальной сети «Вконтакте» 398 волонтеров. В контрольной группе был 201 профиль, а в группе с депрессией – 197.

4. Обзор работ, представленных на соревнованиях AVEC. В рассмотренных далее работах, в основном, исследуется униполярная депрессия или БДР. Значительная часть работ представлена в рамках соревнований по аудиовизуальному определению эмоций AVEC. В общем виде процесс построения автоматической системы для многомодального определения паралингвистических явлений заключается в следующем: имеется многомодальная база данных, содержащая видеозаписи информантов с диагностированной депрессией, и информантов с отсутствием депрессии, а также лингвистическая составляющая их речи. Из данных каждой модальности вычисляются информативные признаки, которые впоследствии являются входными данными для классификатора/регрессора. Выходными же данными классификатора/регрессора являются либо метка класса, либо единственное значение регрессора для классификации и для регрессивной классификации соответственно. Лучшие результаты систем, представленных на соревнованиях 2019 года, отображены в таблице 2. В ней указаны авторы системы, модальности и признаки, которые были использованы для обучения моделей, а также сами модели и результаты классификации, которых удалось добиться авторам по показателям точности CCC (Concordance Correlation Coefficient [27]) и RMSE (Root Mean Squared Error [27]). В качестве сокращений указаны Dev – набор данных для разработки (Development set), а Test – набор данных для тестирования (Test set).

В соревнованиях 2019 года [29] были представлены следующие темы: определение настроения (State-of-Mind Sub-challenge), использование искусственного интеллекта для определения депрессии (Detecting Depression with AI Sub-challenge) и определение эмоций в разных культурах (Cross-cultural Emotion Sub-challenge). В качестве набора данных для определения наличия депрессии был представлен корпус E-DAIC, расширенная версия WOZ-DAIC [21]. Аудио признаки состояли из следующих наборов: MFCC (Mel-frequency Cepstral Coefficients), eGeMAPS, BoAW (Bag of Audio Words), Deep Spectrum. В качестве видео признаков

Таблица 2. Сравнение результатов систем, представленных на соревнованиях AVEC'19

Работа	Модальность	Признаки/ Классификатор	CCC (Dev / Test)	RMSE (Dev / Test)
Базовая работа Ringeval F. et al. [29]	Аудио	MFCCs	0,198 / -	7,28 / -
		eGeMAPS	0,076 / -	7,78 / -
		BoAW-M	0,102 / -	6,32 / -
		BoAW-e	0,272 / 0,045	6,43 / 8,19
		DS-DNet	0,165 / -	8,09 / -
	Видео	DS-VGG	0,305 / 0,108	8,00 / 9,33
		FAUs	0,115 / 0,019	7,02 / 10,0
		BoVW	0,107 / -	5,99 / -
	Аудио + Видео	ResNet	0,269 / 0,120	7,72 / 8,01
Аудио + Видео	VGG	0,108 / -	7,69 / -	
Аудио + Видео	Все признаки	0,336 / 0,111	5,03 / 6,37	
Кава Н. et al. [30]	Аудио + Видео	Объединение разработанных классификаторов	0,481 / 0,344	-
Ray A. et al. [31]	Аудио	Funct MFCC	- / -	5,11 / -
	Видео	BoVW	- / -	5,70 / -
	Текст	Текст	- / -	4,37 / -
Makiuchi M.R. et al. [32]	Аудио	CNN	0,338 / 0,199	5,97 / 7,02
		GCNN-LSTM	0,497 / -	5,70 / -
	Текст	LSTM	0,360 / 0,048	4,97 / 6,88
		8 CNN blocks-LSTM	0,685 / -	4,22 / -
	Видео	GCNN	0,372 / -	5,74 / -
	Аудио + Текст	CNN and LSTM	0,452 / 0,213	5,08 / 6,42
	Аудио + Текст	GCNN-LSTM и 7 CNN blocks-LSTM	0,696 / 0,403	3,86 / 6,11
Аудио + Текст + Видео	GCNN-LSTM, 7 CNN blocks-LSTM и GCNN	0,624 / -	4,86 / -	
Fan W. et al. [33]	Аудио + Видео + Текст	Ансамбль классификаторов	0,466 / 0,430	5,07 / 5,91
Yin S. et al. [34]	Аудио + Видео + Текст	Иерархическая двунаправленная LSTM	0,402 / 0,442	4,94 / 5,50

использовались FAU (Face Action Units), BoVW (Bag of Video Words), ResNet признаки. Базовая модель организаторов представляла собой однослойную 64-d (размерную) сеть с управляемыми рекуррентными блоками (Gated Recurrent Units, GRU) в качестве рекуррентной сети, за которой следует 64-d полносвязный слой для получения единственного значе-

ния оценки регрессии. Лучший результат на отладочном наборе данных для аудио модальности был получен при использовании признаков Deep Spectrum, вычисленных с использованием сети VGG-16, по показателю CCC = 0,289. Для видео модальности лучший результат был получен при использовании ResNet признаков: CCC = 0,269, RMSE = 7,72. При объединении всех репрезентаций результат, полученный на отладочном наборе, был улучшен до значений CCC = 0,336 и 0,111 на отладочном и на тестовом наборах соответственно. Значения RMSE также были улучшены до 5,03 и 6,37 на отладочном и тестовом наборах.

Работа [30] подготовлена нашим коллективом авторов совместно с и зарубежными коллегами. В ней мы аугментировали акустические признаки, предложенные организаторами при помощи транскрипций, полученных при помощи автоматического распознавания речи (Automatic Speech Recognition, ASR). Затем эти транскрипции были использованы для простого представления мешка слов (Bag of Words, BoW), после чего применялся анализ главных компонент для регрессии. Авторы использовали продолжительность реплик из транскрипций для получения общей длительности тишины и дыхания для каждого участника. Для моделирования автоматической сегментации на 7 классов (речь виртуального агента, дыхание, эксплетивные слова, звук губ, смех, тишина, речь субъекта) авторы экспериментировали с признаками eGeMAPS и Deep Spectrum (VGG-16), где применяли оконный метод (100, 200, 400, 500, 1000 мс) с наложением в 1 секунду. Супрасегментные признаки были смоделированы при использовании KELM (Kernel Extreme Learning Machine). Наилучший результат (CCC = 0,344) на тестовом наборе был получен при использовании простых признаков на основе транскрипции ASR (подсчет слов, длительность и BoW). Помимо KELM использовалась ELM с взвешенным ядром (Weighted Kernel ELM), которая присваивает веса высокой важности классу с наименьшим количеством объектов обучения и затем пытается максимизировать невзвешенную среднюю полноту (Unweighted Average Recall, UAR). Лучший результат сегментации на отладочном наборе в 4 из 17 аудио файлов, полученных с 500мс окном без наложения, с использованием функционалов eGeMAPS LLDs и Weighted KELM, был UAR=65,75%. Результаты показали, что, хотя паттерны в части невербальных признаков сигнала важны, объединение их с лингвистической информацией позволяет добиться лучших результатов без использования современных акустических и видео признаков.

Авторы работы [31] объединили признаки трех модальностей и использовали многоуровневую сеть с вниманием, которая обучалась взаимосвязям как между модальностями, так и внутри модальностей. Сеть

использует несколько низкоуровневых и среднеуровневых признаков из аудио- и видео-модальностей, а также векторные представления предложений (sentence embeddings). В архитектуре предложенной сети контекстные признаки каждой модальности проходят через двуслойные сети прямого распространения, а выходные данные этих трех сетей объединяются в другую последовательную двунаправленную LSTM (Long-Short Term Memory), а именно, stacked BLSTM (Bidirectional LSTM). При помощи предложенной системы результат базовой системы был улучшен на 17,52% по показателю RMSE. Отдельные сети с вниманием для аудио и видео модальностей превосходили результат базовой системы по показателю UAR на 20,5%, а текстовая модель с вниманием превзошла лучший результат на 8,95%.

В работе [32] авторы использовали аудио модальность и ее лексическую составляющую для определения депрессии. Для векторизации аудио модальности использовались признаки deep spectrum из предобученной сети VGG-16 и применялась Gated Convolutional Neural Network (GCNN), а затем LSTM слой. Для получения лингвистических представлений были извлечены признаки BERT и применена CNN, за которой следовал LSTM слой. Извлечение лингвистических признаков проводилось при помощи BERT модели, в которой 24 слоя (блоков трансформера), 1024 скрытых слоя и 16 блоков «self-attention». Таким образом, всего получается 340 млн. параметров, которые извлекаются из последнего слоя BERT и преобразуются в единый массив признаков для каждого токена (word token). Аудиомодель на основе GCNN-LSTM состоит из последовательных управляемых сверточных блоков, после которых идет слой LSTM и полносвязный слой. Лингвистическая модель была построена на основе CNN-LSTM. При использовании предложенного подхода удалось получить результаты $CCC = 0,696$ на отладочном наборе и $CCC = 0,403$ на тестовом наборе.

Авторы работы [34] в качестве видео признаков использовали: направление взгляда, 3D положение и ориентацию головы, 17 FAUs для каждого фрейма с весами точности. Для аудио модальности использовались две категории признаков – 4096-размерный вектор, вычисленный из активации второго полносвязного слоя VGG-16, и 1920-размерный вектор, вычисленный из активации последнего слоя со сжатием в DenseNet-201. Для текстовой модальности использовалась RNN с кодировщиком-декодировщиком для создания вектора семантической репрезентации. Для создания вектора эмоциональной репрезентации использовался словарь NRC Emotion Intensity Lexicon. Предложенный подход включает в себя две иерархические двунаправленные LSTM для объединения многомодальных

признаков и предсказания тяжести депрессии. На отладочном и тестовом наборах были достигнуты результаты 0,402 и 0,442 по показателю CCC; 4,94 и 5,50 по показателю RMSE, соответственно.

5. Обзор работ по автоматическому определению депрессии, представленных вне соревнований AVEC. Результаты систем, представленных вне соревнований AVEC, отображены в таблице 3. В ней указаны авторы системы, модальности, которые были использованы для обучения моделей, а также сами модели. Перечислены различные показатели точности, которые были использованы авторами, и результаты, полученные по этим показателям с использованием разработанных моделей. В работе [43] в столбце Классификатор под "Первая модель" подразумевается модель SVM, модели tf-idf со стилистическими и морфологическими признаками, а под "Вторая модель" подразумевается модель SVM, векторные представления слов, стилистические признаки.

В работе [35] для обучения моделей авторы использовали DAIC-WOZ корпус, а именно трехмерные изображения лица и речь информантов. Методика включает в себя использование векторных представлений на уровне предложений (sentence-level "summary" embedding), LSTM и casual-CNN. Предсказывался показатель PHQ и бинарная классификация, о наличии у пациента БДР, обычно со значением PHQ более или равным 10, либо же отсутствие заболевания. Особенность предложенного метода заключается в том, что предложенная система не полагается на контекст интервью. Кроме того, она принимает на вход сырые данные (аудио, трехмерные модели лиц и транскрипция), которые суммируются в один вектор. Стоит отметить, что в подходе используются сделанные вручную и предобученные векторные представления на уровне слов на входе, то же самое и на уровне предложений. Модель показывает среднюю ошибку MAE = 3,67 по опроснику здоровья (PHQ), а также 83,3% чувствительности (Sensitivity [28]), 82,6% специфичности (Specificity) и F1-меру 76,9%.

В работе [36] также использовался DAIC-WOZ корпус. Авторы предложили следующую архитектуру (CombAtt): кодировщики модальностей, которые принимают унимодальные признаки на вход и выдают закодированные данные; сети с механизмом внимания для объединения сетей отдельных модальностей. Также они предложили сети для регрессии, которые принимают на вход объединенные данные из второго компонента и на выходе предсказывают баллы по шкале PHQ-8. Использование предложенного подхода позволило авторам получить результаты RMSE = 4,14 и MAE = 3,07, а также EVS (explained variance score) = 0,62.

Таблица 3. Лучшие результаты систем, представленных вне соревнований AVEC

Работа	Модальность	Классификатор	Показатель	Результат
Haque A. et al. [35]	Аудио + Видео + Текст	Casual CNN	F1-мера,	76,9%,
			Precision,	71,4%,
			Recall,	83,3%,
			Average Error	3,67
Qureshi S.A. et al. [36]	Аудио + Видео + Текст	CombAtt network	RMSE,	4,14,
			MAE,	3,34,
			EVS	0,62
Niu M. et al. [37]	1.7cmАудио	Гибридная сеть (CNN, LSTM и DNN) и l _p -нормированное сжатие	RMSE,	9,66,
			MAE	8,02
Rohanian M. et al. [38]	Аудио + Видео + Текст	LSTM с механизмом окна	F1-мера,	81,0%,
			Precision,	80,0%,
			RMSE	3,61,
			MAE,	4,99
Tao F. et al. [39]	Аудио	SVM	Accuracy	84,5%
Xezonaki D. et al. [40]	Текст	Двухуровневая иерархическая нейронная сеть с механизмом внимания	F1-мера (General Psychotherapy Corpus),	71,6%,
			F1-мера (DAIC-WOZ),	70,3%,
			UAR (DAIC-WOZ)	70,3%
Huang Zh. et al. [41]	Аудио	FVTC-CNN	UAR (SH2-FS),	68,0%,
			UAR (DAIC-WOZ)	88,0%
Zhao Z. et al. [42]	Аудио	Гибридная сеть (сеть с механизмом внутреннего внимания, глубокая сверточная сеть, SVR)	MAE (Corpus 2013),	9,65,
			RMSE (Corpus 2013),	7,38,
			MAE (Corpus 2014),	9,57,
			RMSE (Corpus 2014)	7,94
Seneviratne N. et al. [43]	Аудио	SVM	Accuracy	81,7%
Stankevich N. et al. [44]	Текст	Первая модель; Вторая модель	F1-мера,	63,0%,
			Precision,	65,0%,
			Recall;	61,0%;
			Recall,	84,0%,
Enikolopov S.N. et al. [45]	Текст	Метод случайного леса, психолингвистические признаки и биграммы	F1-мера	73,0%

В работе [37] в качестве данных для обучения использовались два свободно доступных набора данных: AVEC2013 и AVEC2014. После вычисления и сегментирования MFCC авторы использовали l_p -нормированное сжатие, объединенное с LASSO (Least Absolute Shrinkage and Selection Operator), чтобы найти оптимальный параметр сжатия с целью последующей генерации признаков на уровне высказываний для определения депрессии. Эти данные использовались для обучения гибридной модели, которая содержит CNN, LSTM и DNN, а итоговое предсказание уровня депрессии проводилось с использованием SVR (Support Vector Regression). В результате экспериментов было выявлено, что l_p -нормированное сжатие с LASSO (где параметр $p = 4,06$ в AVEC2013 и $p = 2,13$ в AVEC2014) оказалось наилучшим вариантом сжатия. На тестовом наборе AVEC2013 были получены результаты $RMSE = 9,79$, $MAE = 7,48$, а на тестовом наборе AVEC2014 – $RMSE = 9,66$, $MAE = 8,02$.

Авторы работы [38] для обучения моделей использовали многомодальный корпус DAIC-WOZ. Для данного исследования авторы решили прибегнуть к искусственному выравниванию между текстовыми, аудио- и видео признаками, чтобы получить точные временные метки каждого произнесенного слова. На каждой временной отметке они выравнивали слова и соответствующие им отрезки аудио с использованием инструментария Penn Phonetics Lab Forced Aligner (P2FA), который может применяться для сравнения транскрипций с аудиофайлами, фонема за фонемой. Путем ручной проверки искусственное выравнивание было проделано с достаточно высокой точностью для изучения объединения модальностей. В модели использовались highway слои (имплементация стохастического градиентного спуска) прямого распространения с оконным механизмом, которые обучаются регулировать поток информации в сети, присваивая веса видео и аудио входным данным на каждой временной отметке. Результаты обучения LSTM с механизмом окна на признаках трех модальностей: F1-мера = 81,0%, Precision = 80,0%, MAE = 3,61, RMSE = 4,99.

В работе [39] для обучения системы использовали записи 110 человек, из которых 54 никогда не имели психических заболеваний и были обозначены как контрольная группа, а у 56 была диагностирована депрессия, они были определены в группу людей с депрессией. В группе информантов с депрессией у информантов имелись следующие заболевания: большое депрессивное расстройство (19 случаев), биполярное расстройство в депрессивной фазе или с последним депрессивным эпизодом (13 случаев), реактивная депрессия (7 случаев), эндо-реактивная депрессия (6 случаев) и тревожно-депрессивное расстройство (4 случая). Для остальных 7 информантов точный диагноз не был установлен.

Все участники являлись носителями итальянского языка, их просили прочитать вслух басню Эзопа «Ветер и Солнце». В качестве акустических признаков использовался набор Interspeech 2009 Emotion Challenge, который был расширен путем добавления признаков для определения скорости чтения и использования пауз. Обучение классификатора SVM, имплементированного в библиотеке Scikit-learn, происходило при помощи техники leave-one-out. Авторы предположили, что люди с депрессией читают медленнее и чаще используют длинные паузы. Данное предположение подкрепляется исследованиями нейробиологов, которые показывают, что течение процессов в мозге, связанных с языком, занимают больше времени у людей с депрессией. В частности, было показано, что есть связь между депрессией и дисфункцией в некоторых зонах, участвующих в семантической обработке языка, включая фронтальную извилину и префронтальный кортекс. В проведенных экспериментах среднее время на чтение текста и стандартное отклонение 54,92 +/- 2,66 сек и 47,38 +/- 1,20 сек для лю-дей с депрессией и без соответственно, а скорость чтения 202,1 и 234,3 слова в минуту. Так, после добавления этих признаков, авторам удалось улучшить точность распознавания депрессии с 68,2% до 84,5%.

В работе [40] для обучения использовались корпуса General Psychotherapy Corpus и DAIC-WoZ. Авторы предложили подход с использованием иерархической архитектуры нейронной сети с вниманием для определения депрессии по транскрипциям клинических интервью. Авторы предположили, что эмоциональное содержание может быть отличительным фактором между языками людей с депрессией и без. Основываясь на этом, они применили внешние лингвистические знания об эмоциональном содержимом слов, рассмотрев эмоции, тональность, валентность и психолингвистическую аннотацию для слов. Для того, чтобы исследовать использование слов, которые отражают позитивную и негативную тональности, грусть и тревожность, авторы использовали инструментарий LIWC lexicon, в котором представлена психолингвистическая аннотация 18504 слов для 73 различных категорий слов. Эксперименты показали, что дополнительная информация об эмоциях улучшает результат предложенной архитектуры. Авторам удалось добиться F1-меры 71,6% для корпуса General Psychotherapy Corpus, а также F1-меры и невзвешенной средней полноты 70,3%: для корпуса DAIC-WOZ.

Работа [41] посвящена проблеме обобщаемости и предлагает несколько стратегий адаптации, которые модернизируют предобученные модели на основе расширяемых сверточных сетей с целью улучшить точность определения депрессии как в лабораторных, так и в естественных

условиях. Для обучения сетей использовались два корпуса: SH2-FS (Free Speech) и DAIC-WOZ. Авторы использовали четыре набора акустических признаков: 3 форманты, 13 спектральных центроидных частот, 16 MFCC и 16 дельт MFCC. В работе исследуется метод FVTC-CNN (full vocal tract coordination – convolutional neural networks). Он состоит из двух частей: матрицы по типу изображений FVTC и расширяемых CNN. Авторам удалось добиться точности по невзвешенной средней полноте 68,0% для корпуса SH2-FS и 88,0% для корпуса DAIC-WOZ.

В работе [42] изучались преимущества гибридной сети, которая кодирует характеристики речи, относящиеся к депрессии. В работе использовался корпус AViD-Corpus, представленный на соревнованиях по аудиовизуальному определению эмоций в 2013 и 2014 годах. Для вычисления низкоуровневых признаков использовался набор eGeMAPS из openSMILE. Предложенный метод включает в себя сети внутреннего внимания, обученные на низкоуровневых акустических признаках, глубокую сверточную сеть, обученную информации из трехмерных логмел спектрограмм и модуль предсказания степени депрессии. В сверточной сети применяется сжатие по среднему для объединения дополнительных признаков на уровне высказываний, которые являются входными данными для регрессора опорных векторов, который в итоге предсказывает баллы по шкале Бека-2. Для корпуса 2013 года были получены результаты 9,65 и 7,38 по показателям RMSE и MAE соответственно. Для корпуса 2014 года авторам удалось добиться результатов RMSE = 9,57 и MAE = 7,94.

Авторы работы [43] исследовали изменения в речи, которые происходят в результате психомоторной заторможенности, считающейся ключевым признаком БДР. Для этого применялись инверсированные переменные речевого тракта, которые получают путем использования системы инверсии речи, преобразующей акустический сигнал в шесть артикуляторных траекторий. Также были использованы мел-частотные кепстральные коэффициенты и корреляционные признаки. В качестве данных использовался корпус Mundt, а классификатор SVM обучался с использованием LOSO. Авторам удалось добиться точности (Accuracy) определения депрессии 81,7%.

В цикле работ [44–48] авторы используют текстовую модальность для определения депрессии, а именно частеречный анализ, TF-IDF, векторные представления слов, n-граммы, классические текстовые и психолингвистические признаки, а также анализ тональности. Наилучший результат был получен авторами на основе набора данных CLEF/eRisk 2017, в который входит коллекция текстовых сообщений 887 пользователей социальной сети Reddit, из которых 135 текстов помечены как депрессивные.

Использовались методы SVM и случайного леса, имплементированные в библиотеке Scikit-learn. Лучшие результаты определения депрессии были получены при использовании SVM и модели TF-IDF со стилистическими и морфологическими признаками, 63,0% по показателю F1-меры, 61,0% полноты и 65,0% точности. При этом модель на основе SVM, векторного представления слов и стилистических признаков получила наилучший результат полноты, равный 84,0% и F1-меры равной 61,0%. При экспериментах со случайным лесом наилучшей моделью оказалась TF-IDF с морфологическими признаками, при помощи которой был получен результат 79,0% точности и 62,0% F1-меры. Лучший результат на корпусах, собранных авторами, достиг F1-меры в 73,0% при использовании метода случайного леса и набора признаков, включающего психолингвистические признаки и биграммы. Кроме того, авторы собрали два корпуса: корпус эссе «Я, другие, мир» и корпус информации из профилей социальной сети «ВКонтакте».

Стоит отметить, что в последние несколько лет актуальными являются системы, использующие различные архитектуры нейронных сетей, а наилучшие результаты были получены при использовании рекуррентных архитектур и архитектур с механизмом внимания. Дан-ная особенность присуща большинству систем, представленных как на соревнованиях AVEC, так и вне соревнований. Описанные работы показывают эффективность многомодального подхода при определении депрессии. Также, нельзя не заметить, что вербальная информация также играет важную роль и позволяет добиться высоких результатов, наряду с паттернами невербальной информации, что подтверждает опыт терапевтов.

6. Систематизация современных методов, информативных признаков и классификаторов, используемых при разработке многомодальных систем автоматического определения депрессии. В ходе проведенного аналитического обзора был выявлен ряд сложностей при автоматическом определении депрессии: 1) каждый поведенческий сигнал предоставляет только частичную информацию, которая может быть комбинированной формой более реалистичной модели определения поведенческих индикаторов депрессии; 2) некоторая полезная информация может быть недоступна или в принципе скрыта; 3) определение базового поведения может быть осложнено ввиду ограничений в поведенческих данных.

Поскольку сбор данных для обучения моделей в системах автоматического определения депрессии является трудоемким и сложным из-за специфики задачи, существующие на данный момент корпуса имеют относительно небольшое количество данных, а также имеют дисбаланс в

количестве экземпляров в классах обучающих данных. Для решения такой проблемы возможно использование различных показателей точности работы моделей, учитывающих дисбаланс в количестве экземпляров в классах, техник аугментации данных и выбора наиболее информативных признаков. В описанных выше работах для выбора наиболее информативных признаков использовались следующие методы: выбор признаков на основе корреляции, метод главных компонент и метод частных наименьших квадратов.

Для вычисления признаков каждой модальности в рассмотренных работах использовались методы, которые можно классифицировать по трем группам: аудио, видео, текст, среди которых существует разделение на нейросетевые методы и программные инструментари, они представлены на рисунке 2. Также на нем представлены основные информативные признаки депрессии по аудиосигналу, видеоряду и лексике, которые можно получить при использовании методов вычисления признаков.

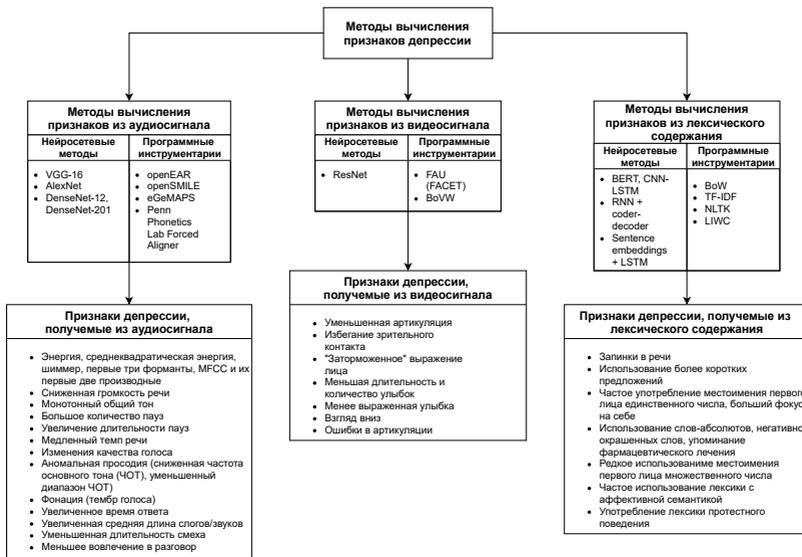


Рис. 2. Систематизация методов для вычисления информативных признаков и основные признаки депрессии, которые используются в современных автоматических системах

Наблюдаются изменения в речеобразовании у людей с депрессией после лечения, заключающиеся в изменениях тона голоса, громкости, частоты, артикуляции, беглости речи. В исследовании [49] показано, что

психологические особенности человека влияют на особенности написанного им текста. Наиболее чувствительным к психологическим особенностям человека оказался показатель частоты лексики с аффективной семантикой. Авторы выявили, что при высоких показателях депрессивности и тревожности, чувстве собственной незначительности и сниженной стратегией самоконтроля чаще употребляется лексика протестного поведения.

Среди рассмотренных работ можно выделить регрессионные и классификационные системы, в которых используются как нейросетевые, так и классические классификаторы, они представлены на рисунке 3. Так, можно отметить, что для задач классификации и регрессии при определении депрессии популярны в основном нейросетевые методы, а именно, сложные архитектуры нейросетевых методов. Вероятно, данная тенденция прослеживается ввиду того, что такие методы обладают большей устойчивостью к переобучению, большей способностью к обобщению, а также способностью к выявлению скрытых корреляций в признаковом пространстве.

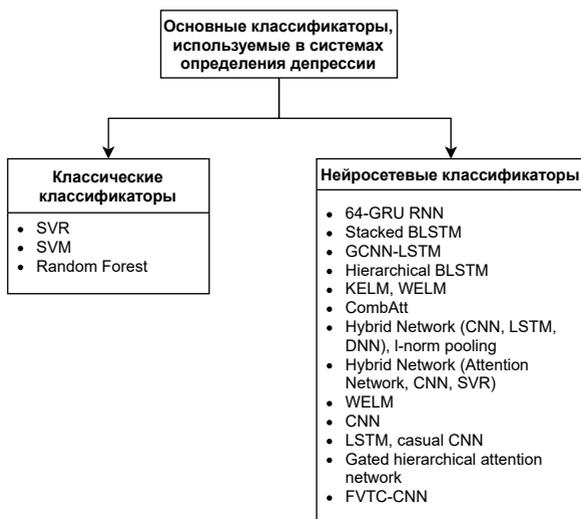


Рис. 3. Систематизация классификаторов, используемых в рассмотренных автоматических системах определения депрессии

На основе проведенного анализа можно сформулировать потенциальные требования, которые могли бы быть выдвинуты к разрабатываемым автоматическим системам определения депрессии, а именно:

1. Использование максимально возможного количества модальностей ввиду того, что специалистами учитываются все модальности при личной беседе. Кроме того, такой подход позволяет расширить возможности применения автоматических систем, так как будет возможность анализа дополнительных паралингвистических явлений.

2. Результат верного распознавания депрессии должен быть как можно выше (на данный момент лучший показатель точности среди автоматических систем находится на уровне 81% по показателю F-меры), так как сфера медицины относится к тем сферам применения, где ложные срабатывания автоматической системы могут быть критическими.

3. Для качественного выполнения второго пункта требований автоматические системы должны проходить тестирование в максимально приближенных к реальной жизни условиях.

4. Апробация в реальной жизни на этапе тестирования должна проходить под контролем специалистов, которые могли бы подтвердить верные предсказания или скорректировать ложные.

7. Заключение. В работе представлен аналитический обзор научных исследований за последние два года, посвященных разработке автоматических систем определения депрессивного состояния у людей. По данным ВОЗ, одним из наиболее распространенных психических расстройств является депрессия. Количество работ по автоматическому определению депрессии подтверждает возросший в последние годы интерес к теме, а также ее актуальность, поскольку депрессивное состояние распространено повсеместно и имеет свойство приводить к ухудшению жизнедеятельности человека и даже инвалидности или смерти. На данный момент используется множество признаков и показателей для объективной диагностики депрессивного состояния. Кроме того, существуют различия по вариантам течения и тяжести заболевания, подтипам. Специалисты определяют наличие депрессии путем беседы и заполнения различного рода опросников, однако такая оценка может варьироваться в зависимости от множества факторов, а потому на данный момент не существует объективного метода диагностики депрессии.

Рассмотрены теоретические и практические работы, представленные как на соревнованиях по аудиовизуальному распознаванию эмоций и определению депрессии, так и вне соревнований. Согласно теоретическим работам, депрессия непосредственно относится к расстройствам настроения, а значит, касается и эмоций в том числе. Это подтверждается тем, что существует сильная корреляция с аффективными величинами валентности, активации и доминанции. При этом наблюдается обратная корреляция между депрессией и агрессией. Работы, в которых прово-

дилось межкультурное сравнение симптомов депрессии, показали, что сочетания признаков депрессии обнаруживаются во всех культурах. Из этого следует, что автоматические системы определения депрессии могут быть универсальны. В практических работах были представлены одно-модальные и многомодальные системы за последние два года, которые решали как задачу регрессивной классификации для определения степени тяжести депрессии, так и задачу бинарной классификации наличия заболевания или его отсутствия. Для создания автоматической системы определения депрессии в рассмотренных работах были предложены различные подходы обработки многомодальных данных и построения моделей машинного обучения. Так, в такой системе могут применяться как линейные методы, так и сложные нейросетевые методы для обучения модели. Также в работе перечислены инструментарии и нейросетевые методы для вычисления признаков в различных модальностях.

В результате выполненного аналитического обзора можно сделать вывод, что объединение модальностей позволяет улучшить результаты определения депрессии (функционирование в сложных условиях при зашумленной или неразборчивой речи, при отсутствии речи и пр.), кроме того, анализ нескольких модальностей снижает вариативность при анализе речи и может значительно улучшить качество распознавания, поскольку анализируются дополнительные паралингвистические аспекты (движения бровей, напряжение губ, направление взгляда, движения рук и др.), которые могут являться информативными характеристиками. Также, анализ лексической составляющей речевого высказывания может позволить обнаружить различные показатели в речи, например, неуверенность, выражающуюся в отдельных словах, паузах хезитации и междометиях и др. В дальнейших работах мы планируем разработать собственный прототип автоматической системы определения депрессии по речи и исследовать на практике признаки депрессии, выявленные в ходе данного аналитического обзора, как со стороны акустических, так и со стороны лексических характеристик.

Литература

1. World Health Organization. 2017. Depression and Other Common Mental Disorders: Global Health Estimates. Technical Report. World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. 2018. DOI: 10.1016/S0140-6736(18)32279-7.
3. *Spitzer R.L.* Patient health questionnaire: PHQ. New York State Psychiatric Institute, 1999.

4. *Beck A.T., Ward C.H., Mock J., et al.* An inventory for measuring depression. Archives of General Psychiatry. 1961. vol. 4. pp. 561–571. DOI: 10.1001/archpsyc.1961.01710120031004.
5. *Rush A.J., Trivedi M.H., Ibrahim H.M., et al.* The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. Biological Psychiatry. 2003. vol. 54. no.5. pp. 573–583. DOI: 10.1016/S0006-3223(02)01866-8.
6. *Gonzalez J.S., Shreck E., Batchelder A.* Hamilton Rating Scale for Depression (HAM-D). In: Gellman MD, Turner JR, editors. Encyclopedia of behavioral medicine. New York: Springer. 2013. pp. 887–888. DOI: 10.1007/978-1-4419-1005-9_198.
7. *Valstar M., Schuller B., Smith K., et al.* AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge (AVEC'13). Association for Computing Machinery, New York, NY, USA. 2013. pp. 3–10. DOI: 10.1145/2512530.2512533.
8. *Valstar M., Schuller B., Smith K., et al.* AVEC 2014 — 3D dimensional affect and depression recognition challenge. Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, Workshop of MM. 2014. pp. 3-10. DOI: 10.1145/2661806.2661807.
9. *Valstar M., Gratch J., Schuller B., et al.* Summary for AVEC 2016: Depression, Mood, and Emotion Recognition Workshop and Challenge. Proceedings of the 24th ACM international conference on Multimedia (MM '16). Association for Computing Machinery, New York, NY, USA. 2016. pp. 1483–1484. DOI: 10.1145/2964284.2980532.
10. *Ringeval F., Schuller B., Valstar M., et al.* AVEC 2017: Real-life Depression, and Affect Recognition Workshop and Challenge. Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (AVEC '17). Association for Computing Machinery, New York, NY, USA. 2017. pp. 3–9. DOI: 10.1145/3133944.3133953.
11. *Ringeval F., Schuller B., Valstar M., et al.* AVEC 2019 Workshop and Challenge: State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 3–12. DOI: 10.1145/3347320.3357688.
12. *Потанова P.K.* Вариативность акустических параметров звучащей речи. Вестник Московского государственного лингвистического университета. Гуманитарные науки. Гуманитарные науки. 2016. т. 740. с. 137-147.
13. *Stahl S.M.* Stahl's essential psychopharmacology: Neuroscientific basis and practical applications. Cambridge: Cambridge University Press (4th ed.). 2013. p. 628.
14. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.). 2013. p. 992. DOI: 10.1176/appi.books.9780890425596.
15. *Franklin J.C., Ribeiro J.D., Fox K.R., et al.* Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull. 2017. vol. 143. no. 2. pp. 187-232. DOI: 10.1037/bul0000084.
16. *Belsher B.E., Smolenski D.J., Pruitt L.D., et al.* Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. JAMA Psychiatry. 2017. vol. 76. no. 6. pp. 642–651.
17. *Singer K.* Depressive disorders from a transcultural perspective. Social Science & Medicine. 1975. vol. 9. pp. 289-301. DOI: 10.1016/0037-7856(75)90001-3.
18. *Beck A.T., Steer R.A., Brown G.* Beck Depression Inventory—II. APA PscTests. 1996. P.38. DOI: 10.1037/t00742-000.
19. *Alghowinem S., Goecke R., Wagner M., et al.* From joyous to clinically depressed: Mood detection using spontaneous speech. Proceedings of FLAIRS Conference, G. M. Youngblood and P. M. McCarthy, Eds. AAAI Press. 2012. pp. 141–146.

20. Yang Y., Fairbairn C., Cohn J. Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*. 2013. vol. 4. no. 2. pp. 142–150.
21. Gratch J., et al. The Distress Analysis Interview Corpus of Human and Computer Interviews. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland. 2014. pp. 3123-3128.
22. Litvinova T., Ryzhkova E., Litvinova O. Features of Written Texts of People with Different Profiles of the Lateral Brain Organization of Functions (on the Basis of RusNeuroPsych Corpus). Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics, ExLing 2016, International Speech Communication Association, Saint Petersburg, Russia. 2016. pp. 107-110.
23. Mundi J.C., Snyder P.J., Cannizzaro M.S., et al. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (ivr) technology. *Journal of Neurolinguistics*. 2007. vol. 20. no. 1. pp. 50 – 64.
24. General Psychotherapy Corpus. URL: <http://alexanderstreet.com>. (дата обращения: 10.12.2020).
25. Huang Z., Epps J., Joachim D., et al. Depression detection from short utterances via diverse smartphones in natural environmental conditions. Proceedings of Interspeech. 2018. pp. 3393–3397.
26. Willmott C.J., Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*. 2005. vol. 30. pp. 79–82. DOI: 10.3354/cr030079.
27. Lin L.I. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989. vol. 45. no. 1. pp. 255-268.
28. Altman D.G., Bland J.M. Diagnostic tests. 1: Sensitivity and specificity. *BMJ (Clinical research ed.)*. 1994. vol. 308. no. 6943. P. 1552. DOI: 10.1136/bmj.308.6943.1552.
29. Ringeval F., Schuller B., Valstar M., et al. AVEC 2019 Workshop and Challenge: State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 3–12. DOI: 10.1145/3347320.3357688.
30. Kaya H., Fedotov D., Dresvyanskiy D., et al. Predicting depression and emotions in the crossroads of cultures, paralinguistics, and non-linguistics. Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 27–35. DOI: 10.1145/3347320.3357691.
31. Ray A., Kumar S., Reddy R., et al. Multi-level Attention Network using Text, Audio and Video for Depression Prediction. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 81–88. DOI: 10.1145/3347320.3357697.
32. Makiuchi M.R., Warnita T., Uto K., et al. Multimodal Fusion of BERT-CNN and Gated CNN Representations for Depression Detection. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 55–63. DOI: 10.1145/3347320.3357694.
33. Fan W., He Z., Xing X., et al. Multi-modality Depression Detection via Multi-scale Temporal Dilated CNNs. In 9th International Audio/Visual Emotion Challenge and Workshop (AVEC '19), Association for Computing Machinery, New York, NY, USA. 2019. pp. 73–80. DOI: 10.1145/3347320.3357695.
34. Yin S., Liang X., Ding H., et al. A Multi-Modal Hierarchical Recurrent Neural Network for Depression Detection. In 9th International Audio/Visual Emotion Challenge and

- Workshop (AVEC '19), Association for Computing Machinery, New York, NY, USA. 2019. pp. 65-71. DOI: 10.1145/3347320.3357696.
35. *Haque A., Guo M., Miner A.S., et al.* Measuring Depression Symptom Severity from Spoken Language and 3D Facial Expressions. Machine Learning for Health (ML4H) Workshop at NeurIPS 2018, Montréal, Canada. 2018. [Online]. Available: <http://arxiv.org/abs/1811.0859>.
 36. *Qureshi S.A., Hasanuzzaman M., Saha S., et al.* The Verbal and Non Verbal Signals of Depression — Combining Acoustics, Text and Visuals for Estimating Depression Level. [Online]. Available: <http://arxiv.org/abs/1904.07656>.
 37. *Niu M., Tao J., Liu B., et al.* Automatic Depression Level Detection via lp-Norm Pooling. Proceedings of Interspeech. 2019. pp. 4559-4563.
 38. *Rohanian M., Hough J., Purver M.* Detecting depression with word-level multimodal fusion. Proceedings of Interspeech. 2019. pp. 1443-1447.
 39. *Tao F., Esposito A., Vinciarelli A.* Spotting the traces of depression in read speech: An Approach Based on Computational Paralinguistics and Social Signal Processing. Proceedings of Interspeech. 2020. pp. 1828-1832.
 40. *Xezonaki D., Paraskevopoulos G., Potamianos A., et al.* Affective Conditioning on Hierarchical Networks applied to Depression Detection from Transcribed Clinical Interviews. Proceedings of Interspeech. 2020. pp. 4556-4560.
 41. *Huang Zh., Epps J., Joachim D., et al.* Domain Adaptation for Enhancing Speech-based Depression Detection in Natural Environmental Conditions Using Dilated CNNs. Proceedings of Interspeech. 2020. pp. 4561-4565.
 42. *Zhao Z., Li Q., Cummins N., et al.* Hybrid Network Feature Extraction for Depression Assessment from Speech. Proceedings of Interspeech. 2020. pp. 4956-4960.
 43. *Seneviratne N., Williamson J.R., Lammert A.C., et al.* Extended Study on the Use of Vocal Tract Variables to Quantify Neuromotor Coordination in Depression. Proceedings of Interspeech. 2020. pp. 4551-4555.
 44. *Stankevich M., Isakov V., Devyatkin D., et al.* Feature Engineering for Depression Detection in Social Media. Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018). 2020. pp. 426-431.
 45. *Ениколопов С.Н., Медведева Т.И., Воронцова О.Ю., и др.* Лингвистические характеристики текстов психически больных и здоровых людей. Психологические исследования. 2018. т. 11. №61. с. 1.
 46. *Kuznetsova Y.M., Kiselnikova N.V., Enikolopov S.N. et al.* Predicting Depression from Essays in Russian. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2019”. 2019. pp. 647-657.
 47. *Stankevich M., Smirnov I., Kiselnikova N., et al.* Depression Detection from Social Media Profiles. Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2019. Communications in Computer and Information Science. 2019. vol. 1223. pp. 181-194.
 48. *Stankevich M., Ignatiev N. Smirnov I.* Predicting Depression with Social Media Images. Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2020). 2020. pp. 235-240.
 49. *Ениколопов С.Н., Кузнецова Ю.М., Пенкина М.Ю., и др.* Особенности текста и психологические особенности: опыт эмпирического компьютерного исследования. Труды Института системного анализа РАН. 2019. т. 69. №3. с. 91-99.

Величко Алёна Николаевна — младший научный сотрудник/аспирант лаборатории речевых и многомодальных интерфейсов, СПб ФИЦ РАН. Область научных интересов: машинное обучение, речевые технологии, компьютерная паралингвистика, определение деструктивных проявлений по речи. Число научных публикаций — 10. velichko.a.n@mail.ru;

<http://hci.nw.ru/ru>; 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7-(812)-328-04-21, +7-(812)-328-70-81.

Карпов Алексей Анатольевич — д-р техн. наук, доцент, главный научный сотрудник/руководитель лаборатории речевых и многомодальных интерфейсов, СПб ФИЦ РАН. Область научных интересов: многомодальные интерфейсы, распознавание речи, речевые технологии, компьютерная паралингвистика. Число научных публикаций — 300+. karpov@iias.spb.su; <http://hci.nw.ru/ru>; 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7-(812)-328-04-21, +7-(812)-328-70-81.

A. VELICHKO, A. KARPOV
**ANALYTICAL REVIEW OF AUTOMATIC SYSTEMS FOR
DEPRESSION DETECTION BY SPEECH**

Velichko A., Karpov A. Analytical Review of Automatic Systems for Depression Detection by Speech.

Abstract. In recent years the interest in automatic depression detection has grown within medical and scientific-technical communities. Depression is one of the most widespread mental illnesses that affects human life. In this review we present and analyze the latest researches devoted to depression detection. Basic notions related to the definition of depression were specified, the review includes both unimodal and multimodal corpora containing records of informants diagnosed with depression and control groups of non-depressed people.

Theoretical and practical researches which present automated systems for depression detection were reviewed. The last ones include unimodal as well as multimodal systems. A part of reviewed systems addresses the challenge of regressive classification predicting the degree of depression severity (non-depressed, mild, moderate and severe), and another part solves a problem of binary classification predicting the presence of depression (if a person is depressed or not). An original classification of methods for computing of informative features for three communicative modalities (audio, video, text information) is presented. New methods used for depression detection in every modality and all modalities in total are defined. The most popular methods for depression detection in reviewed studies are neural networks. The survey has shown that the main features of depression are psychomotor retardation that affects all communicative modalities and strong correlation with affective values of valency, activation and domination, also there has been observed an inverse correlation between depression and aggression. Discovered correlations confirm interrelation of affective disorders and human emotional states. The trend observed in many reviewed papers is that combining modalities improves the results of depression detection systems.

Keywords: Automatic Depression Detection by Speech, Computational Paralinguistics, Speech Technologies, Machine Learning

Velichko Alena — junior researcher/Ph.D. student of the Speech and Multimodal Interfaces Laboratory, SPC RAS. Research interests: machine learning, speech technologies, computational paralinguistics, detection of destructive behaviour by speech. The number of publications — 10. velichko.a.n@mail.ru; <http://hci.nw.ru/ru>; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-(812)-328-04-21, +7-(812)-328-70-81.

Karpov Alexey — Ph.D., Dr. Sci., Associate professor, Head of the Speech and Multimodal Interfaces Laboratory, SPC RAS. Research interests: multimodal interfaces, speech recognition, speech technologies, computational paralinguistics. The number of publications — 300+. karpov@iias.spb.su; <http://hci.nw.ru/ru>; 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone: +7-(812)-328-04-21, +7-(812)-328-70-81.

Acknowledgements. This research was financially supported by RFBR (grant No. 20-37-90144), as well as partially in the framework of the state research № 0073-2019-0005.

References

1. World Health Organization. 2017. Depression and Other Common Mental Disorders: Global Health Estimates. Technical Report. World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases

- and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*. DOI:[https://doi.org/10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7).
3. Spitzer R.L. Patient health questionnaire: PHQ. *New York State Psychiatric Institute*. 1999.
 4. Beck A.T., Ward C.H., Mock J., et al. An inventory for measuring depression. *Archives of General Psychiatry*. 1961. vol. 4. pp. 561–571. DOI:<https://doi.org/10.1001/archpsyc.1961.01710120031004>.
 5. Rush A.J., Trivedi M.H., Ibrahim H.M., et al. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*. 2003. vol. 54. no.5. pp. 573–583. DOI:[https://doi.org/10.1016/S0006-3223\(02\)01866-8](https://doi.org/10.1016/S0006-3223(02)01866-8).
 6. Gonzalez J.S., Shreck E., Batchelder A. Hamilton Rating Scale for Depression (HAM-D). In: Gellman MD, Turner JR, editors. *Encyclopedia of behavioral medicine*. New York: Springer. 2013. pp. 887–888. DOI:https://doi.org/10.1007/978-1-4419-1005-9_198.
 7. Valstar M., Schuller B., Smith K., et al. AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge (AVEC'13). Association for Computing Machinery, New York, NY, USA. 2013. pp. 3–10. DOI:<https://doi.org/10.1145/2512530.2512533>.
 8. Valstar M., Schuller B., Smith K., et al. AVEC 2014 — 3D dimensional affect and depression recognition challenge. Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, Workshop of MM. 2014. pp. 3-10. DOI:<https://doi.org/10.1145/2661806.2661807>.
 9. Valstar M., Gratch J., Schuller B., et al. Summary for AVEC 2016: Depression, Mood, and Emotion Recognition Workshop and Challenge. Proceedings of the 24th ACM international conference on Multimedia (MM '16). Association for Computing Machinery, New York, NY, USA. 2016. pp. 1483–1484. DOI:<https://doi.org/10.1145/2964284.2980532>.
 10. Ringeval F., Schuller B., Valstar M., et al. AVEC 2017: Real-life Depression, and Affect Recognition Workshop and Challenge. Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge (AVEC '17). Association for Computing Machinery, New York, NY, USA. 2017. pp. 3–9. DOI: <https://doi.org/10.1145/3133944.3133953>.
 11. Ringeval F., Schuller B., Valstar M., et al. AVEC 2019 Workshop and Challenge: State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 3–12. DOI: <https://doi.org/10.1145/3347320.3357688>.
 12. Potapova R.K. [Variability of acoustic parameters of sounding speech]. *Variativnost' akusticheskikh parametrov zvuchashhej rechi. Vestnik Moskovskogo gosudarstvennogo lingvisticheskogo universiteta. Gumanitarnye nauki*. [Bulletin of Moscow State Linguistic University. Humanitarian sciences.]. 2016. vol. 740. pp. 137-147. (In Russ.)
 13. Stahl S.M. Stahl's essential psychopharmacology: Neuroscientific basis and practical applications. *Cambridge: Cambridge University Press (4th ed.)*. 2013. P. 628.
 14. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.)*. 2013. P. 992. DOI: <https://doi.org/10.1176/appi.books.9780890425596>.
 15. Franklin J.C., Ribeiro J.D., Fox K.R., et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull*. 2017. vol. 143. no. 2. pp. 187-232. DOI: <https://doi.org/10.1037/bul0000084>.

16. Belsher B.E., Smolenski D.J., Pruitt L.D., et al. Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. *JAMA Psychiatry*. 2017. vol. 76. no. 6. pp. 642–651.
17. Singer K. Depressive disorders from a transcultural perspective. *Social Science & Medicine*. 1975. vol. 9. 289–301. DOI: [https://doi.org/10.1016/0037-7856\(75\)90001-3](https://doi.org/10.1016/0037-7856(75)90001-3).
18. Beck A.T., Steer R.A., Brown G. Beck Depression Inventory–II. *APA PsycTests*. 1996. P.38. DOI: <https://doi.org/10.1037/t00742-000>.
19. Alghowinem S., Goecke R., Wagner M., et al. From joyous to clinically depressed: Mood detection using spontaneous speech. Proceedings of FLAIRS Conference, G. M. Youngblood and P. M. McCarthy, Eds. AAAI Press. 2012. pp. 141–146.
20. Yang Y., Fairbairn C., Cohn J. Detecting depression severity from vocal prosody. *IEEE Transactions on Affective Computing*, 2013. vol. 4. no. 2. pp. 142–150.
21. Gratch J., et al. The Distress Analysis Interview Corpus of Human and Computer Interviews. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland. 2014. pp. 3123–3128.
22. Litvinova T., Ryzhkova E., Litvinova O. Features of Written Texts of People with Different Profiles of the Lateral Brain Organization of Functions (on the Basis of RusNeuroPsych Corpus). Proceedings of 7th Tutorial and Research Workshop on Experimental Linguistics, ExLing 2016, International Speech Communication Association, Saint Petersburg, Russia. 2016. pp. 107–110.
23. Mundt J.C., Snyder P.J., Cannizzaro M.S., et al. Voice acoustic measures of depression severity and treatment response collected via interactive voice response (ivr) technology. *Journal of Neurolinguistics*. 2007. vol. 20. no. 1. pp. 50 – 64.
24. General Psychotherapy Corpus. URL: <http://alexanderstreet.com>. (дата обращения: 10.12.2020).
25. Huang Z., Epps J., Joachim D., et al. Depression detection from short utterances via diverse smartphones in natural environmental conditions. Proceedings of Interspeech. 2018. pp. 3393–3397.
26. Willmott C.J., Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*. 2005. vol. 30. pp. 79–82. DOI: <https://doi.org/10.3354/cr030079>.
27. Lin L.I. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989. vol 45. no. 1. pp. 255–268.
28. Altman D.G., Bland J.M. Diagnostic tests. 1: Sensitivity and specificity. *BMJ (Clinical research ed.)*. 1994. vol. 308. no. 6943. P. 1552. DOI: <https://doi.org/10.1136/bmj.308.6943.1552>.
29. Ringeval F., Schuller B., Valstar M., et al. AVEC 2019 Workshop and Challenge: State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 3–12. DOI: <https://doi.org/10.1145/3347320.3357688>.
30. Kaya H., Fedotov D., Dresvyanskiy D., et al. Predicting depression and emotions in the crossroads of cultures, paralinguistics, and non-linguistics. Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 27–35. DOI: <https://doi.org/10.1145/3347320.3357691>.
31. Ray A., Kumar S., Reddy R., et al. Multi-level Attention Network using Text, Audio and Video for Depression Prediction. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 81–88. DOI: <https://doi.org/10.1145/3347320.3357697>.

32. Makiuchi M.R., Warnita T., Uto K., et al. Multimodal Fusion of BERT-CNN and Gated CNN Representations for Depression Detection. In Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop (AVEC '19). Association for Computing Machinery, New York, NY, USA. 2019. pp. 55–63. DOI: <https://doi.org/10.1145/3347320.3357694>.
33. Fan W., He Z., Xing X., et al. Multi-modality Depression Detection via Multi-scale Temporal Dilated CNNs. In 9th International Audio/Visual Emotion Challenge and Workshop (AVEC '19), Association for Computing Machinery, New York, NY, USA. 2019. pp. 73–80. DOI: <https://doi.org/10.1145/3347320.3357695>.
34. Yin S., Liang X., Ding H., et al. A Multi-Modal Hierarchical Recurrent Neural Network for Depression Detection. In 9th International Audio/Visual Emotion Challenge and Workshop (AVEC '19), Association for Computing Machinery, New York, NY, USA. 2019. pp. 65–71. DOI: <https://doi.org/10.1145/3347320.3357696>.
35. Haque A., Guo M., Miner A.S., et al. Measuring Depression Symptom Severity from Spoken Language and 3D Facial Expressions. Machine Learning for Health (ML4H) Workshop at NeurIPS 2018, Montréal, Canada. 2018. [Online]. Available: <http://arxiv.org/abs/1811.0859>.
36. Qureshi S.A., Hasanuzzaman M., Saha S., et al. The Verbal and Non Verbal Signals of Depression — Combining Acoustics, Text and Visuals for Estimating Depression Level. [Online]. Available: <http://arxiv.org/abs/1904.07656>.
37. Niu M., Tao J., Liu B., et al. Automatic Depression Level Detection via lp-Norm Pooling. Proceedings of Interspeech. 2019. pp. 4559–4563.
38. Rohanian M., Hough J., Purver M. Detecting depression with word-level multimodal fusion. Proceedings of Interspeech. 2019. pp. 1443–1447.
39. Tao F., Esposito A., Vinciarelli A. Spotting the traces of depression in read speech: An Approach Based on Computational Paralinguistics and Social Signal Processing. Proceedings of Interspeech. 2020. pp. 1828–1832.
40. Xezonaki D., Paraskevopoulos G., Potamianos A., et al. Affective Conditioning on Hierarchical Networks applied to Depression Detection from Transcribed Clinical Interviews. Proceedings of Interspeech. 2020. pp. 4556–4560.
41. Huang Zh., Epps J., Joachim D., et al. Domain Adaptation for Enhancing Speech-based Depression Detection in Natural Environmental Conditions Using Dilated CNNs. Proceedings of Interspeech. 2020. pp. 4561–4565.
42. Zhao Z., Li Q., Cummins N., et al. Hybrid Network Feature Extraction for Depression Assessment from Speech. Proceedings of Interspeech. 2020. pp. 4956–4960.
43. Seneviratne N., Williamson J.R., Lammert A.C., et al. Extended Study on the Use of Vocal Tract Variables to Quantify Neuromotor Coordination in Depression. Proceedings of Interspeech. 2020. pp. 4551–4555.
44. Stankevich M., Isakov V., Devyatkin D., et al. Feature Engineering for Depression Detection in Social Media. Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018). 2020. pp. 426–431.
45. Enikolopov S.N., Medvedeva T.I., Voroncova O. Ju., et al. [Linguistic characteristics of texts written by mentally ill and healthy people]. *Lingvisticheskie harakteristiki tekstov psihicheski bol'nyh i zdorovyh ljudej. Psihologicheskie issledovanija*. [Psychological investigations]. 2018. vol. 11. no. 61. pp. 1. (In Russ.).
46. Kuznetsova Y.M., Kiselnikova N.V., Enikolopov S.N. et al. Predicting Depression from Essays in Russian. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2019”. 2019. pp. 647–657.
47. Stankevich M., Smirnov I., Kiselnikova N., et al. Depression Detection from Social Media Profiles. Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL

2019. Communications in Computer and Information Science. 2019. vol. 1223. pp. 181-194.
48. Stankevich M., Ignatiev N. Smirnov I. Predicting Depression with Social Media Images. Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2020). 2020. pp. 235-240.
49. Enikolopov S.N., Kuznecova Ju.M., et al. [Characteristics of text and psychological characteristics: experience of empiric computational research]. *Osobennosti teksta i psihologicheskie osobennosti: opyt jempiricheskogo komp'juternogo issledovanija. Trudy Instituta sistemnogo analiza RAN*. [Proceedings of the institute of system analysis of RAS]. 2019. vol. 69. no. 3. pp. 91-99. (In Russ.).

А.В. ВОРОБЬЕВ, Г.Р. ВОРОБЬЕВА
**МОДЕЛЬ ИНФОРМАЦИОННОГО ВЗАИМОДЕЙСТВИЯ
ЭЛЕМЕНТОВ МНОГОУРОВНЕЙВОЙ СИСТЕМЫ ЦИФРОВЫХ
ДВОЙНИКОВ**

Воробьев А.В., Воробьева Г.Р. Модель информационного взаимодействия элементов многоуровневой системы цифровых двойников

Аннотация. Одним из решений проблемы пространственно-временной анизотропии данных является применение многоуровневой системы цифровых двойников, базирующихся на соответствующих отраслевых моделях и пополняемой базе архивных данных. Применение указанного подхода успешно зарекомендовало себя в информационных системах мониторинга параметров геомагнитного поля и его вариаций, обеспечивая пространственно-временную интерполяцию геомагнитных данных с точностью до 0.81 нТл в магнитоспокойные периоды. При этом нерешенной осталась задача информационного взаимодействия между уровнями системы цифровых двойников, что в значительной степени усугубляется постоянно растущим объемом данных и их неоднородным характером. В работе предлагается решение обозначенной задачи посредством формализованного механизма пакетирования пространственно-временной информации, при котором идентификация источников данных выполняется на базе иерархической системы бинарной токенизации. Кроме того, рассматривается предложенная программная реализация такого подхода, отличительной особенностью которой является комбинирование традиционной клиент-серверной и инновационной бессерверной архитектур для реализации высоконагруженного реактивного веб-приложения для работы с анализируемыми данными. Выделены и программно формализованы основные этапы реализации информационного взаимодействия – от получения исходной информации от ее источников до верификации данных, их анализа, обработки и формирования выходного информационного потока системы. Результаты проведенных вычислительных экспериментов на примере задачи мониторинга параметров магнитного поля Земли и его вариаций подтвердили эффективность предложенных решений, выраженную как в повышении реактивности веб-ориентированных приложений, так и в увеличении вычислительной скорости формирования и заполнения информационных хранилищ, агрегирующих информацию из распределенных гетерогенных источников.

Ключевые слова: цифровые двойники, пространственные данные, информационное взаимодействие, единое информационное пространство, обработка данных, токенизация, геоинформационные системы и технологии

1. Введение. Одним из магистральных направлений развития технологий на протяжении длительного времени является обеспечение эффективной обработки и анализа данных, описывающих сложные процессы и явления различной природы происхождения в контексте их пространственно-временного распределения. Соответствующие системы мониторинга представляют собой комплексы информационно-измерительных устройств, с различной плотностью распределенные по земной поверхности.

Так, к примеру, мониторинг геомагнитного поля и его вариаций традиционно выполняется множеством (порядка 300) магнитных обсерваторий и вариационных станций, которые в режиме реального времени регистрируют соответствующие геомагнитные данные. Данные регистрируются ежеминутно и за время существования сети магнитных обсерваторий накоплено порядка 22,8 Гб геомагнитных данных, представляющих собой временные ряды в CSV-подобном IAGA-формате. Учитывая то, что некоторые магнитные обсерватории предоставляют данные, зарегистрированные с шагом в 1 с, объем накапливаемых при этом данных пропорционально возрастает. При этом практически на каждой магнитной обсерватории в регистрируемых временных рядах ежедневно наблюдаются от 5 до 58% пропущенных значений.

Общей проблемой для подавляющего большинства систем мониторинга процессов и явлений является выраженная пространственная и временная анизотропия, неразрешимая методами интерполяции ввиду специфики анализируемых процессов (как правило, естественной природы происхождения). Не менее серьезной является неполнота соответствующих временных рядов, обусловленная перебоями в системах связи, сбоями в работе оборудования, а также человеческим фактором. Так, к примеру, анализ результатов расчета показателя относительной информационной эффективности для некоторых магнитных обсерваторий сети INTERMAGNET за 2015 г. [1] показал, что в среднем общее время работоспособного состояния для отдельных обсерваторий составляет от 64,08 до 87,95 % времени наработки.

Обозначенные проблемы являются существенным препятствием на пути обработки пространственно-временных данных, что, в свою очередь, негативно сказывается как на процессе принятия решений в отраслевых областях, так и на получении новых знаний в соответствующей прикладной области.

2. Состояние вопроса. Современные подходы к решению задачи пространственно-временной анизотропии данных во многом базируются на применении интеллектуальных геоинформационных технологий. Так, к примеру, широко известны и повсеместно практикуются такие технологии пространственной интерполяции, как крикинг, сплайн, обратно-взвешенные расстояния, метод ближайшего соседа и др. [1]. Однако, к примеру, для геомагнитных данных анализ показал, что применение известных методов пространственной интерполяции сопровождается среднеквадратической ошибкой в среднем от 7,3 до 11,2 нТл при

допустимой стандартами погрешности в 1 нТл. В результате интерполированные данные могут повлечь за собой серьезные ошибки в моделировании и принятии решений, что, в свою очередь, сопряжено с финансовыми и временными затратами на устранение их последствий (например, при инклинометрических исследованиях с применением магниточувствительного оборудования).

Восстановление временных рядов пространственных данных реализуется разнообразными методами как математической, так и геостатистики. Так, на ранних этапах исследований для восстановления данных преимущественно использовались методы линейной интерполяции и кубического сплайна, эффективные для устранения единичных пропусков, но непригодные для импутации более длительных фрагментов ввиду возникновения существенных частотных и амплитудных искажений [2]. Более сложные методы восстановления временных рядов базируются на аналитической обработке известных значений уровня, а также анализе периодических и сезонных составляющих ряда. Так, к примеру, известен метод гравитационного сглаживания временных рядов [3], основанный на понятии гравитационной непрерывности и связанный с квадратичной регрессией. Метод позволяет восстанавливать длительные фрагменты временных рядов, полностью исключая амплитудные и минимизируя частотные искажения. Однако необходимость решения при этом линейных систем большого порядка сопряжена с высокой трудоемкостью применения метода и, как следствие, низкой вычислительной скоростью средств их программной реализации, что недопустимо при оперативной аналитической обработке данных. Для восстановления данных со сложной нерегулярной структурой и многочисленными локальными особенностями применяется метод непрерывного вейвлет-преобразования [4]. Так, к примеру, в [5] показано, что отличие восстановленного по коэффициентам вейвлет-разложения временного ряда от оригинального не превышает 15 %.

Вместе с тем ни один из известных подходов не учитывает особенности пространственной анизотропии интерполируемых данных и принимает одни и те же модели и методы для различных географических областей.

3. Постановка задачи. Одним из вариантов решения проблемы пространственной анизотропии является использование системы цифровых двойников [6], которые в совокупности с их физическими прототипами обеспечивают повышение плотности информационных измерений для соответствующих процессов или явлений различной природы происхождения.

Основанием для выделения уровней системы цифровых двойников является пространственная кластеризация непосредственно двойников и их физических прототипов. На основании заданного параметра разбиение исходного множества источников данных на подмножества осуществляется известными геоинформационными методами, в частности, по результатам расчета индекса Морана-I [7], традиционно являющегося мерой пространственной автокорреляции пространственных объектов. Верификация цифровых двойников осуществляется методами машинного обучения с учетом данных, регистрируемых физическими прототипами.

При объединении цифровых двойников и их физических прототипов в многоуровневую структуру их информационное взаимодействие приобретает особую специфику, обусловленную необходимостью сохранения данных при их интеграции в единое информационное пространство в составе соответствующих систем поддержки принятия решений.

Вместе с тем необходимо формализовать процедуру информационного взаимодействия элементов многоуровневой системы цифровых двойников таким образом, чтобы обеспечить интеграцию, где каждый из элементов систем в отдельности и их агрегации выступает самостоятельным источником данных для систем поддержки принятия решений.

Для решения поставленной задачи необходимо выполнение следующих этапов исследования:

1. Разработка концепции информационного взаимодействия в системе цифровых двойников, учитывающей ее многоуровневую структуру, пространственную кластеризацию источников данных и формирование исходящего информационного потока во внешнюю информационную систему (или систему поддержки принятия решений).
2. Формализация обозначенной концепции в теоретико-множественном базисе для систематизации требований к построению системы информационного межуровневого взаимодействия цифровых двойников и их физических прототипов.
3. Разработка системы самоидентификации источников данных в многоуровневой системе цифровых двойников.
4. Определение и формализация моделей активного поведения программных модулей, реализующих информационное взаимодействие в системе цифровых двойников и

обеспечивающих оперативную передачу информации между уровнями системы в соответствии с заданными критериями.

5. Разработка программной архитектуры системы информационного взаимодействия цифровых двойников, обеспечивающей эргономичный открытый доступ к данным в различных интерфейсах – как на уровне конечных пользователей, так и для применения внешними программными модулями и системами.

В качестве сквозного примера в работе рассматривается система магнитных обсерваторий и вариационных станций, которые в режиме реального времени регистрируют параметры геомагнитного поля и его вариаций – геомагнитные данные. На данном примере был проведен ряд вычислительных экспериментов по оценке целесообразности объединения цифровых двойников и их физических прототипов в многоуровневую систему. В ходе исследований были проанализированы годовые минутные данные более чем 300 магнитных станций, неравномерно распределенных по трем пространственным областям – высокоширотные, среднеширотные и приэкваториальные станции. Анализ результатов экспериментов показал, что дополнение физических прототипов соответствующими цифровыми двойниками позволит снизить количество пропусков в среднем на 54,3 % в среднеширотных и приэкваториальных областях и на 27,4% в высокоширотных областях.

4. Концепция и модель информационного взаимодействия в многоуровневой системе цифровых двойников. Ключевой идеей информационного взаимодействия в многоуровневой системе цифровых двойников является создание информационных узлов на каждом из ее уровней (рис. 1). При этом на нижнем уровне абстракции имеют место информационные потоки, с заданной периодичностью направляемые от цифровых двойников и их физических прототипов в выделенные информационные хранилища. Далее указанные информационные хранилища объединяются внутри соответствующего локального уровня. На более высоком уровне абстракции информационные потоки с локальных уровней поступают в информационные хранилища отдельных кластеров, которые, в свою очередь, передают свои данные в централизованное хранилище данных, например, в выделенное единое информационное пространство в составе системы поддержки принятия решения в отраслевой области.

В общем виде указанную концепцию можно формализовать в теоретико-множественном базисе следующим образом.

Пусть задана пара вида «физический прототип – его цифровой двойник»:

$$\{P, T\}: P \cup T \neq \emptyset; P \neq T, P \not\subset T;$$

$$P = \{p_1, p_2, \dots, p_K\}; T = \{t_1, t_2, \dots, t_K\};$$

где P – физический прототип, T – цифровой двойник, p_1, p_2, \dots, p_K – атомарные данные, зарегистрированные физическим прототипом в моменты времени $1 \dots K$, t_1, t_2, \dots, t_K – атомарные данные, полученные в результате моделирования цифровым двойником за те же моменты времени $1 \dots K$.

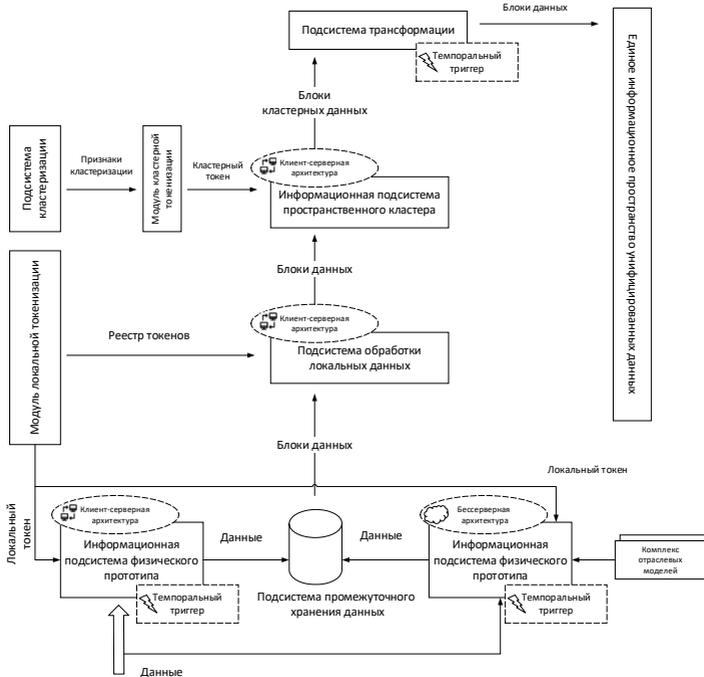


Рис. 1. Обобщенная схема информационного взаимодействия в системе цифровых двойников

В процессе обработки информационных потоков, синхронно поступающих от цифрового двойника и его физического прототипа с заданным временным интервалом, данные агрегируются в специализированном промежуточном хранилище, выделенном для

каждой пары вида «физический прототип – его цифровой двойник». Далее полученный массив анализируется программным модулем, в основе построения которого лежат модели и методы машинного обучения. По результатам анализа из двух входных информационных потоков формируется один набор данных, содержащий результаты верификации наблюдений и выборку наиболее достоверных (с точки зрения используемого модулем математического алгоритма и специфики прикладной предметной области) значений.

В общем виде данный процесс можно описать в виде следующих соотношений. Пусть в результате промежуточной интеграции информационных потоков за период времени K формируется массив данных вида:

$$D = P \times T = \{\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_K, t_K\}\},$$

где D – промежуточный массив как результат попарной интеграции элементов данных, получаемых от физического прототипа P и его цифрового двойника T .

В подсистеме промежуточного хранения выполняется анализ представленных в D пар значений. Применяется процедура трансформации F , которая на основании регрессионного анализа накопленных ретроспективных данных задает однозначное отображение пар значений физического прототипа и двойника в элемент нового множества D' :

$$D' = \{d_1, d_2, \dots, d_K\} \mid \forall i = 1, \dots, K \{p_i, t_i\} \xrightarrow{F} d_i,$$

где d_i – элемент нового множества, полученный в результате применения процедуры трансформации к каждой из K пар данных, полученных от физического прототипа P и его цифрового двойника T .

Все множество пар вида «физический прототип – его цифровой двойник» является информационной базой для хранилища локальных данных, куда направляются все прошедшие аналитическую трансформацию и обработку данные из множества D' .

Пусть для каждого источника данных, например, магнитной обсерватории или вариационной станции, имеется пара вида «физический прототип – цифровой двойник». Таким образом, имеет место соотношение вида:

$$\exists N \in R : \forall i \in N \exists \{P, T\}^i : P^i \cup T^i \neq \emptyset; P^i \neq T^i; P^i \not\subset T^i;$$

$$P^i = \{p_1^i, p_2^i, \dots, p_k^i\}; T^i = \{t_1^i, t_2^i, \dots, t_k^i\};$$

где N – количество доступных источников данных, каждый из которых представлен парами вида «физический прототип – его цифровой двойник».

Тогда в результате трансформации для каждого источника данных из N формируется промежуточное хранилище для K моментов времени следующего вида:

$$D^j = \{d_1^j, d_2^j, \dots, d_k^j\} | \forall i = 1, \dots, K \{p_i^j, t_i^j\} \xrightarrow{F} d_i^j, \forall j = 1, \dots, N,$$

где D^j – массив данных промежуточного хранилища j -го источника данных, для которого существует пара «физический прототип P^j – его цифровой двойник T^j », которые в каждый i -й из K моментов времени предоставляют данные вида p_i^j и t_i^j соответственно.

В результате формирование хранилища локальных данных за период времени K может быть представлено в виде объединения соответствующих множеств D^j . Иными словами, имеет место соотношение вида:

$$Z = D^1 \cup D^2 \cup \dots \cup D^N,$$

где Z – множество данных, поступающих за период K от промежуточных хранилищ N источников данных.

При этом

$$Z = \{d_1^1, d_2^1, \dots, d_k^1, \dots, d_1^j, d_2^j, \dots, d_k^j\} |$$

$$d_i^j \in D^j; i = 1, \dots, K; j = 1, \dots, N,$$

где d_i^j – элемент множества Z , соответствующий данным, полученным от j -го источника данных за i -й момент времени.

В соответствии с требованиями заинтересованных лиц, определяемыми в процессе проектирования информационных систем (в том числе систем поддержки принятия решений), выделяется перечень признаков, на основании которых источники данных разделяются на кластеры. Например, применительно к решению задач геомагнетизма таковым выступает разделение земной поверхности на

высокоширотные, приэкваториальные и среднеширотные области. В соответствии с этим критерием магнитные обсерватории и вариационные станции могут быть в соответствии с их пространственной привязкой отнесены к высокоширотному, приэкваториальному и среднеширотному кластерам.

В этой связи представляется целесообразным ввести следующие соотношения.

Пусть заданы M пространственных кластеров, в каждом из которых представлены от 1 до L источников данных в соответствии с заданным признаком (признаками) кластеризации:

$$\exists G = \{G_1, G_2, \dots, G_M\} : \{S_1, \dots, S_L\} \in G_i,$$

где G – множество пространственных кластеров мощностью M , i -й элемент которого представляет собой отдельный пространственный кластер G_i , каждый из которых, в свою очередь, содержит от 1 до L источников данных S .

Для каждого пространственного кластера G_l должна быть сформирована собственная база данных (база данных l -го пространственного кластера), куда поступают срезы данных из локального хранилища данных Z :

$$Z_{G_l} = Z' \subset Z:$$

$$Z' = \{d_1^1, d_2^1, \dots, d_K^1, \dots, d_1^j, d_2^j, \dots, d_K^j\} |:$$

$$i = 1, \dots, K; j = 1, \dots, R; R \leq N,$$

где Z' – подмножество множества данных из локального хранилища, которое содержит данные вида d_K^j для K моментов времени R источников данных в составе l -го пространственного кластера, при этом число источников данных в источнике, очевидно, не должно превышать числа доступных источников данных.

Важно отметить, что в базе данных пространственного кластера представлены данные не непосредственно физических прототипов источников информации, а те значения, которые были помещены в соответствующие промежуточные хранилища после аналитической трансформации данных пары «физический прототип – его цифровой двойник».

При этом одни и те же источники данных могут быть представлены более, чем в одном кластере. При этом для заданного набора кластеров не исключается существование таких источников данных, которые не принадлежат ни одному из возможных

пространственных кластеров. Такие источники данных будут рекластеризованы в дальнейшем при введении дополнительных признаков.

На завершающем этапе выполняется интеграция баз данных пространственных кластеров в единое информационное пространство в составе информационной системы (системы поддержки принятия решений). Данные подвергаются дополнительной обработке (трансформации) для приведения в унифицированную форму.

5. Система иерархической токенизации элементов системы цифровых двойников. Для повышения эффективности процессов обмена и обработки данных между уровнями системы цифровых двойников предлагается подход, основанный на принципе токенизации информационных ресурсов.

Технология токенизации предполагает замену некоторого реального значения суррогатным, активно используется для защиты и шифрования в современных системах информационной безопасности, а также для реализации систем блокчейн [8, 9].

В настоящей работе предлагается использовать токенизацию для разметки источников данных и их физических прототипов, а также результатов их агрегирования на основании пространственной кластеризации по заданным признакам.

В соответствии с предлагаемой концепцией информационного взаимодействия элементов системы цифровых двойников предполагается маркировка соответствующих наборов данных для идентификации их источника в процессе обработки и анализа информации в составе единого информационного пространства (информационной системы, системы поддержки принятия решений и пр.).

Регистрируемые физическим прототипом или моделируемые его цифровым двойником данные проходят многоэтапную обработку. Так, для физического прототипа на начальном этапе данные попадают в модуль регистрации, где фиксируются принятым в источнике данных способом, например, в виде простых электронных таблиц или файлов csv-подобного формата. Далее выполняется их предварительная обработка – устранение артефактов, обусловленных несовершенством информационно-измерительного оборудования, а также человеческим фактором. На следующем этапе к данным применяются алгоритмы сжатия для экономии дискового пространства, с одной стороны, и упрощения передачи данных на следующий уровень, с другой. На последующем этапе пакетирования данные упаковываются в объект, по структуре подобный HTTP-сообщению (заголовок и тело сообщения).

В заголовке проставляются идентификационные данные источника, направляющего информационный поток. Данные в заголовке являются, в том числе, и авторизационными, поскольку позволяют принимающему их модулю оценить, что является источником информации.

Например, для геомагнитных данных одним из вариантов такой авторизационной информации является заголовок, предусмотренный форматом IAGA-2002 [10]. Заголовок содержит основную информацию о магнитной обсерватории / вариационной станции: название, уникальный код, ведомственная принадлежность, пространственные координаты, характеристики оборудования, регистрируемые параметры и пр. С каждым новым блоком передаваемых данных метаданные, которые и представлены в заголовке, передаются в точку отправления, нагружая как каналы передачи данных (что существенно при насыщенном информационном потоке), так и средства их обработки, которые вынуждены «очищать» полученный набор данных от служебной информации.

Для совершенствования процедуры обработки и передачи данных между уровнями системы цифровых двойников предлагается использовать систему токенизации. При этом каждый токен представляет собой «свертку» метаданных о соответствующем источнике информационного потока, который при необходимости можно декомпонировать и выделить необходимую для работы служебную информацию.

Предлагается подход, основанный на последовательном расширении токена источника данных по мере перемещения с уровня на уровень в системе цифровых двойников. Исходный код токена формируется на начальном уровне – для отдельных физических прототипов и цифровых двойников. Код является глобально уникальным для всех возможных источников данных, задействованных в системе цифровых двойников. Код токена является суррогатным и представляет собой бинарный код, сформированный посредством перевода в двоичный формат уникального десятичного номера.

При этом предлагается следующий формат представления токена на локальном уровне: $L: Type - XXXXX - (ParentID)$, где L – метка локального уровня (неизменный параметр); $Type$ – тип источника данных (неизменный параметр), домен допустимых значений которого представлен в виде (P, T) , где P – соответствует физическому прототипу, T – цифровому двойнику; $XXXXX$ – уникальный двоичный код источника данных, сформированный из его порядкового номера,

представленного в десятичном формате (неизменный параметр); *ParentID* – ссылка на уникальный двоичный код физического прототипа (параметр, обязательный только для тех меток, которые используются для токенизации источников данных на локальном уровне).

Каждый новый источник данных, который будет использован в системе цифровых двойников, проходит обязательную процедуру токенизации, программно реализуемой посредством инструментария подсистемы токенизации, предусмотренной в предложенной архитектуре системы. Каждому источнику данных присваивается порядковый регистрационный номер (чем обеспечивается его уникальность в системе цифровых двойников), значение которого переводится в двоичный формат.

Далее в случае, если выполняется токенизация цифрового двойника в реестре источников данных, осуществляется поиск его физического прототипа, определяется его уникальный двоичный идентификатор, который поставляется в позицию «ParentID» формулы токенизации источника данных на локальном уровне. В результате обеспечивается однозначное соответствие между физическими прототипами и их цифровыми двойниками.

Вопрос ссылочной целостности при кодировании токена цифрового двойника и его прототипа предлагается решать следующим образом. По аналогии с реляционными базами данных предполагается применение строгого правила ссылочной целостности, согласно которому невозможны манипуляции с изменением или удалением токена физического прототипа в случае наличия соответствующего ему цифрового двойника. При этом изменение или удаление токена цифрового двойника никак не сказывается на токене соответствующего ему физического прототипа.

Токен источника данных является его уникальным идентификатором в реестре источников данных системы цифровых двойников, на основании которого можно получить информацию о служебных параметрах (метаданных). В частности, метаданные относительно географического расположения источников данных применяются при формировании пространственных кластеров с соответствующей токенизацией составляющих их источников данных.

В соответствии с заданными признаками выделяются пространственные кластеры, каждый из которых (по аналогии с источниками данных) токенизируется уникальным образом: выполняется преобразование в двоичный формат на основании уникального порядкового номера пространственного кластера, представленного изначально в десятичном формате.

В результате предлагается следующая форма представления токена пространственного кластера: *Cluster: XXXXX*, где *Cluster* – флаг, указывающий на применение токена к идентификации пространственного кластера; *XXXXX* – уникальный двоичный код пространственного кластера, сформированный из его порядкового номера, представленного в десятичном формате (неизменный параметр).

По аналогии с именованим атрибутов в реляционной и свойств в объектной моделях данных полный токен каждого задействованного в системе цифровых двойников источника данных будет складываться из двух токенов: локального и кластерного, разделенных символами «← →».

Иными словами, в общем виде полный токен источника данных можно представить следующим образом: *Cluster: XXXXX -- L: Type – XXXXX – (ParentID)*.

Уникальность полного токена источника данных обеспечивается при этом отсутствием пересечений множеств источников данных, образующих пространственный кластер. Иными словами, должно быть выполнено условие вида:

$$\begin{aligned} & \neg \exists G_i \in \{S_1, \dots, S_{L1}\} \& \& G_j \in \{S_1, \dots, S_{L2}\} : G_i \cap G_j \neq \emptyset; \\ & \forall G_i, G_j \in G = \{G_1, G_2, \dots, G_M\}, \end{aligned}$$

где G – множество пространственных кластеров мощностью M , i -й элемент, который представляет собой отдельный пространственный кластер G_i , каждый из которых, в свою очередь, содержит от 1 до L источников данных S .

При этом предполагается, что пара вида «физический прототип – его цифровой двойник» должна быть представлена в составе одного и только одного пространственного кластера:

$$S_i = (P_i, T_i), \forall S_i \in G_j, \forall G_j \in G = \{G_1, G_2, \dots, G_M\}.$$

Недопустимы ситуации, при которых физический прототип источника данных (например, информационно-измерительного устройства или иного технически сложного объекта/системы) и его цифровой двойник отнесены к разным пространственным кластерам, поскольку сочетание двойника и его прототипа определяет единый информационный поток, исходящий из источника данных.

В результате поэтапного изменения токена имеет место его многоуровневая структура, согласно которой каждый новый уровень системы цифровых двойников добавляет к токену свою метку (так же токен), который позволяет отследить иерархию родительских по отношению к источнику данных (любого типа – физического прототипа или его цифрового двойника) элементов. Например, таким образом можно определить, для какого пространственного кластера используется источник данных, что позволит дополнить анализируемую информацию в том числе и набором метаданных об ее источнике.

Кроме того, отличительной особенностью предлагаемого подхода является двойственное применение токена, что зависит от технических возможностей соответствующей информационной системы, реализующей описанную выше концепцию многоуровневой организации цифровых двойников.

С одной стороны, уникальный токен может быть использован как идентификатор источника данных в соответствующем реестре (базе данных). Передача токена в запросе как один из вариантов фильтрации данных позволит обратиться непосредственно к кортежу, который содержит искомую служебную информацию об источнике данных. Преимущества, связанные с простотой реализации и возможностью интеграции в уже существующую информационную архитектуру, несколько нивелируются тем недостатком, что помимо метаданных необходимо хранить еще и токен.

Другой вариант применения токена источника данных – «свертка» метаданных непосредственно в идентификатор. Применение алгоритма сжатия данных Deflate [11] позволяет выполнить преобразование заданной блочной структуры в компактную «свертку», представленную в шестнадцатеричном формате. Недостаток такого подхода, связанный с необходимостью введения дополнительного программного модуля для реализации процедуры кодирования/декодирования соответствующего токена, компенсируется тем, что метаданные непосредственно встроены в сообщение, передаваемое в модуль обработки или централизованное хранилище от источника данных. Задачей указанного программного модуля, по сути, является простое считывание из токена цепочек бит, упакованных в байты, формирование сжатого блока при кодировании и извлечение оригинального значения при его декодировании.

Еще одним применением токена является решение задачи авторизации источника данных, направляющего информационный поток в модуль обработки или выделенное хранилище данных. При

поступлении данных в программный модуль выполняется процедура детокенизации – выделения токена из поступившего сообщения, его декодирования и выделения метаданных. Далее результаты анализа передаются в тот компонент хранилища, который соответствует заданному токеном пространственному кластеру (или отдельному источнику данных любого типа).

6. Темпоральные триггеры в системе цифровых двойников.

Для реализации процедуры информационного взаимодействия компонентов системы цифровых двойников предложен механизм, получивший название темпорального триггера.

В основе данного подхода лежит фундаментальное понятие триггеров, применимое к транзакционной системе баз данных. Согласно известному термину [12], триггер представляет собой программную процедуру, запускаемую на выполнение в случае наступления в базе данных одного из трех событий – создания, удаления или модификации кортежа (или группы кортежей) данных.

Применительно к системе цифровых двойников событием для привязки к триггеру является достижение определенной временной метки. Предполагается, что темпоральный триггер по принципу своего функционирования подобен программному таймеру и на физическом уровне он реализуется с помощью соответствующих программных процессов-демонов, которые выполняются по принципу Cron [13].

В архитектуре системы цифровых двойников предполагается применение трех темпоральных триггеров. Два из них должны быть задействованы на уровне атомарных источников данных – физических прототипов и их цифровых двойников. По истечению заданного периода времени данные, накопленные в виртуальном буфере, выгружаются в подсистему промежуточного хранения, где выполняется их анализ, обработка и формирование единого информационного блока для передачи на следующий уровень системы.

Еще один темпоральный триггер функционирует на уровне пространственного кластера. По сути, он выполняется по тому же принципу, что и два предшествующих: по истечению заданного временного интервала выполняется выгрузка данных из базы данных пространственного кластера в единое информационное пространство.

Таким образом, темпоральный триггер запускает на выполнение процедуру выгрузки наборов данных из хранилища одного уровня в систему хранения более высокого уровня.

Настройка темпоральных триггеров осуществляется при проектировании системы цифровых двойников и должна учитывать специфику соответствующей прикладной (отраслевой) области. Так, к

примеру, для геомагнитных данных предполагается установка темпорального триггера в 1-минутный интервал относительно заданной временной точки. Другой пример – данные с прогнозом вероятности полярных сияний в области аврорального овала. Здесь темпоральный триггер целесообразно установить во временной интервал, не менее 5-минутного.

7. Программная архитектура информационного взаимодействия системы цифровых двойников. В общем виде процесс информационного взаимодействия в системе цифровых двойников представляет собой сбор данных, их краткосрочное размещение в промежуточных хранилищах с последующей передачей через уровни системы в единое информационное пространство. Каждый подпроцесс информационного взаимодействия программно оформлен в виде подсистемы (совокупности модулей) или отдельных модулей.

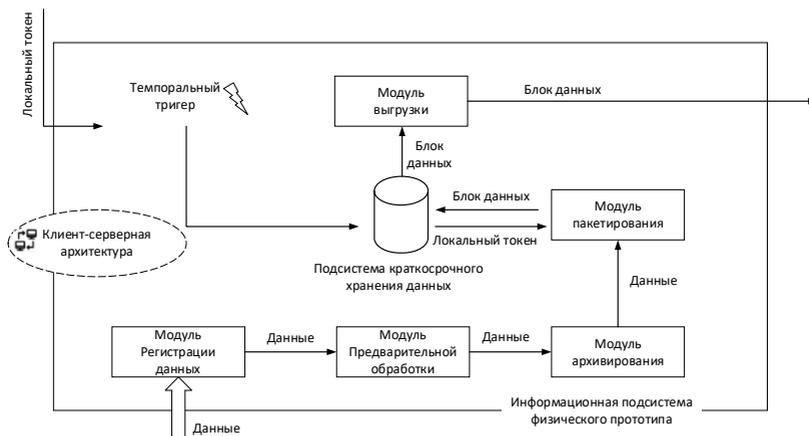
Обобщенная архитектура системы информационного взаимодействия представлена на рисунке 1. Выделены четыре программных уровня, каждый из которых обеспечивает сбор, обработку и хранение информации с учетом уровня иерархии токена источника данных (рис. 2).

На начальном этапе осуществляется процедура получения информации физическими прототипами на основании результатов мониторинга некоторых процессов / явлений / объектов, полученных как непосредственно с измерительных устройств, так и с иных источников информации. Данные поступают в модуль регистрации, результатом работы которого является массив данных унифицированного формата (указываемого при проектировании информационной системы). Далее данные направляются в модуль предварительной обработки, где выполняется устранение ошибок, допущенных несовершенством технического обеспечения, а также имеющих место ввиду человеческого фактора. Далее к данным применяется механизм сжатия (модуль архивирования), в основу работы которого положен алгоритм по типу gzip [14].

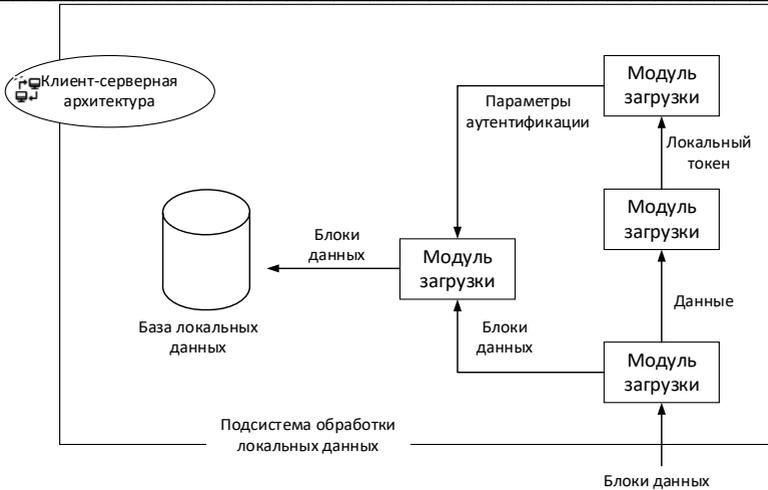
Прошедший архивирование набор данных передается на вход модулю пакетирования, который присоединяет к архивированным данным локальный токен и передает полученный массив данных для последующей отправки в подсистему промежуточного хранения данных. При этом локальный токен физического прототипа размещается в той же подсистеме, которая формирует исходящий набор данных, а поступает туда изначально из внешнего модуля локальной токенизации. Последний, в свою очередь, формирует идентификатор локального токена применительно к источнику данных на основании

его уникального номера и метаданных, полученных в процессе регистрации.

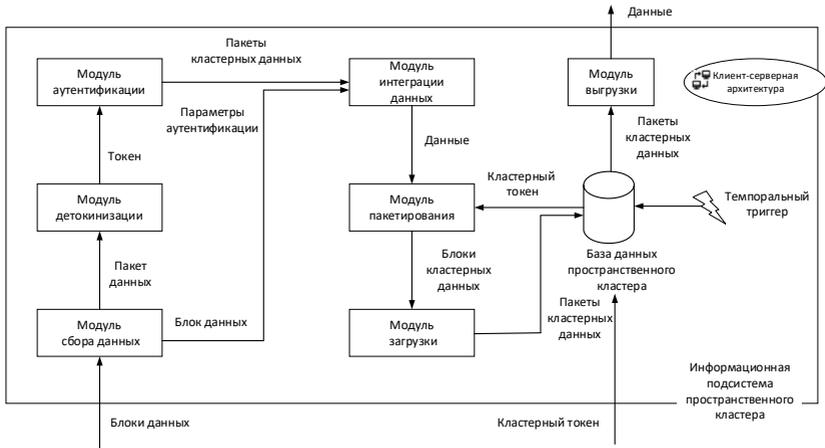
Аналогичным образом функционирует информационная подсистема цифровых двойников в рамках описываемой системы. Поскольку цифровой двойник по определению [15] представляет собой математическую модель соответствующего физического прототипа, основным компонентом его информационной подсистемы является модуль моделирования. Результатом его выполнения является формирование набора данных (в заданном на этапе проектирования формате), который соответствует синхронно зарегистрированным значениям физического прототипа. Данные, как и в случае с физическим прототипом, проходят процедуру архивирования и пакетирования, при этом локальный токен аналогичным образом передается в подсистему из модуля локальной токенизации, поскольку цифровой двойник, также как и его физический прототип, регистрируется в реестре источников данных в составе информационной системы.



А



Б



В

Рис. 2. Структурные схемы подсистем информационного взаимодействия системы цифровых двойников: А – информационная подсистема физического прототипа; Б – подсистема обработки локальных данных; В – информационная подсистема пространственного кластера

Представляется целесообразным отметить, что модуль выгрузки данных в информационных подсистемах как физического прототипа, так и его цифрового двойника запускается на выполнение в фоновом

режиме соответствующим процессом-демоном, имитирующим заданный темпоральный триггер с применением принципов Cron.

В подсистеме промежуточного хранения выполняется верификация синхронных данных, полученных от физического прототипа и его цифрового двойника. На основании уже накопленной информации, выступающей в качестве прецедентов, формируется исходящий из подсистемы информационный поток, в котором представлены данные, наиболее вероятные и достоверные с точки зрения соответствующей регрессионной модели.

Следующим уровнем архитектуры выступает подсистема обработки локальных данных. Точкой входа подсистемы выступает модуль сбора данных, сохраняющий в виртуальный кэш сервера поступающие из распределенных источников данных и прошедшие предварительную обработку информационные потоки.

Далее управление передается модулю детокенизации, основным назначением которого является выделение локального токена из поступившего в составе информационного потока блока данных. Токен считывается из сформированного на предыдущем этапе набора данных (заголовка сформированного пакета).

На следующем этапе выделенный токен передается модулю аутентификации, который сопоставляет его значение с идентификаторами в реестре источников данных. В случае обнаружения соответствия блок данных передается в соответствующую секцию базы локальных данных через программный модуль загрузки. В противном случае источник данных классифицируется как неопознанный, в соответствующую подсистему направляется сообщение о возникшей ошибке, транзакция по записи информации в базу локальных данных отменяется.

Для подсистемы обработки локальных данных выгрузка информационных блоков из базы данных осуществляется в соответствии с заданным на этапе проектирования темпоральным триггером, выполнение которого реализуется соответствующими программными процессами-демонами, функционирующими по принципу Cron.

На следующем этапе управление передается информационной подсистеме пространственного кластера. По аналогии с описанной выше информационной подсистемой поток данных поступает с предшествующего уровня на вход модулю сбора данных. Выделенный в результате работы модуля детокенизации токен представлен в своем полном формате, включая значение токена соответствующего пространственного кластера. На уровне модуля аутентификации

выполняется оценка принадлежности источника данных, от которого поступил информационный блок, соответствующему пространственному кластеру. В случае обнаружения несоответствия источник данных классифицируется как неопознанный, в соответствующую подсистему направляется сообщение о возникшей ошибке, а транзакция по записи информации в базу данных пространственного кластера отменяется. При успешном завершении процедуры аутентификации источника данных соответствующий блок данных передается в модуль интеграции, основным назначением которого является агрегирование всех блоков данных, поступающих от источников данных пространственного кластера в течение заданного временного интервала.

На следующем шаге данные из модуля интеграции поступают для обработки в модуль пакетирования, где к агрегированному набору данных добавляется токен соответствующего пространственного кластера. Последний, в свою очередь, поступает в подсистему в результате работы модуля кластерной токенизации программной подсистемы токенизации синхронно с регистрацией нового пространственного кластера в реестре. Примечательно, что при необходимости присутствующий в блоках данных токен кластера может быть обновлен в соответствии с тем значением, которое присваивается заголовку блока в модуле пакетирования.

Модуль загрузки помещает данные для краткосрочного хранения в базу данных пространственного кластера. По аналогии с предшествующими уровнями системы информационного взаимодействия, для подсистемы обработки пространственного кластера выгрузка информационных блоков из базы данных осуществляется в соответствии с заданным на этапе проектирования темпоральным триггером, выполнение которого реализуется соответствующими программными процессами-демонами, функционирующими по принципу Стоп.

В результате выполнения модулей подсистемы пакеты кластерных данных консолидируются в выделенном хранилище в составе единого информационного пространства для последующего применения в информационных системах (системах поддержки принятия решений).

8. Особенности программной реализации информационного взаимодействия системы цифровых двойников. Одним из технических требований к системе цифровых двойников является обеспечение оперативного доступа к данным, территориально распределенным на физически удаленных вычислительных машинах.

Эффективным способом организации решения, удовлетворяющего указанному критерию, является применение клиент-серверной архитектуры, которая обеспечивает возможность единого доступа к данным большому количеству пользователей (и программ) одновременно.

Результатом программной реализации предложенного в работе решения по информационному взаимодействию системы цифровых двойников является веб-ориентированное клиент-серверное приложение, бизнес-логика которого построена по принципу бессерверных вычислений [16–18]. Преимуществом выбранного подхода является то, что он позволяет избежать необходимости выделения и настройки специализированных серверов.

На физическом уровне разработанная программная система представлена совокупностью информационных подсистем физических прототипов, пространственного кластера и обработки локальных данных. Все компоненты системы реализованы на базе фреймворка Django [19–21], одной из отличительных особенностей которого является поддержка архитектурного паттерна «Модель – Представление – Контроллер». Указанная особенность Django в рамках решения задачи программной реализации системы цифровых двойников позволила в явном виде отделить задачи хранения, обработки и визуализации соответствующих данных.

При этом доступ к отдельным компонентам приложения реализован по принципу инфраструктуры CDN (Content Delivery Network – сеть доставки содержимого) [22], что позволило обеспечить, соответственно, более высокую скорость доступа для клиентского рендеринга по сравнению с другими способами организации веб-инфраструкты.

Важно отметить, что в рамках предложенной реализации для каждого источника данных (независимо от того, является ли он физическим прототипом или его цифровым двойником) выделяется собственная информационная подсистема, построенная по принципу организации клиент-серверной архитектуры и представляющая собой, по сути, компоненту единой информационной многоуровневой системы цифровых двойников. Каждая подсистема оформлена в виде облачной функции, реализованной на основе платформы Yandex.Cloud [24]. Возможность динамического масштабирования выделенной инфраструктуры, поддержка языка программирования Python (используемого и в рамках остальных компонент системы информационного взаимодействия), контейнеруемость – стали определяющими параметрами для выбора данного способа реализации

информационной подсистемы. В результате применения архитектуры бессерверных вычислений для решения поставленной задачи была достигнута высокая реактивность подсистемы, составившая порядка 0,28 с времени отклика на атомарный запрос в противовес 78, 41 с для «классической» клиент-серверной архитектуры.

Для конечных пользователей (например, операторов системы со стороны источников данных или соответствующих программных сервисов) в рамках указанной клиент-серверной архитектуры в соответствии с предложенной программной реализацией предусмотрен единый веб-ориентированный интерфейс доступа, посредством которого осуществляется управление загрузкой регистрируемых или моделируемых наборов данных с заданным временным интервалом и в заданном формате взаимодействия. Дальнейшая обработка данных выполняется исключительно на серверной стороне, в то время как на уровне клиентского приложения (интерфейса) возможно одностороннее наблюдение за процессом смены состояний данных.

Так, согласно предложенной программной реализации, данные, поступившие через клиентский веб-ориентированный интерфейс от физического прототипа (как вариант – магнитной обсерватории или вариационной станции) последовательно передаются от одного серверного модуля другому, проходя соответственно основные этапы преобразования – от предварительной обработки до пакетирования и передачи на следующий уровень (в подсистему промежуточного хранения данных).

Предложен подход, согласно которому передача данных между программными модулями на уровне веб-сервера осуществляется посредством перенаправления результирующих данных с выхода одного модуля на вход другому в виде виртуальных информационных потоков (stream) без дополнительного физического сохранения на дисковом буфере. Результатом выполнения всех операций преобразования на этом уровне в конечном итоге является так называемый пакет или блок данных, сформированный путем добавления уникального токена к архивированному и прошедшему обработке набору исходных данных, полученных непосредственно от технической системы (например, информационно-измерительного устройства магнитной обсерватории или вариационной станции, непрерывно регистрирующего данные о параметрах магнитного поля Земли и его вариациях).

Поскольку одним из определяющих факторов информационного взаимодействия системы цифровых двойников является скорость передачи соответствующих информационных потоков, то в процессе

реализации предложенных решений возникает задача представления данных, передаваемых между уровнями системы. В этой связи в качестве основного был принят формат колоночного представления данных [23], который являет собой бинарное представление агрегированных в столбцы записей и предполагает обязательное архивирование (например, на основе популярного алгоритма сжатия gzip). К основным преимуществам такого подхода относятся высокая реактивность программных средств для обработки данных такого типа, а также относительно низкие вычислительные затраты, связанные с физическим хранением больших наборов данных. Указанные возможности позволяют успешно решить ряд проблем обработки данных, связанных с их большим и непрерывно растущим объемом (кроме того, возможно сокращение затрат на хранение и представление сопровождающей атрибутивные данные пространственной информации).

Также важно отметить, что в соответствии с предложенной программной реализацией в подсистеме промежуточного хранения данных, куда поступают потоки данных от информационных подсистем физических прототипов и их цифровых двойников, инкапсулирован модуль аналитической обработки, реализующий модели и методы статистического анализа, в частности, регрессионного анализа. В результате выполнения операций, предусмотренных данным модулем, осуществляется верификация двух наборов данных – физического прототипа и цифрового двойника, и на основании накапливаемой базы прецедентов определяется тот набор данных, который содержит наиболее вероятные для данных пар значения.

9. Анализ эффективности предложенной инфраструктуры.

Для оценки эффективности предложенных моделей и методов информационного взаимодействия компонентов многоуровневой системы цифровых двойников была разработана информационная веб-ориентированная система, обеспечивающая визуализацию геомагнитных данных, регистрируемых магнитными обсерваториями и вариационными станциями, а также моделируемая соответствующими цифровыми двойниками.

Проект «Geomagnetic Plots» (режим доступа: <https://geomagnetic.ru>) реализован на основе фреймворка Django и предназначен для оперативной веб-ориентированной визуализации пространственных данных, описывающих динамику изменения во времени параметров магнитного поля Земли и его вариаций [25]. В качестве источников данных выступают магнитные обсерватории и вариационные станции, а также их цифровые модели, верификация которых выполняется

программным модулем, реализующим процедуру регрессионного анализа (рис. 3).

Работа веб-проекта «Geomagnetic Plots» инициируется конечным пользователем: вводятся (выбираются посредством элемента управления) дата и время, а также визуализируемый параметр магнитного поля. На основании передаваемых на сервер параметров выполняется поиск соответствующих геомагнитных данных, представленных в едином информационном пространстве. Последние, в свою очередь, передаются в систему из многоуровневой системы цифровых двойников, информационное взаимодействие которых реализовано по предложенному в работе принципу и на основании описанной выше смешанной клиент-серверной и бессерверной архитектур.

Для сравнения был разработан исследовательский прототип веб-ориентированной системы, реализующей функциональность, аналогичную проекту «Geomagnetic Plots» [26]. Однако формирование и заполнение единого информационного пространства здесь реализовано посредством системы прямых запросов к источникам данных (физическим прототипам и их цифровым двойникам) в строгом соответствии с архитектурой «клиент-сервер».

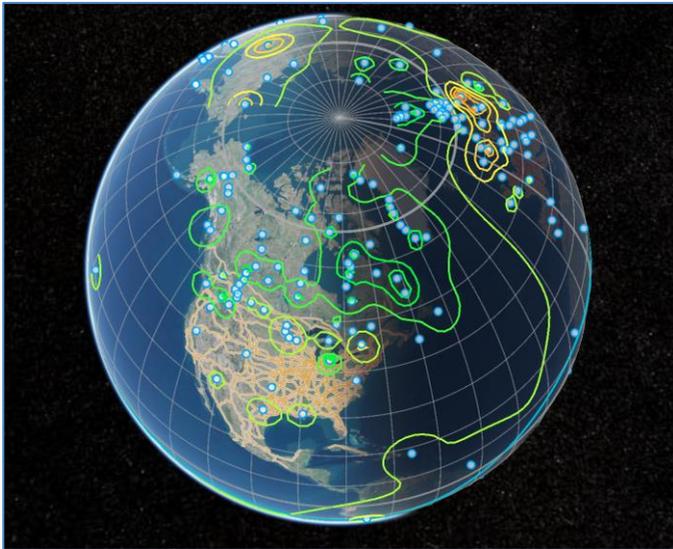


Рис. 3. Экранная форма веб-проекта «Geomagnetic Plots»

Оценка качества обеих программных систем была проведена на основании методики, представленной в ГОСТ 28195-89 «Оценка качества программных средств». В соответствии с этой методикой были оценены показатели надежности, сопровождения, удобства применения, эффективности, универсальности и корректности программной системы. Тестирование было проведено в нормальных и экстремальных условиях, а также исключительных ситуациях. Согласно результатам вычислительных экспериментов, обе программные системы корректно выполняют свои функции. В экстремальных и исключительных ситуациях программы выдают сообщения о соответствующей ошибке и продолжают работу в штатном режиме.

Кроме того, в ходе проведения вычислительных экспериментов была выполнена сравнительная оценка времени отклика соответствующих программных систем. Анализ результатов эксперимента показал, что выполнение запроса к данным и рендеринга их в условиях применения персонального компьютера со средней производительностью (процессор с частотой 1.6 ГГц, 2 ядра, оперативная память 4 Гб, скорость интернет-соединения 342.7 Мбит/с) в обоих случаях занимает в среднем около 6 с, что меньше времени отклика при использовании практикуемого способа, составляющего ~49 с (~12.2 % от исходного времени).

Значительные различия в производительности были обнаружены при исследовании выполнения фоновых серверных процессов каждой из программных систем, связанных с передачей данных между уровнями системы цифровых двойников и их выгрузкой в единое информационное пространство.

Предложенный подход, основанный на применении промежуточных виртуальных хранилищ и комбинировании традиционной клиент-серверной и бессерверной программных архитектур, сравнивался с практикуемым подходом последовательного клиент-серверного опроса каждого источника данных (как физического прототипа, так и его цифрового двойника).

Вычислительные эксперименты были проведены для годовых геомагнитных данных 147 магнитных обсерваторий и цифровых двойников на веб-сервере с характеристиками следующего порядка: 72 * Intel(R) Xeon(R) Gold 6140 CPU @ 2.30GHz.

Результаты проведенных экспериментов показали, что применение предложенного подхода к решению задачи информационного взаимодействия компонентов системы цифровых двойников позволит повысить время выполнения запроса в среднем в

71,3 раза по сравнению с монолитной клиент-серверной архитектурой доступа к физическим прототипам и цифровым двойникам.

10. Заключение. В настоящее время проблема пространственно-временной анизотропии особенно остро проявляется при решении задач изучения объектов и процессов окружающей среды. Методы математической статистики, геостатистики, геоинформационного моделирования и других областей знаний частично решают задачи пространственной и временной интерполяции, но не одинаково эффективно для различных пространственных областей и прикладных задач. К примеру, пространственная интерполяция геомагнитных данных такими методами обеспечивает значение среднеквадратической ошибки, существенно выше допустимой геофизическими стандартами. Похожие результаты наблюдаются и в других областях.

Одним из вариантов решения этой проблемы является применение системы цифровых двойников, которые на основе регрессионных моделей и данных физических прототипов позволяют агрегировать наиболее вероятные значения данных за искомые пространственные и временные интервалы. При этом информационное взаимодействие между уровнями системы значимо усугубляется постоянно растущим объемом и сложностью регистрируемых и обрабатываемых данных.

В работе для решения поставленной задачи была предложена организация информационного взаимодействия в системе цифровых двойников на основе формализованного механизма пакетирования пространственно-временной информации, при котором идентификация источников данных выполняется посредством иерархической системы бинарной токенизации. Отличительной особенностью предложенного подхода является самоидентифицируемость источников данных, обеспечиваемая включением в состав информационного пакета соответствующего токена, который содержит необходимые для анализа метаданные источника данных. Еще одним отличием является предложенная архитектура информационного взаимодействия, предполагающая буферное хранение данных для организации их предварительной обработки перед передачей информационного пакета либо последующему уровню, либо во внешний программный модуль.

Проведенные вычислительные эксперименты (на примере геомагнитных данных) показали, что применение предложенных моделей и методов информационного взаимодействия для реализации системы цифровых двойников позволит:

- 1) повысить скорость сбора, обработки и агрегирования информации с физических прототипов, и их цифровых

двойников за счет минимизации долговременного хранения промежуточных данных посредством применения темпоральных триггеров;

- 2) увеличить реактивность сценариев сбора, обработки и агрегирования информации с физических прототипов, и их цифровых двойников посредством комбинирования клиент-серверной и бессерверной программных архитектур;
- 3) сократить расходы на физическое хранение данных, агрегируемых с физических прототипов и их цифровых двойников посредством формирования специализированных пакетов данных, содержащих идентификационный токен в своем заголовке и архивированные бинарные данные в теле;
- 4) упростить процедуру идентификации источников данных в процессе сбора, обработки и агрегирования информации с физических прототипов, и их цифровых двойников посредством введения многоуровневой системы токенизации.

Ряд сервисов информационной системы размещен на внешнем веб-сервере с открытым физическим доступом, а остальные находятся стадии интеграционного тестирования и отладки на локальном виртуальном сервере. Дальнейшие исследования предполагают полное размещение сервисов на внешнем ресурсе и последующее развитие заявленной функциональности.

Литература

1. Воробьев А.В., Воробьева Г.Р. Подход к оценке относительной информационной эффективности магнитных обсерваторий сети INTERMAGNET // Геомагнетизм и аэронавигация. 2018. Т. 58. № 5. с. 648–652.
2. Гвишиани А.Д., Лукьянова Р.Ю., Соловьёв А.А. Геомагнетизм: от ядра Земли до Солнца. М.: РАН, 2019. 186 с.
3. Гвишиани А.Д., Агаян С.М., Богоутдинов Ш.Р., Каган А.И. Гравитационное сглаживание временных рядов // Труды Института математики и механики УрО РАН. 2011. Т.17. № 2. с. 62–70.
4. Рябова С.А. Геомагнитные вариации и синхронные с ними вариации уровня подземных вод и микросейсмического фона в условиях средних широт: дисс. на соискание ученой степени кандидата физико-математических наук. М., 2018.
5. Kondrashov D., Shprits, Y., Ghil, M. Gap filling of solar wind data by singular spectrum analysis // Geophys. Res. Lett. 2010. no. 37. P. L15101 DOI:10.1029/2010GL044138
6. Аббасова Т.С. Развитие виртуальных инструментов для создания цифровых двойников // Информационно-технологический вестник. 2019. № 2(20). С. 79–88.
7. Chen Y. New Approaches for Calculating Moran's Index of Spatial Autocorrelation // PloS one. 2013. no. 8. P. e68336. DOI: 10.1371/journal.pone.0068336.
8. Freni P., Ferro E., Moncada R. Tokenization and Blockchain Tokens Classification: a morphological framework // IEEE Symposium on Computers and Communications (ISCC), Rennes, France, 2020, pp. 1-6, doi: 10.1109/ISCC50000.2020.9219709.

9. *Marchewka-Bartkowiak K., Nowak K.* Get Tokenized... The Specificity of Personal Tokens in the Context of Tokenization and Axiological Categorization // Proceedings of the 3rd International Conference on Economics and Social Sciences. 2020. pp. 823-831. DOI:10.2478/9788395815072-081.
10. *Love J.* Intermagnet and the Global Community of Magnetic Observatories // AGU Fall Meeting Abstracts. 2006.
11. *Stocchero M.* Iterative deflation algorithm, eigenvalue equations, and PLS2 // Journal of Chemometrics. 2019. no.33. p.10.1002/cem.3144.
12. *Le H. A.* Trigger2B: A Tool Generating Event-B Models from Database Triggers // Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. 2021. vol. 343. pp. 1-7. DOI: 10.1007/978-3-030-67101-38.
13. *Lenz M.* Silent-Cron: A Cron Wrapper // Raku Fundamentals. 2020. pp. 59-80. doi:10.1007/978-1-4842-6109-5_6.
14. *Voss N., Becker T., Mencer O., Gaydadjiev G.* Rapid Development of Gzip with MaxJ // Lecture Notes in Computer Science. 2017. vol. 10216. p. 60-71. DOI: 10.1007/978-3-319-56258-2_6.
15. *Mittal H.* Digital twin: An overview // CSI Communications. 2020. no. 44.
16. *Carver J., Penzenstadler B., Scheuner J., Staron M.* (Research) Insights for Serverless Application Engineering // IEEE Software. 2021. vol. 38. pp. 123-125. DOI: 10.1109/MS.2020.3028659.
17. *Mahmoudi N., Khazaei H.* Performance Modeling of Serverless Computing Platforms // IEEE Transactions on Cloud Computing. 2020. pp. 1-1. DOI: 10.1109/TCC.2020.3033373.
18. *Martins H., Araujo F., Cunha P. R.* Benchmarking Serverless Computing Platforms // Journal of Grid Computing. 2020. vol. 18. DOI: 10.1007/s10723-020-09523-1.
19. *Ali S., Alauldeen R., Ruaa A.* What is Client-Server System: Architecture, Issues and Challenge of Client-Server System (Review) // HBRP Publication. 2020. pp. 1-6. DOI: 10.5281/zenodo.3673071.
20. *Zhu Y., Wu W., Li D.* Efficient Client Assignment for Client/Server Systems // IEEE Transactions on Network and Service Management. 2016. no. 13. pp. 1-10. DOI: 10.1109/TNSM.2016.2597269.
21. *Iskandar T., Lubis M., Kusumasari T., Lubis A.* Comparison between client-side and server-side rendering in the web development // IOP Conference Series Materials Science and Engineering. 2020. no. 801. pp. 012136. DOI: 10.1088/1757-899X/801/1/012136.
22. *Zeng Z., Zhang H.* A Study on Cache Strategy of CDN Stream Media // IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2020. pp. 1424-1429. DOI: 10.1109/ITAIC49862.2020.9338805.
23. *Sharma K., Marjit U., Biswas U.* Efficiently Processing and Storing Library Linked Data using Apache Spark and Parquet // Information Technology and Libraries. 2017. no.37. pp. 29-34. DOI:10.6017/ital.v37i3.10177.
24. *Kostinskiy S., Shaikhudinov D., Narakidze N.* Loss Counter in Power Double Winding Transformers Implementing the Method of Conditionally Constant Coefficients in Online Mode Using the Information Platform // Известия высших учебных заведений. Электромеханика. 2020. no. 63. pp. 79-85. DOI: 10.17213/0136-3360-2020-5-79-85.
25. *Vorobev A.V., Pilipenko V.A., Enikeev T.A., Vorobeva G.R.* Geoinformation system for analyzing the dynamics of extreme geomagnetic disturbances from observations of ground stations // Computer Optics. 2020. no. 44(5). pp. 782-790. DOI: 10.18287/2412-6179-CO-707.
26. *Воробьев А.В., Воробьева Г.Р., Юсупова Н.И.* Концепция единого пространства геомагнитных данных // Труды СПИИРАН. 2019. в. 18, т. 2. С. 390–415.

Воробьев Андрей Владимирович — канд. техн. наук, доцент, кафедра геоинформационных систем факультета информатики и робототехники, Уфимский государственный авиационный технический университет (УГАТУ); старший научный сотрудник Геофизического центра РАН. Область научных интересов: геоинформационные технологии, цифровая обработка сигналов. Число научных публикаций — 147. geomagnet@list.ru; К. Маркса, 12, 450008, Уфа, Российская Федерация; р.т.: +7(917)345-2299.

Воробьева Гульнара Равилевна — канд. техн. наук, доцент, кафедра вычислительной математики и кибернетики факультета информатики и робототехники, Уфимский государственный авиационный технический университет (УГАТУ). Область научных интересов: геоинформационные и веб-технологии, системы хранения и обработки информации. Число научных публикаций — 121. gulnara.vorobeva@gmail.com; К. Маркса, 12, 450008, Уфа, Российская Федерация; р.т.: +7(917)417-4111.

A. Vorobev, G. Vorobeva

MODEL OF INFORMATION INTERACTION BETWEEN ELEMENTS OF MULTILEVEL SYSTEM OF DIGITAL TWINS

Vorobev A., Vorobeva G. Model of Information Interaction Between Elements of Multilevel System of Digital Twins.

Abstract. One of the solutions to the problem of spatio-temporal data anisotropy is the use of a multilevel system of digital twins based on the corresponding industry models and the updated archive data base. The application of this approach has successfully proved itself in information systems for monitoring the parameters of the geomagnetic field and its variations, providing spatio-temporal interpolation of geomagnetic data with an accuracy of 0.81 nT in magnetically quiet periods. At the same time, the problem of information interaction between the levels of the system of digital twins remained unresolved, which is greatly aggravated by the constantly growing volume of data and their heterogeneous nature. The paper proposes a solution to the indicated problem by means of a formalized mechanism for packaging space-time information, in which the identification of data sources is performed on the basis of a hierarchical binary tokenization system. In addition, the proposed software implementation of such an approach is considered, a distinctive feature of which is the combination of traditional client-server and innovative serverless architectures to implement a highly loaded reactive web application for working with analyzed data. The main stages of the implementation of information interaction are highlighted and programmatically formalized - from obtaining initial information from its sources to verifying data, analyzing them, processing and forming the output information flow of the system. The results of the computational experiments carried out on the example of the problem of monitoring the parameters of the Earth's magnetic field and its variations confirmed the effectiveness of the proposed solutions, expressed both in increasing the reactivity of web-based applications and in increasing the computational speed of formation and filling of information storages that aggregate information from distributed heterogeneous sources.

Keywords: Digital Twins, Spatial Data, Information Interaction, Unified Information Space, Data Processing, Tokenization, Geographic Information Systems and Technologies

Vorobev Andrei — Ph.D., Associate Professor, Geoinformation Systems Department of Computer Science and Robotics Faculty, Ufa State Aviation Technical University (USATU); Senior Researcher, Geophysical Center of RAS. Research interests: geoinformation technologies, digital signal processing. The number of publications — 147. geomagnet@list.ru; 12, K. Marx, 450008, Ufa, Russian Federation; office phone: +7(917)345-2299.

Vorobeva Gulnara — Ph.D., Associate Professor, Computational Mathematics and Cybernetics Department of Computer Science and Robotics Faculty, Ufa State Aviation Technical University (USATU). Research interests: geoinformation and web technologies, systems of information storing and processing. The number of publications — 121. gulnara.vorobeva@gmail.com; 12, K. Marx, 450008, Ufa, Russian Federation; office phone: +7(917)417-4111.

References

1. Vorobev A.V., Vorobeva G.R. [An approach to assessing the relative information efficiency of magnetic observatories in the INTERMAGNET network]. *Geomagnetizm i aeronomiya – Geomagnetism and Aeronomy*. 2018. vol. 58. no. 5. pp. 648–652. (In Russ.).

2. Gvishiani A.D., Lukyanova R.Yu., Soloviev A.A. [Geomagnetism: from the core of the Earth to the Sun]. Moscow: RAS, 2019. 186 p. (In Russ.).
3. Gvishiani A.D., Agayan S.M., Bogoutdinov Sh.R., Kagan A.I. [Gravitational smoothing of time series]. *Trudy Instituta matematiki i mekhaniki UrO RAN – Proceedings of the Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences*. 2011. vol. 17. no. 2. pp. 62–70. (In Russ.).
4. Ryabova S.A. [Geomagnetic variations and variations in groundwater level and microseismic background in conditions of middle latitudes synchronous with them]: diss. for the degree of candidate of physical and mathematical sciences. Moscow, 2018 (In Russ.).
5. Kondrashov D., Shprits, Y., Ghil, M. Gap filling of solar wind data by singular spectrum analysis. *Geophys. Res. Lett.* 2010. no. 37. p. L15101 DOI:10.1029/2010GL044138. (In Russ.).
6. Abbasova T.S. [Development of virtual instruments for creating digital twins] // *Informatsionno-tehnologicheskii vestnik – Information and technological bulletin*. 2019. no. 2(20). pp. 79–88. (In Russ.).
7. Chen Y. New Approaches for Calculating Moran’s Index of Spatial Autocorrelation. *PloS one*. 2013. no. 8. p. e68336. DOI: 10.1371/journal.pone.0068336.
8. Freni P., Ferro E., Moncada R. Tokenization and Blockchain Tokens Classification: a morphological framework. *IEEE Symposium on Computers and Communications (ISCC)*, Rennes, France, 2020, pp. 1-6, DOI: 10.1109/ISCC50000.2020.9219709.
9. Marchewka-Bartkowiak K., Nowak K. Get Tokenized.. The Specificity of Personal Tokens in the Context of Tokenization and Axiological Categorization. *Proceedings of the 3rd International Conference on Economics and Social Sciences*. 2020. pp. 823-831. DOI:10.2478/9788395815072-081.
10. Love J. Intermagnet and the Global Community of Magnetic Observatories. *AGU Fall Meeting Abstracts*. 2006.
11. Stocchero M. Iterative deflation algorithm, eigenvalue equations, and PLS2. *Journal of Chemometrics*. 2019. No. 33. p. 10.1002/cem.3144.
12. Le H. A. Trigger2B: A Tool Generating Event-B Models from Database Triggers. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. 2021. vol. 343. pp. 1-7. DOI: 10.1007/978-3-030-67101-38
13. Lenz M. Silent-Cron: A Cron Wrapper. *Raku Fundamentals*. 2020. pp. 59-80. DOI: 10.1007/978-1-4842-6109-56.
14. Voss N., Becker T., Mencer O., Gaydadjiev G. Rapid Development of Gzip with MaxJ. *Lecture Notes in Computer Science*. 2017. vol. 10216. pp. 60-71. DOI: 10.1007/978-3-319-56258-2_6.
15. Mittal H. Digital twin: An overview. *CSI Communications*. 2020. no. 44.
16. Carver J., Penzenstadler B., Scheuner J., Staron M. (Research) Insights for Serverless Application Engineering. *IEEE Software*. 2021. vol. 38. pp. 123-125. DOI: 10.1109/MS.2020.3028659.
17. Mahmoudi N., Khazaei H. Performance Modeling of Serverless Computing Platforms. *IEEE Transactions on Cloud Computing*. 2020. pp. 1-10. DOI: 10.1109/TCC.2020.3033373.
18. Martins H., Araujo F., Cunha P. R. Benchmarking Serverless Computing Platforms. *Journal of Grid Computing*. 2020. vol. 18. DOI: 10.1007/s10723-020-09523-1.
19. Ali S., Alauldeen R., Ruaa A. What is Client-Server System: Architecture, Issues and Challenge of Client-Server System (Review). *HBRP Publication*. 2020. pp. 1-6. DOI: 10.5281/zenodo.3673071.
20. Zhu Y., Wu W., Li D. Efficient Client Assignment for Client/Server Systems. *IEEE Transactions on Network and Service Management*. 2016. no. 13. pp. 1-10. DOI: 10.1109/TNSM.2016.2597269.

21. Iskandar T., Lubis M., Kusumasari T., Lubis A. Comparison between client-side and server-side rendering in the web development. *IOP Conference Series Materials Science and Engineering*. 2020. no. 801. pp. 012136. DOI: 10.1088/1757-899X/801/1/012136.
22. Zeng Z., Zhang H. A Study on Cache Strategy of CDN Stream Media. *IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Chongqing, China, 2020. pp. 1424-1429. DOI: 10.1109/ITAIC49862.2020.9338805.
23. Sharma K., Marjit U., Biswas U. Efficiently Processing and Storing Library Linked Data using Apache Spark and Parquet. *Information Technology and Libraries*. 2017. no. 37. pp. 29-34. DOI: 10.6017/ital.v37i3.10177.
24. Kostinskiy S., Shaikhutdinov D., Narakidze N. Loss Counter in Power Double Winding Transformers Implementing the Method of Conditionally Constant Coefficients in Online Mode Using the Information Platform. *Izvestiya vysshikh uchebnykh zavedeniy. Elektromekhanika - Proceedings of higher educational institutions. Electromechanics*. 2020. no. 63. pp. 79-85. DOI: 10.17213/0136-3360-2020-5-79-85. (In Russ.).
25. Vorobev A.V., Pilipenko V.A., Enikeev T.A., Vorobeva G.R. Geoinformation system for analyzing the dynamics of extreme geomagnetic disturbances from observations of ground stations. *Computer Optics*. 2020. no. 44(5). pp. 782-790. DOI: 10.18287/2412-6179-CO-707.
26. Vorobev A.V., Vorobeva G.R., Yusupova N.I. [The concept of a single space of geomagnetic data]. *Trudy SPIRAN - SPIRAS Proceedings*. 2019. no. 18, vol. 2. pp. 390-415. (In Russ.).

М.В. ВОРОНОВ, В.И. ПИМЕНОВ
ФОРМАЛИЗАЦИЯ РЕГУЛЯТИВНЫХ ТЕКСТОВ

Воронов М.В., Пименов В.И. Формализация регулятивных текстов.

Аннотация. Современные информационные технологии с высокой эффективностью обеспечивают процессы манипулирования текстами. В первую очередь это хранение, редактирование и форматирование текстов и их составляющих. Добившись значительных успехов в разработке средств бессодержательной компьютерной обработки текстов, исследователи столкнулись с проблемами их содержательной обработки. Поэтому дальнейшие шаги в этом направлении связаны с созданием, в том числе, методов автоматизированного целенаправленного манипулирования текстами с учетом заложенного в них содержания. Проведен анализ работ, посвященных исследованию проблем формального представления текстов и их последующего использования. Несмотря на ряд успешных проектов, задачи по разрешению проблемы соотношений между содержанием текста и его смыслом остаются актуальными. По-видимому, формализация текста общего назначения при сохранении его семантики на данном этапе развития науки не осуществима. Вместе с тем существуют типы текстов, для которых формализация с сохранением существенной части их семантики представляется возможной. Одним из них является регулятивный тип текста, который по существу представляет собой словесно выраженный алгоритм последовательности целенаправленных действий. Его отличают логичность и точность (отсутствие иносказаний), связность и цельность, ясность, понятность (за счет отсутствия эмоциональной окраски и образных средств), доступность (за счет использования специфической терминологии). Иначе говоря, при разработке регулятивных текстов обычно стараются как можно понятнее отображать механизмы описываемых действий. Цель: разработка метода формализации регулятивного текста с сохранением его семантики. Методы: структурная лингвистика, представление объектов в виде онтологии, конструктивные алгоритмы. Использование данного метода демонстрируется на примере описания решения системы алгебраических уравнений. Результаты: метод построения математической модели регулятивного текста. Практическая значимость: применение разработанного метода обеспечивает возможность разработки программных комплексов для построения библиотек отдельных предметных областей, разработки средств оценки регулятивных текстов на предмет их определенности, полноты, связности и других характеристик, а также тренажеров и средств самообучения.

Ключевые слова: регулятивный текст, понятие, структурная лингвистика, искусственный интеллект, математическая модель, конструктивный алгоритм, онтология, фрейм

1. Введение. Результаты своей мыслительной деятельности человек стремится зафиксировать на каком-либо материальном носителе. Для этого чаще всего используются определенные последовательности символов, называемые текстами. Изображение текста представляет собой определенную форму (упорядоченную последовательность символов), отображающую вложенное в него автором содержание и является попыткой автора фиксировать и объективировать последнее.

Представители многих областей знаний рассматривают тексты в качестве объектов своих исследований, поскольку они предназначены, в первую очередь, для переноса в пространстве и во времени заложенных в них автором результатов его речемыслительной деятельности. Иными словами, текст исполняет роль почтальона, цель которого – передать получателю вложенное автором содержание, причем без искажений.

Вычленив же из текста находящееся в нем содержание сегодня способен только человек, ибо содержание текста не существует вне процесса его восприятия и понимания. Эти же процессы «включаются» под воздействием сигналов, формирующихся при выявлении составляющих текст символов. При этом успех процессов восприятия и передачи смысла (того, что может быть понято) обусловлен тем, в какой мере возможности восприятия информации данного контекста у автора и читателя совпадают, что в свою очередь зависит и от того, как построен сам текст, то есть из каких компонентов и связей между ними он состоит.

Современные информационные технологии, оперируя с текстами в основном как с бессодержательной структурой, с высокой эффективностью обеспечивают процессы такого манипулирования ими. В первую очередь это хранение, редактирование и форматирование текстов и их составляющих. Но не только. Отдельные программы в определенной мере уже «умеют», например, выделять факты из текста, распознавать сарказм и т.п. Однако они «не понимают слова» как это делают люди, а используя, как правило, вероятностные модели и нечеткие алгоритмы, попросту угадывают значения слов, словесных оборотов и даже отдельных предложений. Дальнейшие шаги в этом направлении связаны с созданием, в том числе, методов автоматизированного целенаправленного манипулирования текстами с учетом заложенного в них содержания (исходя из этого содержания и руководствуясь им), что является предметом нашего исследования. Таким образом, несмотря на прогресс в области информационных технологий, задачи по разрешению проблемы соотношений между содержанием текста и его смыслом остаются актуальными [1].

Ставя задачу формализации текста со стремлением сохранить исходное содержание при передаче, делается попытка открыть новые возможности формирования баз знаний для перспективных систем искусственного интеллекта.

2. Анализ проблем формализации текстов. Использование в текстах слов, представляющих собой имена некоторых сущностей денотата, обеспечивает читателю возможность соотносить с

конкретным словом свое представление о соответствующих объектах. Например, отдельно взятое слово «стол» практически у каждого читателя будет ассоциироваться с соответствующим элементом мебели. Однако о каком конкретно столе идет речь и, следовательно, можно ли получить о нем какие-либо дополнительные сведения, данная совокупность букв не сообщает. Иначе говоря, считанное изолированное слово передает весьма обобщенные и, как правило, крайне скудные сведения. Более того, интерпретация слова в существенной мере зависит от воспринимающего его субъекта. Дело в том, что передаваемое данным словом содержание для конкретного лица в значительной мере обусловлено его индивидуальными представлениями об этом объекте (свойствах, связях и отношениях с другими объектами и явлениями). В таких случаях посредником между использованным словом и действительностью выступает концепт, включающий в себя некоторый потенциал данного слова [2]. Концепт, как выраженная словом сущность, и выступает основной единицей, передающей определенный компонент содержания текста. В этой связи в процессе формализации для каждого использованного в тексте слова целесообразно фиксировать отображаемый им концепт.

Для конкретизации и уточнения передаваемой отдельными словами информации используются их определенные сочетания. Именно добавление структуры в данную совокупность отдельных слов обеспечивает передачу существенных с позиции автора законченных мыслей. В таких случаях эту упорядоченную совокупность слов называют предложением. Важно заметить, что образующие предложения слова, объединенные определенной лексической семантикой, обеспечивают эффект так называемой семантической эмерджентности: для данной ситуации это объединение более информативно значимо, чем все слова этого предложения, взятые в отдельности. Именно поэтому предложение считают основной единицей содержания в тексте.

Содержательно обусловленное соединение предложений образует текст, представляющий собой некую целостность, причем подмножество составляющих его предложений не может в полной мере отображать эту целостность. Иначе говоря, из смыслов, передаваемых отдельными словами и даже предложениями, невозможно в полной мере воспринять смысл данного текста.

Автор, закладывая в формируемый им текст определенное содержание, стремится к тому, чтобы восприятие читателя текста было в достаточной мере адекватно его мыслям. К сожалению, на пути от автора через текст к читателю (адресату) зафиксированные результаты

мыслительной деятельности автора подвергаются различного рода искажениям. В какой мере результаты деятельности читающего текст будут соответствовать результатам мыслительной деятельности автора, зависит, как минимум, от трех составляющих.

Во-первых, от личностных особенностей автора, ведь часто думая одно, он отображает в тексте несколько иное. Так, в работе В.Д. Орехова, например, отмечается, что точность такой передачи составляет порядка 14% [3].

Во-вторых, от читателя. Основная цель читающих данный текст людей заключается в извлечении из него той информации, которую в него заложил автор. Автор, как правило, один. Читателями же могут выступать самые различные люди и каждый из них может по-разному воспринимать заложенное в данном тексте содержание, причем даже при условии, что автор и читатель текста, что называется, «говорят на одном языке». Влияние множества разнородных факторов на восприятие текста столь существенно, что позволяет ряду авторов утверждать: передача смысла в принципе не может быть автоматизирована [4].

В-третьих, от того, каким образом трансформировался текст на пути от автора к читателю. Дело в том, что после своего появления текст начинает свою «самостоятельную жизнь», переноса в пространстве и во времени зафиксированные в нем мысли автора, обеспечивает потенциальную возможность воспользоваться ими другим людям. При этом на своем «жизненном пути» текст может подвергаться различного рода трансформациям, причем как случайным, так и целенаправленным.

Стремление использовать компьютерные технологии для работы с текстами вызывает необходимость разработки методов, позволяющих трансформировать исходный вербальный текст в текст, представленный на искусственном (машинном) языке. Задача при этом состоит в том, чтобы при прочих равных условиях смысл, воспринимаемый при чтении текста в формализованном виде, в максимальной мере соответствовал смыслу, воспринимаемому при чтении этого же текста в вербальной форме.

Примечание. В данной работе авторы, стремясь минимизировать повторение фраз, используют категорические суждения типа «основной», «сохранить», «не может быть» и др., понимая, что более корректно дополнять их словами «как правило», «в основном» и т.п.

Тематика исследований, связанных с проблемами формального представления текстов и их последующего использования, весьма широка [5, 6]. При этом важное место в них обычно занимают вопросы

передачи смыслового содержания текстов [7]. Несмотря на широкий спектр подходов, лидирующие места здесь занимает, пожалуй, подход, позволяющий на основе выявления концептуального состава текста определять его смысловую структуру [8].

Одним из заслуживающих внимания подходов к решению поставленной выше задачи является методика, суть которой заключается в замене лексико-грамматической структуры текста денотативной структурой графа, который призван отображать присутствующие в тексте семантические связи [9]. В процессе его построения выделяются объекты (денотаты), о которых идет речь в тексте (будущие вершины графа) и связи между ними (будущие ребра графа). При этом все операции построения графа обусловлены передаваемым текстом смыслом и представляют собой некий способ материализации результата понимания [10]. К сожалению, методика построения денотатного графа (как и методики, построенные на использовании триединой конструкции «Узел-Функция-Объект» [11]) включает большое число процедур творческого характера, которые каждый индивидуум способен реализовывать по-своему, что может приводить к неоднозначности результата. Необходима же разработка методов, включающих минимум творческой составляющей субъектов.

Во многих современных приложениях, посвященных вопросам автоматизированного сравнения текстов, используется идея их представления в виде векторов (построение Vector Space Model) [12]. Координаты этих векторов формируются на основе частот вхождений каждого термина в данный текст. Тем самым исходная задача сводится к задаче, решаемой средствами векторной алгебры. Однако имеющие место несовпадения семантической и лексикографической схожести далеко не всегда гарантируют удовлетворительное решение поставленной задачи [13].

При решении вопросов формализации передаваемых текстами смыслов все большее значение приобретают достижения лингвистики. Дело в том, что объектом изучения лингвистики является язык, который самым тесным образом связан со всеми аспектами мыслительной и познавательной деятельности человека. Поэтому лингвистика в целом и когнитивная лингвистика в особенности контактируют с широким спектром наук и во все большей мере с проблемами информатики [14]. В этом плане представляет интерес анализ исследований по вычислительной лингвистике В.З. Демьянкова [15].

Со многими трудностями, характерными для задач формализации текстов, встречаются исследователи сферы переводов текстов с языка на язык (ведь задачу формализации текста в известной

мере можно сопоставлять с задачами его перевода с языка на язык). В теории перевода формулируются лексические и грамматические правила, которые помогают анализировать текст на входном языке и синтезировать его на выходном. Однако переводчики, как правило, воспринимают текст в виде совокупности отдельных предложений. В итоге применение этих правил не всегда обеспечивает сохранение смысла, заложенного его автором в первоисточнике [16].

Информацию о представленных в тексте фактах лингвисты обычно связывают с его содержанием, вводя в рассмотрение семантическое пространство, которое наделяют определенной структурой [17, 18]. Для нас важно, что основными компонентами этой структуры выступают пропозиции, как объективизированные средства содержания мысли [19]. Введение в анализ текстов структуры – существенный шаг в направлении решения вопросов их формализации. Следующим шагом в этом направлении было введение пресуппозиции, трактуемой как суждение, выводимое из данного высказывания по правилам истинности или уместности [20]. Казалось бы, здесь лингвисты вплотную приблизились к возможности применения математической логики. Однако основные их усилия направлены на развитие теории речевой коммуникации и дискурсивной психологии, поэтому полученные при этом результаты не всегда находят прямое применение при разработке формализованных методов передачи содержания текстов. Вместе с тем результаты лингвистических исследований являются базой, на которой строятся многие методы формализации текстов в различных сферах деятельности людей.

Проблематикой формализации передаваемых текстами смыслов занимаются и со стороны информатики. Здесь, в рамках рассматриваемой тематики, следует обратить внимание, например, на работу [21], где формирование и использование ассоциативно-онтологического представления данных обеспечивает получение средств автоматического реферирования научных и технических документов. Особо следует отметить работы, проведенные в рамках когнитивной информатики, которые связаны с изучением отображения в информационных системах процессов эволюции знаковых систем, в частности построения взаимно однозначных отношений между концептами и компьютерными кодами [22].

3. Основания построения моделей регулятивных текстов.

Представляется перспективным вести поиск путей разрешения проблемы, возникающей при формализации текста с сохранением передаваемого им содержания, базируясь на следующем утверждении: содержание любых формальных преобразований заключено в

причинности [23]. Причина же становится ясной, если понятен механизм ее действия. Отсюда следует, что при построении модели текста максимум внимания целесообразно сосредоточить на выявлении и передаче содержащихся в нем причинно-следственных механизмов.

К сожалению, различные тексты в неодинаковой мере содержат явные и четко описанные механизмы действий. Именно поэтому многие авторы считают, что разработать методику формализации произвольного текста при достаточно высоком уровне сохранения его содержания вряд ли возможно [24]. По-видимому, на данном этапе развития науки это так.

Однако существуют тексты, само назначение которых заключается в выполнении координационной (регулятивной) функции для обеспечения определенной деятельности людей, что обуславливает их специфическое построение и, как следствие, открывает возможности, достаточные для их формализации с сохранением исходной семантики.

Вместе с тем замечено, что любой текст в той или иной мере обладает свойством регулятивности, под которым понимается способность текста управлять познавательной и интерпретационной деятельностью читателя [25]. Оказалось, что этот феномен регулятивности благодаря целенаправленной организации текста позволяет более эффективно реализовывать его авторскую интенцию. Поскольку намерения автора всегда направлены на адекватное восприятие его текста читателем.

Сопоставление различных типов текстов позволяет выявить образцы, отличающиеся друг от друга примененными в них неодинаковыми регулятивными стратегиями и спектрами используемых регулятивных средств (лексических, морфологических, синтаксических, композиционных, логических и др.).

Наибольшую активность в разработке теории и приложений регулятивности текстов проявляют лингвисты. Так, реализована весьма успешная попытка построения эмпирической теории регулятивности художественного текста [26]. В рамках этой теории описываются механизмы воздействия текста на читателя, а также формирования в сознании адресата смысла текста, адекватного замыслу автора. Подчеркнуто, что в художественных текстах регулятивные функции выполняют лексические и композиционные регулятивные средства, причем проявляются они не жестко и часто обусловлены образной направленностью текста. Авторы этой монографии отмечают фактор смысловой избыточности этого типа, заключающийся в многократной актуализации одного и того же смысла, представляемого с помощью

разных лексических средств, правда, с некоторыми вариациями. Заметим, эти факты, как и некоторые другие специфические черты художественных текстов, препятствуют разрешению проблем их формализации.

Регулятивность текстов достаточно часто является объектом рассмотрения психологов, юристов, социологов и педагогов [27–29]. К сожалению, работ в сфере информатики, базирующихся на использовании регулятивных свойств текстов, нами не отмечено.

Вместе с тем существует обширное множество текстов, формализация которых, с одной стороны, крайне актуальна, а с другой – они обладают характерным набором явно выраженных регулятивов, способствующих разработке методов их формализации. К таковым относятся тексты, описывающие технологии, различные распорядительные документы (приказы, инструкции, распоряжения и т.п.), в значительной мере это и учебные тексты, особенно математические. От текстов других типов их отличают логичность и точность (отсутствие иносказаний), связность и цельность, ясность, понятность (за счет отсутствия эмоциональной окраски и образных средств), доступность (за счет использования специфической терминологии). Они априори ориентированы на отображения регулятивных действий, интерпретация которых должна ориентировать читателей на адекватную организацию и воспроизведение описываемых в них процессов. Для приобретения таких качеств эти тексты должны содержать все необходимые для их интерпретации слова (при необходимости дополненные нужными уточнениями) и иметь схемы построения предложений, способствующие их быстрому и однозначному пониманию. Кроме того, в этих текстах широко используются специальные регулятивы (регулятивные структуры), такие как предписания (надо), запреты (нельзя), разрешения и рекомендации (можно) [30]. Тексты, структура построения и содержание которых явно ориентированы на регулирование поведения людей, будем называть регулятивными. Именно в силу своей специфики такие тексты в наибольшей мере подходят для разработки методов их записи на формальном языке, обеспечивая при этом достаточно высокий уровень сохранения содержания текста. Почему?

Основу лингвистической структуры регулятивных текстов образует упорядоченная последовательность побудительных предложений. В них каждое предложение (или логически агрегированная их группа) описывает конкретное обособленное (частичное) действие, приводящее к определенному результату,

которое можно трактовать как описание элементарной технологии [31]. В этой связи их часто называют императивными [32].

Как правило, такое действие подготовлено предшествующими действиями (по контексту понятно, что его следует начинать только после окончания (или начала) всех предшествующих ему частичных действий). После описания (читай выполнения) процедуры данного частичного действия соответствующая локальная цель предполагается достигнутой и переходят к описанию следующих частичных действий, реализация которых полагается возможной только в связи с выполнением определенного ряда предворяющих их частичных действий. Фактически идет описание алгоритма действия читателя, что является еще одним проявлением регулятивности такого рода текстов.

Процессная сущность регулятивных текстов обуславливает целесообразность более внимательного взгляда на синтаксис составляющих их предложений с позиций структурной лингвистики, в основу которой положена вербоцентрическая теория структурного синтаксиса [33]. Согласно этой теории, единственным главным членом предложения является глагол (глагольный предикат), при котором указаны актанты – участники описываемого действия, обстоятельства и условия его реализации (сирконстанты), а также отношения к этому действию (через модальность). Большинство предложений наиболее характерных представителей регулятивных текстов построены именно так: имеется предикативное ядро, обычно представляемое в виде императивных глагольных словоформ (включить, проверить, повернуть и т.п.), непосредственно участвующие в соответствующем действии объекты и ряд уточняющих обстоятельств и дополнений.

В качестве примера приведем фрагмент описания технологии пошива накладных карманов [34]:

- 1) на изнаночной стороне, на расстоянии 3 см от среза припуска на обработку верхнего края кармана наметить линию перегиба верхнего края кармана;
- 2) накладывая клеевую прокладку клеевой стороной на изнаночную сторону припуска, располагая край по намеченной линии, продублировать припуск на обработку верхнего края кармана;
- 3) вместе с прокладкой, выполняя строчку с лицевой стороны детали, обметать срез припуска на обработку верхнего края кармана.

Позиционирование в рамках структурного синтаксиса глагола как единственного главного члена предложения устраняет ряд трудностей, возникающих при попытках построения формальных моделей предложения. В частности, предложение регулятивного текста может быть представлено упорядоченной гроздью слов, где

формирующим центром выступает глагольный предикат [35]. К нему через «входящие» связи сходятся все участвующие в данном действии объекты с указанием дополнений и обстоятельств, а также «исходящие» из центра связи к результатам этого действия (см. рисунок 1). Это еще один результат лингвистики, который использован нами при решении поставленной задачи.

Примечание. В регулятивных текстах встречаются и так называемые описательные предложения. У них более простая, чем у предложений иного типа, структура. Обычно они представлены в форме предложения о состоянии, где в роли предиката выступают слова категории «состояния». При этом подразумевается наличие отношения, неявно выражаемое глаголами: быть, иметь, является. Например, «расстояние от кромки припуска до пуговицы 3 см». В этой связи их формализация при помощи излагаемого ниже метода реализуется значительно проще, чем в случае предложений иного типа, а именно: указанные в тексте значения вводятся в соответствующие места онтологии, относящейся к описываемому объекту.

Все вышеизложенное позволяет сформулировать суть предлагаемого метода формализации регулятивных текстов: обеспечение передачи предметной сущности используемых слов и механизмов причинно-следственных связей между ними, отображаемых соответствующим конструктивным алгоритмом.

Опираясь на специфический характер построения регулятивных текстов и ряд идей структурного синтаксиса, можно предположить нижеследующий метод формализованного представления такого рода текстов.

По существу, регулятивный текст можно рассматривать как вербальную запись алгоритма действий, направленных на обеспечение перехода некоего объекта из заданного исходного фазового состояния $S(X_0)$ в определенное конечное состояние $S(Y_f)$. Задача состоит в том, чтобы представить данный алгоритм в формализованном виде:

$$W : S(X_0) \rightarrow S(Y_f)$$

Этот переход осуществляется в результате реализации упорядоченной последовательности ряда конкретных действий, рассматриваемых в данном тексте как элементарные. В этой связи алгоритм построения формальной модели регулятивного текста предлагается представлять в виде конструктивного процесса упорядочивания моделей отдельных действий. Тем самым решение поставленной задачи распадается на два этапа: построение моделей

отдельных действий и конструирование из них соответствующего исходному тексту семантического графа.

4. Модель предложения регулятивного текста. Как правило, в регулятивных текстах каждое действие описывается одним предложением (условно назовем его содержательным), которое может дополняться и уточняться. Структурно такие части текста часто имеют нумерацию или выделены в виде абзацев.

Первый этап метода формализации регулятивного текста представляет собой автоматизированную процедуру построения модели одного содержательного предложения (описывающего одно элементарное действие как одноактное действие [36]). Для ее реализации по мере рассмотрения (чтения) исходного текста каждому очередному содержательному предложению приписывается свой номер q ($q \in [1, n]$). Пусть описываемое в этом предложении действие d_q по переводу объекта рассмотрения из фазового состояния $S(X_q)$ в $S(Y_q)$ описывает оператор W_q

$$W_q : S(X_q) \rightarrow S(Y_q). \quad (1)$$

Требуется сформировать формализованное представление компонентов, входящих в (1).

В рассматриваемом предложении выявляется описываемое действие (d_q) и фиксируется обозначающий его глагол (глагольный предикат). Для него формируется соответствующий фрейм-экземпляр с именем $FD(q)$. В этот фрейм, имя которого соответствует выявленному действию, заносятся имена участвующих в данном действии акторов и сведения об описывающих его сирконстантах. По существу, он в конструктивной форме описывает реализацию операций, осуществляемых оператором W_q .

Примечание. Если при анализе данного предложения выявлены отношения с акторами и обстоятельства его реализации, не зафиксированные на предыдущих шагах, то они вводятся и в формируемый протофрейм данного глагола, где нарастающим итогом формируется структурный облик последнего.

Выявленные акторы данного действия разбиваются на два множества: X_q и Y_q . К первому, как показано на рисунке 1, относятся объекты $x_{iq} \in X_q$, участвующие в осуществлении этого действия (назовем их входными), ко второму – результирующие объекты $y_{qj} \in X_q$ (назовем их выходными).

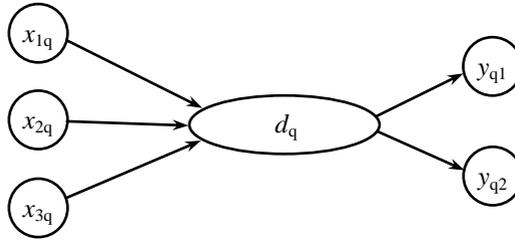


Рис. 1. Структура модели предложения регулятивного текста

Ряд авторов полагает, что текст семантически эквивалентен данному отображению, если совпадают их предикатные выражения [37]. Несомненно, в целом последовательность предикатных выражений обеспечивает возможность уяснить суть данного текста. Однако в регулятивных текстах весьма важны представленные в них описания условий, в которых описываемые действия протекают. В этой связи, помимо фиксации в формализованной форме сведений о собственно действиях, требуется отображать все сведения об участвующих в этих действиях объектах, причем по принципу «здесь и сейчас», и вот почему.

При изложении одних и тех же мыслей разные авторы используют разные слова, применяют синонимию, омонимию, иноказания, другие возможности живого языка. Эти обстоятельства обуславливают значительные трудности при формализации текстов. Сложностей добавляет использование знаков препинания и различных структур построения предложений, способных изменить их семантику. Отметим, что регулятивные тексты, как правило, используются в конкретных предметных областях, где уделяется внимание соответствующей устоявшейся лексике. Однако в излагаемом подходе предлагается решать вопросы совпадения смыслов рассматриваемых слов не столько по написанию имен сущностей, сколько по использованию интенционального метода построения совокупности формализующих их фреймов в виде родовидовой структуры [38].

Формируемая онтология понятий представляется в виде кортежа

$$Q = \langle C, M, R \rangle,$$

где $C = \{c_j\}$ – множество понятий, образующих онтологию Q , $i = \overline{1, I}$, то есть $|C| = I$;

$M_i = \{m_{d_1}, \dots, m_{d_i}\}$ – множество атрибутов понятия c_i , d_i – количество атрибутов, описывающих данное понятие;

$R \subseteq C \times C$ – отношение непосредственного наследования. Отношение R может быть задано матрицей размером $I \times I$. Если понятие c_k непосредственно наследует понятию c_i , то есть $(c_j, c_k) \in R$, то вводится элемент матрицы наследования r_{ik} , который принимает значение единица ($r_{ik}=1$), в противном случае, если $(c_i, c_k) \notin R$, то $r_{ik}=0$.

Онтология в общем случае поддерживает множественное наследование понятий, то есть:

$$\exists c_i \exists c_k \exists c_l : ((r_{il}=1) \wedge (r_{kl}=1)).$$

Отсюда нахождение двух понятий в отношении непосредственного наследования означает, что между ними нет ни одного промежуточного понятия:

$$\exists c_i \exists c_k : ((c_i, c_k) \in R \Rightarrow \bar{\exists} c_h : ((c_i, c_h) \in R) \wedge ((c_h, c_k) \in R)).$$

Дочерние понятия наследуют все атрибуты, входящие в состав родительского понятия, расширяя его состав собственными атрибутами:

$$\forall c_i \forall c_k : (r_{jk}=1) \Rightarrow (M_i \subset M_j).$$

Если множества атрибутов двух понятий совпадают, то понятие c_i эквивалентно понятию c_j , то есть

$$M_i = M_j \Rightarrow c_i \sim c_j.$$

Данный метод автоматизированного построения онтологии предметной области позволяет, в частности, автоматически определить отношение, в котором находится данная пара понятий.

Такой подход в значительной мере позволяет снизить актуальность предварительной нормализации исходного текста.

В этой связи описываемое в данном предложении состояние каждого i -го входного актора x_{iq} фиксируется в соответствующем фрейме-экземпляре этого объекта $FX(x_{iq}) \equiv FX_{iq}$, а для каждого j -го выходного актора y_{qj} во фрейме-экземпляре $FY(x_{qj}) \equiv FY_{qj}$. Подобные

фреймы строятся для всех входных акторов данного действия, что обеспечивает формирование множества $\{FX_{iq}\}$, описывающего состояние рассматриваемого объекта $S(X_q)$ к началу действия d_q . Аналогичным образом формируется множество $\{FY_{iq}\}$, описывающее состояние этого объекта на момент окончания действия $S(Y_q)$.

В рассматриваемом тексте данный объект, находясь в разных состояниях и отношениях с другими объектами и действиями, может упоминаться несколько раз. Следовательно, содержание фреймов-экземпляров для одного и того же объекта, упомянутого в разных предложениях, обычно не совпадает. Для обеспечения же возможностей по использованию модели текста, сведения о каждом объекте данной предметной области целесообразно накапливать, что осуществляется в соответствующих протофреймах.

Таким образом, моделью каждого содержательного (описывающего отдельное действие регулятивного текста) предложения выступает именованная тройка

$$M_q = \langle FD(q), \{FX_{iq}\}, \{FY_{iq}\} \rangle, \quad (2)$$

где $FD(q)$ – фрейм-экземпляр, описывающий концепт-действие данного предложения d_q ; $\{FX_{iq}\}$ – множество фреймов-экземпляров, описывающих участвующих в данном действии акторов $X_q = \{x_{iq}\}$; $\{FY_{iq}\}$ – множество фреймов, описывающих результирующих акторов $Y_q = \{y_{iq}\}$. Применение регулятивного текста предполагает, что читателю известно исходное состояние описываемых в рассматриваемом тексте объектов $X_0 = (x_{01}, x_{02}, \dots, x_{0n})$, либо дается их описание. Такого рода сведения фиксируются в виде соответствующих фреймов-экземпляров $\{FX_0\}$.

Примечание. Основным отличительным моментом модели (2) является то, что ее ядром является выступающий в роли оператора W_q фрейм, описывающий отображение отдельного элементарного перехода рассматриваемого объекта из одной точки его фазового состояния $S(X_q)$ в другую $S(Y_q)$. При этом, поскольку каждое содержательное действие исходного текста рассматривается как элементарное, построенная описанным способом модель адекватна содержанию, передаваемому соответствующим предложением исходного текста (фиксируется только то, что непосредственно представлено в тексте, ибо задача заключается в сохранении смысла именно данного текста).

Все сформированные таким образом модели вида (2) образуют множество вершин описывающего текст семантического графа $M = \{M_q\}$.

5. Формализация поиска связей между действиями. Целью следующего этапа на пути построения математической модели регулятивного текста является установление на множестве $\{M_q\}$ формальных связей, отражающих описанные в тексте отношения между действиями. Ключевым моментом для достижения этой цели является следующее обстоятельство: в регулятивном тексте выходные объекты каждого действия (кроме так называемых заключительных, то есть $\{Y_f\}$) выступают в качестве входных для других, последующих действий. Тем самым, что крайне важно подчеркнуть, эти связи неявно уже присутствуют во множествах $\{FX_{iq}\}$ и $\{FY_{qj}\}$. Задача заключается в том, чтобы выявить связи непосредственного следования и зафиксировать их в явном виде. С этой целью на данных множествах необходимо сформировать систему переходов Келлера, иными словами, построить бинарную матрицу, отображающую искомые отношения в явном виде.

Непосредственное предшествование действия d_k ($k \in [1, n]$) действию d_q означает, что среди выходных объектов действия d_k существует аналогичный входному объекту действия d_q . Если исходный текст с точки зрения описываемых процессов корректен, такого рода отношения должны находиться во множестве $\{M_q\}$. Иначе говоря, для непосредственно следующих друг за другом действий d_k и d_q во множестве $\{M_q\}$ найдется пара фреймов $FY_{ki}, FX_{qj} \in \{M_q\}$, описывающих один и тот же объект, причем находящийся в одинаковых фазовых состояниях. Методы автоматического сравнения фреймов представлены в работе [39]. Найденная на этой основе пара фреймов (результат одного действия и актора другого действия) находится в отношении эквивалентности $FY_{ki} \leftrightarrow FX_{qj}$ (рисунок 2).

Примечание. Для удобства выполнения процедуры проверки условия эквивалентности фазовых состояний данного объекта структуры всех его фрейм-экземпляров должны быть одинаковыми. Носителями этой структуры и призваны быть соответствующие фрейм-прототипы упомянутых в тексте объектов.

Построенное множество фреймовых моделей типа (2) позволяет формальным образом (следовательно, автоматически) сформировать бинарную квадратную матрицу порядка $n \times n$ непосредственно предшествующих действий $GD = \parallel g_{k,q} \parallel$.

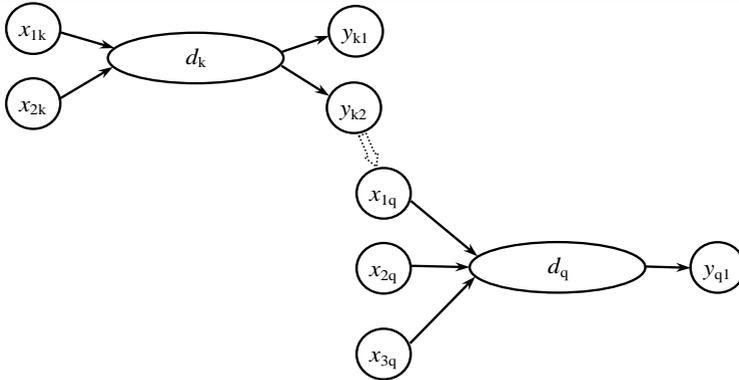


Рис. 2. Установление связей непосредственного следования

Для этих целей необходимо, рассматривая все возможные пары действий, установить те из них, которые имеют эквивалентные пары акторов и зафиксировать эти факты:

$$\forall q, k, \exists(ki, qj) : FY_{ki} \leftrightarrow FX_{qj} \rightarrow g_{kq} = 1. \quad (3)$$

Поскольку каждое действие может иметь несколько выходных $\{y_{ki}\}$ и входных объектов $\{x_{qj}\}$, требуется зафиксировать именно те из них, для которых выполнено условие (3). Эту фиксацию удобно осуществлять непосредственно в модифицированной матрице GDM , в которую вместо символа «1» вводить, например, вектор, описывающий пары номеров фреймов, описывающих эквивалентные понятия:

$$\forall q, k, \exists(ki, qj) : FY_{ki} \leftrightarrow FX_{qj} \rightarrow g_{kq} := \{(k; i), (q; j)\}.$$

Построенную модель рассматриваемого текста

$$T = \langle M, GDM \rangle,$$

где $GDM = \|(k, j), (q, j)\|$, можно интерпретировать как ориентированный граф (семантическую сеть). Вершинами построенного графа выступают фреймы, описывающие отдельные действия, и участвующие в них акторы. В этом графе присутствуют три вида дуг: дуги, связывающие акторов с действием, дуги, связывающие действие с его результатом и дуги, связывающие результаты и акторы, являющиеся эквивалентными.

Построенный граф обладает одной особенностью, он состоит из ярко выраженных подграфов, каждый из которых отображает отдельные содержательные предложения исходного текста (рис. 2).

В качестве примера рассмотрим следующий фрагмент текста из учебника по математике, где приводится описание решения линейных матричных уравнений. Пусть задана система линейных алгебраических уравнений в матричной форме $A \cdot X = B$. Если матрица коэффициентов системы A обратима, то, используя свойство ассоциативности умножения матриц и свойства единичной матрицы, можно показать, что решение имеет вид $X = A^{-1}B$. Для этого необходимо умножить левую и правую части уравнения слева на обратную матрицу A^{-1} , а затем применить свойства единичной матрицы: $A^{-1} \cdot A = E$ и $E \cdot A = A$.

Описанный выше подход формализованного представления такого рода текстов позволяет формализовать это описание следующим образом.

Пусть в ходе формализации предыдущих параграфов данного учебника были зафиксированы входящие в условия задачи свойства матриц, представленные на рисунке 3 в виде последовательности стилизованных равенств.

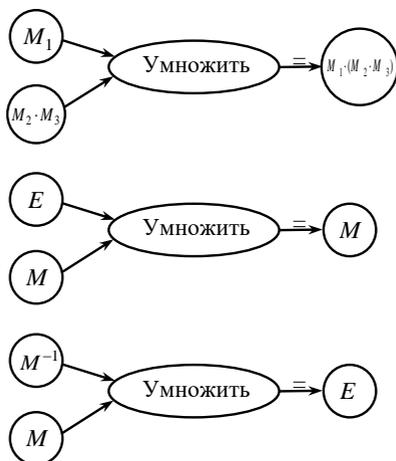


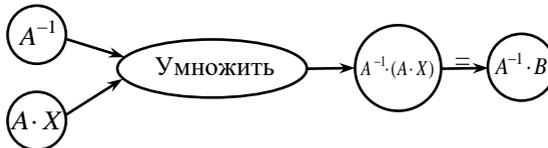
Рис. 3. Структурное представление использованных свойств

Сформируем фреймы акторов A , X , B и представим исходное уравнение $A \cdot X = B$ в виде следующей структуры (рис. 4).



Рис.4. Структура исходного уравнения

Отобразим факт умножения левой и правой части исходного уравнения слева на обратную матрицу A^{-1} в виде структуры, представленной на рисунке 5.

Рис.5. Структура уравнения после его умножения слева на A^{-1}

Расположенный в левой части уравнения результат есть уже представленная по условию структура, описывающая свойство ассоциативности умножения матриц $[M_1 \cdot (M_2 \cdot M_3) \equiv (M_1 \cdot M_2) \cdot M_3]$. Найдя совпадение соответствующих структур, алгоритм, как показано на рисунке 6, меняет левую часть уравнения.

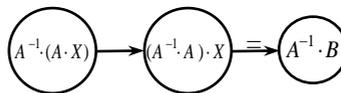


Рис. 6. Использование свойства ассоциативности умножения матриц

Поскольку искомый результат не получен, алгоритм продолжает поиск структур, аналогичных тем, что стоят теперь в левой и правой части уравнения. Структура левой части может активизировать имеющееся в условии основное свойство (часто рассматриваемое как определение) единичной матрицы $M^{-1} \cdot M = E$. Его применение приводит к новому виду уравнения, представленного на рисунке 7.

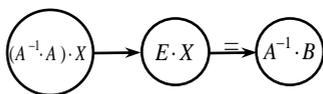


Рис. 7. Использование определения единичной матрицы

Полученная в левой части уравнения структура аналогична той, что отвечает другому свойству единичной матрицы ($E \cdot M = M$) – свойству умножения единичной матрицы на матрицу. Активация этого свойства, как показано на рисунке 8, изменяет левую часть уравнения, приводя к виду, полностью аналогичному искомому результату.

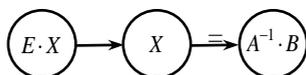


Рис. 8. Использование свойства умножения единичной матрицы на матрицу

Объединение всех описанных шагов позволяет получить формализованную структуру, которая отражает содержание изложенного в тексте фрагмента учебника.

Примечание. Анализ изложения описанного фрагмента свидетельствует, что в разных учебниках при решении описанной задачи не акцентируется внимание на том, какие конкретно свойства матриц используются. Чаще всего опускается явное указание на необходимость использования свойства ассоциативности умножения матриц. Авторы текстов считают это необязательным, надеясь на память читателя. Это обстоятельство приводит к невозможности построения формального представления данного текста изложенным методом, что свидетельствует об изъянах рассматриваемого текста.

Следует понимать, что на практике рассматриваемые тексты всегда конечны. Следовательно, они содержат слова, которые именуют понятия, предполагаемые априорно общеизвестными и в достаточной мере одинаково трактуемые читателями. Иначе говоря, любой конкретный текст формально не является замкнутым относительно использованных в них понятий.

В этой связи, во-первых, всегда следует рассматривать определенную предметную область с достаточно устоявшейся лексикой и не требующей разъяснений. Во-вторых, каждый регулятивный текст предполагает, что читатель имеет определенный уровень подготовки

(образования). Таким образом, для использования изложенного метода исходное состояние описываемых в рассматриваемом тексте объектов читателю должно быть известно и зафиксировано в виде упомянутых выше соответствующих фреймов-экземпляров $\{FX_0\}$.

6. Заключение. Регулятивный текст по существу представляет собой словесно выраженный алгоритм последовательности целенаправленных действий, что открывает возможность построения математической модели, адекватно отображающей семантику этого текста.

Процедура построения такой модели осуществляется в два этапа. Вначале в автоматизированном режиме производится формирование и наполнение слотов фреймов, адекватно отображающих описанные в тексте действия, и участвующих в них акторов. При этом применяется интенсиональный способ построения совокупности формализующих их фреймов в виде родо-видовой структуры, что в значительной мере позволяет снизить актуальность предварительной нормализации исходного текста. Затем в автоматическом режиме формируется соответствующий тексту семантический граф.

Важным преимуществом представленного метода является возможность существенного сокращения количества актов принятия творческих решений при построении формальной модели рассматриваемого текста. По существу, субъект, работающий над формализацией текста, выполняет в основном рутинные действия: выделяя из текста отдельные слова (устоявшиеся фразы) он разносит их по соответствующим фреймам, если необходимо, вводя в них при этом новые пары слотов. Тем самым результат построения формализованного представления исходного текста становится слабо зависящим от личностных характеристик исполнителя.

Рассмотренный подход обеспечивает возможность разработки программных комплексов для построения баз знаний отдельных предметных областей, предприятий и организаций, активное использование которых обеспечит возможность эффективного решения расширенного списка задач. Поясним сказанное на следующем примере.

На ряде производств своевременное обновление ассортимента выпускаемой продукции является одним из главных залогов их успешной деятельности. Так, например, владимирская обувная фабрика в год обновляет до 80 наименований своей продукции. При этом полный цикл смены одного наименования (включая маркетинговые исследования, работу конструкторов и технологов, отладку собственно производства, элементов снабжения и сбыта) занимает до полутора лет.

В этой связи формирование системы, обеспечивающей поддержку процессов обновления ассортимента – весьма актуальная трудоемкая задача.

Для каждой выпускаемой модели технология ее производства на данном предприятии, описание используемых материалов и оборудования зафиксировано в соответствующих информационных базах в виде текстов. Поскольку применяемые технологические приемы, свойства материалов и подготовка новых моделей к выпуску занимают существенно различное время, а спектр возможных решений весьма широк, предприятие заинтересовано в создании автоматизированных средств поддержки процессов разработки технологий производства новых моделей, которые в сложившихся условиях были бы технологически и ресурсно возможны и отвечали бы интересам предприятия. По существу, речь идет о создании современного инструментария решения крайне сложной задачи синтеза производственных процессов в условиях функционирования конкретно данного предприятия.

Эффективное решение этой задачи можно обеспечить при возможности целенаправленно, в автоматическом режиме оперировать представленными в формализованном виде описаниями всех возможных технологий изготовления продукции. Подготовку этих формализмов с учетом актуальных для предприятия ограничений и допущений обеспечивает изложенный метод.

Соответствующие алгоритмы синтеза новых технологий удобно формировать в автоматическом режиме, комбинируя имеющиеся фрагменты существующих технологий. Для этого целесообразно наличие специальным образом сформированных библиотек, компоненты которых позволяет формировать изложенный подход.

В итоге рассмотренный подход обеспечивает возможность разработки программных комплексов для построения библиотек регулятивных текстов отдельных предметных областей, разработки средств оценки такого рода текстов на предмет их определенности, полноты, связности и других характеристик, а также тренажеров и средств, используемых при самообучении.

Литература

1. *Пешкова Н.П.* Семантика и смысл текста: >, <, =, #? (экспериментальный подход к теоретическим проблемам // Вестник Челябинского государственного университета. 2015. № 15 (370). Филология. Искусствоведение. Вып. 96. С. 69–77.

2. *Аранова Г.У.* Концепт, понятие и значение слова // Международный журнал прикладных и фундаментальных исследований. 2016. № 1. С. 591–593.
3. *Орехов В.Д.* Прогностика: от прошлого к будущему человечества // М., Сканбук, Е-Books. 2016. 210 с.
4. *Баринова И.А.* К вопросу о соотношении смысла и содержания текста при переводе // Вестник Вятского государственного университета. 2015. № 10. С. 69–72.
5. *Lochter Johannes V., Pires Pedro R., Bossolani Carlos, Yamakami Akebo, Almeida Tiago A.* Evaluating the impact of corpora used to train distributed text representation models for noisy and short texts // International Joint Conference on Neural Networks (IJCNN). 2018. pp. 315–322.
6. *Двойникова А.А., Карпов А.А.* Аналитический обзор подходов к распознаванию тональности русскоязычных текстовых данных // Информационно-управляющие системы. 2020. № 4(107). С. 20–30.
7. *Ali Ismael, Al Madi Naser, Melton Austin.* Using Text Comprehension Model for Learning Concepts, Context, and Topic of Web Content // IEEE 11th International Conference on Semantic Computing (ICSC). 2017. pp. 101–104.
8. *Li Jie, Chen Siming, Chen Wei, Andrienko Gennady, Andrienko Natalia.* Semantics-Space-Time Cube: A Conceptual Framework for Systematic Analysis of Texts in Space and Time // IEEE Transactions on Visualization and Computer Graphics. 2020. vol. 26. iss. 4. pp. 1789–1806.
9. *Sheridan Tom.* Models of Human-System Interaction [Book Review] // IEEE Systems, Man, and Cybernetics Magazine. 2017. vol. 3. iss. 2. pp. 56–С3.
10. *Митрофанова О.И., Романенко О.Н.* Работа с научным текстом на занятиях по русскому языку с иностранными военнослужащими // Научный альманах. 2016. № 8–1(22). С. 473–478.
11. *Levinkov Evgeny, Uhrig Jonas, Tang Siyu, Omran Mohamed, Insafuldinov Eldar, Kirillov Alexander etc.* Joint Graph Decomposition & Node Labeling: Problem, Algorithms, Applications // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1904–1912.
12. *Xu LiHong, Sun ShuTao, Wang Qi.* Text similarity algorithm based on semantic vector space model // IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). 2016. pp. 1193–1196.
13. *Бондарчук Д.В.* Алгоритмы интеллектуального поиска на основе метода категориальных векторов: дис. ... к-та физ.-мат. наук. Екатеринбург: УрГУПС. 2016. 141 с.
14. *Gaspers Judith, Cimiano Philipp, Rohlfing Katharina, Wrede Britta.* Constructing a Language From Scratch: Combining Bottom-Up and Top-Down Learning Processes in a Computational Model of Language Acquisition // IEEE Transactions on Cognitive and Developmental Systems. 2017. vol. 9. iss. 2. pp. 183–196.
15. Когнитивные аспекты лексикографии. Список публикаций В. З. Демьянкова. Статьи и монографии. URL: www.infolex.ru/Dem_Publ.html (дата обращения: 25.07.2020).
16. *Демьянков В.З.* Трансфер знания и перевод // Язык. Культура. Перевод. Коммуникация: сб. научн. тр. М.: КДУ, Университетская книга. 2018. Вып. 2. С. 18–21.
17. *Pavlovskiy I.S.* The semantic models of large terminological texts // Tenth International Conference Management of Large-Scale System Development (MLSD). 2017. pp. 1–5.
18. *Kang Sungku, Patil Lalit, Rangarajan Arvind, Moitra Abha, Jia Tao, Robinson Dean, Ameri Farhad, Dutta Debasish.* Extraction of Formal Manufacturing Rules from Unstructured English Text // Computer-Aided Design. 2021. vol. 130: 102990.

19. *Fernández-Duque David, Nepomuceno-Fernández Ángel, Sarrión-Morrillo Enrique, Soler-Toscan Fernando, Velázquez-Quesada Fernando R.* Forgetting complex propositions // *Logic Journal of the IGPL*. 2015. vol. 23. iss. 6. pp. 942–965.
20. *Sauerland Uli, Yatsushiro Kazuko.* Remind-Me Presuppositions and Speech-Act Decomposition: Evidence from Particles in Questions // *Linguistic Inquiry*. 2017. vol. 48. iss. 4. pp. 651–678.
21. *Кулешов С. В., Зайцева А.А., Марков В.С.* Ассоциативно-онтологический подход к обработке текстов на естественном языке // *Интеллектуальные технологии на транспорте*. 2015. № 4. С. 40–45.
22. *Зацман И.М.* Семiotическая модель взаимосвязей концептов, информационных объектов и компьютерных кодов // *Информатика и ее применение*. 2009. Том 3. №2. С. 65–81.
23. *Марева С.Н.* Диалектика содержания и формы и проблемы формализации // *Научные ведомости Белгородского государственного университета. Серия «Философия. Социологи. Право»*. 2017. № 42 (24). С. 15–26.
24. *Walter Benedikt, Hammes Jakob, Piechotta Marco, Rudolph Stephan.* A Formalization Method to Process Structured Natural Language to Logic Expressions to Detect Redundant Specification and Test Statements // *IEEE 25th International Requirements Engineering Conference (RE)*. 2017. pp. 263–272.
25. *Болотнова Н.С.* О типологии регулятивных структур в тексте как форме коммуникации // *Вестник Томского государственного педагогического университета*. 2011. Вып. 3 (105). С. 34–40.
26. *Болотнова Н.С., Бабенко И.И., Бакланова Е.А.* и др. Коммуникативная стилистика текста: лексическая регулятивность в текстовой деятельности / Под ред. Н.С. Болотновой // *Томск: ТГПУ*. 2011. 492 с.
27. *Троцук И.В.* Анализ текстовых данных в социологии: основания систематизации концептуальных моделей, методологических принципов и методических решений: дис. д-ра социол. наук. Москва: РУДН. 2014. 338 с.
28. *Мясникова О.В., Фроловская М.Н.* Педагогические условия понимания текстов в процессе обучения // *Известия АлтГУ*. 2014. Вып. 2 (82). Том 2. С. 43–48.
29. *Дьякова А.А.* Регулятивность текстов законов, адаптированных к условиям медийного дискурса // *Грани познания*. 2015. №1. С. 95–100.
30. *Кармин А.С.* Культурология. Краткий курс // СПб.: Питер. 2009. 240 с.
31. *Krupp Lars, Gruenerbl Agnes, Bahle Gernot, Lukowicz Paul.* Towards Automatic Semantic Models by Extraction of Relevant Information from Online Text // *IEEE International Conference on Smart Computing (SMARTCOMP)*. 2019. pp. 481–483.
32. *Zimmer Melanie, Al-Yacoub Ali, Ferreira Pedro, Lohse Niels.* Understanding Human Decision-making during Production Ramp-up using Natural Language Processing // *IEEE 17th International Conference on Industrial Informatics (INDIN)*. 2017. pp. 337–342.
33. *Zakharov Victor, Krassovitskiy Alexander, Meirambekkyzy Zhazirakhanyim, Ualiyeva Irina, Khoroshilov Aleksander, Khoroshilov Alexey.* Automatic Creation Technologies of Declarative Tools for Clustering Media Documents // *International Conference on Engineering Technologies and Computer Science (EnT)*. 2019. pp. 39–42.
34. *Рослякова Т.А.* Школа шитья // *Ростов-на-Дону: Феникс*. 2000. 445 с.
35. *Sithole Vusi, Marshal Linda.* Building Concept Hierarchies for the Internet of Things Patterns Using Domain-specific Dependency Knowledge // *Open Innovations (OI)*. 2019. pp. 1–7.

36. *Воронов М.В.* Модель технологического действия // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1. Естественные и технические науки. 2016. № 1. С. 47–52.
37. *Богатырев М.Ю., Коржук Н.Л.* Исследование семантической инвариантности концептуальных моделей текстовых данных // Известия ТулГУ. Технические науки. 2019. Вып. 12. С. 499–503.
38. *Антонов И.В., Воронов М.В.* Методы анализа данных в задачах автоматизации построения онтологии предметной области // Дистанционное и виртуальное обучение. 2011. № 8 (50). С. 19–35.
39. *Воронов М.В.* Система активной поддержки самоподготовки студентов // Образовательные технологии. 2018. № 3. С. 107–117.

Воронов Михаил Владимирович – д-р техн. наук, профессор, заведующий кафедрой, кафедра прикладной математики, Московский государственный психолого-педагогический университет. Область научных интересов: теория систем и системный анализ, управление организационными системами. Число научных публикаций – 400. mi.voronov@yandex.ru, mgppu.ru/project/156; Открытое ш., д. 24, стр. 27, г. Москва, 107143, РФ; р.т.: +7(499)1676674.

Пименов Виктор Игоревич – д-р техн. наук, старший научный сотрудник, заведующий кафедрой, кафедра информационных технологий, Санкт-Петербургский государственный университет промышленных технологий и дизайна. Область научных интересов: распознавание образов, интеллектуальный анализ данных, когнитивные технологии. Число научных публикаций – 200. v_pim@mail.ru, sutd.ru/universitet/staff/person/0_0_2398; Вознесенский пр., д. 46, СПбГУПТД, каф. ИТ, каб. 475, г. Санкт-Петербург, 190068, РФ; р.т.: +7(812)3122567.

M. VORONOV, V. PIMENOV
FORMALIZATION OF REGULATORY TEXTS

Voronov M.V., Pimenov V.I. Formalization of Regulatory Texts.

Abstract. Modern information technologies provide text manipulation processes with high efficiency. First of all, this means storing, editing, and formatting texts and their components. Having achieved significant success in developing tools for content-free computer text processing, researchers faced problems with their content processing. Therefore, further steps in this direction are associated with the creation, among other things, of methods for automated purposeful manipulation of texts, taking into account their content. The analysis of works devoted to the study of the problems of formal presentation of texts and their subsequent use is carried out. Despite a number of successful projects, the challenges of solving the problem of the relationship between the content of the text and its meaning remain relevant. It seems that formalization of a General-purpose text while preserving its semantics is not feasible at this stage. However, there are types of texts that can be formalized while preserving their semantics. One of them is a regulatory text type, which is essentially a verbally expressed algorithm for a sequence of targeted actions. It is distinguished by logic and accuracy (lack of allegories), coherence and integrity, clarity, understandability (due to the lack of emotional coloring and figurative means), accessibility (due to the use of specific terminology). In other words, when developing regulatory texts, they usually try to display the mechanisms of the described actions as clearly as possible. Purpose: development of a method for formalizing a regulatory text while preserving its semantics. Methods: structural linguistics, representation of objects in the form of an ontology, constructive algorithms. The use of this method is demonstrated by describing the solution of a system of algebraic equations. Results: method for constructing a mathematical model of a regulatory text. Practical relevance: the application of the developed method makes it possible to develop software systems for building libraries of individual subject areas, develop tools for evaluating regulatory texts for their certainty, completeness, connectivity and other characteristics, as well as simulators and self-learning tools.

Keywords: Regulatory Text, Artificial Intelligence, Mathematical Model, Constructive Algorithm, Ontology, Frame

Voronov Mihail – Dr. Sci., Tech., professor, head of the Department of applied mathematics, Moscow State University of Psychology & Education. Research interests: systems theory and system analysis, management of organizational systems. The number of publications – 400. mivoronov@yandex.ru, mgppu.ru/project/156; 24, Otkrytoye Hwy., Moscow, 107143, Russia; office phone +7(499)1676674.

Pimenov Viktor – Dr. Sci., Tech., senior researcher, head of the Department of information technologies, Saint-Petersburg State University of Industrial Technologies and Design. Research interests: image recognition, data mining, cognitive technologies. The number of publications – 200. v_pim@mail.ru, sutd.ru/universitet/staff/person/0_0_2398; 46, B. Morskaya St., St. Petersburg, 190068, Russia; office phone +7(812) 3122567.

References

1. Peshkova N.P. [Text semantics and text sense: >, <, =, #? (experimental approach to theoretical problems)]. *Vestnik Cheljabinskogo gosudarstvennogo universiteta – Bulletin of Chelyabinsk State University*. 2015. no. 15 (370). Philology. Arts. iss. 96. pp. 69–77. (In Russ.).

2. Arapova G.U. [Concept, comprehension and meaning of word]. *Mezhdunarodnyj zhurnal prikladnyh i fundamental'nyh issledovanij – International Journal of Applied and Fundamental Research*. 2016. no. 1. pp. 591–593. (In Russ.).
3. Orehov V.D. *Prognostika: ot proshlogo k budushchemu chelovechestva* [Prognostics: from the past to the future of humanity]. Moscow, Skanbuk, E-Books, 2016. 210 p. (In Russ.).
4. Barinova I.A. [On relation between sense and content in text translating]. *Vestnik Vyatskogo gosudarstvennogo gumanitarnogo universiteta – Bulletin of Vyatka state University*. 2015. no. 10. pp. 69–72. (In Russ.).
5. Lochter Johannes V., Pires Pedro R., Bossolani Carlos, Yamakami Akebo, Almeida Tiago A. Evaluating the impact of corpora used to train distributed text representation models for noisy and short texts. International Joint Conference on Neural Networks (IJCNN). 2018. pp. 315–322.
6. Dvojnjkova A.A., Karpov F.F. [Analytical review of approaches to Russian text sentiment recognition]. *Informatsionno-upravliaiushchie sistemy – Information and Control Systems*. 2020. no. 4. pp. 20–30. (In Russ.).
7. Ali Ismael, Al Madi Naser, Melton Austin. Using Text Comprehension Model for Learning Concepts, Context, and Topic of Web Content. IEEE 11th International Conference on Semantic Computing (ICSC). 2017. pp. 101–104.
8. Li Jie, Chen Siming, Chen Wei, Andrienko Gennady, Andrienko Natalia. Semantics-Space-Time Cube: A Conceptual Framework for Systematic Analysis of Texts in Space and Time. *IEEE Transactions on Visualization and Computer Graphics*. 2020. vol. 26. iss. 4. pp. 1789–1806.
9. *Sheridan Tom*. Models of Human-System Interaction [Book Review]. *IEEE Systems, Man, and Cybernetics Magazine*. 2017. vol. 3. iss. 2. pp. 56–C3.
10. Mitrofanova O.I., Romanenko O.N. [Working with academic text in the Russian language classes with the foreign soldiers]. *Nauchnyj al'manah – Science Almanac*. 2016. no. 8–1(22). pp. 473–478. (In Russ.).
11. Levinkov Evgeny, Uhrig Jonas, Tang Siyu, Omran Mohamed, Insaftudinov Eldar, Kirillov Alexander etc. Joint Graph Decomposition & Node Labeling: Problem, Algorithms, Applications. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1904–1912.
12. Xu LiHong, Sun ShuTao, Wang Qi. Text similarity algorithm based on semantic vector space model. IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). 2016. pp. 1193–1196.
13. Bondarchuk D.V. *Algoritmy intellektual'nogo poiska na osnove metoda kategorial'nyh vektorov. Dis. kand. fiz.-mat. nauk* [Intelligent search algorithms based on the categorical vector method. PhD. phys. and math. sci. diss.]. Yekaterinburg, UrGUPS Publ. 2016. 141 p. (In Russ.).
14. Gaspers Judith, Cimiano Philipp, Rohlfing Katharina, Wrede Britta. Constructing a Language from Scratch: Combining Bottom–Up and Top–Down Learning Processes in a Computational Model of Language Acquisition. *IEEE Transactions on Cognitive and Developmental Systems*. 2017. vol. 9. iss. 2. pp. 183–196.
15. Kognitivnye aspekty leksikografii. Spisok publikacij V. Z. Dem'yankova. Stat'i i monografii [Cognitive aspects of lexicography. List of publications by V. Z. Demyankov. Articles and monographs]. Available at: www.infolex.ru/Dem_Publ.html (accessed 25.07.2020). (In Russ.).
16. Dem'jankov V.Z. [Knowledge transfer and translation]. *Yazyk. Kul'tura. Perevod. Kommunikaciya: Sb. nauchn. tr.* [Language. Culture. Translation. Communication: Collected papers]. Moscow: KDU, Universitetskaya kniga Publ, 2018. iss. 2. pp. 18–21. (In Russ.).

17. Pavlovskiy I.S. The semantic models of large terminological texts. Tenth International Conference Management of Large-Scale System Development (MLSD). 2017. pp. 1–5.
18. Kang Sungku, Patil Lalit, Rangarajan Arvind, Moitra Abha, Jia Tao, Robinson Dean, Ameri Farhad, Dutta Debasish. Extraction of Formal Manufacturing Rules from Unstructured English Text. *Computer-Aided Design*. 2021. vol. 130: 102990.
19. Fernández–Duque David, Nepomuceno–Fernández Ángel, Sarrión–Morrillo Enrique, Soler–Toscan Fernando, Velázquez–Quesada Fernando R. Forgetting complex propositions. *Logic Journal of the IGPL*. 2015. vol. 23. iss. 6. pp. 942–965.
20. Sauerland Uli, Yatsushiro Kazuko. Remind-Me Presuppositions and Speech-Act Decomposition: Evidence from Particles in Questions. *Linguistic Inquiry*. 2017. vol. 48. iss. 4. pp. 651–678.
21. Kuleshov C.V., Zajceva A.A., Markov V.C. [Associative-ontological approach to natural language texts processing]. *Intellektual'nye tehnologii na transporte – Intellectual Technologies on Transport*. 2015 no. 4. pp. 40–45. (In Russ.).
22. Zacman I.M. [Semiotic model of interrelations of concepts, information objects and computer codes]. *Informatika i ee primeneniye – Informatics and its application*. 2009. vol. 3. no. 2. pp. 65–81. (In Russ.).
23. Mareev S.N. [The dialectic of content and form, and the problem of formalization]. *Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta. Seriya «Filosofiya. Sociologi. Pravo» – Belgorod State University Scientific bulletin. Series' philosophy. Sociologists. Right*. 2017. no. 42 (24). pp. 15–26. (In Russ.).
24. Walter Benedikt, Hammes Jakob, Piechotta Marco, Rudolph Stephan. A Formalization Method to Process Structured Natural Language to Logic Expressions to Detect Redundant Specification and Test Statements. IEEE 25th International Requirements Engineering Conference (RE). 2017. pp. 263–272.
25. Bolotnova N.S. [Typology of regulative structures in text as a form of communication]. O tipologii reguljativnyh struktur v tekste kak forme kommunikacii. *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta – Tomsk state pedagogical university bulletin*. 2011. iss. 3 (105). pp. 34–40. (In Russ.).
26. Bolotnova N.S., Babenko I.I., Baklanova E.A. i dr. *Kommunikativnaja stilistika teksta: leksicheskaja reguljativnost' v tekstovoj dejatel'nosti* [Communicative stylistics of the text: lexical regularity in text activity] Tomsk: TSPU, 2011. 492 p. (In Russ.).
27. Trocuk I.V. *Analiz tekstovoyh dannyh v sociologii: osnovaniya sistematizacii konceptual'nyh modelej, metodologicheskikh principov i metodicheskikh reshenij. Dis. dokt. sociol. nauk* [Text data analysis in sociology: the basis for systematization of conceptual models, methodological principles and methodological solutions]. Moscow: RUDN, 2014. 338 p. (In Russ.).
28. Mjasnikova O.V., Frolovskaja M.N. [Pedagogical Conditions of Text Understanding in the Process of Studying]. *Izvestiya of Altai State University*. 2014. iss. 2 (82). vol. 2. pp. 43–48. (In Russ.).
29. D'jakova A.A. [Regulativeness of law texts adapted to media discourse conditions]. *Grani poznaniya – Facets of knowledge*. 2015. №1. pp. 95–100. (In Russ.).
30. Karmin A.S. *Kul'turologija. Kratkij kurs* [Cultural studies. Short course]. Saint Petersburg: Piter, 2009. 240 p. (In Russ.).
31. Krupp Lars, Gruenerbl Agnes, Bahle Gernot, Lukowicz Paul. Towards Automatic Semantic Models by Extraction of Relevant Information from Online Text. IEEE International Conference on Smart Computing (SMARTCOMP). 2019. pp. 481–483.
32. Zimmer Melanie, Al-Yacoub Ali, Ferreira Pedro, Lohse Niels. Understanding Human Decision-making during Production Ramp-up using Natural Language Processing. IEEE 17th International Conference on Industrial Informatics (INDIN). 2017. pp. 337–342.

33. Zakharov Victor, Krassovitskiy Alexander, Meirambekkyzy Zhazirakhany, Ualiyeva Irina, Khoroshilov Aleksander, Khoroshilov Alexey. Automatic Creation Technologies of Declarative Tools for Clustering Media Documents. International Conference on Engineering Technologies and Computer Science (EnT). 2019. pp. 39–42.
34. Rosljakova T.A. *Shkola shit'ja* [School of sewing]. Rostov-on-Don: Feniks, 2000. 445 p. (In Russ.).
35. Sithole Vusi, Marshal Linda. Building Concept Hierarchies for the Internet of Things Patterns Using Domain-specific Dependency Knowledge. Open Innovations (OI). 2019. pp. 1–7.
36. Voronov M.V. [Modeling of technological processes]. *Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta tekhnologii i dizajna. Seriya 1. Estestvennye i tekhnicheskie nauki – Bulletin of the Saint Petersburg state University of technology and design. Series 1. Natural and technical Sciences*. 2016. no. 1. pp. 47–52. (In Russ.).
37. Bogatyrev M.Y., Korzhuk N.L. [Study of semantic invariance of conceptual models of text data]. *Izvestiya TulGU. Tekhnicheskie nauki – News of TulSU. Technical science*. 2019. iss. 12. pp. 499–503. (In Russ.).
38. Antonov I.V., Voronov M.V. [Data analysis methods for automating the construction of domain ontology]. *Distancionnoe i virtual'noe obuchenie*. 2011. no. 8 (50). pp. 19–35. (In Russ.).
39. Voronov M.V. [Active support system for students self-training]. *Obrazovatel'nye tekhnologii – Educational technology*. 2018. no. 3. pp. 107–117. (In Russ.).

А.М. КОРСАКОВ, А.В. БАХШИЕВ, Л.А. АСТАПОВА, Л.А. СТАНКЕВИЧ
**РЕАЛИЗАЦИЯ ПОВЕДЕНЧЕСКИХ ФУНКЦИЙ НА
СПАЙКОВЫХ НЕЙРОННЫХ СЕТЯХ**

Корсаков А.М., Бахшиев А.В., Астапова Л.А., Станкевич Л.А. **Реализация поведенческих функций на спайковых нейронных сетях.**

Аннотация. Рассматривается моделирование поведенческих функций животных, в частности, условного рефлекса. Проводится анализ современного состояния нейронных сетей с возможностью структурного реконфигурирования. В основу модели рефлекса положена нейронная сеть, базирующаяся на сегментной спайковой модели нейрона с возможностью структурной адаптации дендритного дерева к входному паттерну импульсов. Модель нейрона позволяет описывать и изменять онлайн структуру (размер тела клетки, количество и длина дендритов, количество синапсов) в зависимости от поступающего на её входы паттерна импульсов. Приведено краткое описание сегментной спайковой модели нейрона, отмечены её основные особенности с точки зрения возможности её структурного реконфигурирования. Описывается способ структурной адаптации модели к входному паттерну импульсов. Для исследования работы предложенной модели нейрона в сети, в качестве примера обосновывается выбор условного рефлекса, как частного случая механизма формирования ассоциативных связей. Приведено описание структурной схемы и алгоритма формирования условного рефлекса как с положительным, так и с отрицательным подкреплением. Представлено пошаговое описание экспериментов по формированию ассоциативных связей и условного рефлекса. Сделан вывод о перспективности применения спайковых сегментных моделей нейронов для повышения биоправдоподобности реализации поведенческих функций в нейроморфных системах управления. Рассмотрены дальнейшие перспективные направления развития нейроморфных систем, основанных на спайковых сегментных моделях нейрона.

Ключевые слова: нейроморфные системы, спайковый нейрон, спайковые нейронные сети, условный рефлекс, ассоциативные связи, управление поведением

1. Введение. В настоящее время активно разрабатываются системы управления динамическими объектами, по своей структуре, функциям и способам обработки информации подобных нервной системе живых организмов. Это связано с тем, что при решении многих трудноформализуемых задач классические вычислительные методы оказываются неэффективными. Так, при распознавании изображений и речевых сообщений часто невозможно создать и использовать точные математические модели и соответствующие методы решения. В то же время мозг достаточно эффективно решает такие задачи за счет массивного параллелизма своих нейронных сетей и способности к их структурной и параметрической настройке. Поэтому мозг считается непревзойденным инструментом восприятия и анализа больших потоков информации и принятия решений. Однако принципы обработки сигналов и их передачи между нейронными образованиями

мозга и нервной системы в целом остаются малоизученными, поскольку их структура чрезвычайно сложна и с трудом поддается анализу существующими методами.

Современные искусственные нейронные сети в определенной степени облегчают решение трудноформализуемых задач [1]. Однако использование в них формальных нейронов с функциями взвешенного суммирования входных сигналов и нелинейной обработки выходного сигнала лишь в самых общих чертах соответствуют своим биологическим прототипам, что существенно ограничивает их возможности по сравнению с мозгом и нервной системой.

Нейроморфный подход, связанный с созданием математических моделей компонентов нервной системы и их аппаратных реализаций, в которых обработка сигналов адекватна биологическим прототипам, открывает новые возможности решения трудноформализуемых задач в технических системах. Предполагается, что разработка на основе нейроморфного подхода новых компонентов и архитектур глубоких нейронных сетей и специальных алгоритмов их обучения позволит создавать универсальные системы, эффективно решающие множество трудноформализуемых задач в пространствах высокой размерности и в условиях существенных неопределенностей [2]. Такой подход также перспективен в плане раскрытия механизмов переработки информации в мозге и нервной системе, что дает возможность обосновать использование выявляемых принципов структурной организации нейронов и выполняемых ими функций в сети при построении нейроморфных систем.

В общем случае под нейроморфными системами понимаются искусственные нейронные сети, архитектура и дизайн которых основаны на особенностях структуры и принципах работы реальных нейробиологических систем. Их моделирование стимулировано желанием понять и технически воплотить такие ключевые особенности нейронных структур мозга, как высокая чувствительность, адаптивность, обучаемость, устойчивость к повреждениям, способность иметь дело с нечеткой, избыточной, зашумленной информацией и, наконец, параллельный и распределенный способ обработки информации [3-5].

В настоящее время нейроморфные системы в основном создаются на спайковых (импульсных) нейронах. Спайковый нейрон – это искусственный нейрон, оперирующий с непрерывно поступающим многомерным потоком спайков как с потоком точечных событий. Простейшей моделью спайкового нейрона является модель порогового интегратора, IF (Integrate-and-Fire) [6], описываемая

дифференциальным уравнением первого порядка. В общем случае каждый входной спайк приводит к отклику нейрона как динамической системы на импульсное входное воздействие. Эти отклики называются постсинаптическими потенциалами и меняют потенциал мембраны нейрона во времени. Как и формальные модели нейронов, спайковый нейрон интегрирует несколько входных сигналов с некоторыми весами и способен генерировать выходные спайки, идущие на входные синапсы других нейронов на единственном выходном канале. Однако в отличие от бинарных и частотных моделей, в спайковом нейроне ведется интегрирование не только пространственной, но и временной информации.

Нейроморфные системы, построенные на базе простых спайковых нейронов, настраиваются на задачи (адаптируются), в основном, путем параметрической настройки нейронов и сетей, то есть изменением весов связей и порогов нейронов [7]. Однако известно, что в нервной системе осуществляется не только параметрическая, но и структурная адаптация нейронов и сетей под решаемую задачу. Это вдохновило исследователей на создание спайковых динамически развивающихся нейронных сетей eSNN (Evolving Spiking Neural Network), способных работать в реальном времени. Для большинства спайковых нейронных сетей создана традиционная стратегия развития, которая изменяет их структуру путем регулирования количества нейронов в процессе обучения, а также в динамическом режиме адаптирует сеть под задачу, настраивая веса нейронов в реальном времени.

Например, в ряде исследований [8-12] для каждого нового вектора входных данных, который требуется распознавать, динамически выделяется новый выходной нейрон, который подключается к входным нейронам. Инициализация, обновление весов и установление связей нейронов, чтобы настроить сеть на распознавание входного вектора, осуществляется с использованием правила RO (Rank Order). В некоторых реализациях нейроны с одинаковыми весовыми векторами объединяются на основе евклидова расстояния между ними. Это позволяет достичь очень быстрого обучения (может быть достаточно только одного прохода) как при обучении с учителем, так и без учителя [8]. Тем не менее, подход требует решения ряда проблем, таких как автоматическая настройка нескольких параметров обучения и повышение масштабируемости обучения для больших наборов данных.

Подход, описанный в [11], расширяет модель eSNN и предлагает новую динамическую модель спайковой нейронной сети deSNN [13].

Предлагаемый метод основан на таких механизмах обучения, как RO и обучение с помощью временных спайков – правила синаптической пластичности, управляемой спайками SDSP (Spike Driven Synaptic Plasticity). На этих же механизмах основан метод, предложенный в [14]. Нейрон LIF (Leaky Integrate-and-Fire) учится распознавать весь пространственно-временной паттерн, используя только одну итерацию в режиме онлайн-обучения. В [14] авторы отмечают возможность реализации нейрона и сети на нейроморфном оборудовании. Однако все еще нерешенным остается вопрос оптимизации многочисленных параметров нейронной сети.

В работе [15] представлен метод инкрементного обучения спайковой нейронной сети, структура скрытых слоев которой изменяется динамически с использованием алгоритма кластеризации (обучения без учителя). Однако входной и выходной слой классификации динамическими не являются и обучаются с учителем. Теми же авторами алгоритм модифицирован до алгоритма обучения спайковой нейронной сети с учителем SpikeComp [16], входной слой которой не является динамическим, а выходной слой состоит из динамически добавляемых нейронов, причем имеет место динамическое обновление весов или центров кластеров. При этом алгоритм SpikeComp, основанный на точном времени спайка, добавляет гораздо меньше выходных нейронов, чем метод eSNN, основанный на ранжировании. Кроме того, в [17] авторы представили усовершенствование метода обучения на основе ранжирования [12, 13], называемое SpikeTemp. Для преобразования входных данных используются гауссовские рецептивные поля и методы кодирования популяции с квадратным косинусом.

Архитектура спайковой нейронной сети с динамически изменяемыми синапсами представлена в [18]. Показано, что такая сеть способна выполнять классификацию образов с использованием временного кодирования. Используется обучение с учителем, основанное на методе эволюционной стратегии, ES (evolutionary strategy) [19], который используется для поиска оптимальных параметров, лежащих в основе процесса обучения без вычисления производных. Сеть поддерживает топологию прямой связи и может использоваться в качестве системы обнаружения признаков.

Динамическое формирование кластеров в популяциях импульсных нейронов, а также пластичность, зависящая от времени спайка – STDP (Spike-Timing Dependent Plasticity) использовано в [20]. Структура спайковой нейронной сети постоянно корректируется путем добавления и сокращения популяций нейронов. Кроме того, количество

нейронов в каждой популяции постоянно изменяется по мере поступления новых данных. Алгоритм STDP используется для регулировки силы синаптических связей.

В [21] предложена саморегулирующаяся развивающаяся нейронная сеть SRESN (Self-Regulating Evolving Spiking Neural), в которой используется специальная схема кодирования популяции IF нейронов. Во время обучения саморегулирующийся механизм выбирает подходящую стратегию обучения (добавление нейронов, обновление параметров или пропуск выборки) на основе знаний, хранящихся в сети, и новых знаний, присутствующих в обучающей выборке. Веса нового нейрона инициализируются с использованием модифицированной схемы обучения в порядке ранжирования.

Применение модели нейрона с учетом описания структуры дендритного дерева предложено в [22], однако модель не рассматривает сигналы в сети как спайковые. Подход к созданию пластичных нейронных сетей на основе эволюционного подхода рассматривается в [23]. Такой подход, несомненно, имеет ряд преимуществ при решении задач поиска оптимальной структуры нейронной сети на этапе проектирования.

Анализ приведенных примеров, демонстрирующих спайковые нейронные сети и методы их структурного и параметрического обучения, позволяет сделать вывод о том, что для существенного увеличения адаптивной способности таких сетей необходимо рассматривать не только параметрическую, но и структурную адаптацию нейронных сетей, как на уровне изменения топологии связей между узлами (нейронами) сети, так и на уровне усложнения модели каждого узла, например, путем перехода к сегментным (компаратментным) моделям нейронов. По мнению авторов, это даст возможность более эффективной реализации поведенческих функций в нейроморфных системах управления, например, в робототехнике.

В качестве альтернативы традиционным стратегиям развития, регулирующим количество нейронов в процессе обучения и адаптирующим веса нейронов в динамическом режиме, в настоящей работе предлагается подход, предлагающий, помимо возможности динамического изменения структуры нейронной сети и связей между отдельными нейронами, возможность структурной онлайн перестройки самой модели нейрона.

В настоящей работе рассматривается возможность реализации поведенческих функций в нейроморфных системах, построенных на спайковых нейронных сетях с сегментными нейронами. Для этого предлагается специальный алгоритм структурной адаптации нейронов

ко входным паттернам импульсов, а также алгоритм формирования ассоциативных связей нейронной сети, для настройки ее на реализацию определенных поведенческих функций. Работа алгоритмов демонстрируется на примере реализации простых поведенческих функций типа безусловного и условного рефлекса.

Далее в разделе 2 приведена постановка задачи, определены основные термины, представлена используемая при моделировании структурная схема модели нейрона, приведены её важнейшие свойства с точки зрения возможностей структурного реконфигурирования дендритного и синаптического аппаратов. Раздел 3 описывает принцип структурной адаптации спайковой сегментной модели нейрона, решается задача адаптации модели нейрона к произвольному паттерну импульсов. В разделе 4 представлены структурные схемы организации нейронной сети для моделирования ассоциативных связей и, как частный случай, для моделирования условного рефлекса (как с положительным, так и с отрицательным подкреплением). Также в разделе 4 приведён алгоритм для моделирования процесса формирования ассоциативных связей. В разделе 5 представлено пошаговое описание экспериментов по формированию ассоциативных связей и условного рефлекса. В заключительном разделе 6 приведено обсуждение результатов работы, отмечены основные недостатки предлагаемых моделей, а также намечены перспективные пути развития данного направления работ.

2. Постановка задачи. Цель данной работы – показать возможность реализации поведенческих функций на сети из сегментных спайковых нейронов. Для этого требуется решить задачу разработки алгоритмов структурной адаптации нейрона к входному паттерну импульсов и формирования ассоциативных связей нейронной сети, отображающей заданную поведенческую функцию.

В основу настоящей работы положена разработанная авторами ранее модель спайкового нейрона, подробно описанная в [24]. В данной модели предполагается, что на входы поступают импульсные потоки, которые преобразуются в синапсах в аналоговые величины, описывающие динамические процессы выделения и распада медиатора в синаптической щели. Входные и выходные сигналы нейрона равны нулю в отсутствии импульса и константе на время действия импульса.

Структурная схема рассматриваемой модели показана на рисунке 1.

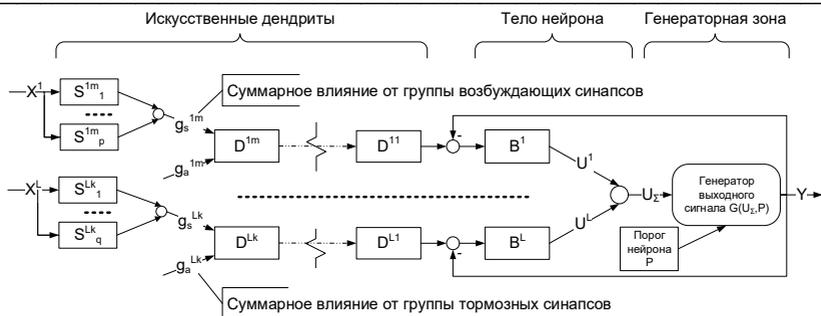


Рис. 1. Структурная схема модели сегментного спайкового нейрона для решения задачи распознавания паттерна импульсов

Здесь B^1-B^L – участки мембраны, описывающие тело нейрона. D^{ij} – участки мембраны, описывающие i -й дендрит. S^{ij_l} – модели синаптической передачи. Их суммарное влияние оказывает либо возбуждающее $g_s^{ij_l}$, либо тормозное $g_a^{ij_l}$ воздействие на соответствующий участок мембраны.

Такая структура модели соответствует решению задачи структурного обучения нейрона паттерну импульсов, представляемому входным вектором X . Отметим особенности предложенной модели нейрона, играющие существенную роль при решении задачи разработки алгоритмов структурного обучения модели нейрона:

- возможность создания произвольного числа сегментов тела нейрона (участков мембраны с обратной связью), что означает возможность изменять размерность входного паттерна импульсов;
- возможность независимого наращивания длин дендритов (участков мембраны без обратной связи); увеличение длины дендрита приводит к временной задержке сигнала и к уменьшению его амплитуды;
- возможность добавления произвольного числа синапсов на произвольном участке мембраны (теле, дендрите); добавление синапса приводит к увеличению амплитуды сигнала;
- возможность произвольной организации связей (как возбуждающих, так и тормозных) между отдельными элементами.

Всё вышеперечисленное означает возможность структурного реконфигурирования модели нейрона, что, в свою очередь, позволяет изменять время прихода и амплитуду сигнала с каждого из дендритов на сому нейрона.

Под паттерном импульсов будем понимать совокупность сигналов, поступающих на входы модели нейрона (дендриты) за некоторый (выбранный) промежуток времени. Отдельный входной сигнал представляет собой одиночный импульс единичной амплитуды. Соответственно, отдельные сигналы в паттерне имеют одинаковую (единичную) амплитуду, но, в общем случае, различное время поступления на соответствующий вход (дендрит) модели нейрона. Предполагается, что за выбранный промежуток времени поступления паттерна на отдельном входе (дендрите) может быть зафиксировано не более одного входного импульса.

Выходной сигнал генерируется в низкопороговой зоне модели нейрона (генератор потенциала действия) и представляет собой одиночный импульс, либо серию импульсов единичной амплитуды, формируемых при превышении величиной потенциала некоторого порога. Само же значение потенциала внутри низкопороговой зоны нелинейно связано с суммой сигналов, пришедших на вход нейрона.

Таким образом, для преодоления порога и генерации потенциала действия необходимо максимизировать сумму вкладов в потенциал внутри низкопороговой зоны от всех дендритов, что может быть достигнуто синхронизацией действия входных сигналов в нейроне.

Как будет показано далее, это возможно достичь путём добавления отдельных сегментов дендритов и добавлением синапсов на конечный сегмент дендрита.

Процедуру наращивания длины дендрита будем называть синхронизацией, а процедуру добавления синапсов – нормализацией. Процессы синхронизации и нормализации для конкретного паттерна будем называть обучением модели нейрона данному паттерну импульсов. Генерацию потенциала действия при фиксированном паттерне импульсов и соответствующей ему конфигурации обученной модели нейрона будем называть распознаванием данного паттерна импульсов.

Общий подход к реализации поведенческой функции может быть сведён к минимизации отличия целевой поведенческой реакции сети

f_{beh} на заданный паттерн импульсов от реальной реакции f_{beh} , то есть:

$$\left\{ \begin{array}{l} f_{beh} = F(\bar{x}, t) \\ f_{beh} = \hat{F}(\bar{x}, n_1, \dots, n_M, c_{i,j}, t_1, \dots, t_T), \\ f_{beh} - f_{beh} \rightarrow \min, \end{array} \right. \quad (1)$$

где \bar{x} – вектор входного воздействия; F – поведенческий функционал, связывающий входы и выходы биологического прототипа во времени; \hat{F} – нейросетевое отображение поведенческой функции; n_k – число сегментов в k -м нейроне ($k = 1, \dots, M$); $c_{i,j}$ – связи между i -м и j -м нейронами; t_k – моменты времени прихода сигналов на вход k -го сегмента.

В процессе такой минимизации производится структурная настройка нейронной сети для установления нужной последовательности выходных импульсов.

3. Структурная адаптация модели нейрона. Как было указано выше, обучение модели нейрона заданному паттерну импульсов сводится к решению двух подзадач – синхронизации и нормализации.

Задача синхронизации сигналов заключается в максимальном временном сопоставлении экстремумов кривых потенциалов всех дендритов при суммировании их вкладов внутри низкопороговой зоны нейрона.

Способом решения данной задачи будет минимизация модуля суммы разностей между временем фиксации экстремума кривой потенциала базового дендрита и временами фиксации экстремумов кривых потенциалов остальных дендритов (рис. 2), что достигается путём добавления отдельных сегментов на дендриты нейрона (наращивание дендритов). Под базовым дендритом будем понимать такой дендрит, в котором начальный экстремум кривой потенциала фиксируется с наибольшим запаздыванием. Следует отметить, что базовый дендрит не наращивается.

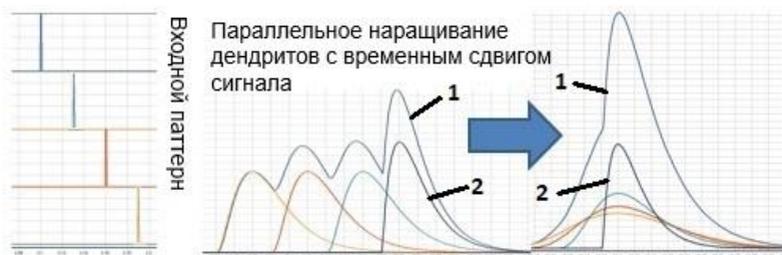


Рис. 2. Параллельное наращивание дендритов с временным сдвигом сигналов

На рисунке 2 цифрой «1» обозначен суммарный вклад дендритов в потенциал внутри низкопороговой зоны, цифрой «2» – потенциал базового дендрита. Остальные кривые описывают потенциал наращиваемых дендритов.

Следующим шагом является решение задачи нормализации сигналов путём добавления синапсов. Задача нормализации заключается в компенсации потери амплитуды потенциала на отдельном дендрите при наращивании его длины и решается после задачи синхронизации, таким образом, наращивание длины дендритов уже произошло, а экстремумы кривых потенциалов отдельных дендритов синхронизированы (рис. 2).

Поскольку две эти задачи (синхронизация и нормализация) связаны, для решения задачи нормализации необходимо на первой итерации алгоритма синхронизации (до того, как произошло наращивание дендритов) запомнить эталонные значения экстремумов кривых потенциалов отдельных дендритов.

Имея эталонные значения экстремумов кривых потенциалов отдельных дендритов, для каждого дендрита будем добавлять синапс на последний сегмент этого дендрита до тех пор, пока максимум потенциала данного дендрита не превысит эталонного значения.

После превышения эталонного значения требуется удалить один синапс с последнего сегмента данного дендрита. Таким образом, задача нормализации сигнала на данном дендрите может быть переформулирована следующим образом: требуется найти максимальное количество синапсов на последнем сегменте каждого дендрита, но так, чтобы максимальное значение потенциала на данном дендрите не превысило эталонного (рис. 3).



Рис. 3. Параллельное увеличение числа синапсов на дендритах

На рисунке 3 цифрой «1» обозначен суммарный вклад дендритов в потенциал внутри низкопороговой зоны, остальные кривые описывают потенциалы отдельных дендритов.

Следует отметить, что данная процедура может совершаться параллельно (одновременно) для всех дендритов.

Все значения максимальной суммарной амплитуды потенциалов дендритов, при которых происходит генерация выходного импульса,

независимо от числа дендритов в модели нейрона, укладываются в один и тот же коридор значений. Это позволяет подобрать универсальный порог срабатывания модели нейрона на паттерн импульсов, которому она ранее обучалась.

4. Структурная схема формирования ассоциативных связей.

Для комплексной проверки свойств модели спайкового нейрона с возможностью структурной адаптации и взаимодействия таких нейронов в сети требовалось подобрать пример организации естественной нейронной сети для его воспроизведения посредством описанной выше модели нейрона. Требованиями к такой естественной сети являлись, с одной стороны, достаточная её исследованность в рамках нейрофизиологии, а с другой - простота в реализации. В качестве такого примера была выбрана схема формирования условного рефлекса по Э.А. Асратяну [25] (рис. 4).

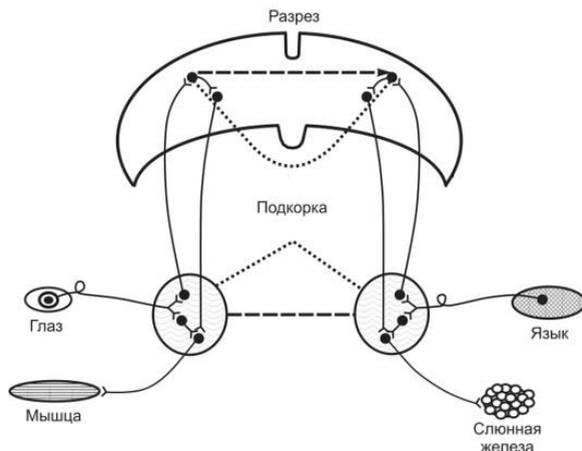


Рис. 4. Схема нейронных путей формирования слюноотделительного рефлекса на действие света (по Э.А. Асратяну)

Согласно этой модели, условный раздражитель, например звонок, вызывает безусловно-рефлекторную ориентировочную реакцию. Звонок, являясь индифферентным раздражителем для пищевого рефлекса, пока не вызывает отделения слюны. В то же время безусловный раздражитель – пища – вызывает появление слюноотделения. Таким образом, в коре (и стволе мозга) одновременно возникают очаги возбуждения, и если их одновременное возбуждение повторяется многократно, то между этими центрами происходит формирование пути проведения нервных импульсов, обеспечивающее

замыкание условно-рефлекторной временной связи. После установления такой связи условный звуковой сигнал звонка приобретает для животного новое значение – сигнала или информации о предстоящем кормлении. С этого момента для вызова безусловного пищевого рефлекса, проявляющегося слюноотделением, двигательным беспокойством, другими вегетативными реакциями, становится достаточно подачи только условного раздражителя – включения звонка [26].

В биологии рефлексы с положительным подкреплением – это такие рефлексы, при выработке которых применяется мотивационно благоприятное или желаемое подкрепление. Главным свойством положительного подкрепления является то, что оно увеличивает частоту соответствующей реакции или побуждает животное приближаться к источнику подкрепления. Те же рефлексы, в которых применяется мотивационно неблагоприятное подкрепление (то есть подкрепление, которого хочется избежать), например удары током, называются рефлексами с отрицательным подкреплением. Оно, напротив, снижает частоту соответствующей реакции или заставляет удаляться от источника подкрепления и избегать ситуации в целом, если этот источник с ней неразрывно связан.

В рамках настоящего исследования на моделях сегментного спайкового нейрона следовало воспроизвести условный рефлекс и формирование ассоциативных связей.

Описанная выше модель нейрона с возможностью структурной адаптации позволяет нам предложить схему организации нейронной сети, которая воспроизводит условный рефлекс (рис. 5).

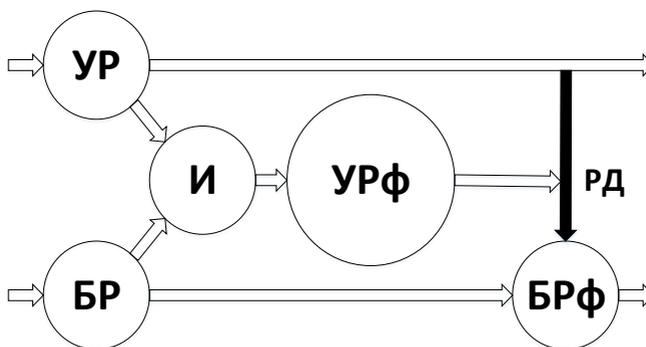


Рис. 5. Структурная схема нейронной сети, воспроизводящей условный рефлекс

На рисунке 5 «УР» – это нейрон с возможностью структурной адаптации, отвечающий условному раздражителю («условный раздражитель»); «БР» – нейрон с возможностью структурной адаптации, отвечающий безусловному раздражителю («безусловный раздражитель»); «И» – нейрон, реализующий функцию «И» (нейрон «И»); «УРф» – нейрон, формирующий условный рефлекс («большой нейрон»); «БРф» – нейрон, реализующий функцию «ИЛИ» и отвечающий безусловному рефлексу (нейрон «ИЛИ»).

На момент начала работы системы из всех связей, изображённых на рисунке 5, создаётся лишь связь «БР – БРф», что постулирует изначальное существование безусловного рефлекса.

При старте системы происходит обучение моделей нейронов «УР» и «БР» входным, произвольно выбранным, паттернам импульсов (способом, описанным в пункте 3). По завершении процесса обучения формируются связи «УР – И» и «БР – И», что соответствует моменту начала формирования рефлекторной дуги «РД».

Нейрон «И», состоящий из двух сегментов сомы, обучается сигналам, поступающим от нейронов «УР» и «БР». Обучение возможно лишь до определённого, заданного параметром, предела. Если рассинхронизация входных сигналов больше заданного этим параметром порога, воздействие условного и безусловного раздражителей не считается одновременным и дальнейшее формирование рефлекторной дуги не происходит. В противном случае, по окончании обучения нейрон «И» начинает генерировать выходные импульсные сигналы и создаёт связь «И – УРф».

Нейрон «УРф» представляет из себя «большой» нейрон, то есть количество сегментов его сомы больше единицы. Конкретное значение размера «большого» нейрона выбирается пользователем и фактически означает скорость, с которой система сформирует условный рефлекс при одновременном воздействии условного и безусловного раздражителей. Нейрон «И» формирует связь лишь с одним из сегментов сомы «большого» нейрона, что является недостаточным для генерации моделью нейрона «УРф» выходного импульсного сигнала. Однако при поступлении на вход нейрона «УРф» импульсов от нейрона «И», происходит добавление синапсов на входной участок сомы, что повышает значение потенциала внутри низкопороговой зоны «большого» нейрона. Когда порог генерации потенциала действия в низкопороговой зоне нейрона оказывается превышенным, происходит генерация выходного импульса моделью нейрона «УРф», что приводит к формированию связи «УР – БРф». Таким образом рефлекторная дуга замыкается.

Особенностью модели нейрона «БРф», состоящего лишь из одного сегмента сомы, является наличие на этом сегменте двух синапсов, что позволяет данной модели реализовывать функцию «ИЛИ». К одному из этих синапсов, как указывалось выше, произведено подключение связи от нейрона «БР» ещё при инициализации. Ко второму подключается связь от нейрона «УР» при начале генерации выходных импульсов нейроном «УРф». При формировании этой связи и при наличии возбуждения на входе нейрона «УР», модель нейрона «БРф» оказывается способной генерировать выходные импульсы даже при отсутствии активности со стороны нейрона «БР».

Другими словами, для вызова безусловной реакции становится достаточно подачи только условного раздражителя, что и означает завершение формирования условного рефлекса.

Следует отметить, что все описанные ранее сформированные связи между моделями нейронов подключались к возбуждающим моделям синапсов. Кроме того, данная схема реализует условный рефлекс с положительным подкреплением. Но для реализации условного рефлекса с отрицательным подкреплением нам достаточно на последнем шаге алгоритма подключить связь от нейрона «УР» не к возбуждающему, а к тормозному синапсу на соме нейрона «БРф».

Как будет показано в эксперименте, в результате такого подключения нейрон «БРф» не будет генерировать выходных импульсов при активности нейрона «УР» даже в том случае, если поступает входной сигнал от нейрона «БР».

Отметим также, что входные паттерны импульсов для условного и безусловного раздражителей в ходе описанных ниже экспериментов выбирались произвольно, выходные же импульсы просто фиксировались без какой-либо их дальнейшей обработки. Однако следует понимать, что со стороны входов на каком-либо этапе обработки должны располагаться модели афферентных нейронов, преобразующие внешние воздействия (зрительные, слуховые и т.п.) в импульсные потоки. Со стороны же выходов системы на каком-то этапе должны располагаться модели эфферентных нейронов, преобразующих импульсные потоки, например, в мышечные сокращения. Вариант такого взаимодействия на примере модели «сетчатка – афферентный нейрон – глазная мышца» был описан авторами в [27].

Рассмотрим теперь схему формирования ассоциативных связей в общем случае (рис. 6). В качестве его биологической интерпретации можно предложить формирование взаимных ассоциаций на различные раздражители при их продолжительном одновременном воздействии на субъект.

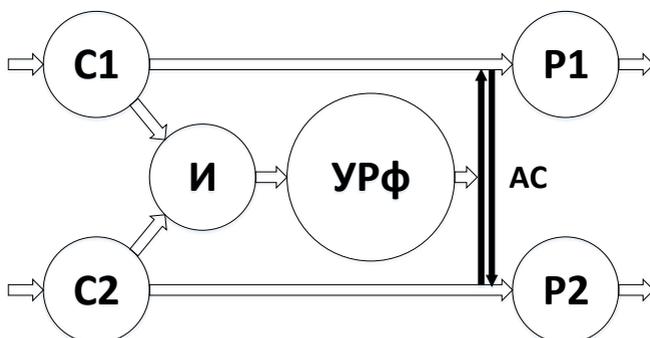


Рис. 6. Обобщённая структурная схема нейронной сети, формирующей ассоциативные связи

На рисунке 6 «С1» – это нейрон, на который поступает паттерн импульсов, который можно условно считать первым стимулом («стимул №1»). Нейрон «P1» имеет лишь один сегмент сомы, на котором, однако, расположено две модели синапса. Связь «С1 – P1» образуется при инициализации системы. Таким образом, модель нейрона «P1» реализует безусловную реакцию на стимул «С1».

Аналогично, нейроны «С2» и «P2» реализуют безусловную реакцию «P2» на стимул «С2» («стимул №2»). Связь «С2 – P2» также создаётся при инициализации системы.

Нейроны «И» и «УРФ» моделируют образование рефлекторной дуги способом, полностью совпадающим с описанным выше для условного рефлекса, за одним исключением: при начале генерации сигналов нейроном «УРФ» образуется сразу две ассоциативных связи («АС») – «С1 – P2» и «С2 – P1». Таким образом, нейроны «P1» и «P2» приобретают способность генерировать выходные сигналы (импульсы) при активности лишь одного входа: «С1» либо «С2».

Формирование ассоциативных связей реализовано в соответствии с алгоритмом на рисунке 7. Алгоритм формирования условного рефлекса выглядит аналогичным образом с заменой условных сокращений за исключением того, что на последнем шаге формируются не две, а одна связь «УР – БРф».

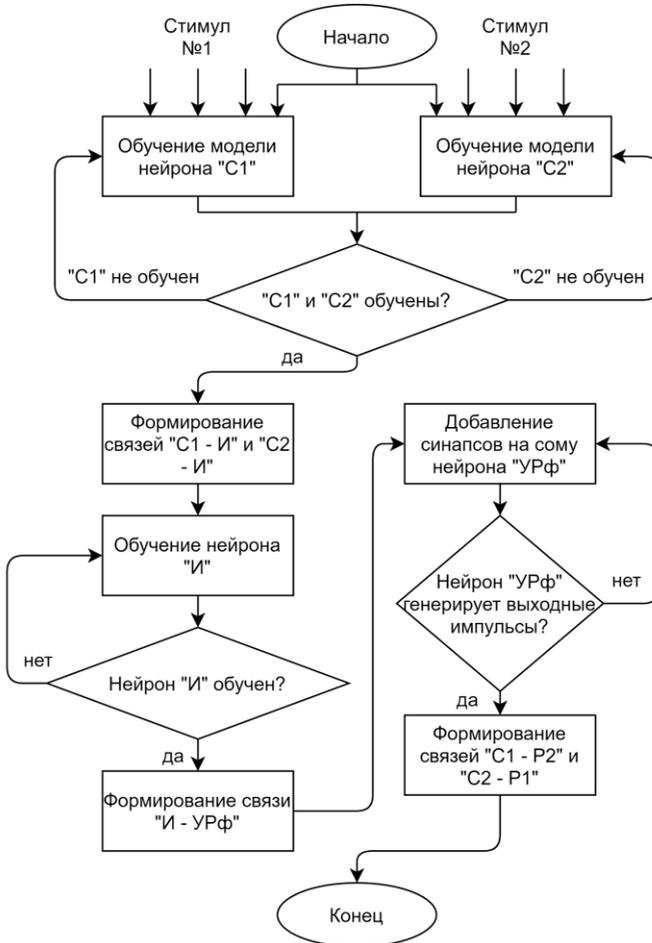


Рис. 7. Алгоритм формирования ассоциативных связей

5. Эксперименты по формированию ассоциативных связей.

Создание моделей нейронных сетей с возможностью структурного реконфигурирования как самой сети, так и её элементов, подразумевает необходимость обеспечить возможность описывать системы, в которых заранее не определена топология связей между элементами системы и их структурным описанием.

Для решения поставленной задачи использовалась программная платформа, описанная в [28, 29].

Результатом работы алгоритма в случае модели условного рефлекса являлся ответ модели нейрона безусловной реакции при активности модели нейрона условного раздражителя даже при отсутствии активности модели нейрона безусловного раздражителя. В случае модели формирования ассоциативных связей результатом работы являлась способность пары нейронов, моделирующих реакцию на стимулы, генерировать выходные сигналы (импульсы) при активности лишь одного произвольного стимула.

На рисунке 8 представлена исходная модель для воспроизведения условного рефлекса, реализованная в упомянутой выше программной платформе.

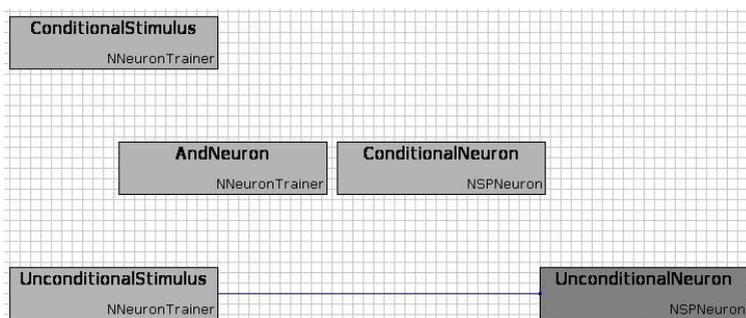


Рис. 8. Инициализированная модель формирования условного рефлекса

На рисунке 8 «ConditionalStimulus» – это нейрон, отвечающий условному раздражителю (здесь и ниже см. рис. 5); «UnconditionalStimulus» – нейрон, отвечающий безусловному раздражителю; «AndNeuron» – нейрон, реализующий функцию «И»; «ConditionalNeuron» – нейрон, формирующий условный рефлекс; «UnconditionalNeuron» – нейрон, реализующий функцию «ИЛИ» и отвечающий безусловному рефлексу.

После начала работы системы установлена лишь одна связь, отражающая изначально наличие безусловного рефлекса.

В качестве входных паттернов импульсов в описываемом эксперименте были выбраны следующие паттерны импульсов: (0,03; 0; 0,09; 0,06) для нейрона, отвечающего условному раздражителю, и (0,09; 0,06; 0,03; 0) для нейрона, отвечающего безусловному раздражителю. Отдельные сигналы в паттерне заданы как величины задержки относительно начала генерации очередного паттерна. Величина измерения – секунда. Частота генерации паттернов составляла 0,67 Гц.

Входные паттерны импульсов и инициализированная модель нейрона с возможностью обучения представлены на рисунке 9.

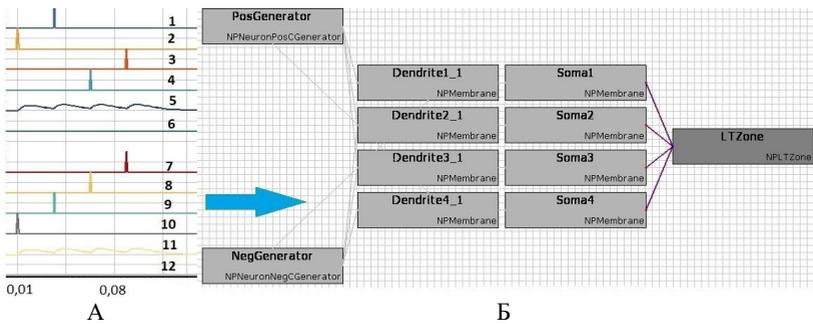


Рис. 9. А – входные паттерны импульсов; Б – инициализированная модель нейрона с возможностью обучения

На рисунке 9А цифрами 1–4 обозначен входной паттерн импульсов, отвечающий условному раздражителю; 5,11 – суммарный потенциал в низкопороговой зоне нейронов; 6 – выход нейрона, отвечающего условному раздражителю; 7-10 – входной паттерн импульсов, отвечающий безусловному раздражителю; 12 – выход нейрона, отвечающего безусловному раздражителю.

На рисунке 9Б: PosGenerator и NegGenerator – вспомогательные блоки, задающие начальное состояние активности участка мембраны; DendriteN_I–I-й сегмент N-го дендрита; SomaN – N-й сегмент сомы; LTZone –низкопороговая зона нейрона.

В результате процесса обучения нейронов «ConditionalStimulus» и «UnonditionalStimulus» соответствующим им входным паттернам импульсов, в определённый момент времени эти нейроны начинают генерировать выходной сигнал. Этот момент соответствует окончанию обучения нейронов и образованию связей «ConditionalStimulus – And-Neuron» и «UnonditionalStimulus – AndNeuron». На рисунке 10 представлена структурная организация обученного нейрона, отвечающего безусловному раздражителю.

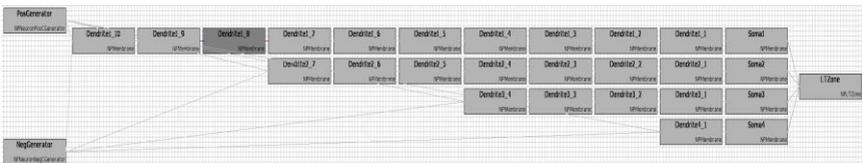


Рис. 10. Структурная организация нейрона, отвечающего безусловному раздражителю

Как видно из рисунка 10, в результате обучения нейрона, представленного на рисунке 9Б, длины четырёх дендритов составили десять, семь, четыре и один сегмент.

На следующем шаге, после образования связей «ConditionalStimulus – AndNeuron» и «UnonditionalStimulus – AndNeuron», начинается обучение нейрона, реализующего функцию «И». Структурная организация этого нейрона приведена на рисунке 11.

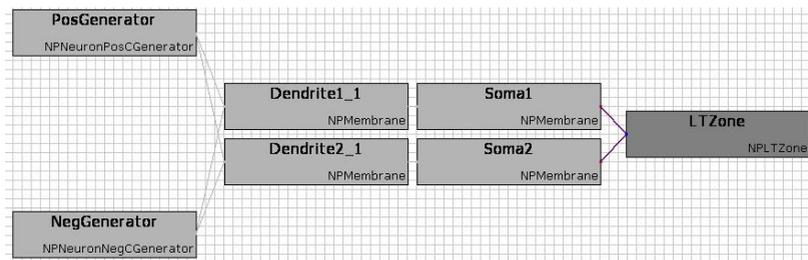
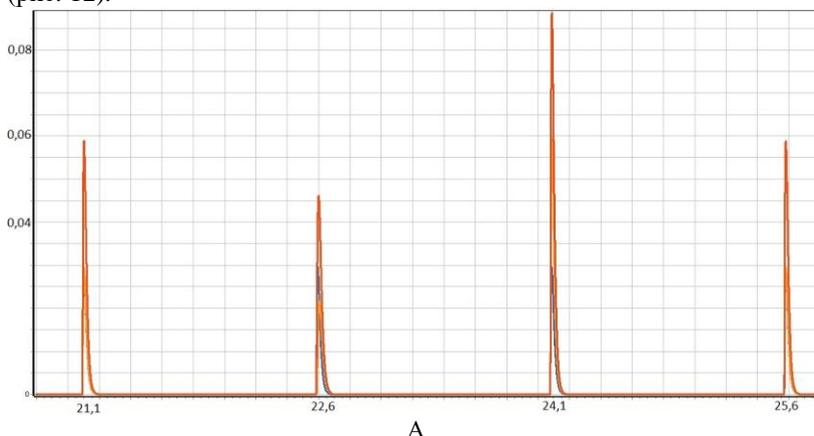


Рис. 11. Структурная организация нейрона, реализующего функцию «И»

Блоки на рисунке 11 имеют то же значение, что и на рисунке 9Б. Важной особенностью этого нейрона является генерация им выходных импульсов лишь в случае одновременной активности двух его входов.

В нашем случае входные импульсы приходят на вход данного нейрона одновременно, что в качестве простого примера хорошо иллюстрирует алгоритм структурной адаптации модели нейрона (рис. 12).



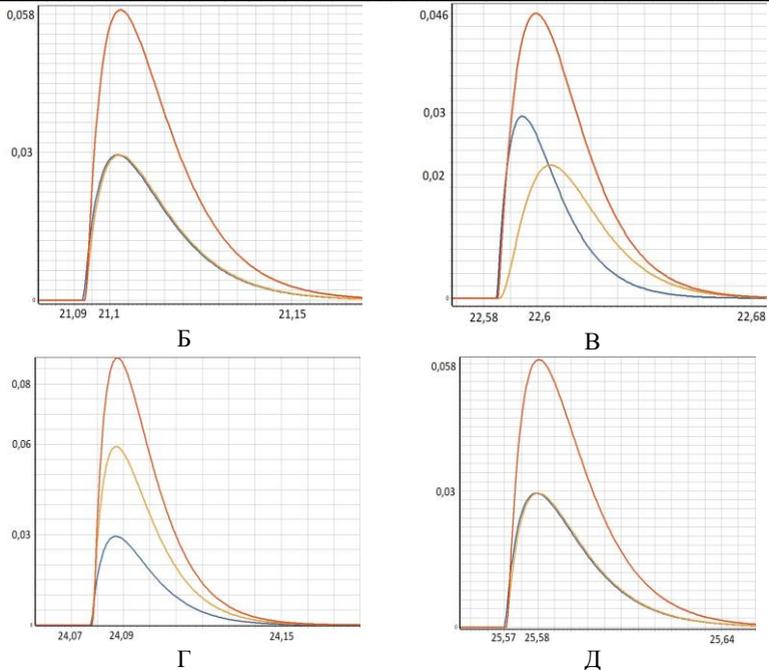


Рис. 12. Обучение нейрона, реализующего функцию «И»: А - всплеск суммарного потенциала внутри низкопороговой зоны нейрона, каждый из которых соответствует отдельной итерации; Б-Д – детальная иллюстрация всплесков

На рисунках 12 Б-Д кривые с наибольшей амплитудой соответствуют суммарному значению потенциала внутри низкопороговой зоны нейрона, остальные кривые – отдельным дендритам.

Рисунок 12Б соответствует первому всплеску – началу обучения модели нейрона, когда на его вход одновременно пришло два импульса. Одновременность означает отсутствие необходимости структурно адаптировать нейрон, тем не менее, необходимо в этом убедиться алгоритмически.

На рисунке 12В произошло наращивание одного из дендритов на один сегмент. Это соответствует процедуре синхронизации. Как видно из рисунка, экстремум потенциала, соответствующего данному дендриту, получил временное смещение (запаздывание) относительно базового дендрита и, вместе с тем, уменьшил свою амплитуду.

В результате фиксации увеличения рассинхронизации сигналов с двух дендритов лишний сегмент дендрита был удалён, и начался процесс нормализации (рис. 12Г). В результате был добавлен синапс на последний сегмент дендрита. Как видно на рисунке, рассинхронизация пропадает, но увеличивается амплитуда, соответствующая данному дендриту.

В результате фиксации превышения экстремумом этой амплитуды эталонного значения, лишний синапс удаляется, и обучение заканчивается с конфигурацией структуры нейрона, эквивалентной той, что и была до начала обучения (рис. 12Д).

После завершения обучения нейрона, реализующего функцию «И», создаётся связь «AndNeuron – ConditionalNeuron», и «большой» нейрон начинает наращивать синапсы (рис. 13).

На рисунке 13А показана структура модели «большого» нейрона. В данном эксперименте «большой» нейрон состоял из десяти сегментов модели сомы «SomaN». «LTZone» - модель генератора выходного импульса (потенциала действия). Каждый новый входной импульс от модели нейрона, реализующего функцию «И» (рис. 13В), добавлял на данный сегмент сомы новый синапс. Это отражено на рисунке 13Б, где ExcSynapseN – N-я модель возбуждающего синапса; ExcChannel – модель возбуждающего ионного канала; InhSynapse1 – модель тормозного синапса; InhChannel – модель тормозного ионного канала.

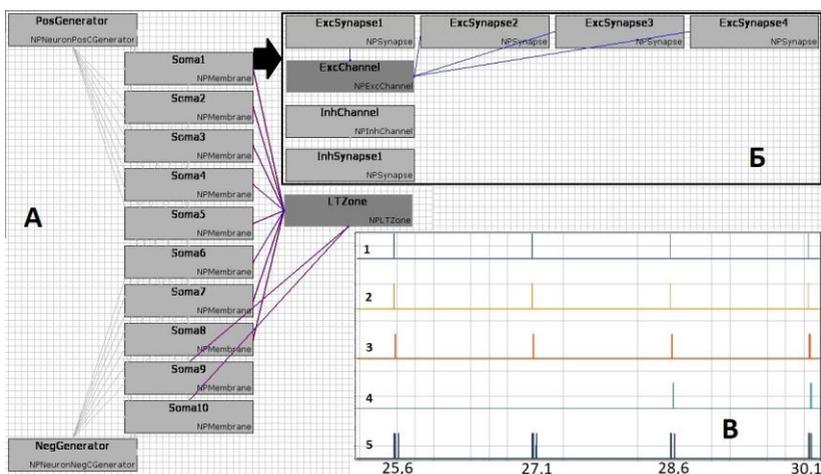


Рис. 13. Обучение нейрона, формирующего условный рефлекс

На рисунке 13В обозначены: 1 – выход нейрона, отвечающего условному раздражителю; 2 – выход нейрона, отвечающего

безусловному раздражителю; 3 – выход нейрона, реализующего функцию «И»; 4 – выход нейрона, формирующего условный рефлекс («большой» нейрон); 5 – выход нейрона, реализующего функцию «ИЛИ» и отвечающего безусловному рефлексу.

Как видно из рисунка 13В, в определённый момент «большой» нейрон начинает генерировать выходные импульсы, после чего формируется связь «ConditionalStimulus – UnconditionalNeuron», и, таким образом, рефлекторная дуга «РД» замыкается.

На рисунке 14 изображена схема единственного сегмента сомы нейрона, реализующего функцию «ИЛИ» и отвечающего безусловному рефлексу.

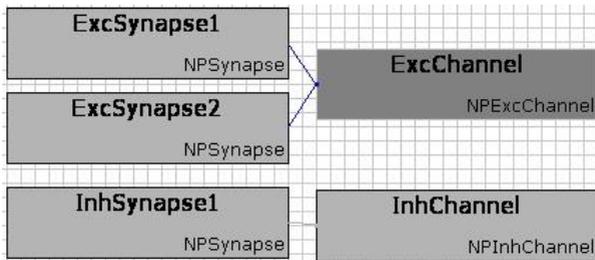


Рис. 14. Схема сомы нейрона, реализующего функцию «ИЛИ» и отвечающего безусловному рефлексу

Блоки на рисунке 14 имеют то же значение, что и на рисунке 13Б.

В момент инициализации системы формируется отвечающая безусловному рефлексу связь «UnconditionalStimulus – UnconditionalNeuron», которая соединена с блоком «ExcSynapse1» на рисунке 14. После замыкания рефлекторной дуги связь «ConditionalStimulus – UnconditionalNeuron» соединяется с блоком «ExcSynapse2». Теперь достаточно активности на входе лишь одной из моделей синапсов для того, чтобы нейрон, реализующий функцию «ИЛИ», генерировал выходные импульсы.

Результат работы модели, сформировавшей условный рефлекс, приведён на рисунке 15.



Рис. 15. Работа системы при активности лишь условного раздражителя

Графики на рисунке 15 расположены в том же порядке, что и на рисунке 13В. Как видно из рисунка 15, отключение в какой-то момент времени безусловного раздражителя привело к прекращению активности выходов блоков «AndNeuron» и «ConditionalNeuron», однако, выход блока «UnconditionalNeuron» продолжает быть активным.

Таким образом, условный рефлекс был сформирован.

Отметим, что всё вышесказанное относилось лишь к условному рефлексу с положительным подкреплением. Для того, чтобы сформировать условный рефлекс с отрицательным подкреплением, нам необходимо на последнем шаге алгоритма при организации связи «ConditionalStimulus – UnconditionalNeuron» соединить выход низкопороговой зоны блока «ConditionalStimulus» с блоком «InhSynapse1» (тормозный синапс), а не с блоком «ExcSynapse2» (возбуждающий синапс) (рис. 14). В таком случае, после завершения формирования рефлекса, при активности условного раздражителя, на выходе блока «UnconditionalNeuron» генерация выходных импульсов не будет производиться вне зависимости от наличия или отсутствия активности безусловного раздражителя (рис. 16).

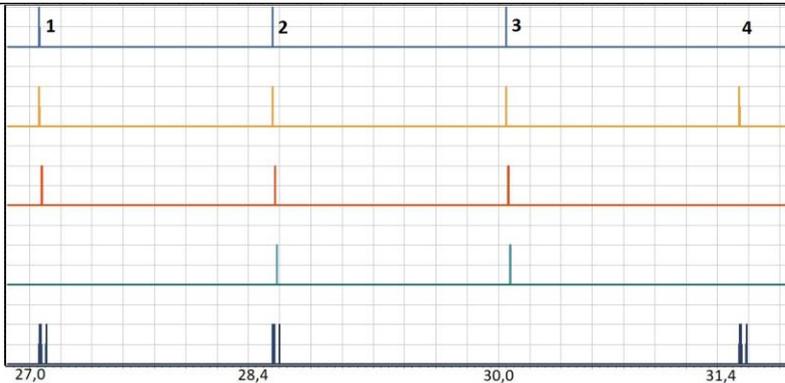


Рис. 16. Формирование условного рефлекса с отрицательным подкреплением

Графики на рисунке 16 расположены в том же порядке, что и на рисунке 13В. На рисунке 16 на итерации «1» подходит к завершению обучение модели «большого» нейрона, рефлекторная дуга «РД» не замкнута. Блок «UnconditionalNeuron» генерирует выходной импульс из-за активности выхода нейрона, отвечающего безусловному раздражителю.

На итерации «2» обучение «большого» нейрона завершается. «Большой» нейрон генерирует выходной импульс, что служит сигналом к формированию связи «ConditionalStimulus – UnconditionalNeuron».

На итерации «3» связь «ConditionalStimulus – UnconditionalNeuron» сформирована. Активность на выходе нейрона «UnconditionalNeuron» пропадает, несмотря на активность выхода на модели безусловного раздражителя.

На итерации «4» стимуляция модели условного раздражителя отключается, что приводит к восстановлению активности на выходе нейрона «UnconditionalNeuron».

Таким образом, условный рефлекс с отрицательным подкреплением был сформирован.

Рассмотрим теперь формирование ассоциативных связей в спайковых сегментных нейронных сетях в общем случае. На рисунке 6 была приведена структурная схема нейронной сети, формирующей ассоциативные связи. На рисунке 17 представлена инициализированная модель для формирования ассоциативных связей, отвечающая этой структурной схеме.

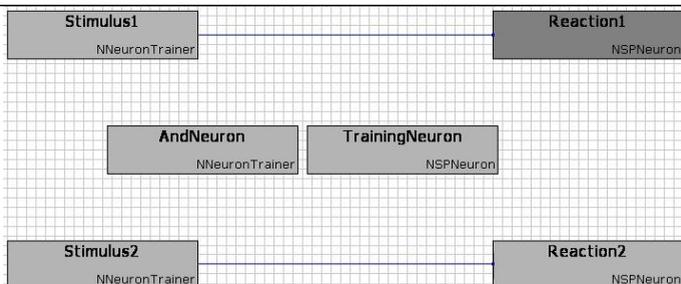


Рис. 17. Инициализированная модель для формирования ассоциативных связей

На рисунке 17 «Stimulus1» – это нейрон, на который поступает паттерн импульсов (здесь и ниже см. рис. 6). Нейрон «Reaction1» имеет лишь один сегмент сомы, на котором, однако, расположено два синапса. Связь «Stimulus1 – Reaction1» образуется при инициализации системы. Нейроны «Stimulus2» и «Reaction2» реализуют безусловную реакцию нейрона «Reaction2» на стимул «Stimulus2». Связь «Stimulus2 – Reaction2» также создаётся при инициализации системы.

Блоки «AndNeuron» и «TrainingNeuron» моделируют образование рефлекторной дуги способом, полностью совпадающим с описанным выше для условного рефлекса, за одним исключением: при начале генерации сигналов нейроном «TrainingNeuron» образуется сразу две ассоциативных связи – «Stimulus1 – Reaction2» и «Stimulus2 – Reaction1». Таким образом, нейроны «Reaction1» и «Reaction2» приобретают способность генерировать выходные сигналы (импульсы) при активности лишь одного входа: «Stimulus1» либо «Stimulus2» (рис. 18).

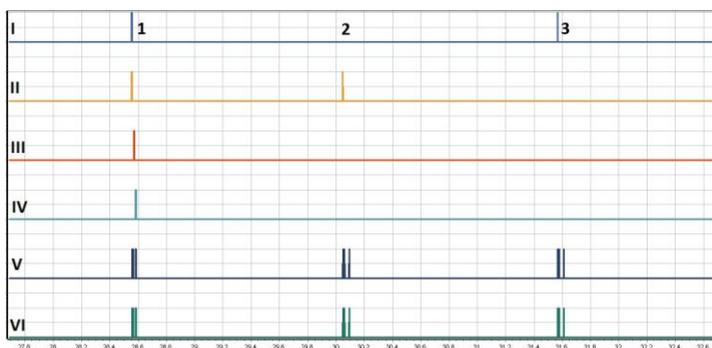


Рис. 18. Формирование ассоциативных связей

На рисунке 18 обозначены: I – выход нейрона, обученного стимулу №1; II – выход нейрона, обученного стимулу №2; III – выход

нейрона, реализующего функцию «И»; IV – выход нейрона, формирующего ассоциативные связи («большой» нейрон); V – выход нейрона, отвечающего за реакцию на стимул №1; VI – выход нейрона, отвечающего за реакцию на стимул №2.

На первой итерации, изображённой на рисунке 18 («1»), все нейроны обучены и ассоциативные связи сформированы, что видно по наличию активности на выходе нейрона, формирующего ассоциативные связи (график «IV»). На этой итерации активны оба стимула и выходы обоих нейронов, отвечающих за реакцию.

На второй итерации («2»), генерация первого стимула была отключена, тем не менее, активны оба нейрона, отвечающие за реакцию. И аналогично для итерации №3 («3»): генерация второго стимула отключена, тем не менее, активны оба нейрона, отвечающие за реакцию.

Как видим, ассоциативные связи были сформированы.

6. Заключение. Анализ существующих архитектур спайковых нейронных сетей показал, что в рамках структурной адаптации рассматриваются в основном вопросы регулирования количества нейронов и синапсов. В работе показано, что адаптивная способность спайковой нейронной сети может быть увеличена, если использовать модели нейронов с изменяемой структурой дендритного дерева и синаптического аппарата. Использование наряду с настройкой параметров нейронов и структуры нейронной сети дополнительной возможности настройки внутренней структуры самих нейронов, по мнению авторов этой работы, может значительно повысить эффективность реализации поведенческих функций в нейроморфных системах управления.

Представленные в работе алгоритмы структурной адаптации для решения задач распознавания паттернов импульсов, а также алгоритм формирования ассоциативных связей нейронной сети, показали возможность реализации поведенческих функций в нейроморфных системах управления, построенных на спайковых нейронных сетях с сегментными нейронами.

Предложенный метод и реализующий его алгоритм структурной адаптации модели нейрона к произвольному паттерну входных импульсов фактически обеспечивает обучение нейрона положительной реакции на данный паттерн. После завершения обучения нейрон приобретает возможность распознавать паттерн входных импульсов, которому он был ранее обучен. Однако следует отметить, что предлагаемый метод структурной адаптации модели нейрона не дает

возможности переобучения отдельного нейрона в автоматическом режиме.

Эффективность предложенной схемы и реализующего ее алгоритма формирования ассоциативных связей в сети сегментных спайковых нейронов показана на примере реализации простых поведенческих функций типа безусловного и условного рефлекса с положительным и отрицательным подкреплением. Однако следует отметить, что созданные на настоящий момент времени ассоциативные связи не обладают способностью со временем ослабевать либо полностью разрушаться.

Поиск решения для устранения недостатков предложенных методов структурной адаптации сегментных нейронов к входному паттерну, а также формирования ассоциативных связей при настройке спайковой нейронной сети на заданную поведенческую функцию, является перспективным направлением развития исследований по данной тематике. Остаётся также открытым вопрос автоматического выбора сигналов, которые должны быть ассоциированы друг с другом в нейроморфной системе. В настоящее время пара связываемых друг с другом стимулов задаётся оператором. Переход на верхний уровень моделирования ассоциативных процессов представляется сложной, но перспективной задачей [30]. В планах дальнейших исследований значится также создание моделей, не ограниченных лишь парой входных паттернов импульсов, и способных создавать сложные взаимосвязи (как возбуждающие, так и тормозные) между составляющими их слоями спайковых нейронов.

Литература

1. *Shrestha A., Mahmood A.* Review of deep learning algorithms and architectures. IEEE Access. Institute of Electrical and Electronics Engineers Inc. 2019. vol. 7. pp. 53040–53065.
2. *Tavanaei A., Ghodrati M., Kheradpisheh S. R., Masquelier T., Maida A.* Deep learning in spiking neural networks // *Neural Networks*. 2019. vol. 111. pp. 47–63.
3. *James C. D. et al.* A historical survey of algorithms and hardware architectures for neural-inspired and neuromorphic computing applications // *Biologically Inspired Cognitive Architectures*. 2017. vol. 19. pp. 49–64.
4. *Schuman C.D., Potok T.E.* A Survey of Neuromorphic Computing and Neural Networks in Hardware // *arXiv preprint arXiv:1705.06963*. 2017.
5. *Мельников Г.С., Мельникова Э.И., Самков В.М.* Нейроморфные системы анализа изображений (обзор) // Труды Международного научно-технического конгресса «Интеллектуальные системы и информационные технологии-2020» («IS&IT'20»). Т.2. Таганрог: Изд-во Ступина С.А. 2020. С. 120–148.
6. *Feng J.* Is the integrate-and-fire model good enough a review // *Neural Networks*. 2001. vol. 14. no. 6. pp. 955–975.
7. *Gaier A., Ha. D.* Weight Agnostic Neural Networks // 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). 2019. pp. 1–19.

8. *Kasabov N. K.* Evolving Connectionist Systems: The Knowledge Engineering Approach // London, Springer Science & Business Media. 2007. 451 p.
9. *Soltic S., Kasabov N. K.* Knowledge extraction from evolving spiking neural networks with rank order population coding // International Journal of Neural Systems. 2010. vol. 20. no. 06. pp. 437–445.
10. *Wysoski S. G., Benuskova L., Kasabov N.* Fast and adaptive network of spiking neurons for multi-view visual pattern recognition // Neurocomputing. 2008. vol. 71. no. 13. pp. 2563–2575.
11. *Wysoski S. G., Benuskova L., Kasabov N. K.* Evolving spiking neural networks for audiovisual information processing // Neural Networks. 2010. vol. 23. no. 7. pp. 819–835.
12. *Wysoski S. G., Benuskova L., Kasabov N.* On-Line Learning with Structural Adaptation in a Network of Spiking Neurons for Visual Pattern Recognition // Artificial Neural Networks. 2006. vol. 4131. pp. 61–70.
13. *Dhoble K., Nuntalid N., Indiveri G., Kasabov N.* Online spatio-temporal pattern recognition with evolving spiking neural networks utilising address event representation, rank order, and temporal spike learning // The 2012 International Joint Conference on Neural Networks (IJCNN). 2012. pp. 1–7.
14. *Kasabov N. K., Dhoble K., Nuntalid N., Indiveri G.* Dynamic evolving spiking neural networks for on-line spatio- and spectro-temporal pattern recognition // Neural Networks. 2013. vol. 41. pp. 188–201.
15. *Wang J., Belatreche A., Maguire L., McGinnity T. M.* An online supervised learning method for spiking neural networks with adaptive structure // Neurocomputing. 2014. vol. 144. pp. 526–536.
16. *Wang J., Belatreche A., Maguire L. P., McGinnity T. M.* SpikeComp: An Evolving Spiking Neural Network with Adaptive Compact Structure for Pattern Classification // Neural Information Processing. 2015. vol. 9490. pp. 259–267.
17. *Wang J., Belatreche A., Maguire L. P., McGinnity T. M.* SpikeTemp: An Enhanced Rank-Order-Based Learning Approach for Spiking Neural Networks With Adaptive Structure // IEEE Transactions on Neural Networks and Learning Systems. 2017. vol. 28. no. 1. pp. 30–43.
18. *Belatreche A., Maguire L. P., McGinnity M.* Advances in Design and Application of Spiking Neural Networks // Soft Computing. 2017. vol. 3. no. 11. pp. 239–248.
19. *XinYao, Yong Liu, Guangming Lin.* Evolutionary programming made faster // IEEE Transactions on Evolutionary Computation. 1999. vol. 3. no. 2. pp. 82–102.
20. *Belatreche A., Paul R.* Dynamic cluster formation using populations of spiking neurons // The 2012 International Joint Conference on Neural Networks (IJCNN). 2012. pp. 1–6.
21. *Dora, S., Subramanian, K., Suresh, S., Sundararajan, N.* Development of a Self-Regulating Evolving Spiking Neural Network for classification problem // Neurocomputing. 2016. vol. 171. pp. 1216–1229.
22. *Khan G. M., Miller J. F.* In search of intelligence: evolving a developmental neuron capable of learning // Connection Science. 2014. vol. 26. no. 4. pp. 297–333.
23. *Soltoggio A., Stanley K. O., Risi S.* Born to learn: The inspiration, progress, and future of evolved plastic artificial neural networks // Neural Networks. 2018. vol. 108. pp. 48–67.
24. *Bakhshiev A.V., Gundelakh F.V.* Mathematical Model of the Impulses Transformation Processes in Natural Neurons for Biologically Inspired Control Systems Development // CEUR Workshop Proceedings. 2015. vol. 1452. pp. 1–12.
25. *Асратян Э.А.* Учение академика И.П. Павлова о высшей нервной деятельности // Серия: Знание, Серия III №1 М.: Знание. 1956. 32 с.
26. *Кубарко, А. И.* Нормальная физиология. В 2 ч. Ч. 2: учебник / А. И. Кубарко, А. А. Семенович, В. А. Переверзев, Д. А. Александров, Л. М. Лобанок, А. Н. Харламов // Минск: Выш. шк. 2014. 604 с.

27. *Korsakov, A., Bakhshiev, A.* The Neuromorphic Model of the Human Visual System // Studies in Computational Intelligence. 2021. vol. 1452. pp. 339–346.
28. *Бахшиев А. В.* Библиотека средств разработки моделей нейронных сетей со сложной и динамически меняющейся архитектурой – NMSDK // Нейроинформатика, ее приложения и анализ данных: материалы XVIII Всероссийского семинара, 8–10 октября 2010 г. / под ред. А. Н. Горбаня, Е. М. Миркеса. ИВМ СО РАН. Красноярск. 2010. С. 26–30.
29. *Bakhshiev A. V., Fomin I. S., Gundelakh F. V., Demcheva A. A., Korsakov A. M.* The architecture of a software platform for growing spiking neural networks simulator developing // Journal of Physics: Conference Series. 2020. vol. 1679. pp. 042001.
30. *Bakhshiev A., Korsakov A., Stankevich L.* The hierarchical memory based on compartmental spiking neuron model // 13th International Conference on Artificial General Intelligence, AGI 2020. 2020. vol. 12711. pp. 34–43.

Корсаков Антон Михайлович — старший научный сотрудник, 913 лаборатория «Технологий искусственного интеллекта», ЦНИИ РТК. Область научных интересов: нейро-морфные системы и спайковые нейронные сети, математические основы управления робототехническими системами, системы технического зрения. Число научных публикаций — 27. anton_korsakov@mail.ru; Тихорецкий проспект, д. 21, г. Санкт-Петербург, 194064, РФ; м.т.: +7(911)7842747.

Бахшиев Александр Валерьевич — канд. техн. наук, доцент, высшая школа автоматизации и робототехники, ИММиТ, СПбПУ Петра Великого. Область научных интересов: применение методов искусственного интеллекта в решении задач технического зрения, развитие биологически-инспирированных подходов к созданию систем управления и обработки информации, разработка новых архитектур спайковых нейронных сетей. Число научных публикаций — 90. palexab@gmail.com; ул. Политехническая, д.29, г. Санкт-Петербург, 195251; м.т.: +7(921)748-0010.

Астапова Любовь Алексеевна — студент-магистр, высшая школа автоматизации и робототехники, ИММиТ, СПбПУ Петра Великого, математик 2к, 913 лаборатория "Технологий искусственного интеллекта", ЦНИИ РТК. Область научных интересов: нейроморфные системы и спайковые нейронные сети, математические основы управления робототехническими системами, системы технического зрения. Число научных публикаций — 2. astapova.la@yandex.ru; Тихорецкий проспект, д. 21, г. Санкт-Петербург, 194064, РФ; м.т.: +7(981)7852931.

Станкевич Лев Александрович — канд. техн. наук, доцент, начальник отдела, отдел разработки ООО «АйБрейн», ведущий научный сотрудник, ЦНИИ РТК. Область научных интересов: искусственный интеллект, когнитивные системы, антропоморфная робототехника, интерфейсы «мозг-компьютер». Число научных публикаций — 265. Stankevich_lev@inbox.ru; РФ, г. Санкт-Петербург, Тихорецкий пр. 21, 194064, м.т.: +7(812)297-08-73.

Поддержка исследований. Работа проводилась в рамках выполнения в ЦНИИ РТК государственного задания Минобрнауки России на 2021 год: “Разработка и исследование новых архитектур реконфигурируемых растущих нейронных сетей, методов и алгоритмов их обучения”

A. KORSAKOV, A. BAKHSHIEV, L. ASTAPOVA, L. STANKEVICH
BEHAVIORAL FUNCTIONS IMPLEMENTATION ON
SPIKING NEURAL NETWORKS

Korsakov A., Bakhshiev A., Astapova L., Stankevich L. **Behavioral Functions Implementation on Spiking Neural Networks.**

Abstract. The question of behavioral functions modeling of animals (in particular, the modeling and implementation of the conditioned reflex) is considered. The analysis of the current state of neural networks with the possibility of structural reconfiguration is carried out. The modeling is carried out by means of neural networks, which are built on the basis of a compartmental spiking model of a neuron with the possibility of structural adaptation to the input pulse pattern. The compartmental spike model of a neuron is able to change its structure (the size of the cell body, the number and length of dendrites, the number of synapses) depending on the incoming pulse pattern at its inputs. A brief description of the compartmental spiking model of a neuron is given, and its main features are noted in terms of the possibility of its structural reconfiguration. The method of structural adaptation of the compartmental spiking model of the neuron to the input pulse pattern is described. To study the work of the proposed model of a neuron in a network, the choice of a conditioned reflex as a special case of the formation of associative connections is justified as an example. The structural scheme and algorithm of formation of a conditioned reflex with both positive and negative reinforcement are described. The article presents a step-by-step description of experiments on the associative connection formation in general and conditioned reflex (both with positive and negative reinforcement), in particular. The conclusion is made about the prospects of using spiking compartmental models of neurons to improve the efficiency of the implementation of behavioral functions in neuromorphic control systems. Further promising directions for the development of neuromorphic systems based on spiking compartmental models of the neuron are considered.

Keywords: Neuromorphic Systems, Spike Neuron, Spike Neural Networks, Classical Conditioning, Associative Connections

Korsakov Anton — Senior Researcher, Laboratory of Artificial Intelligence Technologies, RTC. Research interests: neuromorphic systems and spike neural networks, mathematical foundations of robotic systems control, computer vision. The number of publications — 27. anton_korsakov@mail.ru; 21 Tikhoretsky pr., Saint-Petersburg, 194064, Russia; phone +7(911)7842747.

Bakhshiev Aleksandr — Ph.D., Docent, Peter the Great St.Petersburg Polytechnic University (SPbPU). Research interests: application of artificial intelligence methods in solving computer vision problems, development of biologically inspired approaches to the creation of information management and processing systems, development of new spike neural network architectures. The number of publications — 90. palexab@gmail.com; 29 Politechnicheskaya ul., Saint-Petersburg, 195251, Russia; office phone: +7(921)748-0010.

Astapova Lyubov — Student, Peter the Great St.Petersburg Polytechnic University (SPbPU), mathematician, Laboratory of Artificial Intelligence Technologies, RTC. Research interests: neuromorphic systems and spike neural networks, mathematical foundations of robotic systems control, computer vision. The number of publications — 2. astapova.la@yandex.ru; 21 Tikhoretsky pr., Saint-Petersburg, 194064, Russia; office phone: +7(981)7852931.

Stankevich Lev — Ph.D., Docent, Head of Development, Department of “AiBrain”, Senior researcher, Laboratory of Artificial Intelligence Technologies, RTC. Research interests: artificial intelligence, cognitive systems, anthropomorphic robotics, brain-computer interfaces.

References

1. *Shrestha A., Mahmood A.* Review of deep learning algorithms and architectures // IEEE Access. Institute of Electrical and Electronics Engineers Inc. 2019. vol. 7. pp. 53040–53065.
2. *Tavanaei A., Ghodrati M., Kheradpisheh S. R., Masquelier T., Maida A.* Deep learning in spiking neural networks // Neural Networks. 2019. vol. 111. pp. 47–63.
3. *James C. D. et al.* A historical survey of algorithms and hardware architectures for neural-inspired and neuromorphic computing applications // Biologically Inspired Cognitive Architectures. 2017. vol. 19. pp. 49–64.
4. *Schuman C.D., Potok T.E.* A Survey of Neuromorphic Computing and Neural Networks in Hardware // arXiv preprint arXiv:1705.06963. 2017.
5. *Mel'nikov G.S., Mel'nikova Je.I., Samkov V.M.* Nejromorfnye sistemy analiza izobrazhenij (obzor) // Trudy Mezhdunarodnogo nauchno-tehnicheskogo kongressa «Intel'kual'nye sistemy i informacionnye tehnologii-2020» («IS&IT'20»). T.2. Taganrog: Izd-vo Stupina S.A. 2020. pp. 120-148. (In Russ.).
6. *Feng J.* Is the integrate-and-fire model good enough a review // Neural Networks. 2001. vol. 14. no. 6. pp. 955–975.
7. *Gaier A., Ha. D.* Weight Agnostic Neural Networks // 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). 2019. pp. 1–19.
8. *Kasabov N. K.* Evolving Connectionist Systems: The Knowledge Engineering Approach // London, Springer Science & Business Media. 2007. 451 p.
9. *Soltic S., Kasabov N. K.* Knowledge extraction from evolving spiking neural networks with rank order population coding // International Journal of Neural Systems. 2010. vol. 20. no. 6. pp. 437–445.
10. *Wysoski S. G., Benuskova L., Kasabov N.* Fast and adaptive network of spiking neurons for multi-view visual pattern recognition // Neurocomputing. 2008. vol. 71. no. 13. pp. 2563–2575.
11. *Wysoski S. G., Benuskova L., Kasabov N. K.* Evolving spiking neural networks for audiovisual information processing // Neural Networks. 2010. vol. 23. no. 7. pp. 819–835.
12. *Wysoski S. G., Benuskova L., Kasabov N.* On-Line Learning with Structural Adaptation in a Network of Spiking Neurons for Visual Pattern Recognition // Artificial Neural Networks. 2006. vol. 4131. pp. 61–70.
13. *Dhoble K., Nuntalid N., Indiveri G., Kasabov N.* Online spatio-temporal pattern recognition with evolving spiking neural networks utilising address event representation, rank order, and temporal spike learning // The 2012 International Joint Conference on Neural Networks (IJCNN). 2012. pp. 1–7.
14. *Kasabov N. K., Dhoble K., Nuntalid N., Indiveri G.* Dynamic evolving spiking neural networks for on-line spatio- and spectro-temporal pattern recognition // Neural Networks. 2013. vol. 41. pp. 188–201.
15. *Wang J., Belatreche A., Maguire L., McGinnity T. M.* An online supervised learning method for spiking neural networks with adaptive structure // Neurocomputing. 2014. vol. 144. pp. 526–536.
16. *Wang J., Belatreche A., Maguire L. P., McGinnity T. M.* SpikeComp: An Evolving Spiking Neural Network with Adaptive Compact Structure for Pattern Classification // Neural Information Processing. 2015. vol. 9490. pp. 259–267.
17. *Wang J., Belatreche A., Maguire L. P., McGinnity T. M.* SpikeTemp: An Enhanced Rank-Order-Based Learning Approach for Spiking Neural Networks With Adaptive Structure // IEEE Transactions on Neural Networks and Learning Systems. 2017. vol. 28. no. 1. pp. 30–43.
18. *Belatreche A., Maguire L. P., McGinnity M.* Advances in Design and Application of Spiking Neural Networks // Soft Computing. 2017. vol. 3. no. 11. pp. 239–248.

19. *XinYao, Yong Liu, Guangming Lin.* Evolutionary programming made faster // IEEE Transactions on Evolutionary Computation. 1999. vol. 3. no. 2. pp. 82–102.
20. *Belatreche A., Paul R.* Dynamic cluster formation using populations of spiking neurons // The 2012 International Joint Conference on Neural Networks (IJCNN). 2012. pp. 1–6.
21. *Dora, S., Subramanian, K., Suresh, S., Sundararajan, N.* Development of a Self-Regulating Evolving Spiking Neural Network for classification problem // Neurocomputing. 2016. vol. 171. pp. 1216–1229.
22. *Khan G. M., Miller J. F.* In search of intelligence: evolving a developmental neuron capable of learning // Connection Science. 2014. vol. 26. no. 4. pp. 297–333.
23. *Soltoggio A., Stanley K. O., Risi S.* Born to learn: The inspiration, progress, and future of evolved plastic artificial neural networks // Neural Networks. 2018. vol. 108. pp. 48–67.
24. *Bakhshiev A.V., Gundelakh F.V.* Mathematical Model of the Impulses Transformation Processes in Natural Neurons for Biologically Inspired Control Systems Development // CEUR Workshop Proceedings. 2015. vol. 1452. pp. 1–12.
25. *Asratjan Je.A.* Uchenie akademika I.P. Pavlova o vysšej nervnoj dejatel'nosti // Serija: Znanie, Serija III №1 M.: Znanie. 1956. 32 p. (In Russ.).
26. *Kubarko, A. I.* Normal'naja fiziologija. V 2 ch. Ch. 2: uchebnik / A. I. Kubarko, A. A. Semenovich, V. A. Pereverzev, D. A. Aleksandrov, L. M. Lobanok, A. N. Harlamov // Minsk: Vysh. shk. 2014. 604 p. (In Russ.).
27. *Korsakov, A., Bakhshiev, A.* The Neuromorphic Model of the Human Visual System // Studies in Computational Intelligence. 2021. vol. 1452. pp. 339–346.
28. *Bakhshiev A. V.* Biblioteka sredstv razrabotki modelej nejronnyh setej so slozhnoj i dinamičeski menjajushhejsja arhitekturoj – NMSDK // Nejroinformatika, ee prilozhenija i analiz dannyh: materialy XVIII Vserossijskogo seminaru, 8–10 oktjabrja 2010 g. / pod red. A. N. Gorbanja, E. M. Mirkesa. IVM SO RAN. Krasnojarsk. 2010. pp. 26–30. (In Russ.).
29. *Bakhshiev A. V., Fomin I. S., Gundelakh F. V., Demcheva A. A., Korsakov A. M.* The architecture of a software platform for growing spiking neural networks simulator developing // Journal of Physics: Conference Series. 2020. vol. 1679. pp. 042001.
30. *Bakhshiev A., Korsakov A., Stankevich L.* The hierarchical memory based on compartmental spiking neuron model // 13th International Conference on Artificial General Intelligence, AGI 2020. 2020. vol. 12711. pp. 34–43.

S. Altaf, S. Iqbal, M. W. Soomro
**EFFICIENT ALGORITHM FOR NATURAL LANGUAGE
CLASSIFICATION TO DETECT DUPLICATE UNSUPERVISED
FEATURES**

Altaf S., Iqbal S., Soomro M.W. Efficient Natural Language Classification Algorithm for Detecting Duplicate Unsupervised Features

Abstract. This paper focuses on capturing the meaning of Natural Language Understanding (NLU) text features to detect the duplicate unsupervised features. The NLU features are compared with lexical approaches to prove the suitable classification technique. The transfer-learning approach is utilized to train the extraction of features on the Semantic Textual Similarity (STS) task. All features are evaluated with two types of datasets that belong to Bosch bug and Wikipedia article reports. This study aims to structure the recent research efforts by comparing NLU concepts for featuring semantics of text and applying it to IR. The main contribution of this paper is a comparative study of semantic similarity measurements. The experimental results demonstrate the Term Frequency–Inverse Document Frequency (TF-IDF) feature results on both datasets with reasonable vocabulary size. It indicates that the Bidirectional Long Short-Term Memory (BiLSTM) can learn the structure of a sentence to improve the classification.

Keywords: Clustering, Information Retrieval, TF-IDF Feature, Par2Vec, Natural Language Texts, Lexical Approaches

1. Introduction. Humans can exchange all kinds of information through language. For instance, language is used to talk about activities, discuss abstract concepts, and even helps to determine the sentimental state of other human beings. Consequently, language is used frequently in all kinds of communication, like e-mails, reports, conversations, and scientific papers. This usage generates large language datasets, which contain a lot of information. Approximately 80% of all relevant corporate data is text-heavy unstructured data [1]. Unfortunately acquired insights out of the text are a major problem for big data approaches. Currently, the main part of that information cannot be used for further automatic analysis. The lack of such methods makes it difficult to find relevant information in large text datasets.

Without being able to find relevant information already existing information is likely reproduced. The reproduction of information causes regularly unnecessary work. One example of this is the reimplementing of an algorithm. In software development, it is well known, that this duplication creates additional problems for maintaining the implementations and even causes inconsistencies based on slight differences in the several implementations. The produced heterogeneous data is a problem in other areas as well and makes the search itself more difficult. Other consequences of information shortage are even more devastating. Managers could make

wrong business decisions, lawyers would not be able to defend their clients, engineers could design buggy products and the research of scientists would be more difficult without accessing the results of colleagues. The information has become by far the most valuable resource for almost everyone. Therefore, the required access through Information Retrieval (IR) systems to manage large amounts of data is one of today's key challenges. IR recognizes the implicit patterns contained in collections of unstructured data to find data that satisfy information needs [2]. This defines a variety of IR algorithms for a broad application field. Depending on the task particular datasets, patterns, and formulations of information needs are considered.

The main objective of IR is making information accessible through document search based on an information requirement. It organizes large amounts of data and structures the data according to implicit patterns contained in the data and the information required. IR tasks consist of two distinct sub-goals: understanding the input data and analyzing it concerning this understanding by classification or clustering methods. Understanding a text according to Wittgenstein means creating a picture based on this text and a priori knowledge [3]. Transferring this concept into the domain of machine learning the creation of a picture is the generation of representation within a vector space model based on several features of the text. Present available systems depend on several domain-specific features and so recognize certain features [4]. The research moves towards a general understanding of texts to create a task-independent representation of the meaning similar to the pictures humans create in their minds.

After the NLU algorithms created a representation for all texts, it is possible to analyze them regarding specific patterns. This analysis step provides a subset of documents, which comply with information needs: the retrieved documents [3]. Therefore, the analysis focuses on grouping and potentially ranking documents according to this need. Information needs can either be formulated directly by transforming them into an NLU feature and compared them to the document representations or implicitly by classifying all documents [4]. This text classification determines information needs based on different classes and requires labeled data for each class. In contrast to grouping, all documents according to labels, the direct comparison of NLU features does not depend on labeled data. Instead, the most similar documents compared to information needs are retrieved by clustering them. Clustering gathers similar documents and information needs with a data-driven approach. It does not need a priori knowledge about the partitioning of the document collection [5]. However, the main goal of retrieving documents, which satisfy a specific information need, is shared by both supervised and unsupervised models. The retrieval task determines whether classification or

clustering is applied. The following Figure 1 depicts the fundamental IR process with the described NLU, and analysis steps is given by.

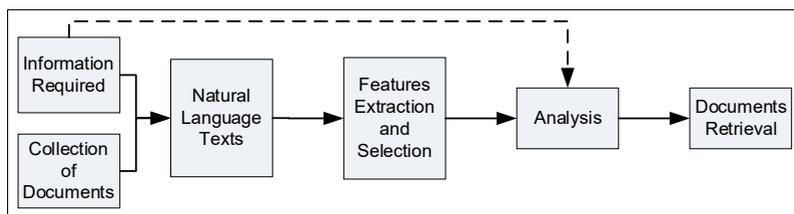


Fig. 1. Schematic diagram of the transformation of text into the feature extraction and selection with NLU

This work aims to structure the recent research efforts by comparing NLU concepts for featuring semantics of text and applying them to IR [6]. Following this goal, the main contribution of this paper is a comparative study of semantic similarity measurements. The focus of this study is to identify useful patterns and heuristics for specific IR tasks with varying datasets.

The rest of the paper is organized in six sections as follows. Section 2 provides an overview of the state-of-art related work that leads to motivation and problem identification in section 3. Section 4 presents the Natural Language Texts Features and their methodologies leading towards the Lexical Approaches in Section 5. Section 6 presented on Transfer-Learning approach and experimental results based on different discussed algorithms in section 7 and the conclusion of this paper are presented in the following section 8.

2. Literature Review. The field of Information Retrieval is a cornerstone of the current information age and has a significant value for corporations and individuals. This value becomes apparent by looking at the success of web search engines, which represent just one application of the whole IR field. The diverse field of IR systems has led to many publications around this topic. The described state-of-the-art in the following sections focuses on the creation of a vector space model to represent the meaning of texts and the related classification and clustering analysis tasks.

2.1. Natural Language Understanding. As described the first step towards finding relevant information understands the document collection and the information needs. The field of NLU addresses this problem for general NLP tasks by extracting features, which represent the meaning [6]. Those features can be extracted either by rule-based systems or with data driven statistical approaches. Similar to other fields in Artificial Intelligence (AI) recent algorithms of NLU focus on statistical deep learning models,

which provide numerical features instead of simple nominal features provided by most rule-based approaches [7].

2.2. Rule-based Approaches. First NLU approaches were published between 1960 and 1975. The examples of those early systems are solving algebra text problems from schoolbooks [8] or answering the question about the position of building blocks in a model world [5] to prove the understanding of text input. Those early systems use rule-based methods to understand language commands and perform actions according to understanding. For example, the block model of the SHRDLU system [9] moves blocks and answers questions based on English keywords. The rules used in NLP systems are described with context-free grammars and checked by parsers. For example, SHRDLU uses systemic functional grammar to process language commands [10]. The idea of rule-based systems is supported by the linguistic concept of generativist. This concept was introduced by Noam Chomsky and claims that general laws and principles govern all NL, therefore NL can be described with a generative grammar [10].

Rule-based systems focus on a specific domain, like a restricted world of colored blocks. In this domain, the possible inputs and outputs can be processed with few rules or simple grammar and the system can fully understand the reduced language input. Improvements in the following years showed that those systems are very difficult to scale to a wider domain or adapt to another domain [11]. Rule-based systems, like SHRDLU [9], require knowledge about the different utterances and how they can be connected. Therefore, such systems grow exponentially with an increasing number of utterances. Nevertheless, rule-based approaches are still considered in NLP applications with specific domains, especially data preparation relies on the simplistic approaches provided by rule-based systems.

2.3. Statistical approaches. The increase in computational power and available language data was responsible for the shift from rule-based methods to data-driven statistical methods in the early 1990s. The statistical revolution also marks the shift from generative grammars to more structural linguistics-based approaches. Structural linguistics focus on probabilities assembled over a large utterance corpus [12]. Instead of trying to find complex rules, which describe the language, statistical methods analyze the occurrence of words and phrases and try to find patterns to understand the text. An example of this analysis task is language models. Those models try to predict a word with the help of given context words. That so-called word embedding is proven to be useful for many NLP tasks. The language models are used as inputs for supervised NLU tasks, like Question Answering (QA) [54], sentiment analysis [13], or Semantic Textual Similarity (STS) [14]. Those

systems use Convolutional Neural Networks (CNNs) or Recurrent Neural Network Training (RNNs) with large training datasets and handcrafted additional features like Part-of-Speech (POS) tags. Other NLP tasks like machine translation, text summarization use similar encoder-decoder architectures to convert the text. Those tasks also require a semantical understanding of the input text; therefore, the encoder is also trained to produce an abstract representation of the meaning [13]. Statistic methods can process the complete corpus of a language. They are not restricted to specific domains with reduced vocabulary [15]. However, statistical methods focus on specific subtasks of NLP, like language models instead of understanding the whole language.

Some research related to the transformer models has been presented [9] to show the overall representation of network layers and they are based on Transformers Architecture. Most of QA models in the field of NLP are encoder-decoder based architectures. QA has applications in a vast array of tasks including information retrieval, entity extraction, chatbots, and dialogue systems to name but a few. While QA can be done in various ways, perhaps the most common flavour of QA is selecting the answer from a given context. In other words, the system will pick a span of text from the context that correctly answers the question. If a correct answer cannot be found from the context, the system will merely return an empty string. BERT [22], XLNET [23], CoQA [24] and QuAC [25] are all commonly used models for QA. Sur et al. [26], proposed a model on google based BERT model by adopting the BERT linear output layer including decoding framework for overall analysis of IR.

2.4. Thesauri-based Approaches. Thesauri-based approaches formulate specific rules based on the a priori relations. The features are nominal, and it is only possible to analyze words within a thesaurus with predefined relations. Naturally, words are nominal [16] and distributional semantics introduces a notion of similarity based on the differences of the word co-occurrences. This is used by distributional models to generate numerical features. Co-occurrences are not able to distinguish word relations and therefore use vague relations. The thesauri-based approaches mostly rely on manual feature engineering. Therefore, it does not use feature learning at all. Subsequently, no additional training data is required. The landscape of distributional features is more divided. Traditional models use unsupervised methods, like Principal Component Analysis (PCA) or Singular Value Decomposition (SVD) to reduce the dimensionality. Word embeddings use language models and thereby are also unsupervised. Finally, typical deep learning approaches, like CNNs or RNNs integrate the feature learning into the classification task [14]. Essentially each layer of those networks

represents features. To learn those features labeled data is required. The features are closely related to the classification task they are trained on and do not provide any theoretical insight. Thesauri-based approaches are not considered, because differentiating between special word relations is not required and distributional models evolved as quasi-standard for the vast majority of NLP tasks [4]. Also, creating a suitable thesaurus based on the specific domain data is time-consuming, costs expensive, and would require an extensive domain and linguistic knowledge [3]. As a consequence, the subsequently presented methods are restricted to distributional semantics.

2.5. Classification. In IR, every search can be seen as a classification problem. The classification separates all documents into a group, which matches an information need, and another group of documents, which does not match the information, need [16]. Consequently, it groups the collection constantly into multiple groups. Each group satisfy one specific information need. This makes classification suitable for common search terms and split a collection into several known topics. This is also the reason why classification in IR is referred to topic classification. Typical IR classification tasks are spam filtering or finding news stories about a certain topic. However, tasks like sentimental analysis can be considered as retrieval tasks as well. Most algorithms use word-level features to classify text. Term frequency extensions are used to create feature invariant to text length and take the different information entropy of words into account. Chali et al. [16] introduced tree-structured RNNs according to dependency graphs to group the sentiment of movie reviews. Aside from word-level representation CNNs, which works for many tasks well on raw signals, are used to classify the collection on character level [10]. This approach could recognize the meaning of affixes.

2.6. Clustering. In IR, clustering is based on the so-called cluster Theoretical hypopaper testing. Aswani et al. [17] describe the hypopaper as follows: "Documents in the same cluster behave similarly concerning relevance to information needs"[17]. Accordingly, if one document of the cluster satisfies an information need, all other documents of the cluster most likely satisfy it as well. This hypopaper can be exploited for many use-cases in IR. One example is applying hierarchical clustering to stepwise refine the information need. At each level, it is possible to select the cluster most suitable for the intended information need. This scatter-gather search avoids the ambiguity of search queries [18]. Another example of clustering is speeding up the search process by dividing the collection into clusters and searching just the nearest clusters to the information need. The used clustering algorithms are determined by the specific, demanded characteristics of the clustering task, therefore the clustering approaches are

highly task specific. Commonly k-means models are used for topic clustering and scatter-gather searches often use single-linkage approaches [18]. However, most applications require specialized clustering algorithms to take unknown cardinalities, partial or fuzzy cluster, and prototype representations of clusters into account.

3. Problems Identification. Section 2 reveals a few limitations of the NLU feature and IR. Based on the following depicted limitations the contributions of this work are going to be explained here.

3.1. Natural Language Texts Feature. The current research on the semantic feature space for IR tasks has two important shortcomings, which will have been addressed by this paper. Firstly, the end-to-end approaches applied by many supervised NLP tasks like STS, sentiment analysis, or QA rely on training data and do not provide an explicit sentence or even text representations. Therefore, unsupervised tasks have to fall back on simpler features [19]. The recently proposed general-purpose representations based on transfer learning tasks could solve this problem, but as of today, those features were not used for IR tasks. Secondly, feature taking the word order into account are trained and used mostly on grammatically correct sentences. In IR, many documents do not contain just continuous text. Bullet points, keywords, and other interruptions of text with a distinct grammatical structure are common. Consequentially the current research does not report results for those noise data.

3.2. Datasets Analysis. Current text classification approaches use the large dataset with a structure aligned to English grammar rules. Therefore, those models lack evaluation capabilities for less strictly structured text with smaller training datasets. Due to the predominantly unsupervised setting, most IR approaches still rely on the simple feature, which is obtained in an unsupervised fashion [20]. Those term-frequency or word embedding models fail to represent the combined meaning of words in a sentence.

4. Natural Language Texts Feature. Natural Language Understanding is a large research field with an even wider range of tasks, each of them with a unique set of requirements and varying definitions of understanding. Translating a text requires another form of understanding than identifying the overall topic of a text or answering questions. Based on those differences each task also requires specific NLU features. This is the reason why many features with divided characteristics exist. To select, evaluate, and even understand those features is useful to categorize them. Distributional models require instead of a priori defined relations between certain words a large text corpus to obtain the implicit relations. The features are nominal, and it is only possible to analyze words within a thesaurus with pre-defined relations. Naturally, words are nominal. Distributional semantics introduces

a notion of similarity based on the differences of the word co-occurrences. This is used by distributional models to generate numerical features. Co-occurrences are not able to distinguish word relations and therefore use vague relations.

Apart from the linguistic categorization one important factor for all features is the required data to learn them. Initially, all methods use one-hot encoded, concatenated vectors, or conceptually similar representation as a feature. The feature learning transforms those simple features into more sophisticated ones with lower dimensionality to avoid overfitting and to reduce training time. The thesauri-based approaches mostly rely on manual feature engineering. Therefore, it does not use feature learning at all. Subsequently, no additional training data is required. The landscape of distributional features is more divided. Traditional models use unsupervised methods, like PCA or SVD to reduce the dimensionality. Word embedding use language models and thereby are also unsupervised. Finally, typical deep learning approaches, like CNNs or RNNs integrate the feature learning into the classification task. Essentially each layer of those networks represents features. To learn those features labeled data is required. The features are closely related to the classification task they are trained on and do not provide any theoretical insight.

4.1. Data Pre-processing. The pre-processing for extracting the NLU feature has two main objectives: Separating the text into features units and cleaning the input data. The separation is arguably required for extracting any feature based on the introduced notion of hierarchical semantics. This tokenization is independent of the actual features. It provides the elements to calculate those features. On the contrary data-cleansing relies on the actual extracted features, because it changes key characteristics of the dataset to increase the generalizability and adapts to the requirements of the feature extraction models.

- **Tokenization:** Splitting the input text yields the parts used to calculate features. The introduced general pipeline to extract NLU features uses words as atomic units and groups them into sentences and finally into a larger utterance. Thereby tokenization must divide the text into words, sentences, and in some cases into larger utterances like paragraphs. Since the separation into paragraphs is not required for the short texts analyzed during this study it is not described in detail. Additionally, paragraphs depend heavily on the used convention, e.g. two-line feeds as a separator. Those conventions lack ambiguity and are consequently easy to separate.
- **Data Cleansing:** One crucial part of all practical machine learning tasks is data quality. Typical real-world data has a low quality due to

missing measurements, faulty measurements or simply inaccurate measurements. These errors introduce unwanted noise. Therefore, data cleansing is required to improve the quality and boost the outcome. Textual data is no exception to this general rule. In addition to data specific cleaning steps, textual data faces commonly a problem with dimensionality. Two properties of the text contribute to this problem. On one side some words occur frequently in almost any text without adding information on the other side some words are so unique that they do not occur in other texts. Both aspects lead to a larger vocabulary and consequently to more dimensions in the initial BoW model. Therefore, the curse of dimensionality is supplementary handled by removing and generalizing specific tokens.

- **Token Removal:** Removing irrelevant or redundant features is frequently used for most ML methods. At this point, tokens are the most basic features of the text. Consequently, irrelevant tokens are removed before more complex features are calculated. The removal of tokens depends on the actual task and is thereby a domain-specific task. The three token types (Numbers, Special Characters, and Stop words) are commonly removed. Regular expressions are used to remove number tokens and tokens containing special characters. Stop words are removed based on an NLTK corpus containing 179 stop words.
- **Token Generalization:** Comparable to the aggregation of data objects it is possible to reduce the distinguished words by grouping them. Tempus, numbers, and other grammatical details are in most cases not important to determine the overall meaning of a paragraph accordingly those lemmas can be transformed into the corresponding lexeme. The recovering of lexemes is called lemmatization. The word usage and the context are analyzed to determine the lexeme. For example, POS tags have to distinguish between meeting as a verb or a noun because the lexeme of the verb is met but the noun should not be changed at all. This kind of morphological analysis is a complex and time-consuming task, therefore stemming is often considered as a heuristic alternative. Stemming only analyses the word itself and tries to reduce the word to its stem. This stem differs from the lexeme. Lemmatization focuses on converting lemmas to lexemes by removing inflections while stemming focuses on grouping the words with a similar meaning and therefore removes also derivational suffixes. Stemming analyses single words with rule-based approaches to determine the stem.

5. Lexical Approaches. The TF-IDF model is used as a baseline due to its frequent usage in topic clustering/classification for many IR tasks. Furthermore, it does not measure any semantic information aside from the information entropy of each occurring word. This conceptual difference to all other models makes TF-IDF suitable to provide conclusions about the advantages or disadvantages of semantic models in general. This baseline model is accompanied by word2Vec [5] and Global Vectors (GloVe) [6] models on the lexical level. Undoubtedly both models involve semantic knowledge within the word representation and use dense vectors. Similar to the TF-IDF approach both models are popular for many NLP tasks, especially ANN methods apply frequently Word2Vec due to the possibility to use backpropagation [5]. Both approaches are considered because the influence of the global optimization used by GloVe and the local context window of word2Vec is not foreseeable. This word embedding also builds the fundament for the other two presented models, because those advanced models use the same lexical representation and expand it by taking compositional semantics into account.

As mentioned earlier, lexical methods do not use the word order of a text. According to this sentence and paragraph representation consider the word representations respectively sentence representations as an unsorted list [9]. Since no additional dependencies must be considered the main task for both steps is normalization. This maps the sets with different cardinalities to a fix-sized vector. Therefore, the models become invariant to text lengths. TF-IDF uses averaging, a particularly often used way to normalize data. This is frequently used for word embedding as well. Cer et al. [8] provide an example of this on the STS task. Consequently, averaging the word representation to calculate the sentence and paragraph representation is also used in this work. The resulting process pipeline for all lexical approaches is depicted in (Figure 2). The sentence and paragraph representations are calculated by averaging the word representation.

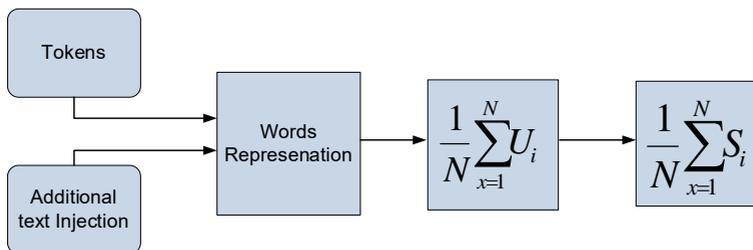


Fig. 2. Process pipeline for lexical approaches

where N denotes the total number of words in the paragraph and U_i is the i^{th} word representation. And S_i averaging the word representation to calculate the sentence.

5.1. Word Representation. Naturally, the word representation is the defining element of pure lexical approaches. The three considered models are default approaches. Those basic concepts are not altered, but to be improved, the representations aspects regarding the vocabulary and training are discussed in the following paragraphs.

- **Vocabulary:** TF-IDF uses a weighted one-hot encoding for words. This requires a fixed vocabulary size otherwise each word would be represented by a vector with a different length. This vocabulary can be built by analyzing the corpus. Including all words creates, on the one hand, the most accurate representation but on the other hand, the representations get sparse. As mentioned, data sparsity is an undesirable property for further analysis [12]. Based on this trade-off the definition of the vocabulary is a crucial part of the word representation for TF-IDF. Considering the most frequent words will decrease the dimensionality and overall, the least number of words are ignored. On the other side frequent words do not carry the same information value, therefore stop words are also removed from the vocabulary. The dimensionality of word embeddings is independent of the vocabulary size. Accordingly, no fixed size vocabulary is required. The trade-off between vocabulary size and vector dimensionality does not exist. Consequently, the removal of stop words or less frequent words is unnecessary. This allows adding new words to the vocabulary. The previous weights are simply appended and trained again.
- **Training:** In opposition to the one-hot representation word embedding's are defined by their occurrence. Therefore, the corpus must contain a sufficient occurrence of each word to define the lexical-semantic of it. A large amount is more likely to fulfill this requirement, but it also contains word ambiguity. Improving the training without introducing more ambiguity has to append the corpus with domain-related data. This work gathers additional data by identifying domain-related Wikipedia articles based on the nouns of the corpus. The extraction utilizes the POS tagger of Natural Language Toolkit (NLTK) to identify those nouns and the Python library to gather articles of those identified nouns.

5.2. Paragraph Representation. The next processing step generates a representation of larger words [8]. Pure lexical approaches normalize the

word representations to generate length invariant features. As mentioned, this normalization step calculates for all three approaches the average value defined as in (equation 1):

$$\vec{w}_d = \frac{1}{n} \sum_{i=0}^n \vec{u}_i \quad (1)$$

where n denotes the total number of words in the paragraph and \vec{u}_i is the i^{th} word representation. This calculation creates especially for lexical-semantic encoded in word embedding's two problems.

By averaging contradicting words, the meaning of each word gets annihilated. Since oxymora are used to emphasize certain aspects, this would not be a desirable result. Due to the un-specific notion of word relations used by word embedding's an oxymoron is just an artificial example for two appropriate reverse vectors. Avoiding this behavior requires a more complex model that considers compositional semantics.

Another aspect that is not accurately modeled by simple averaging word representations is the individual influence of words. Depending on the POS, word position, and the occurrence the meaning of each word has to be weighted differently. For example, nouns and verbs contribute more to the overall meaning than adjectives. Partially the IDF factor applied in TF-IDF addresses this problem. Consequently, the averaging of word embedding could be improved by including similar weights as well [5]. This extension is defined in (equation 2).

$$\vec{w}_d = \frac{1}{n} \sum_{i=0}^n IDF(i, D) \vec{u}_i \quad (2)$$

where D is the word representations.

5.3. Implementation. Two lexical approaches (TF-IDF, GloVe) are implemented from scratch with Python and in the case of GloVe with TensorFlow. The Word2Vec representation is calculated by the Gensim Python library. All paragraph representations are calculated with a trivial NumPy averaging function. Therefore, just the word level implementation is discussed briefly.

- GloVe: The Global Vector (GloVe) model is implemented with TensorFlow. Similar to the one-hot encoding a word is mapped to one index which refers to the actual embedding. The TensorFlow graph receives a matrix of indices for context words, center words, and the corresponding co-occurrence counts as input [5]. The center words behave like context words, on that account they are omitted. The

corresponding embedding²s are generated and randomly initialized with a value between -1 and 1. Based on the index matrix a pointer to the related embedding is generated for each matrix element.

- Par2Vec: One model using the word order is the unsupervised Par2Vec model introduced by Pajak et. al [3]. This method extends the language model of Word2Vec to additionally optimize a paragraph representation. Instead of just using the word embedding²s to predict words for each paragraph a vector is added to the word embedding²s and is subsequently also altered during the optimization. The Par2Vec models assume with this additional vector for each paragraph that the accurate prediction of a word does not only depend on the corresponding context/center words but also the overall topic of this paragraph. Therefore, the vector distinguishes between paragraphs with different topics. Due to the simultaneous calculation, both representations influence on each other. Therefore, the word embedding of this model differs from the simple lexical approaches. Par2Vec considers the word order based on the context windows, but it does not use any syntactical structure other than unordered neighboring words. According to this property, it is controversial whether Par2Vec can be considered a sequential mode as shown in Figure 3.

Similar to the Word2Vec model the Par2Vec representation is implemented by the Gensim Python library. Following the usage of this library is described, which does not differ greatly between Word2Vec and Par2Vec. Gensim renames the Par2Vec model to Doc2Vec, which established it as a synonym for Par2Vec. The Gensim approach is oriented on the Scikit-learn classifier. One classifier object is created with all hyper-parameters, trained and finally, the classifier predicts the results.

Both models (Word2Vec, Par2Vec) require a special input format of tagged words/paragraphs. This pre-processing step creates trainText as a list of named tuples. For each element, the words and unique tags are required to train the embedding. The invert-vector function does not need this tag later on.

6. Transfer-Learning Approach: BiLSTM. The last reviewed feature is the most elaborated model concerning the syntactical structure of a text. Instead of using context windows similar to Par2Vec the presented model includes the full word order and represents thereby a typical sequential model on the sentence level. As such it can enclose the linguistic structure of a sentence, like dependencies to gather compositional semantics. a sentence, like dependencies to gather compositional semantics.

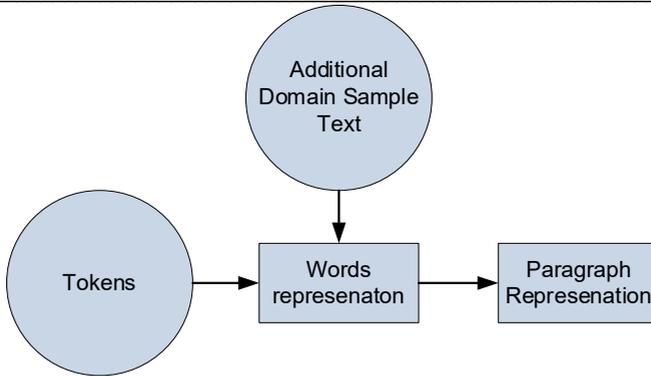


Fig. 3. Process pipeline for the Par2Vec model

This approach extends the lexical model (word2Vec or GloVe) by replacing the word averaging on sentence-level with an RNN which is trained to provide compositional semantics. Likewise, a brief look at the processing pipeline (Figure 4) shows this difference.

The sentence representation or to be more precise the ANN has to be trained to extract compositional semantics. This internal training task must include semantic knowledge. Through the training on such a task, the parameters of the ANN learn to create an accurate sentence representation that captures the sentences meaning. Afterwards, the trained model can generate a representation of all kinds of sentences. The learned knowledge from a specific task is transferred. In opposition to the Par2Vec and word, embedding approaches this training data is not provided by large unlabeled corpora based on language models. Hence the model is supervised [14].

The two key factors of this transfer-learning approach are the model and the initial task. Especially the initial task must rely on semantic knowledge. One example of such a task is the STS task. It provides due to the SemEval workshop sufficient labeled data and benchmark models. Furthermore, the task of classifying sentences due to overall similarity resembles the later contemplated problem of detecting duplicates. Based on the training task the BiLSTM model proposed in [4] for STS is considered as classifier during this work. This BiLSTM model shows on the STS benchmark a reasonably good performance. Supplementary LSTM approaches are commonly used for many NLP tasks due to the variable input length. However, the focus of this work lays on the principal application of transfer-learning, therefore evaluating a great number of different models is out of scope for this paper.

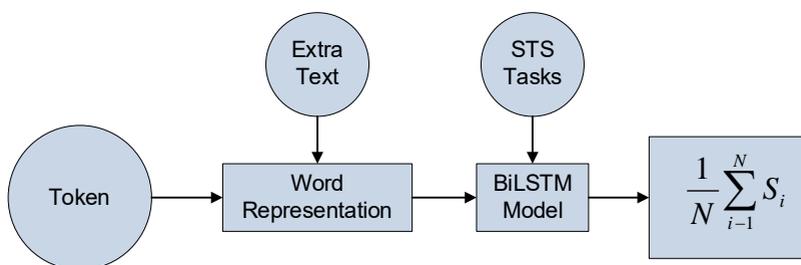


Fig. 4. Process pipeline for the BiLSTM model

6.1. Sentence Representation. The sentence representation is generated by training the weights and biases of a BiLSTM model to predict the semantic similarity between two sentences. The intermediate representation of this training task constitutes an NLU feature for each sentence, which includes compositional semantics and can be applied to tasks with less or none training data. The process of the whole training task as proposed in [4] that is illustrated in (Figure 5). The intermediate representation of the two sentences is compared with a simple feedforward network with one hidden layer. The resulting probability vector is compared to the actual label to calculate the loss and adapt the parameter of the feedforward model and the BiLSTM. Both components are in detail introduced in the following paragraph.

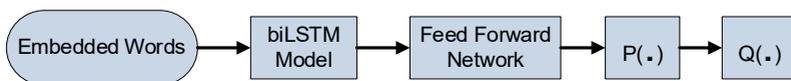


Fig. 5. The network architecture of all components for the STS training task

The input sequence of word embeddings is considered in the forward direction and backward direction to avoid influencing the hidden state excessively by the end or beginning of the sequence. Each direction is fed into one single layer LSTM cell. This constitutes the BiLSTM. The last hidden state of both cells is concatenated and forms the final output of the BiLSTM component.

The feedforward network compares two sentences regarding the similarity of the semantic annotated representations. To avoid training the feedforward network to append the semantic sentence representation, the inputs of this component are two simple difference measurements instead of the actual vectors [16]. The semantic representation should be exclusively

generated by the BiLSTM component because the feedforward network is exchanged later on. Consequently, all linguistic knowledge incorporated into it would be lost. In theory, using a different value as input separates the semantic representation and the similarity classification. Due to the difficult interpretation of ANN transformations, this cannot be proven trivially. Therefore, the sentence vectors calculated by the BiLSTM are compared regarding the absolute difference in each dimension and the angular difference. Both measurements are defined as:

$$\vec{h}_+ = |\vec{h}_L - \vec{h}_R| \quad (3)$$

$$\vec{h}_+ = \vec{h}_L \theta \vec{h}_R \quad (4)$$

where \vec{h}_L respectively \vec{h}_R denotes the sentence representations and θ denotes elementwise multiplication.

According to Tai et al. a combination of both comparisons is empirically proven to be superior compared to either of them alone [13]. Thereby, two hidden layer combines \vec{h}_x and \vec{h}_i by adding both transformed vectors W as formalized in (equation 5).

$$\vec{h}_s = \sigma[W_{h_s} \vec{h}_x + W_{h_i} \vec{h}_i + b_{h_s}] \quad (5)$$

Finally, the output layer calculates the probabilities for each similarity class of the STS task by applying a *softmax* nonlinearity, therefore the result vector is defined as $pin[0, 1]$ and the output layer is shown in (equation 6):

$$p = \text{soft max}[W_{h_s} \vec{h}_s + b_y] \quad (6)$$

Logically the predicted class is defined as $y = [0, 1, 2, 3, \dots, 5] \times p$. By this definition, the probabilities are used as weights to sum the classes and average them. This allows the average label a continuous result instead of a strict classification into the six classes.

The actual loss is calculated between those two probabilities as cross-entropy and optimized by simple gradient descent. Due to many parameters of the model, it tends to overfit, therefore, an L2 regularization term is added to the cost function. This term penalizes large weight values by adding the sum of quadratic weights. Additionally, the BiLSTM applies dropout to introduce noise to avoid overfitting. Merx et al. [4] demonstrated that no

significant improvement for the STS task, but due to the focus on transfer-learning and the reported improvements for the sentimental task dropout regularization is applied in this model.

The BiLSTM approach is implemented like GloVe with TensorFlow. The BiLSTM component is assembled out of standard TensorFlow classes for LSTMs, calculating dropouts and passing a bidirectional sequence to RNN cells. Those abstractions allow it to implement it within a few lines.

The cross-entropy losses are optimized with a standard TensorFlow gradient descent algorithm. The required probability for the actual label is calculated. The two equal comparisons mask the two nearest classes to the actual label and split the floating-point part between those classes.

The detection of duplicates as a binary classification problem can be evaluated by the Receiver Operating Characteristic (ROC) curve. This metric is commonly applied to evaluate a binary classifier. The basics of the ROC are briefly described in the following section. This default model is complemented with a view into the distribution of similarity between duplicates and non-duplicates.

The binary classification can be divided into four possible situations for each tuple:

- True Positive (TP): The text tuple is labeled as duplicate and classified as such.
- False Positive (FP): It is not labeled as duplicate but classified as one.
- True Negative (TN): The tuple is not labeled nor classified as duplicate.
- False Negative (FN): It is labeled as duplicate but classified as non-duplicate.

7. Evaluation and Results. The following section discusses the evaluation details for each NLU feature. All features utilize here established threshold classifier in combination with cosine similarity. The ROC curves adjust the threshold value in 0:01 steps starting at $T = 0$ till $T = 1$. The feedforward network is not used as a classifier due to the small amount of labeled data.

7.1. Pre-processing. The preprocessing steps stemming and stop word removal aim to reduce the sparsity of the word representations. Therefore, just models that directly utilize BoW word representations can be improved by those methods. Consequently, just the TF-IDF representation is evaluated regarding the preprocessing steps. The result of the Bosch dataset is presented in Figure 6. In the following (Figure 6), TF-IDF representation does not show any improvement for stemming and the removal of stop words. The reason can be seen by comparing it to a simple TF model. The stemming does not significantly boost the performance of this model either. Most likely

the vocabulary size is too small to observe any effect of generalization. Additionally, the reported improvements for other English NLP tasks are quite small [8]. In opposition to the modest results regarding stemming the removal of stop words improves the TF model. Stop words are by definition frequent words which means that the TF-IDF model already reduces the influence of those words with the IDF factor. Hence the expected positive influence of the preprocessing steps to reduce the data sparsity cannot be observed due to the small vocabulary and corpus size. The Wikipedia dataset which is even smaller shows similar behavior. To align the TF-IDF representation with the other evaluated NLU features both preprocessing steps are skipped in other comparisons.

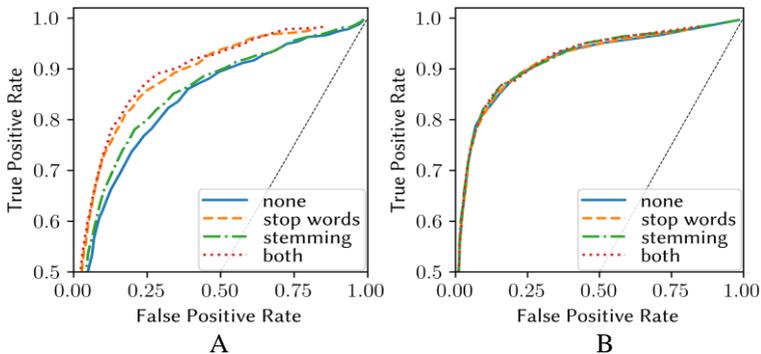


Fig. 6. The result of the Bosch dataset. A - ROC curves for TF and TF-IDF; B - representations with different preprocessing steps on the Bosch dataset. The y-axis is shifted by 0:5

7.2. Training data. One important aspect of word embedding is the training corpus. Naturally, embedding's benefit from a large training corpus, since more text provides more information for the word co-occurrences required to define the actual embeddings. On the other side due to word ambiguity and text-specific word usage, the training dataset has to represent the word distribution of the analyzed text as well. Those two requirements often contradict each other. Therefore, a closer look into the influences of training data for detecting duplicates is conducted on the GloVe embeddings. This evaluation requires domain-specific language. Hence it uses the Bosch dataset. Additionally, the retrained GloVe embedding is already trained on the Wikipedia corpus. Therefore, comparing domain-specific training and pre-trained models contain a bias on the dataset.

The used collection of bug reports for the evaluation is provided by Robert Bosch Engineering and Business Solutions Private Limited. These

bug reports are collected but due to the nature of those bug reports the limited dataset is made publicly available. It contains precise descriptions of sensitive data about Bosch software. These key characteristics of the used dataset are presented in Table 1.

The pre-trained embedding is trained on a large corpus of Wikipedia articles and the fifth Gigaword corpus which contains text from several news agencies. The overall corpus contains up to 6 billion tokens [11]. The original Bosch dataset contains just 118,835 tokens. Also, a small amount of domain-specific data is extracted from suitable Wikipedia articles. This collection contains additional 29,318 tokens. The corresponding results of the threshold classification are illustrated in (Figure 7).

The dataset contains 13,095 bug reports of which 1,515 reports are marked as a duplicate of at least one other ticket of the dataset. Each report contains a short one-sentence summary, a longer description, and a list of duplicate tickets. For the evaluation, the NLU features are extracted solely from the description. Additionally, there are the duplicate lists used to generate a labeled list with two tickets and a Boolean value to identify those tickets either as duplicate tickets or not. Overall, the 1,515 duplicate reports are transformed into 780 duplicate tuples. The small difference between a tuple and a duplicate number shows that a few reports have more than one duplicate. The remaining tuples are considered as non-duplicates. The tuples are subsampled to improve the ration between duplicates and non-duplicates. This is described as part of the data preparation, after taking a closer look into the properties of the actual descriptions.

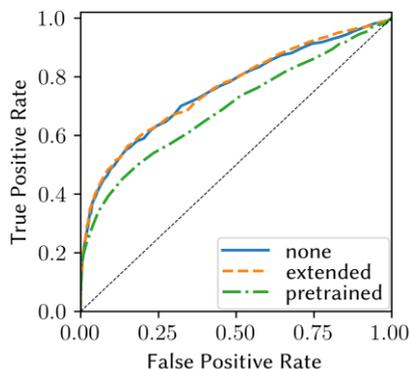


Fig. 7. ROC curves for GloVe embedding trained on different datasets

Table 1. The most important dataset properties

Property	Bosch bug reports	Wikipedia summaries
Total size	13095 reports	200 summaries
Tuple duplication	780	110
No duplication	3900	550
Data origin	Real	Artificial using methods
Structure	Several paragraphs with Headings, bullet points, etc.	One paragraph with no Structural elements
Language	Keywords, English sentences	Plain English sentences
Vocabulary	Specific domain	No specific domain

The Wikipedia dataset contains the summaries of different articles. The summary or lead section of a Wikipedia article is the text before the table of contents. The summaries are gathered from the English, German, French, and Spanish Wikipedia. All the summaries are translated into English using Google Translate except the English texts. This creates several summaries about the same topic which are consequently labeled as duplicates. To avoid large differences, the text size differs at most 15% between the duplicates and is also manually filtered. The Wikipedia dataset contains just 200 summaries of which each summary is a duplicate. Therefore, it contains 110 duplicate tuples and 550 randomly selected non-duplicate tuples. The dataset is smaller than the bug reports but provides more reliable Duplicate labels and it is more homogenous in language and text structure.

The domain-specific training shows a consistently better result than the pre-trained embedding. Taking a closer look, it indicates two aspects that contribute to this result. Firstly, the domain-specific training captures the actual meaning even with fewer data more accurately. Secondly, from the 11,309 different words in all Bosch tickets, just 8,0085 are also words in the pre-trained dataset. Over 3,000 words remain randomly initialized for the pre-trained embedding. However, the word embedding trained on the extended dataset shows a slight improvement over the plain dataset. The additional domain data just contains around 30,000 tokens. Consequently, the expected improvement is also small. The observed behavior is in line with the expectation and hints that a larger amount of domain data improves the model.

7.3. Comparison of Lexical Methods. Finally, the three introduced word representations are directly compared. The results for both datasets are illustrated (Figure 8). All features utilized to established threshold classifiers are in combination with cosine similarity. The ROC curves adjust the threshold value in 0.01 steps starting at $T = 0$ till $T = 1$. For the evaluating data, the list of all lexical, Par2Vec and BiLSTM hyper-parameters except for learning parameter is shown in following Tables 2-4 given by respectively:

Table 2. List of all lexical hyper-parameters except for the learning parameter

Model	Parameters	Values
TF-IDF	vocabulary size	9000
Word2Vec	vector size	400
	context window length	5
	subsampling	0.01
	negative samples	5
GloVe	vector size	400
	context window length	5
	co-occurrence maximum	100
	Scaling factor	0.80

Table 3. List of Par2Vec hyper-parameters except for the learning parameter

Model	Parameters	Values
Distributed BoW	vector size	400
	context window length	5
	subsampling	0.01
	negative samples	5
GloVe	vector size	400
	context window length	5
	subsampling	0.001
	negative samples	5

Table 4. List of Par2Vec hyper-parameters with the exception of the learning parameter

Sub-Model	Parameters	Values
Word embedding	GloVe	
BiLSTM	output size	100
	maximum sentence length	60

Продолжение таблицы 1

Feed Forward Network	hidden layers size	40
Regularization	strength λ	10^{-4}
	keep probability dropout	0.05

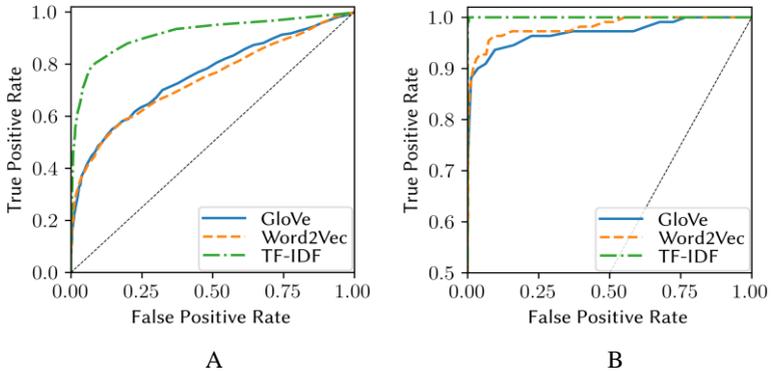


Fig. 8. ROC curves for the mean lexical models. A - the results of the Bosch dataset; B - the results for the Wikipedia dataset. The y-axis of B is shifted by 0.5

The divergence between embeddings and baseline is almost an order of magnitude greater for the Bosch dataset than for the Wikipedia collection. One key difference between the datasets on the level of lexical semantics is the heterogenic style of the Bosch dataset. The Bosch histogram shows in (Figure 9) for non-duplicates a slightly bigger variation.

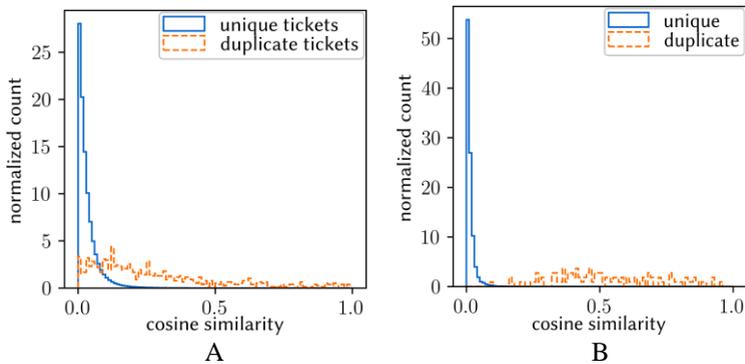


Fig. 9. Histogram of TF-IDF: A - Bosch dataset; B - Wikipedia dataset

Instead of full sentences bullet points and keywords occur frequently in the text. In opposition to this, the Wikipedia dataset only consists of complete English sentences. Possibly this lack of structure deteriorates the performance of the embeddings. The TF-IDF features are even able to classify the Wikipedia dataset almost perfectly. For $T = 0.1$ the recall is 0.99 with a precision of 0.98. Naturally the classification of the Wikipedia dataset is easier, because of the lack of an overall topic between all documents like the software application by bug reports. The low threshold indicates that most of duplicates of the Wikipedia collection are not similar at all. Due to the high variance in both duplicate groups the prediction accuracy decreases significantly based on this difference.

7.4. Par2Vec. The first NLU feature taking compositional semantics into account is Par2Vec. The ROC curve of those unsupervised features is depicted in Figure 10.

The Figure 10 shows for both models of Par2Vec that are an improvement compared to the simple averaging of the lexical approaches. They are on a similar level than the baseline TF-IDF although both features are still slightly worse. Interestingly the simpler distributed BoW model for the Bosch dataset is even better than the more complex distributed memory feature. This emphasizes the text structure again. The distributed BoW feature does not use a window to define context words. The representation is optimized to predict all the words of one paragraph. Consequently, the feature is more suitable for the keyword-oriented language of the Bosch dataset. This also indicates word embedding with larger context windows is more robust against ungrammatical sentences.

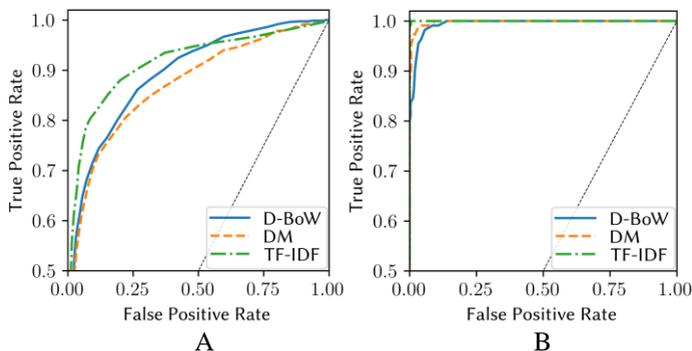


Fig. 10. ROC curves for the distributed BoW (D-BoW) and the Distributed Memory (DM) approach of the Par2Vec model: A - the results of the Bosch dataset; B - the results for the Wikipedia dataset. The y-axis is shifted by 0:5

The ROC curve of the Wikipedia dataset does not provide any information about the models due to the principal simple classification task. But the similarity histogram allows based on the variance which is the variance of unique documents from the distributed memory model bigger and the duplicate variance smaller as shown in the following Figure 11. This Histogram result confirms that the slightly better ROC curve as well from the previous method.

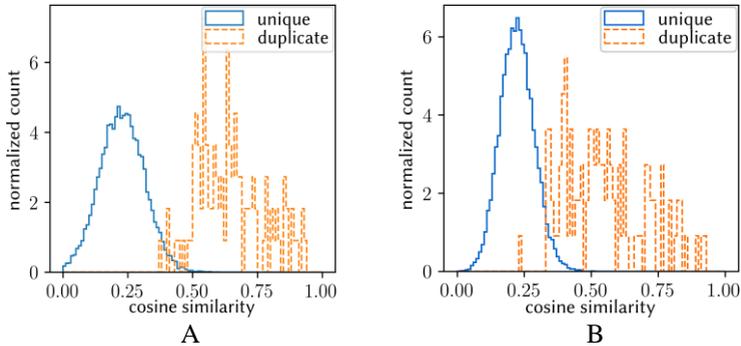


Fig 11. Histogram of the distributed: A – memory; B - BoW model

7.5. BiLSTM. The last evaluated feature is the transfer-learning approach. The resulting ROC curves are illustrated in (Figure 12). The BiLSTM features show overall the worst performance on the Bosch dataset and improve the mean word-embedding model insignificantly on the Wikipedia dataset. The sentence representations are learned with STS sentences. The text structure of the Bosch dataset does not resemble those sentences. Logically all possible learned linguistic rules cannot be applied. This example shows the obvious limitations of transfer learning in a completely unsupervised setting. However, the performance on the dataset collection, which has a similar sentence structure of the STS training task, shows that the feature learned just to classify the STS sentences. The model can generalize the learned representation to solve other tasks. However, the small difference between both datasets shows that at least a few linguistic rules have been learned. This indication that BiLSTM model with more training data could extract a larger amount of generalized linguistic rules.

After evaluating the ROC curve for the BiLSTM model, the next step is to evaluate BiLSTM features concerning overall performance on the Bosch and Wikipedia datasets and calculate the error rate for the wrong training of data sets. To create the confusion matrix, test highlights datasets information is displayed in Figure 13, that possess the data about analysis amid anticipated

and fixated group classes in both datasets. Figure 13 presented the confusion matrix for 3 procedural phases of training and testing separately.

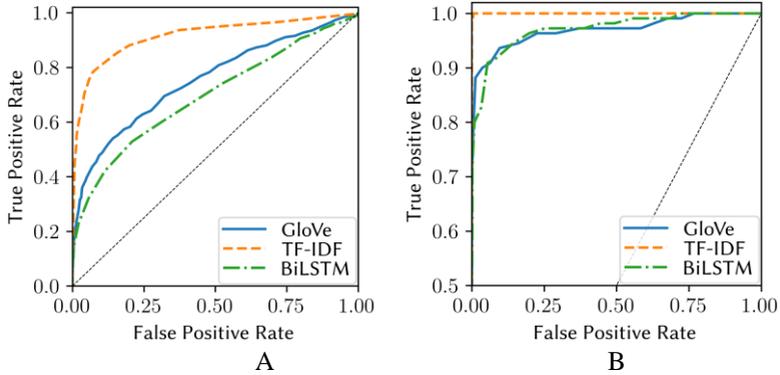


Fig. 12. ROC curves for the BiLSTM model: A - the results of the Bosch dataset; B - the results for the Wikipedia dataset. The y-axis of (b) is shifted by 0.5

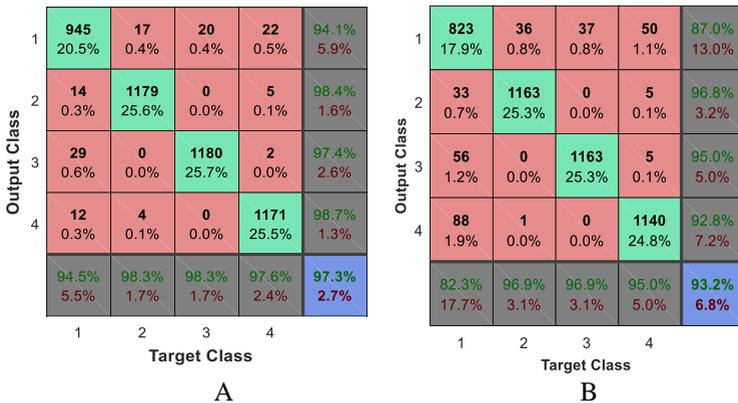


Fig. 13. Performance confusion matrix for the BiLSTM model: A - Bosch datasets; B - Wikipedia datasets

As an anticipated class, it shows all sample feature classification sets. As per the Figure 13, the green cells show the successfully classified datasets with patterns and red cells show the unsuccessful classified datasets, and grey cells show the cumulative result of result row and column wise. Finally, blue cell shows the accuracy parentage and error rate that are found during training, test, and classification phases.

Referring to Table 1, the blue cell indicates the general ratio of tried cases that were characterized accurately in green and the other way around in red which can learn the structure of a sentence to improve the classification with this knowledge. It also shows that BiLSTM model with more training data could extract a larger amount of generalized linguistic rules with less error rate compared with other methods.

8. Conclusion and Future Work. This paper focused on comparing NLU concepts for featuring semantics of text and applying it to IR. Following this goal, the main contribution of this paper is a comparative study for semantic similarity measurements to identify useful patterns and heuristics for specific IR tasks with varying datasets. This paper compared NLU features for the usage in IR systems. Therefore, several features, which encounter semantic knowledge, were identified, categorized, and described. Those features were evaluated by detecting duplicates in two datasets with different text structures. Due to direct comparison between documents the detection of duplicates is a perfect example for the grouping and order commonly used in IR. Lexical features and Par2Vec were not able to outperform the TF-IDF baseline. Additionally, both models were influenced by the text structure. For well-defined English sentences, both models performed well. The additional compositional semantic incorporated in Par2Vec also improved the performance. The transfer-learning feature was not able to improve the classification compared to word embedding approaches on documents consisting of English sentences and performed even worse on documents with a mixed text structure. This indicates that the BiLSTM can learn the structure of a sentence to improve the classification with this knowledge. BiLSTM model with more training data could extract a larger amount of generalized linguistic rules. In summary, the results demonstrate the TF-IDF feature, an impressive result on both datasets with reasonable vocabulary size. The evaluation results show that further work is necessary to improve NLU features for the field of IR to achieve better performances than the TF-IDF feature.

The most obvious further work is gathering of more data which is labeled as duplicates. More labeled data allows better classification models like the drafted feedforward network. Additionally, it will be possible to optimize the hyper-parameter of all models. Due to the skew groups, the labeling of randomly picked tuples out of a collection is difficult. Most likely the creation of a new document as a duplicate is easier. However, labeled data is the bottleneck for this concrete detection task and is not going to improve NLU features in general. Furthermore, other methods like Bidirectional Encoder Representations from Transformers (BERT) and Embeddings from

Language Model (ELMo), etc. related algorithms can also be utilized to improve the classification with given sentences knowledge.

References

1. Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. Task Effects on Linguistic Complexity and Accuracy: A Large-Scale Learner Corpus Analysis Employing Natural Language Processing Techniques. *Language Learning*, 2017. 67(S1), pp. 180–208.
2. Keersmaekers, A. Creating a richly annotated corpus of papyrological Greek: The possibilities of natural language processing approaches to a highly inflected historical language. *Digital Scholarship in The Humanities*, 2019.
3. Pajak, B., Fine, A., Kleinschmidt, D., & Jaeger, T. Learning Additional Languages as Hierarchical Probabilistic Inference: Insights from First Language Processing. *Language Learning*, 2016. 66(4), pp. 900-944.
4. Merx, D., & Frank, S. Learning semantic sentence representations from a visually grounded language without lexical knowledge. *Natural Language Engineering*, 2019. 25(4), pp. 451–466.
5. Huang, F., Ahuja, A., Downey, D., Yang, Y., Guo, Y., & Yates, A. Learning Representations for Weakly Supervised Natural Language Processing Tasks. *Computational Linguistics*, 2014. 40(1), pp. 85-120.
6. Kozachok, A. V., Kopylov, S. A., Meshcheryakov, R. V., Evsutin, O. O., & Tuan, L. M. An approach to a robust watermark extraction from images containing text. *SPIRAS Proceedings*, 2018. 5(60), 128 p.
7. Nazari, P., Khorram, E., & Tarzanagh, D. Adaptive online distributed optimization in dynamic environments. *Optimization Methods and Software*, 2019. pp. 1–25.
8. Altaf, S., Waseem, M., & Kazmi, L. IDCUP Algorithm to Classifying Arbitrary Shapes and Densities for Center-based Clustering Performance Analysis. *Interdisciplinary Journal of Information, Knowledge, And Management*, 2020. 15, pp. 091–108.
9. Chen, R., Dai, R., & Wang, M. Transcription Factor Bound Regions Prediction: Word2Vec Technique with Convolutional Neural Network. *Journal Of Intelligent Learning Systems and Applications*, 2020. 12(01), pp. 1–13.
10. Mitra, B., & Craswell, N. An Introduction to Neural Information Retrieval t. *Foundations And Trends In Information Retrieval*, 2018. 13(1), pp. 1-126.
11. Savyanavar, P., & Mehta, B. Multi-Document Summarization Using TF-IDF Algorithm. *International Journal of Engineering and Computer Science*. 2016.
12. Liang, P. Learning executable semantic parsers for natural language understanding. *Communications Of the ACM*, 2016. 59(9), pp. 68–76.
13. Berant, J., & Liang, P. Imitation Learning of Agenda-based Semantic Parsers. *Transactions of the Association for Computational Linguistics*, 2015. 3, pp. 545–558.
14. Merx, D., & Frank, S. Learning semantic sentence representations from a visually grounded language without lexical knowledge. *Natural Language Engineering*, 2019. 25(4), pp. 451-466.
15. Roberts, L. Individual Differences in Second Language Sentence Processing. *Language Learning*, 2012. 62, pp. 172–188.
16. Dontsov, D. O. Algorithm of thesaurus extension generation for enterprise search. *SPIRAS Proceedings*, 2014. 7(30), 189.
17. Aswani Kumar, C., Radvansky, M., & Annapurna, J. Analysis of a Vector Space Model, Latent Semantic Indexing and Formal Concept Analysis for Information Retrieval. *Cybernetics And Information Technologies*, 2012. 12(1), pp. 34–48.
18. Ch, A. Latent Semantic Indexing based Intelligent Information Retrieval System for Digital Libraries. *Journal Of Computing and Information Technology*. 2006.

19. Susanto, G., & Purwanto, H. Information Retrieval Menggunakan Latent Semantic Indexing Pada Ebook. *SMATIKA JURNAL*, 2018. 8(02), pp. 74–79.
20. Blynova, N. Latent semantic indexing (LSI) and its impact on copywriting. *Communications And Communicative Technologies*, 2019. (19), pp. 4–12.
21. Rataj, Karolina. “Electrophysiology of Semantic Violations and Lexical Ambiguity Resolution in Bilingual Sentence Processing.” *Bilingual Lexical Ambiguity Resolution*, 2020. pp. 250–72.
22. Qu, C., Yang, L., Qiu, M., Croft, W. B., Zhang, Y., & Iyyer, M. BERT with History Answer Embedding for Conversational Question Answering. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2019.
23. Ghavidel, H., Zouaq, A., & Desmarais, M. Using BERT and XLNET for the Automatic Short Answer Grading Task. *Proceedings of the 12th International Conference on Computer Supported Education*. 2020.
24. Choi, E., He, H., Iyyer, M., Yatskar, M., Yih, W., Choi, Y., . . . Zettlemoyer, L. QuAC: Question Answering in Context. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018.
25. Reddy, S., Chen, D., & Manning, C. D. CoQA: A Conversational Question Answering Challenge. *Transactions of the Association for Computational Linguistics*, 7, 2019. pp. 249–266.
26. Sur, C. RBN: Enhancement in language attribute prediction using global representation of natural language transfer learning technology like Google BERT. *SN Applied Sciences*, 2019. 2(1).

Saud Altaf — Ph. D., Pir Mehr Ali Shah Arid Agriculture University. Research interests: the area of wireless sensor networks, artificial intelligence, HCI, and visible light communication for research, teaching, and learning. The number of publications — 35. SAUD@uuar.edu.pk, www.uuar.edu.pk; Office 21, UIIT, Main Murree Road, Rawalpindi, Pakistan; office phone: +923130009013.

Sofia Iqbal — MPhil., Pakistan Space and Upper Atmosphere Research Commission (SUPARCO), Pakistan. SOFIAIQBAL.SUPARCO@gmail.com; www.suparco.gov.pk; Sector-H, DHA Phase II, Islamabad, Pakistan; office phone: +923213399831.

Muhammad Waseem Soomro — Professional Engineer and Member of Engineering New Zealand. The number of publications — 22. MWASEEM@manukau.ac.nz, www.manukau.ac.nz; Gate 1, L Block Main Reception, Newbury Street, Otara MIT Manukau: Ground floor, MIT Manukau, Auckland, New Zealand; office phone: +64221323717.

С. АЛЬТАФ, С. ИКБАЛ, М. СООМРО
**ЭФФЕКТИВНЫЙ АЛГОРИТМ КЛАССИФИКАЦИИ
ЕСТЕСТВЕННОГО ЯЗЫКА ДЛЯ ОБНАРУЖЕНИЯ
ПОВТОРЯЮЩИХСЯ НЕКОНТРОЛИРУЕМЫХ ПРИЗНАКОВ**

Альтаф С., Икбал С., Соомро М. Эффективный алгоритм классификации естественного языка для обнаружения повторяющихся неконтролируемых признаков

Аннотация. Эта статья фокусируется на том, чтобы уловить смысл значения текстовых функций понимания естественного языка (NLU) для обнаружения дубликатов неконтролируемых признаков. Особенности NLU сравниваются с лексическими подходами для доказательства подходящей методики классификации. Подход трансфертного обучения используется для обучения извлечению признаков в задаче семантического текстового сходства (STS). Все функции оцениваются с помощью двух типов наборов данных, которые принадлежат отчетам об ошибках Bosch и статьям Википедии. Цель данного исследования – структурировать последние исследовательские усилия путем сравнения концепций NLU для описания семантики текста и применения их к IR. Основным вкладом данной работы является сравнительное исследование измерений семантического сходства. Экспериментальные результаты демонстрируют результаты функции Term Frequency – Inverse Document Frequency (TF-IDF) для обоих наборов данных с разумным объемом словаря. Это указывает на то, что двунаправленная долговременная кратковременная память (BiLSTM) может изучать структуру предложения для улучшения классификации.

Ключевые слова: кластеризация, информационный поиск, функция TF-IDF, Par2Vec, тексты на естественном языке, лексические подходы

Сауд Альтаф — Ph. D., Сельскохозяйственный университет Пир Мехр Али Шах Арид. Область научных интересов: беспроводные сенсорные сети, искусственный интеллект, человеко-машинное взаимодействие и связи видимого света. Число научных публикаций – 35. SAUD@uaar.edu.pk, www.uaar.edu.pk; офис 21, Равалпинди, Пакистан; р.т.: +923130009013.

София Икбал — M. Phil, Пакистанская Комиссия по исследованию космоса и верхних слоев атмосферы (SUPARCO), Пакистан. Область научных интересов: беспроводные сенсорные сети, искусственный

интеллект, человеко-машинное взаимодействие и связи видимого света. sofiaiqbal.suparco@gmail.com; www.suparco.gov.pk; Сектор-Н, фаза II ДНА, Исламабад, Пакистан; р.т.: +923213399831.

Мухаммад Соомро — член инженерного общества Новой Зеландии. Область научных интересов: беспроводные сенсорные сети, искусственный интеллект, человеко-машинное взаимодействие и связи видимого света. Число научных публикаций — 22. MWASEEM@manukau.ac.nz, www.manukau.ac.nz; выход 1, главная стойка регистрации блока L, Ньюбери-стрит, Оклендский технологический университет, Окленд, Новая Зеландия; р.т.: +64221323717.

Литература

1. *Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D.* Task Effects on Linguistic Complexity and Accuracy: A Large-Scale Learner Corpus Analysis Employing Natural Language Processing Techniques. *Language Learning*, 2017. 67(S1), pp. 180–208.
2. *Keersmaekers, A.* Creating a richly annotated corpus of papyrological Greek: The possibilities of natural language processing approaches to a highly inflected historical language. *Digital Scholarship in The Humanities*. 2019.
3. *Pajak, B., Fine, A., Kleinschmidt, D., & Jaeger, T.* Learning Additional Languages as Hierarchical Probabilistic Inference: Insights from First Language Processing. *Language Learning*, 2016. 66(4), pp. 900–944.
4. *Merkx, D., & Frank, S.* Learning semantic sentence representations from a visually grounded language without lexical knowledge. *Natural Language Engineering*, 2019. 25(4), pp. 451–466.
5. *Huang, F., Ahuja, A., Downey, D., Yang, Y., Guo, Y., & Yates, A.* Learning Representations for Weakly Supervised Natural Language Processing Tasks. *Computational Linguistics*, 2014. 40(1), pp. 85–120.
6. *Kozachok, A. V., Kopylov, S. A., Meshcheryakov, R. V., Evsutin, O. O., & Tuan, L. M.* An approach to a robust watermark extraction from images containing text. *SPIIRAS Proceedings*, 2018. 5(60), 128 p.
7. *Nazari, P., Khorram, E., & Tarzanagh, D.* Adaptive online distributed optimization in dynamic environments. *Optimization Methods and Software*, 2019. pp. 1–25.
8. *Altaf, S., Waseem, M., & Kazmi, L.* IDCUP Algorithm to Classifying Arbitrary Shapes and Densities for Center-based Clustering Performance Analysis. *Interdisciplinary Journal of Information, Knowledge, And Management*, 2020. 15, pp. 091–108.
9. *Chen, R., Dai, R., & Wang, M.* Transcription Factor Bound Regions Prediction: Word2Vec Technique with Convolutional Neural Network. *Journal Of Intelligent Learning Systems and Applications*, 2020. 12(01), pp. 1–13.
10. *Mitra, B., & Craswell, N.* An Introduction to Neural Information Retrieval t. *Foundations And Trends In Information Retrieval*, 2018. 13(1), pp. 1–126.
11. *Savyanavar, P., & Mehta, B.* Multi-Document Summarization Using TF-IDF Algorithm. *International Journal of Engineering and Computer Science*. 2016.
12. *Liang, P.* Learning executable semantic parsers for natural language understanding. *Communications Of the ACM*, 2016. 59(9), pp. 68–76.
13. *Berant, J., & Liang, P.* Imitation Learning of Agenda-based Semantic Parsers. *Transactions of the Association for Computational Linguistics*, 2015. 3, pp. 545–558.

14. *Merkx, D., & Frank, S.* Learning semantic sentence representations from a visually grounded language without lexical knowledge. *Natural Language Engineering*, 2019. 25(4), pp. 451-466.
15. *Roberts, L.* Individual Differences in Second Language Sentence Processing. *Language Learning*, 2012. 62, pp. 172-188.
16. *Dontsov, D. O.* Algorithm of thesaurus extension generation for enterprise search. *SPIIRAS Proceedings*, 2014. 7(30), 189.
17. *Aswani Kumar, C., Radvansky, M., & Annapurna, J.* Analysis of a Vector Space Model, Latent Semantic Indexing and Formal Concept Analysis for Information Retrieval. *Cybernetics And Information Technologies*, 2012. 12(1), pp. 34-48.
18. *Ch, A.* Latent Semantic Indexing based Intelligent Information Retrieval System for Digital Libraries. *Journal Of Computing and Information Technology*. 2006.
19. *Susanto, G., & Purwanto, H.* Information Retrieval Menggunakan Latent Semantic Indexing Pada Ebook. *SMATIKA JURNAL*, 2018. 8(02), pp. 74-79.
20. *Blynova, N.* Latent semantic indexing (LSI) and its impact on copywriting. *Communications And Communicative Technologies*, 2019. (19), pp. 4-12.
21. *Rataj, Karolina.* Electrophysiology of Semantic Violations and Lexical Ambiguity Resolution in Bilingual Sentence Processing. *Bilingual Lexical Ambiguity Resolution*, 2020. pp. 250-72.
22. *Qu, C., Yang, L., Qiu, M., Croft, W. B., Zhang, Y., & Iyyer, M.* BERT with History Answer Embedding for Conversational Question Answering. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2019.
23. *Ghavidel, H., Zouaq, A., & Desmarais, M.* Using BERT and XLNET for the Automatic Short Answer Grading Task. *Proceedings of the 12th International Conference on Computer Supported Education*. 2020.
24. *Choi, E., He, H., Iyyer, M., Yatskar, M., Yih, W., Choi, Y., Zettlemoyer, L.* QuAC: Question Answering in Context. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018.
25. *Reddy, S., Chen, D., & Manning, C. D.* CoQA: A Conversational Question Answering Challenge. *Transactions of the Association for Computational Linguistics*, 7, 2019. pp. 249-266.
26. *Sur, C.* RBN: Enhancement in language attribute prediction using global representation of natural language transfer learning technology like Google BERT. *SN Applied Sciences*, 2019. 2(1).

Кротов К. В., Скатков А.В.
**ПОСТРОЕНИЕ КОМПЛЕКСНЫХ РАСПИСАНИЙ
ВЫПОЛНЕНИЯ ПАКЕТОВ ЗАДАНИЙ ПРИ ФОРМИРОВАНИИ
КОМПЛЕКТОВ В ЗАДАНИЕ ДИРЕКТИВНЫЕ
СРОКИ**

Кротов К. В., Скатков А.В. Построение комплексных расписаний выполнения пакетов заданий при формировании комплектов в заданные директивные сроки

Аннотация. Современное состояние с решением задачи комплексного планирования выполнения пакетов заданий в многостадийных системах характеризуется отсутствием универсальных способов формирования решений по составам пакетов, наличием ограничений на размерность задачи и невозможностью гарантированного получения эффективных решений при различных значениях ее входных параметров, а также невозможностью учета условия формирования комплектов из результатов. В статье авторами реализовано решение задачи планирования выполнения пакетов заданий в многостадийных системах при формировании комплектов результатов в заданные для них директивные сроки. Для решения задачи планирования выполнена декомпозиция обобщенной функции системы на совокупность иерархически взаимосвязанных подфункций. Применение декомпозиции позволило использовать иерархический подход для планирования выполнения пакетов заданий в многостадийных системах, предусматривающий определение решений по составам пакетов на верхнем уровне иерархии и построение расписаний выполнения пакетов на нижнем уровне иерархии. Для оптимизации решений по составам пакетов и расписаниям их выполнения в соответствии с иерархическим подходом применена теория иерархических игр. Построена математическая модель иерархической игры для определения эффективных составов пакетов заданий и расписаний их выполнения, представляющая собой систему критериев на уровнях принятия решений. Оценка эффективности решений по составам пакетов на верхнем уровне иерархии обеспечивается распределением результатов выполнения заданий по комплектам в соответствии с сформированным расписанием. Для оценки эффективности решений по составам пакетов сформулирован метод упорядочивания идентификаторов типов комплектов с учетом директивных сроков и метод распределения результатов выполнения заданий по комплектам, реализующий вычисление моментов времени окончания формирования комплектов и запаздываний с их формированием относительно заданных директивных сроков. Выполнены исследования планирования процесса выполнения пакетов заданий в многостадийных системах при условии формирования комплектов в заданные директивные сроки. На их основе сформулированы выводы, касающиеся зависимости эффективности планирования от входных параметров задачи.

Ключевые слова: многостадийная система, пакеты заданий, расписание, комплекты результатов, директивные сроки, иерархическая игра

1. Введение. В производственных и вычислительных системах конвейерного типа (в многостадийных системах (МС)) возникают задачи планирования выполнения пакетов заданий (ПЗ). Для их решения используется иерархический подход [1]. При планировании заданными являются наборы однотипных заданий (однотипными

являются задания, длительности выполнения которых на приборах МС одинаковы, длительности переналадок приборов на выполнение этих заданий также одинаковы). При значительных неоднородностях длительностей выполнения заданий и неоднородностях длительностей переналадок приборов на выполнение заданий действия с ними в МС реализуются в составе пакетов. Пакетом является набор однотипных заданий, выполняемых на приборах МС без их переналадки на выполнение заданий других типов. Количество заданий в пакетах определяется с учетом временных характеристик процесса их выполнения в МС. Если составы пакетов не оптимизируются и в пакеты входят все задания из их наборов, то такие пакеты называются фиксированными. Оптимизация составов ПЗ, реализуемая с учетом временных параметров их выполнения, позволяет повысить эффективность планирования реализации действий с ними на приборах МС.

Решение задачи планирования выполнения ПЗ в МС предполагает определение составов пакетов и расписаний их выполнения на приборах МС [1]. Определение этих решений реализуется на иерархически взаимосвязанных уровнях системы планирования [1]. Развитием указанной задачи [1] является задача планирования выполнения ПЗ в МС при формировании комплектов результатов (КР) (в частности, с условием задания для КР директивных сроков окончания их формирования). Под комплектом подразумевается набор заданного состава результатов выполнения в МС заданий разных типов.

Рассмотренный в предлагаемой работе подход применим, в частности, к планированию выполнения ПЗ в МС при обработке данных дистанционного зондирования Земли (ДЗЗ) в Web-ориентированных сервисах мониторинга окружающей среды (ОС). Эти сервисы реализуют идентификацию наличия на земной поверхности различных видов негативных природных явлений (НПЯ), техногенных воздействий (ТВ) на ОС, а также (для областей заданных размеров) условий их развития и распространения. Наборами данных ДЗЗ разных типов являются совокупности массивов значений отражательной способности подстилающей поверхности, полученные из разных каналов различных измерительных приборов искусственных спутников Земли (на основе этих наборов массивов значений формируются пакеты данных, обрабатываемые в МС). Для каждого типа НПЯ и ТВ на ОС определен набор параметров (характеризующих как состояние подстилающей поверхности, так и растительности на ней), соответствующих условиям их распространения (развития). Для

каждого из этих параметров формируются в результате обработки данных ДЗЗ в МС массивы значений. В итоге идентифицированного на поверхности Земли НПЯ и ТВ на ОС формируется соответствующий ему набор массивов значений параметров, характеризующих условия его распространения (развития). Тогда комплектами являются наборы результатов обработки данных ДЗЗ в МС (массивов значений), соответствующие условиям распространения (развития) НПЯ и ТВ на ОС различных типов, идентифицированных на земной поверхности.

С целью разработки методов оптимизации решений при комплексном планировании выполнения ПЗ в МС при формировании КР выполнена декомпозиция функции системы, которая позволила определить набор иерархически взаимосвязанных подфункций. Каждая из подфункций реализуется на соответствующем уровне иерархии системы планирования и обеспечивает определение следующих решений:

- первый уровень – решений по составам ПЗ;
- второй уровень – расписаний выполнения ПЗ, решение по составам которых сформировано на первом уровне.

Планирование выполнения ПЗ в МС с учетом формирования КР предполагает также решение задачи распределения результатов выполнения заданий, входящих в пакеты, по комплектам для определения значений критерия оптимальности решений по их (пакетов) составам. Реализация распределения результатов выполнения ПЗ по комплектам сопровождается определением моментов времени их (комплектов) формирования и значений запаздываний с их формированием относительно директивных сроков. Определение решений по составам ПЗ и расписаниям их выполнения с использованием иерархического подхода связано с применением теории иерархических игр.

Методы комплексного планирования выполнения ПЗ в многостадийных системах, предложенные зарубежными авторами, рассмотрены в работах [2–20]. В работах [2–8] рассмотрены методы планирования выполнения ПЗ в потоковых системах, реализующих непрерывные процессы в химической и фармацевтической промышленности. Под выполнением ПЗ в этом случае подразумевается обработка в системе партий материалов разных типов для получения требуемой продукции. Обрабатывающая система состоит из различных приборов, с использованием которых формируются маршруты движения материалов, определены объемы партий материалов, а также порядок выпуска продукции. Количество продукции, которое должно быть получено, определяется спросом. Маршруты движения партий

определяются в процессе оптимизации. За счет оптимизации маршрутов реализуется совмещение выпуска продукции различных видов. Планирование реализации операций с партиями предполагает определения решений, состоящих из [2–4]: бинарных переменных (партия материалов i -го типа обрабатывается во временном интервале t , обработка партий материалов i -го типа начинается в момент времени n и заканчивается в момент времени n' и т.д.) и переменных, соответствующих размерам партий (размер партии материалов i -го типа, обработка которой начинается во временном интервале t на j -ом элементе некоторого ресурса r , размер партии материалов i -го типа, обработка которой активизируется в момент времени n , размер партии материалов i -го типа, которая обрабатывается в интервала времени t). Определение решений предполагает применение аппарата смешанного целочисленного программирования (СЦЧП). Однако с его использованием определяются решения задач только ограниченной размерности [2, 4].

Аналогичный подход к решению задачи планирования в непрерывном производстве изложен в работах [5, 6]. Решение формируется в виде бинарных переменных, соответствующих назначению для выполнения партий материалов i -ых типов на j -е ресурсы, и переменных, соответствующих размерам партий обрабатываемых материалов. Особенностью рассматриваемых в [5, 6] задач является учет неудовлетворенного спроса на продукцию и введение критерия, учитывающего штрафы. Развитием работ [2–6] является работа [7], в которой рассматривается метод планирования обработки партий материалов в непрерывном производстве, учитывающий неопределенности во входных данных (в длительностях обработки партий). Задача переформулируется как детерминированная и для ее решения применена многошаговая адаптивная робастная оптимизация. В силу того, что задача в рассматриваемой постановке является NP-трудной, реализовано ее упрощение, предполагающее применение к временным переменным задачи аффинных правил, позволяющих при решении интерпретировать их как параметры. Это обеспечивает возможность определения решения задачи за ограниченное время.

Одним из вариантов решения задач планирования обработки партий материалов в непрерывных производствах является применение метаэвристических алгоритмов (генетических алгоритмов (ГА)), рассмотренное в [8]. Размеры партий материалов определяются спросом на продукцию, а построение расписания формируемых партий реализуется с использованием ГА. Определены особенности задачи

планирования: обработка партий материалов проводится непрерывно, между приборами нет промежуточных буферов для накопления материалов; требуется определить такую последовательность запуска партий материалов на обработку в систему, чтобы минимизировать общее время выполнения всех операций со всеми партиями. В соответствии с выполненной постановкой задачи разработан ГА, реализующий формирование расписаний обработки партий материалов заданного состава, исследованы его особенности, определены значения параметров ГА, которые позволяют получить лучшие результаты.

Метод планирования обработки пакетов заданий на параллельно функционирующих устройствах, использующий целочисленное линейное программирование (ЧЦЛП), рассматривается в [9]. Решаемая в [9] задача предполагает, что в систему для выполнения поступают заказы, каждый заказ состоит из некоторых продуктов (заданий), а продукт (задание) состоит из компонентов. Определение составов пакетов осуществляется в [9] не на уровне компонентов, а на уровне заданий (пакеты образуются не из компонентов продуктов, входящих в заказ, а непосредственно из продуктов (заданий)). Реализация планирования в данной задаче предусматривает определение решений по составам пакетов и расписаниям их выполнения. Решение задачи планирования предполагает определение значений переменных, соответствующих как составам пакетов, так и порядкам их выполнения на приборах. Так как для решения указанных задач применено ЧЦЛП, то их размерность является ограниченной [9].

В работе [10] рассматривается метод планирования выполнения ПЗ на параллельно функционирующих приборах на основе локального поиска совместно с методом муравьиной колонии при использовании эвристических правил. Включение каждого задания в пакеты предполагает определение значений плотности следа феромона, характеризующей качество решений. При распределении заданий по пакетам используется эвристическое правило, предполагающее, что задания добавляются в первую очередь в пакет, выполняемый на параллельной машине с наибольшей емкостью. Также при распределении заданий по пакетам учитываются значения параметра, в котором учтены длительности выполнения самого задания и соответствующего пакета. Использование метаэвристических алгоритмов для планирования выполнения ПЗ на параллельных приборах рассматривается в работах [11, 12]. Для определения составов ПЗ в [11] применены генетические алгоритмы (ГА). Формируемые при реализации ГА хромосомы соответствуют решениям по составам ПЗ. Приближенно оптимальное решение обеспечивает минимизацию

простоев приборов при выполнении заданий в ПЗ. В [12] при распределении заданий по пакетам используется алгоритм муравьиной колонии (МК), позволяющий группировать задания, имеющие приближенно одинаковую длительность выполнения. В силу стохастического характера рассмотренных алгоритмов они не гарантируют получение эффективных решений при разных значениях входных параметров задачи.

В [13] реализуется решение задачи планирования реализации действий с ПЗ на параллельных приборах при указании директивных сроков для выполняемых заданий. В состав пакетов включаются задания, у которых близкими являются значения директивных сроков окончания их формирования. Упорядочивание ПЗ на параллельных приборах также осуществляется с учетом директивных сроков окончания их выполнения.

В работе [14] рассматривается решение задачи определения составов пакетов и очередности их выполнения на параллельно действующих приборах с использованием эвристических правил. При формировании пакетов для каждого задания реализуется вычисление значения параметра, учитывающего вес задания, длительность его выполнения, среднюю длительность выполнения заданий, не распределенных по приборам, момент времени поступления и т.д. В соответствии со значениями этого параметра формируется последовательность заданий, после чего количество заданий, соответствующее размеру пакета, закрепляется за определенным не занятым прибором. Рассмотрены различные способы вычисления указанного параметра. Пакеты, закрепленные за определенными приборами, упорядочиваются для выполнения с использованием ГА.

Применению эвристических правил при определении составов ПЗ, выполняемых на параллельных приборах, посвящена работа [15]. В соответствии с этим правилом в пакет включаются задания с близкими значениями длительностей их выполнения. Формирование ПЗ приводит к уменьшению общего времени выполнения всех заданий.

Использование различных эвристических правил при формировании чередующихся окрестностей текущего локально оптимального решения по составам ПЗ рассмотрено в работе [16]. В формируемых окрестностях реализуется определение нового локально оптимального решения по составам ПЗ и расписаниям их выполнения. Формулировки правил, позволяющих формировать решения по составам ПЗ и расписаниям их выполнения, имеют вид: объединить два ПЗ с фиксацией позиции нового пакета в последовательности выполнения; разделить один ПЗ и вставить один из полученных пакетов

в новую позицию; менять местами ПЗ в последовательности их выполнения; переместить ПЗ в новую позицию в последовательности; менять местами два задания разных пакетов. Рассмотренный способ реализует не направленный поиск локально оптимальных решений, а стохастический поиск, что не гарантирует получение эффективного результата при разных значениях входных параметров.

Работы [17, 18] также посвящены рассмотрению эвристических правил формирования ПЗ. В [17] рассматривается правило пакетной совместимости (BCLPT), которое предполагает:

- сортировку заданий по не возрастанию длительностей;
- в полученной последовательности определяются задания i_1 и i_1+1 , длительности выполнения которых связаны введенным условием с длительностью выполнения первого задания; если задания из последовательности до i_1+1 -го включаются в пакет;
- для оставшихся в последовательности заданий вновь проверяются введенные условия и формируется новый пакет;
- если условия не выполняются, то все рассматриваемые задания включаются в один пакет.

Последовательность выполнения ПЗ определяется либо с использованием динамического программирования, либо с использованием предложенного в [17] полиномиально сходящегося алгоритма.

В работе [18] в качестве критерия используется сумма момента времени окончания выполнения всех ПЗ и штрафа за отказ в обслуживании заданий. Отказ в выполнении задания связан с ограничениями мощности параллельных приборов и срока выполнения заданий. Эвристические правила, рассмотренные в [18], предусматривают формирование различных множеств отклоняемых заданий и для каждого множества не отклоненных заданий формирования пакетов с использованием соответствующих условий.

В [19] рассматривается метод планирования выпуска нескольких изделий в многомашинных системах с учетом заданных директивных сроков. В соответствии с указанным подходом планирование реализуется с использованием эвристического правила, предполагающего, что в пакеты объединяются изделия, у которых директивные сроки окончания их выпуска не превышают момент времени окончания выполнения соответствующего пакета. При построении расписаний используется метод вставки. В работе [20] развивается подход к планированию выполнения ПЗ на основе имитационного моделирования. Рассматривается имитационная модель

технологического процесса, в которой учтены его особенности (вводятся дополнительные операции, которые учитываются при построении расписаний). Размеры партий материалов являются задаваемыми в качестве входных параметров модели. Определение маршрутов движения партий материалов и расписаний обработки реализуется с использованием правил. Каждый вариант решения (размеры партий, маршруты их движения в системе, расписание обработки) характеризуется своим значением критерия. Использование имитационной модели позволяет выполнить выбор составов партий и эвристических правил, обеспечивающих эффективные маршруты их передачи и расписания их запуска в систему.

В работах отечественных авторов, таких как Танаев В.С., Сотсков Ю.Н., Ковалев М.Я., Лазарев А.А., Кобак В.Г., Нейдорф Р.А. (в том числе, основополагающих, являющихся базой теории расписаний) рассматриваются методы построения расписаний выполнения единичных заданий, не объединяемых в пакеты. При этом рассматриваются различные виды обрабатываемых систем (с одинаковыми маршрутами выполнения заданий, с различными маршрутами выполнения заданий, с нефиксированными маршрутами), учитываются различные виды упрощающих предположений (ограниченное количество приборов, одинаковые последовательности выполнения заданий на приборах, упорядочивание значений длительностей выполнения заданий на разных приборах и т.д.). Ограниченное количество работ (в частности, [21]) посвящены рассмотрению использования эвристических правил при определении составов пакетов и методам построения расписаний фиксированных пакетов. Методы оптимизации составов пакетов в работах упомянутых авторов не рассматриваются.

Таким образом, планирование выполнения ПЗ, включающее определение их составов, осуществляется путем применения:

- смешанного целочисленного программирования (для партий материалов, обрабатываемых в непрерывном производстве) при решении задач ограниченной размерности;
- имитационного моделирования технологического процесса обработки партий материалов в непрерывном производстве для оценки эффективности применения эвристик при определении маршрутов и расписаний выполнения операций с ними;
- целочисленного линейного программирования для определения составов ПЗ и расписаний их выполнения (для задач ограниченной размерности);

- метаэвристических алгоритмов (ГА либо метода МК) при определении составов ПЗ; применение этих алгоритмов не гарантирует приближения к оптимальным решениям при различных значениях входных параметров;
- эвристических правил при формировании составов ПЗ (с учетом директивных сроков выполнения заданий), что позволяет получать решения, лучшие, чем без оптимизации, но не приближающиеся к оптимальным.

Таким образом, отсутствуют методы оптимизации составов ПЗ, лишенные ограничений на размерность задач, использование которых не предполагает введения дополнительных условий (в частности, директивных сроков окончания выполнения отдельных заданий). В рассмотренных методах планирования выполнения ПЗ в МС не учитывается условие построения КР (в том числе при заданных директивных сроках). В связи с этим совершенствование методов комплексного планирования выполнения ПЗ в МС, учитывающих условие формирования комплектов из результатов и предполагающих совместную (комплексную) оптимизацию составов пакетов и расписаний реализации действий с ними на приборах МС, является актуальным.

Комплексное планирование выполнения ПЗ в МС при условии формирования комплектов в заданные для них директивные сроки реализуется в составе двух–уровневой системы принятия решений. На первом уровне определяются составы ПЗ, на втором– расписания их (ПЗ) выполнения в МС, а также с целью определения значений критерия оптимальности решений по составам ПЗ на первом уровне реализуется распределение результатов выполнения заданий в пакетах по комплектам. При этом для лучшего (локально оптимального) решения суммарное отклонение сроков окончания формирования комплектов от директивных сроков, заданных для них (суммарное запаздывание с формированием комплектов относительно директивных сроков), является минимальным. Определение решений на первом уровне реализуется с использованием метода оптимизации составов ПЗ, рассмотренного в [1], определение расписаний выполнения ПЗ в МС реализуется с использованием метода оптимизации решений, рассмотренного в [22]. В тоже время реализация алгоритма распределения результатов выполнения заданий из пакетов по комплектам предваряется упорядочиванием идентификаторов типов комплектов в соответствии с заданными для этих комплектов директивными сроками (то есть предварительное упорядочивание

типов комплектов в соответствии с директивными сроками обеспечивает распределение результатов выполнения заданий по комплектам и определение суммарного запаздывания с их формированием). Поэтому для достижения цели работы – совершенствования методов комплексного планирования выполнения ПЗ в МС, – в ней требуется решить следующие задачи:

- 1) построения математической модели иерархической игры оптимизации решений по составам ПЗ и расписаний их выполнения в МС, учитывающей условие формирования комплектов результатов в заданные для них директивные сроки;
- 2) построения метода упорядочивания идентификаторов типов комплектов в соответствии с заданными для комплектов директивными сроками;
- 3) построения метода распределения результатов выполнения в МС заданий, входящих в пакеты, по комплектам и определения значений суммарного запаздывания с формированием комплектов относительно заданных для них директивных сроков.

2. Математическая модель иерархической игры для определения составов ПЗ и расписаний их выполнения в МС. В предлагаемом методе планирования выполнения ПЗ в МС, реализующем иерархический подход к оптимизации [1], теория иерархических игр применена в качестве способа определения комплексных (взаимосвязанных) локально-оптимальных решений по составам ПЗ и расписаниям реализации действий с ними на приборах МС. Как способ определения локально-оптимальных решений на каждом уровне иерархии системы планирования теория иерархических игр предполагает, что заданным является следующий порядок ходов:

- формирование решения по составам ПЗ ведущим игроком (на первом уровне) и передачу этого решения ведомому игроку (на втором уровне);
- формирование игроком второго уровня расписания выполнения ПЗ в МС, соответствующего решению по их составу, и передачу расписания на первый уровень с целью вычисления значения критерия, характеризующего решение по составам ПЗ;
- переопределение решений по составам ПЗ ведущим игроком и определение лучшего среди них (локально-оптимального) с учетом решений по порядкам выполнения пакетов на приборах МС.

Таким образом, игра предусматривает передачу ведущим игроком ведомому информации о сформированном им решении. Игроки реализуют указанный порядок ходов до тех пор, пока не будет получено комплексное решение по составам ПЗ и расписаниям их выполнения в МС, состоящее из локально-оптимальных решений, сформированных на каждом из уровней.

В результате теория иерархических игр предоставляет механизм совместной (комплексной) оптимизации решений по составам ПЗ и расписаниям их выполнения в МС.

Решения, формируемые ведомым игроком (на нижнем уровне), благоприятствуют достижению ведущим игроком его цели – минимизации общего запаздывания с формированием комплектов относительно их директивных сроков (максимизации соответствия моментов времени окончания формирования комплектов заданным для них директивным срокам). Игрок нижнего уровня является благожелательным по отношению к игроку верхнего уровня, то есть выбирает такие решения, которые обеспечивают минимизацию этим игроком суммарного запаздывания с формированием КР. При этом решения, формируемые ведущим игроком, являются не зависимыми от решений ведомого игрока. Реализуемая в системе планирования выполнения ПЗ в МС игра является неантагонистической игрой в оптимистических стратегиях. При построении модели иерархической игры определения составов ПЗ и расписаний их выполнения в МС использованы обозначения, которые представлены в Таблице 1.

Таблица 1. Обозначения параметров процесса выполнения ПЗ в МС

Обозначение	Назначение параметра
1	2
i	Идентификатор типа заданий ($i = \overline{1, n}$)
n^i	Количество заданий i -го типа
l	Номер прибора ($l = \overline{1, L}$)
m^i	Количество ПЗ i -го типа ($i = \overline{1, n}$)
M	Вектор количества пакетов i -ых типов ($i = \overline{1, n}$)

Продолжение таблицы 1

1	2
a_{ih}	Количество заданий i -го типа в h -ом пакете ($h = \overline{1, m_i}$)
A	Матрица составов комплектов
P	Матрица порядка выполнения заданий
R	Матрица количества заданий i -ых типов в пакетах, занимающих в последовательностях их выполнения на приборах МС j -е позиции
π^l	Последовательность выполнения ПЗ на l -ых приборах МС
t_{li}	Длительность выполнения задания i -го типа на l -ом приборе МС ($i = \overline{1, n}, l = \overline{1, L}$)
T	Матрица длительностей выполнения заданий на приборах МС
t_{ij}^l	Длительность переналадки l -го прибора с выполнения заданий i -го типа на выполнение заданий j -го типа
T^l	Матрицы переналадки l -ых приборов с выполнения заданий одного типа на выполнение заданий другого типа ($l = \overline{1, L}$)
t_{jq}^{0l}	Моменты времени начала реализации действий с q -ыми заданиями в пакетах, занимающих j -е позиции в последовательностях их выполнения в МС ($q = \overline{1, n_j}$)
T^{0l}	Матрицы начала выполнения заданий в пакетах в j -ых позициях в последовательностях реализации действий с ними на l -ых приборах МС ($l = \overline{1, L}$)
$t_{jn_j}^{0l}$	Моменты времени начала выполнения последнего n_j -го задания в пакете, занимающем в последовательностях их выполнения на приборах МС j -ю позицию

С учетом введенного в рассмотрение вектора количества пакетов i -ых типов M и матрицы составов пакетов A решение на первом уровне имеет следующий вид: $[M, A]$. Для формализации вида последовательностей π^l расписания выполнения ПЗ на приборах МС в рассмотрении введена матрица порядка P . Элемент $p_{ij}=1$, ПЗ i -го типа занимает в последовательностях π^l j -ю позицию, $p_{ij}=0$ в случае, если ПЗ i -го типа не занимает в последовательности π^l j -ю позицию, размерность матрицы $n \times n_p$, где n – количество типов заданий, n_p – количество ПЗ в

последовательностях $\pi^l (n_p = \sum_{i=1}^n m_i)$. Порядок выполнения ПЗ на всех

приборах одинаков, поэтому достаточно определения одной матрицы порядка P . Матрица R – матрица количества заданий i -ых типов в пакетах, занимающих в последовательностях π^l j -е позиции (элемент r_{ij} – количество заданий i -го типа в пакете, занимающем j -ую позицию в π^l , размерность матрицы $n \times n_p$). Значения t_{jq}^{0l} вычисляются с использованием математической модели процесса выполнения заданий в МС [1]. Эти значения используются при определении значений критерия оптимизации решений по порядкам выполнения ПЗ в МС на нижнем уровне, а также при распределении результатов выполнения заданий в пакетах по комплектам. Моменты времени окончания выполнения ПЗ в j -ых позициях в этих последовательностях

определяются выражением вида: $t_{j,n_j}^{0L} + \sum_{q=1}^n t_{Lq} \cdot p_{q,j}$. Решение,

формируемое на втором уровне, имеет вид $[P, R, \{T^{0l} / l = \overline{1, L}\}]$. Для формализации модели процесса выполнения ПЗ в МС при формировании КР в заданные сроки использованы дополнительные обозначения, которые представлены в Таблице 2.

Таблица 2. Обозначения, используемые при построении модели выполнения ПЗ в МС при формировании КР в директивные сроки

Обозначение	Назначение параметра
1	2
g	Идентификатор типа КР
g^{KOM}	Общее количество типов комплектов
n_g^{KOM}	Количество комплектов g -го типа ($g = \overline{1, g^{KOM}}$)
h_g	Индекс комплекта g -го типа ($h_g = \overline{1, n_g^{KOM}}$)
w_{ig}	Количество результатов i -го типа, которые входят в один комплект g -го типа (образуют i -ю компоненту комплекта g -го типа, $i = \overline{1, n}$, $g = \overline{1, g^{KOM}}$)

Продолжение таблицы 2

1	2
W	Матрица составов комплектов (w_{ig} – элемент матрицы W)
t_{i,h_g}^g	Моменты времени окончания формирования i -ых компонент h_g -ых комплектов g -ых типов ($i = \overline{1, n}$, $h_g = \overline{1, n_g^{KOM}}$, $g = \overline{1, g^{KOM}}$)
T^g	Матрицы моментов времени окончания формирования i -ых компонент h_g -ых комплектов g -ых типов
d_{g,h_g}^{KOM}	Директивный срок окончания формирования h_g -го комплекта g -го типа ($h_g = \overline{1, n_g^{KOM}}$, $g = \overline{1, g^{KOM}}$)
D^{KOM}	Матрица директивных сроков окончания формирования h_g -го комплектов g -ых типов ($h_g = \overline{1, n_g^{KOM}}$, $g = \overline{1, g^{KOM}}$)
K_y^{KOM}	Упорядоченное в соответствии со значениями d_{g,h_g}^{KOM} множество идентификаторов типов комплектов
D_y^{KOM}	Упорядоченное множество значений d_{g,h_g}^{KOM} для h_g -ых комплектов g -ых типов ($h_g = \overline{1, n_g^{KOM}}$, $g = \overline{1, g^{KOM}}$)
g_y^{KOM}	Мощность множеств K_y^{KOM} и D_y^{KOM} .
p	Номер позиции типа комплекта в множестве K_y^{KOM} и директивного срока d_{g,h_g}^{KOM} в множестве D_y^{KOM} ($p = \overline{1, g_y^{KOM}}$)
g_p	Идентификатор типа комплекта, занимающий в множестве K_y^{KOM} p -ю позицию
$t_{g_p}^{KOM}$	Момент времени окончания формирования комплекта g -го типа, идентификатор которого занимает p -ю позицию в множестве K_y^{KOM}
$d_{g_p}^{KOM}$	Директивный срок окончания формирования комплекта g -го типа в p -й позиции в множестве K_y^{KOM}

С учетом введенных параметров n_g^{KOM} и w_{ig} значения n^i ($i = \overline{1, n}$) вычисляются с использованием выражения вида [22]:

$$n^i = \sum_{g=1}^{KOM} n_g^{KOM} \cdot w_{ig} \cdot t_{i, h_g}^g$$

Значения t_{i, h_g}^g соответствуют значениям моментов времени окончания выполнения ПЗ, результаты которых позволяют сформировать i -е компоненты h_g -ых комплектов g -ых типов. Они определяются при реализации алгоритма распределения результатов выполнения заданий по комплектам. С использованием значений t_{i, h_g}^g

для каждого комплекта g -го типа ($g = \overline{1, g^{KOM}}$) формируется матрица T^g ($g = \overline{1, g^{KOM}}$) моментов времени окончания формирования их компонент.

Задача планирования выполнения ПЗ в МС с учетом формирования КР предполагает задание директивных сроков d_{g, h_g}^{KOM} для каждого h_g -го комплекта g -го типа ($h_g = \overline{1, n_g^{KOM}}$), ($g = \overline{1, g^{KOM}}$). Значения директивных сроков d_{g, h_g}^{KOM} образуют матрицу D^{KOM} .

В соответствие с постановкой задачи формирование комплектов требуется реализовать к назначенным для них директивным срокам. Лучшим решением по составам ПЗ и расписаниям их выполнения в МС является решение, гарантирующее минимальные отклонения моментов времени формирования комплектов от их директивных сроков. По этой причине для решений по составам ПЗ на первом уровне требуется определять соответствие между моментами времени формирования комплектов и директивным срокам.

Распределение результатов выполнения ПЗ по комплектам предваряется упорядочиванием их типов в множестве K^{KOM}_y в соответствии с директивными сроками d_{g, h_g}^{KOM} , а также директивных сроков в множестве D_y^{KOM} . Мощность g_y^{KOM} множеств K^{KOM}_y и D_y^{KOM}

определяется выражением $g_y^{KOM} = \sum_{g=1}^{KOM} n_g^{KOM}$. Между множествами K^{KOM}_y и

D_y^{KOM} установлено соответствие. То есть g -ому типу комплекта, занимающему в множестве K^{KOM}_y p -ю позицию (обозначенному как g_p),

соответствует директивный срок окончания его формирования в множестве D_y^{KOM} , обозначенный как $d_{g_p}^{KOM}$.

В результате реализации алгоритма распределения результатов выполнения ПЗ в МС по комплектам в матрице $(T^g (g = 1, g^{KOM}))$ для h_g -го комплекта g -го типа будет сформирован столбец значений t_{i,h_g}^g , на основе которых реализуется вычисление значения $t_{g_p}^{KOM}$ в соответствие с выражением вида $t_{g_p}^{KOM} = \max_{1 \leq i \leq n} [t_{i,h_g}^g]$. В случае, если для вычисленного значения $t_{g_p}^{KOM}$ и соответствующего ему значения $d_{g_p}^{KOM}$ из множества D_y^{KOM} выполняется условие $t_{g_p}^{KOM} > d_{g_p}^{KOM}$, то определяется запаздывание $t_{g_p}^{KOM}$ с формированием комплекта по сравнению с его директивным сроком $d_{g_p}^{KOM}$. Для каждого комплекта идентификатор g типа которого занимает в K_y^{KOM} p -ю позицию, значение $t_{g_p}^{KOM}$ определяется выражением: $t_{g_p}^{KOM} = \max \{0, t_{g_p}^{KOM} - d_{g_p}^{KOM}\}$. Для всех комплектов g -ых типов в p -ых позициях в множестве K_y^{KOM} (обозначенных как $g_p, p = 1, g_y^{KOM}$) суммарное запаздывание с их формированием

определяется выражением $\sum_{p=1}^{g_y^{KOM}} t_{g_p}^{KOM}$, которое является критерием оптимизации решений по составам ПЗ.

В работе [22] рассмотрен метод построения расписаний выполнения ПЗ в МС. Порядок выполнения ПЗ на приборах МС оптимизируется с учетом требования минимизации суммарных простоев приборов МС [22]. Этот критерий ([22]) также используется при оптимизации расписаний в задаче планирования выполнения ПЗ в МС с учетом формирования КР в заданные директивные сроки. Вид иерархической игры комплексного планирования выполнения ПЗ в МС (при формировании КР в заданные директивные сроки), следующий:
 – первый уровень:

$$\min f_1([M, A]), \tag{1}$$

где $f_1 = \sum_{p=1}^{g_y^{ком}} t z_{g_p}^{ком}$;

– второй уровень [1, 22]:

$$\min f_2([M, A], [P, R, \{T^{0l} / l = \overline{1, L}\}]), \quad (2)$$

$$\begin{aligned} \text{где } f_2 = & \sum_{l=2}^L t_{11}^{0l} + \sum_{l=1}^L \sum_{j=2}^{n_p} \left[t_{j1}^{0l} - \left[t_{j-1, n_{j-1}}^{0l} + \sum_{h=1}^n t_{lh} \cdot P_{h, j-1} \right] \right] + \\ & + \sum_{l=2}^L \sum_{j=1}^{n_p} \sum_{q=2}^{n_j} \left[t_{jq}^{0l} - \left[t_{j, q-1}^{0l} + \sum_{h=1}^n t_{lh} \cdot P_{hj} \right] \right]. \end{aligned}$$

Критерий (1) оптимизации решений по составам ПЗ на верхнем уровне соответствует суммарному запаздыванию с формированием комплектов относительно их директивных сроков (соответственно, выигрышем ведущего игрока является соответствие моментов времени окончания формирования комплектов заданным для них директивным срокам, которое будет максимизироваться). Оптимизируемым в соответствии с критерием (1) решением является решение $[M, A]$ по количеству и составам ПЗ (определяется оптимальное количество и составы пакетов). Критерий (2) оптимизации решений на нижнем уровне соответствует суммарным простоям приборов МС при выполнении ПЗ, составы которых сформированы на верхнем уровне. Оптимизируемым (с учетом решения $[M, A]$ по составам ПЗ) решением является решение по порядку выполнения ПЗ в МС.

Задача оптимизации составов ПЗ и расписаний их выполнения в МС сформулирована следующим образом. Заданными являются: количество типов комплектов $g^{ком}$; количество комплектов каждого типа $n_g^{ком}$ ($g = \overline{1, g^{ком}}$); матрица W составов комплектов; матрица $D^{ком}$ директивных сроков окончания формирования комплектов; количество n типов заданий, выполняемых в МС; матрица T длительностей выполнения заданий на l -ых приборах МС ($l = \overline{1, L}$); матрицы T^l ($l = \overline{1, L}$) длительностей переналадок приборов с выполнения заданий i -го типа на выполнение заданий j -го типа. Требуется определить: количество m_i ПЗ i -ых типов ($i = \overline{1, n}$), которые следует сформировать,

составы этих пакетов (значения элементов a_{ih} ($i = \overline{1, n}, h = \overline{1, m_i}$) матрицы A составов пакетов), расписание выполнения пакетов на приборах МС (матрицу P порядка выполнения, матрицу R количества заданий в пакетах, матрицы T^{0l} ($l = \overline{1, L}$) начала выполнения заданий в пакетах в последовательностях реализации действий с ними на приборах МС). Оптимизация составов ПЗ реализуется с использованием критерия вида (1). Оптимизация расписания выполнения ПЗ в МС реализуется с использованием критерия вида (2), характеризующего суммарные простои приборов МС при выполнении ПЗ.

Планирование выполнения ПЗ в МС при условии формирования КР в заданные директивные сроки реализуется с привлечением методов оптимизации: составов ПЗ [1] для формирования решений $[M, A]$ на первом уровне; расписаний $[P, R, \{T^{0l} / l = \overline{1, L}\}]$ выполнения ПЗ в МС [22] на втором уровне. С целью определения значений критерия

$f_1([M, A])$ на верхнем уровне требуется реализовать распределение по комплектам результатов, получаемых в соответствии с расписанием $[P, R, \{T^{0l} / l = \overline{1, L}\}]^*$, определение моментов времени окончания формирования комплектов и значений суммарного запаздывания с

формированием комплектов (значений критерия $f_1([M, A])$). Для распределения результатов по комплектам требуется предварительно реализовать упорядочивание идентификаторов типов комплектов с учетом директивных сроков. В соответствии с выполненными рассуждениями сформирована обобщенная схема системы планирования выполнения ПЗ в МС при формировании комплектов в заданные директивные сроки, представленная на рисунке 1.

3. Метод упорядочивания типов комплектов и метод распределения результатов выполнения ПЗ в МС по комплектам. В алгоритмах упорядочивания идентификаторов типов комплектов в соответствии с их директивными сроками и распределения результатов выполнения ПЗ по комплектам использованы обозначения, представленные в Таблице 3.

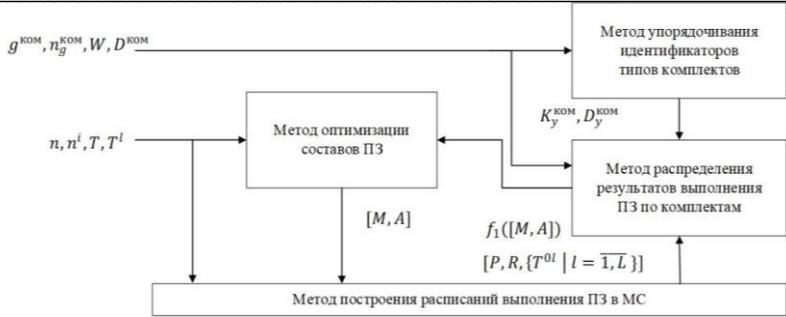


Рис. 1. Схема системы планирования выполнения ПЗ в МС при формировании комплектов в директивные сроки

Таблица 3. Обозначения, используемые в алгоритме упорядочивания идентификаторов типов комплектов и алгоритме распределения результатов выполнения заданий по комплектам

Обозначение	Назначение параметра
g'	Идентификатор типа комплекта, добавляемый в K_y^{KOM}
I	Множество типов результатов ($I = \{1, 2, \dots, n\}$)
i'	Тип результатов, на основе которых формируется i' -я компонента комплекта g' -го типа
$h_{i'}$	Номер ПЗ i' -го типа, результаты выполнения которого добавляются в комплект g' -го типа
$st_{i', g'}$	Количество результатов i' -го типа, добавленных в комплект g' -го типа
$tz_{g', p}^{KOM}$	Запаздывание с формированием комплекта g' -го типа в p -ой позиции в множестве K_y^{KOM}

Рассматриваемый алгоритм реализует упорядочивание типов комплектов в множестве K_y^{KOM} , а в множестве D_y^{KOM} упорядочивание директивных сроков их формирования. Алгоритм формирования множеств K_y^{KOM} , D_y^{KOM} имеет следующий порядок шагов:

- 1) инициализировать единичным значением номер позиции p для рассматриваемого типа g' комплектов в множестве K_y^{KOM} и для его директивного срока $d_{g', h_{g'}}^{KOM}$ в множестве D_y^{KOM} ($p = 1$);

- 2) определить наличие в матрице D^{KOM} таких элементов d_{g,h_g}^{KOM} ($h_g = \overline{1, n_g^{KOM}}$, $g = \overline{1, g^{KOM}}$), что $d_{g,h_g}^{KOM} \neq 0$; при отсутствии в матрице D^{KOM} элементов d_{g,h_g}^{KOM} ($h_g = \overline{1, n_g^{KOM}}$, $g = \overline{1, g^{KOM}}$), что $d_{g,h_g}^{KOM} \neq 0$, перейти на пункт 6;
- 3) определить элемент $d_{g',h_{g'}}^{KOM} = \min_{1 \leq g \leq g^{KOM}} \min_{1 \leq h_g \leq n_g^{KOM}} (d_{g,h_g}^{KOM})$ (номер строки g' соответствует типу комплекта, номер столбца $h_{g'}$ – номеру комплекта g' -го типа);
- 4) добавить идентификатор g' в множество K_y^{KOM} в p -ой позиции ($K_y^{KOM} = K_y^{KOM} \cup \{g'\}$), добавить значение $d_{g',h_{g'}}^{KOM}$ в множество D_y^{KOM} в p -ой позиции ($D_y^{KOM} = D_y^{KOM} \cup \{d_{g',h_{g'}}^{KOM}\}$);
- 5) присвоить рассмотренному элементу $d_{g',h_{g'}}^{KOM}$ матрицы D^{KOM} нулевое значение ($d_{g',h_{g'}}^{KOM} = 0$); модификация номера позиции p элементов в множествах K_y^{KOM} и D_y^{KOM} : $p = p + I$; перейти на пункт 2;

б) окончание алгоритма.

Формирование с использованием рассмотренного алгоритма множеств K_y^{KOM} и D_y^{KOM} предвдывает распределение результатов выполнения ПЗ в МС по комплектам, которое соответствует расписанию выполнения пакетов $[P, R, \{T^{0l} / l = \overline{1, L}\}]^*$, вычисление значений $t_{g_p}^{KOM}$ ($p = \overline{1, g_y^{KOM}}$) и в случае $t_{g_p}^{KOM} > d_{g_p}^{KOM}$ определение значений $t_{g_p}^{KOM}$ для комплектов g -ых типов (в p -ых позициях в множестве K_y^{KOM}). Реализация алгоритма распределения результатов выполнения ПЗ в МС по комплектам позволяет вычислить значения $t_{g_p}^{KOM}$ и $t_{g_p}^{KOM}$ ($p = \overline{1, g_y^{KOM}}$) в случае запаздываний с формированием комплектов.

Входными данными для алгоритма распределения результатов по комплектам являются: множества K_y^{KOM} и D_y^{KOM} , расписание $[P, R, \{T^{0l} / l = \overline{1, L}\}]^*$, множество I типов результатов (ему

соответствует множество I' , изменяемое в процессе реализации алгоритма). Значение p номера позиции типа комплекта и его директивного срока в множествах K_y^{KOM} и D_y^{KOM} задается равным 1.

Порядок шагов алгоритма распределения результатов по комплектам, определения значений $t_{g_p}^{KOM}$ ($p = 1, g_y^{KOM}$) и $t_{g_p}^{KOM}$ (в случае $t_{g_p}^{KOM} > d_{g_p}^{KOM}$) следующий:

- 1) положить $I' = I, p = 1, h_g = 1$ ($g = 1, g_y^{KOM}$);
- 2) для каждого i -го типа результатов положить $h_i = 0$ ($i = \overline{1, n}$);
- 3) на основе номера позиции p в множестве K_y^{KOM} определить тип комплекта g' , в который будут распределяться результаты выполнения заданий: $g' = g_p$; на основе номера позиции p в множестве D_y^{KOM} определить значение директивного срока формирования комплекта g' -го типа: $d_{g'}^{KOM} = d_{g_p}^{KOM}$;
- 4) определить тип i' результатов выполнения ПЗ, добавляемых в комплект g' -го типа: $i' = \min\{i / i \in I'\}$; модифицировать I' : $I' = I \setminus \{i'\}$;
- 5) положить значение $st_{i',g'}$ равным 0 ($st_{i',g'} = 0$);
- 6) определить первый элемент $r_{i',h}$ в i' -ой строке матрицы R такой, что $r_{i',h} \neq 0$ (при $h = \overline{h_{i'} + 1, n_p}$);
- 7) если $r_{i',h} \geq (w_{i',g'} - st_{i',g'})$, то модифицировать значение $r_{i',h}$: $r_{i',h} = r_{i',h} - (w_{i',g'} - st_{i',g'})$; выполнить переход на пункт 8; если $r_{i',h} < (w_{i',g'} - st_{i',g'})$, то модифицировать значение $st_{i',g'}$: $st_{i',g'} = st_{i',g'} + r_{i',h}$; модифицировать значение $h_{i'}$: $h_{i'} = h$; выполнить переход на пункт 6;
- 8) вычислить значение $t_{i',h_{g'}}^{g'} = t_{h,n_h}^{0L} + \sum_{q=1}^n t_{L,q} \cdot p_{q,h}$;

- 9) если $I' = \emptyset$, то рассчитать значение $t_{g'p}^{KOM}$ момента времени окончания формирования комплекта g' -го типа: времени запаздывания:
- $$t_{g'p}^{KOM} = \max_{1 \leq i \leq n} (t_{i,h_k}^{g'}),$$
- $t_{g'p}^{KOM} = \max \{0, t_{g'p}^{KOM} - d_{g'p}^{KOM}\}$; если $t_{g'p}^{KOM} > 0$, то добавить значение $t_{g'p}^{KOM}$ в критерий f_1 : $f_1 = f_1 + t_{g'p}^{KOM}$; модифицировать значение $h_{g'} : h_{g'} = h_{g'} + 1$; выполнить переход на пункт 10; если $I' \neq \emptyset$, то выполнить переход на пункт 4;
- 10) модифицировать значение p номера позиции типа комплекта g в K_y^{KOM} и директивного срока в D_y^{KOM} : $p = p + 1$; если $p > g_y^{KOM}$, то выполнить переход на пункт 11; если $p \leq g_y^{KOM}$, то положить $I' = I$, переход на пункт 3;
- 11) останов алгоритма.
1. Результатом реализации рассмотренного алгоритма является значение критерия f_1 , которое характеризует текущее решение $[M, A]$ по составам ПЗ на первом уровне иерархии системы планирования.

Решение задачи планирования выполнения ПЗ в МС при формировании КР обеспечивается применением метода локальной оптимизации составов пакетов [1]. Поэтому результаты комплексного планирования выполнения ПЗ в МС являются зависимыми от формируемых на верхнем уровне решений по составам пакетов. При исследовании иерархического подхода к планированию выполнения ПЗ в МС в рассматриваемой задаче использованы следующие параметры: неоднородность длительностей выполнения заданий i -ых типов ($i = \overline{1, n}$) на l -ых приборах МС ($l = \overline{1, L}$) – $\max(t_i^l) / \min(t_i^l)$; неоднородность длительностей переналадок приборов МС на выполнение заданий j -ых типов – $\max(t_{ij}^l) / \min(t_{ij}^l)$; неоднородность составов комплектов – $\max(w_{ig}) / \min(w_{ig})$. Задаваемые значения параметров: $n \in \{5, 10\}$, $L \in \{5, 10\}$; $g^{KOM} = 2$; $n_g^{KOM} \in \{2, 4\}$. Значения

$\max(w_{ig}) / \min(w_{ig'})$ задавались равными 1, 2, 3; значения $\max(t_{ij}^l) / \min(t_{ij}^l)$ – равными 1, 2, 4, 8, 16; значения $\max(t_i^l) / \min(t_i^l)$ – равными 1, 2, 4, 8, 16. Также при исследовании эффективности планирования выполнения ПЗ в МС с условием формирования КР в заданные директивные сроки использованы обозначения: $d_{1,h_g}^{KOM} / d_{g^{KOM},h_g}^{KOM}$ – отношение значений директивных сроков h_g -ых комплектов первого и g^{KOM} -го типов; отношение $d_{g,1}^{KOM} / d_{g,n_g^{KOM}}^{KOM}$ ($d_{g,1}^{KOM}$ – директивный срок формирования первого комплекта g -го типа, $d_{g,n_g^{KOM}}^{KOM}$ – директивный срок формирования последнего n_g^{KOM} -го комплекта этого же g -го типа). Значение $d_{g^{KOM},1}^{KOM}$ задавалось равным 100. Значения отношений $d_{1,h_g}^{KOM} / d_{g^{KOM},h_g}^{KOM}$ и $d_{g,1}^{KOM} / d_{g,n_g^{KOM}}^{KOM}$ задавались равными 1.5, 2. В соответствии с указанными значениями параметра $d_{g^{KOM},1}^{KOM}$, отношений $d_{1,h_g}^{KOM} / d_{g^{KOM},h_g}^{KOM}$, $d_{g,1}^{KOM} / d_{g,n_g^{KOM}}^{KOM}$ для каждой задачи сформированы матрицы D^{KOM} .

Задачи планирования выполнения ПЗ в МС характеризуются значениями пары параметров (g^{KOM}, n_g^{KOM}) . Рассматривались задачи планирования, значения параметров (g^{KOM}, n_g^{KOM}) в которых равны (2, 2) и (2, 4). Для задачи планирования, характеризуемой значениями параметров (g^{KOM}, n_g^{KOM}) вида (2, 2), матрица D^{KOM} в соответствии со значениями отношений $d_{1,h_g}^{KOM} / d_{g^{KOM},h_g}^{KOM} = 1.5$ и $d_{g,1}^{KOM} / d_{g,n_g^{KOM}}^{KOM} = 2$, значением $d_{g^{KOM},1}^{KOM} = 100$ определена в виде:

$$D^{KOM} = \begin{pmatrix} 150 & 300 \\ 100 & 200 \end{pmatrix}. \quad (3)$$

Для задачи планирования, характеризуемой значениями параметров (g^{KOM}, n_g^{KOM}) вида (2, 4), матрица D^{KOM} в соответствии со

значениями отношений $d_{1,h_g}^{KOM} / d_{g,h_g}^{KOM} = 2$ и $d_{g,1}^{KOM} / d_{g,h_g}^{KOM} = 2$,

значением $d_{g^{KOM},1}^{KOM} = 100$ определена в виде:

$$D^{KOM} = \begin{pmatrix} 200 & 250 & 350 & 400 \\ 100 & 125 & 175 & 200 \end{pmatrix}. \quad (4)$$

Значения элементов d_{g,h_g}^{KOM} матриц (3) и (4) фиксировались в качестве директивных сроков формирования h_g -ых комплектов g -ых типов при проведении исследований метода планирования выполнения ПЗ в МС. Для других параметров задачи планирования (отношений $\max(t_i^l) / \min(t_i^l)$, $\max(t_{ij}^l) / \min(t_{ij}^l)$, $\max(w_{ig}) / \min(w_{ig})$, g^{KOM} , n_g^{KOM} и т.д.) значения указаны выше. Для каждой задачи планирования результаты фиксировались в виде значения критерия f_i , получаемого для локально оптимального решения по составам ПЗ.

В качестве параметра, позволяющего продемонстрировать снижение суммарного запаздывания с формированием комплектов при использовании для планирования метода оптимизации составов ПЗ [1] в сравнении с планированием для фиксированных пакетов, использовалось отношение $f_{эмоп} = (f_1^{фикс} - f_1^{моп}) / f_1^{фикс}$, где $f_1^{фикс}$ – значение критерия f_i для решения по фиксированным пакетам, $f_1^{моп}$ – значение критерия f_i для решения с оптимизированными составами ПЗ. Этот параметр характеризует степень снижения суммарного запаздывания с формированием комплектов относительно директивных сроков для фиксированных пакетов и пакетов, сформированных в соответствии с методом оптимизации их составов [1].

Графики, отображающие зависимости снижения суммарного запаздывания с формированием комплектов по отношению к директивными срокам для оптимизируемых составов ПЗ в сравнении с фиксированными ПЗ от входных параметров задачи (при $n=5$, $L=5$), представлены на рисунках 2–7.

На этих же рисунках сопоставлены степени снижения суммарных запаздываний, полученных для решений по составам ПЗ, сформированных с использованием метода оптимизации [1], и полученных при использовании генетических алгоритмов для оптимизации составов пакетов [23] (МО СП– метод оптимизации составов пакетов, ГА – генетические алгоритмы).

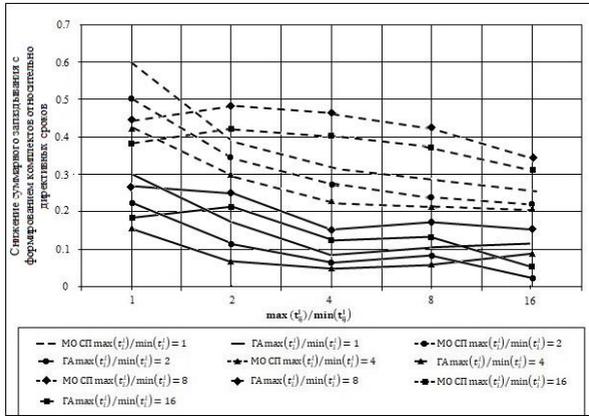


Рис. 2. Снижение суммарного запаздывания с формированием комплектов относительно директивных сроков (при $g^{KOM} = 2, n_g^{KOM} = 2,$

$$\max(w_{ig}) / \min(w_{ig}) = 1)$$

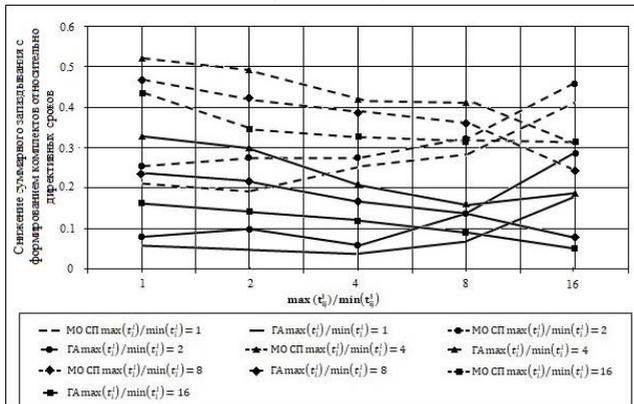


Рис. 3. Снижение суммарного запаздывания с формированием комплектов относительно директивных сроков (при $g^{KOM} = 2, n_g^{KOM} = 2,$

$$\max(w_{ig}) / \min(w_{ig}) = 2)$$

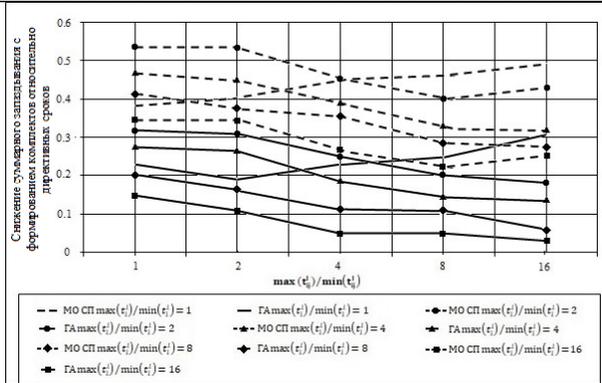


Рис. 4. Снижение суммарного запаздывания с формированием комплектов относительно директивных сроков (при $g^{KOM} = 2, n_g^{KOM} = 2,$

$$\max(w_{ig'}) / \min(w_{ig'}) = 3)$$

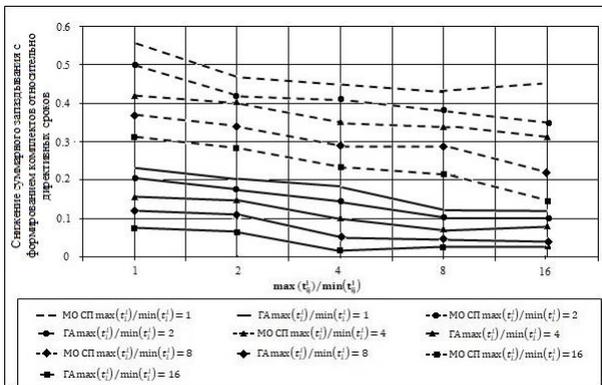


Рис. 5. Снижение суммарного запаздывания с формированием комплектов относительно директивных сроков (при $g^{KOM} = 2, n_g^{KOM} = 4,$

$$\max(w_{ig'}) / \min(w_{ig'}) = 1)$$

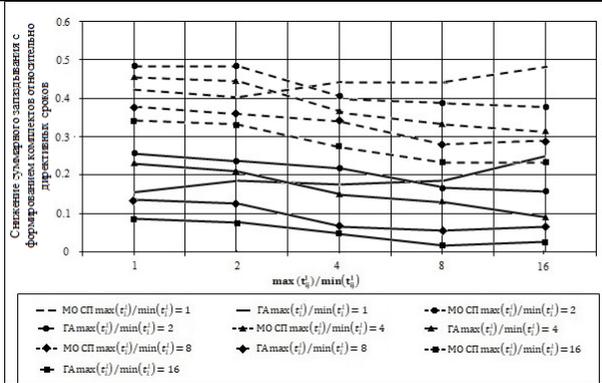


Рис. 6. Снижение суммарного запаздывания с формированием комплектов относительно директивных сроков (при $g^{KOM} = 2$, $n_g^{KOM} = 4$,

$$\max(w_{ig}) / \min(w_{ig'}) = 2)$$

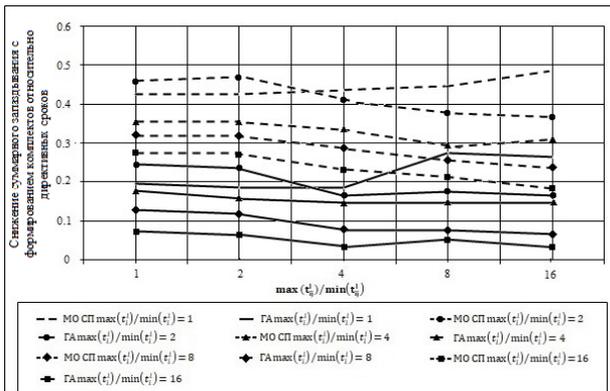


Рис. 7. Снижение суммарного запаздывания с формированием комплектов относительно директивных сроков (при $g^{KOM} = 2$, $n_g^{KOM} = 4$,

$$\max(w_{ig}) / \min(w_{ig'}) = 3)$$

Анализ полученных зависимостей показал:

- если неоднородность составов комплектов характеризуется значением 1 ($\max(w_{ig}) / \min(w_{ig'}) = 1$), то максимальное снижение суммарного запаздывания с формированием комплектов относительно директивных сроков (при

оптимизации составов ПЗ по сравнению с фиксированными ПЗ) фиксируется при $\max(t_i^l) / \min(t_i^l) = 1$;

- при увеличении неоднородности составов комплектов (значение $\max(w_{ig}) / \min(w_{ig})$ равно 2, 3) максимальное снижение суммарного запаздывания с формированием комплектов фиксируется при значениях $\max(t_i^l) / \min(t_i^l)$ равных 2, 4; таким образом, если неоднородность составов комплектов больше, то при увеличении значений отношения $\max(t_i^l) / \min(t_i^l)$ с использованием метода оптимизации ПЗ будут получены решения, обеспечивающие большее снижение запаздываний с формированием комплектов.

Для значений параметров $n=5$, $L=5$ при $g^{KOM} = 2$, $n_g^{KOM} = 2$ определены следующие особенности планирования выполнения ПЗ:

- степень снижения суммарного запаздывания с формированием комплектов относительно директивных сроков уменьшается при увеличении значений отношения $\max(t_{ij}^l) / \min(t_{ij}^l)$ для каждого значения отношения $\max(t_i^l) / \min(t_i^l)$ (то есть рост значений параметра неоднородности длительностей переналадок приборов МС приводит к ухудшению результатов планирования выполнения ПЗ в МС);
- максимальная степень снижения суммарного запаздывания с формированием комплектов, равная 0.6, зафиксирована при значениях параметров $n=5$ и $L=5$; максимальное значение степени снижения суммарного запаздывания с формированием комплектов уменьшается до 0.5 при значениях параметров $n=5$ и $L=10$.

Для значений входных параметров $n=10$ ($L=5$ или $L=10$) при $g^{KOM} = 2$, $n_g^{KOM} = 2$ определены следующие особенности планирования выполнения ПЗ:

- при малой неоднородности длительностей выполнения заданий (значения отношения $\max(t_i^l) / \min(t_i^l)$ равны 1, 2) рост

- значений параметра $\max(t_{ij}^l) / \min(t_{ij}^l)$ обуславливает увеличение степени снижения суммарного запаздывания с формированием комплектов относительно директивных сроков;
- при значительной неоднородности длительностей выполнения заданий (значения отношения $\max(t_i^l) / \min(t_i^l)$ равны 4, 8, 16) увеличение неоднородности длительностей переналадок приборов МС (значений отношения $\max(t_{ij}^l) / \min(t_{ij}^l)$) обуславливает уменьшение степени снижения значений суммарного запаздывания с формированием комплектов относительно директивных сроков.

Увеличение количества комплектов каждого типа ($g^{KOM} = 2$, $n_g^{KOM} = 4$) обуславливает следующие особенности планирования выполнения заданий:

- при $n=5$, $L=5$, малой неоднородности длительностей выполнения заданий (значения $\max(t_i^l) / \min(t_i^l)$ равны 1, 2), малой неоднородности длительностей переналадок $\max(t_{ij}^l) / \min(t_{ij}^l)$ фиксируется максимальное снижение суммарного запаздывания с формированием комплектов относительно директивных сроков, равное 0.58, которое уменьшается при росте значений $\max(t_{ij}^l) / \min(t_{ij}^l)$;
- при $n=5$, $L=5$, увеличении значений отношения $\max(t_i^l) / \min(t_i^l)$ до 4, 8 и 16 снижение суммарного запаздывания с формированием комплектов относительно директивных сроков составляет 0.31-0.42, снижение суммарного запаздывания составляет 0.2 при росте значений параметра $\max(t_{ij}^l) / \min(t_{ij}^l)$;
- при количестве приборов МС $L=10$ (при $n=5$ и малой неоднородности длительностей выполнения заданий $\max(t_i^l) / \min(t_i^l) = 1$) зависимость степени снижения

суммарного запаздывания $f_{эмосп}$ от неоднородности длительностей переналадок приборов МС аналогична этой же зависимости для случая $n=5, L=5$; степень снижения суммарных запаздываний с формированием комплектов максимальна при $\max(t_i^l) / \min(t_i^l) = 1$ (равна 0.45) и убывает до 0.35 при увеличении значений $\max(t_{ij}^l) / \min(t_{ij}^l)$;

- при количестве комплектов $n_g^{ком} = 4$ и количестве приборов МС $L=10$ максимальное снижение суммарного запаздывания фиксируется для малой неоднородности длительностей выполнения заданий (значения отношения $\max(t_i^l) / \min(t_i^l)$ равны 1, 2), при увеличении неоднородности длительностей выполнения заданий (значения отношения $\max(t_i^l) / \min(t_i^l)$ равны 4, 8, 16) степень снижения суммарного запаздывания с формированием комплектов уменьшается;
- при увеличении количества типов заданий n до 10 (при $L=5$ или $L=10$), при $g^{ком} = 2, n_g^{ком} = 4$ снижение суммарного запаздывания с формированием комплектов является максимальным при $\max(t_i^l) / \min(t_i^l) = 1$ (равно 0.33 при $\max(t_{ij}^l) / \min(t_{ij}^l) = 1$ и уменьшается до 0.25 при увеличении значения отношения $\max(t_{ij}^l) / \min(t_{ij}^l)$), при увеличении значения отношения $\max(t_i^l) / \min(t_i^l)$ до 4, 8, 16 степень снижения суммарного запаздывания с формированием комплектов уменьшается: при $\max(t_i^l) / \min(t_i^l) = 16$ она равна 0.22 для $\max(t_{ij}^l) / \min(t_{ij}^l) = 1$ и уменьшается до 0.15 по мере роста значений отношения $\max(t_{ij}^l) / \min(t_{ij}^l)$.

Реализация планирования выполнения ПЗ в МС с использованием изложенного подхода позволила в среднем на 40% снизить суммарное запаздывание с формированием комплектов по

сравнению с фиксированными пакетами (составы которых не оптимизируются).

Рассматриваемый в работе подход предполагает совместную (комплексную) оптимизацию решений на уровнях иерархии системы планирования с применением теории иерархических игр. Это позволяет выполнить решение большого количества задач теории расписаний, получение результатов в которых существующими способами (ЧЦЛП, эвристические правила, генетические алгоритмы) является затруднительным. Использование в рассматриваемой задаче планирования выполнения ПЗ в МС метода локальной оптимизации составов пакетов [1] показало возможность его применения при решении следующих задач: 1) непосредственно планирования выполнения ПЗ в МС ([1]), 2) планирования выполнения ПЗ в МС при оперативном формировании комплектов [22], 3) планирования выполнения ПЗ в МС при формировании комплектов в заданные для них директивные сроки (рассматриваемая работа), 4) планирования выполнения ПЗ в МС при задании ограничений на длительность интервалов времени ее функционирования [24]. В тоже время использование рассмотренного иерархического подхода, предусматривающего, в том числе, оптимизацию составов ПЗ, позволяет получить результаты, являющиеся лучшими, чем результаты, формируемые с использованием метаэвристических алгоритмов (в частности, ГА).

Рассматриваемая работа, являясь развитием опубликованных ранее работ [1], [22], [24], обладает следующими отличиями от них: разработана новая модель иерархической игры, позволяющая оптимизировать составы ПЗ и расписания их выполнения в МС при формировании КР в заданные директивные сроки; разработан новый метод упорядочивания идентификаторов типов комплектов с учетом директивных сроков их формирования, который используется при распределении результатов выполнения ПЗ в МС; разработан новый метод распределения результатов выполнения ПЗ в МС, позволяющий определять моменты времени окончания формирования комплектов, значения запаздываний с их формированием и, как следствие, значения критерия оптимизации, характеризующие решения по составам ПЗ.

4. Заключение. Существующие методы планирования выполнения ПЗ в МС не обеспечивают определение решений для задач большой размерности, не предусматривают определения решений по составам пакетов при условии формирования КР в заданные директивные сроки. Предложен иерархический подход к комплексному планированию выполнения ПЗ в МС. Построена модель иерархической

игры, обеспечивающая принятия решений в системе планирования, позволяющая оптимизировать составы ПЗ и расписания их выполнения с учетом формирования КР в заданные директивные сроки. Для вычисления значений критерия оптимальности решений по составам ПЗ разработан метод упорядочивания идентификаторов типов комплектов и значений директивных сроков их формирования, а также метод распределения результатов по комплектам, позволяющий определять моменты времени окончания их формирования и значения запаздывания с их формированием, входящие в аддитивный критерий на первом уровне оптимизации. Исследования эффективности комплексного планирования выполнения ПЗ в МС при условии формирования КР в заданные директивные сроки позволили идентифицировать особенности планирования, связанные с временными характеристиками этого процесса. Полученные результаты исследований показали, что используемый подход позволяет в среднем на 40% снизить запаздывания с формированием комплектов по сравнению с фиксированными пакетами.

Литература

1. *Кротов К.В.* Комплексный метод определения эффективных решений по составам партий данных и расписаниям их обработки в конвейерных системах // Журнал «Вычислительные технологии», Новосибирск, Изд-во Института вычислительных технологий СО РАН, 2018. № 3, с. 58–76.
2. *Mendez C.A.* State-of-the-art review of optimization methods for short-term scheduling of batch processes // *Computers and Chemical Engineering*, 2006, vol. 30, pp. 913–946.
3. *Adonyi R., Romero J., Puigjaner L., Friedler F.* Incorporating heat integration in batch process scheduling // *Applied Thermal Engineering*, 2003, vol. 23, pp. 1743–1762.
4. *Agha M.* Integrated management of energy and production: scheduling of batch process and Combined Heat & Power plant // University of Toulouse. National Polytechnic Institute of Toulouse (FRANCE), 2009.
5. *Zeballos L.J., Heming G.P.* A CP approach to the scheduling of resource-constrained multiproduct continuous facilities // *Latin American Applied Research*, 2006. № 36, pp. 205–212.
6. *Díaz-Ramírez J., Huertas J.I.*: A continuous time model for a short-term multiproduct batch process scheduling // *Ingeniería e Investigación*, 2018, vol. 38, 1, pp. 96-104. DOI: 10.15446/ing.investig.v38n1.66425.
7. *Ning Ch., You F.* Batch Process Scheduling under Uncertainty using Data-Driven Multistage Adaptive Robust Optimization // *Chemical engineering transactions*, 2017, vol. 61, pp. 1567-1572. DOI: 10.3303/CET1761259.
8. *Chaudhry I.A., Elbadawi I. A-Q., Usman M., Chughtai M. T.* Minimising Total Flowtime in a No-Wait Flow Shop (NWFS) using Genetic Algorithms // *Ingeniería e Investigación*, 2018, vol. 38, № 3, pp. 68-79. DOI: 10.15446/ing.investig.v38n3.75281.

9. *Ogun B., Cigdem A.-U.* Mathematical Models for a Batch Scheduling Problem to Minimize Earliness and Tardiness // *Journal of Industrial Engineering and Management*. JIEM, 2018, № 11(3), pp. 390–405. DOI: 10.3926/jiem.2541.
10. *L. X. L., Wang Y.* Scheduling Batch Processing Machine Using Max–Min Ant System Algorithm Improved by a Local Search Method // *Mathematical Problems in Engineering*. 2018, Volume 2018, Article ID: 3124182, 10 pages. DOI: 10.1155/2018/3124182.
11. *Tan Y., Huangi W., Sun Y., Yue Y.* Comparative Study of Different Approaches to Solve Batch Process Sheduling and Optimisation Problems // *Proceedings of the 18th International Conference on Automation & Computing*. Loughborough University. Leicestershire. UK. 2012, pp. 424–444.
12. *Cheng B.-Y., Chen H.-P., Wang S.-S.* Improved ant colony optimization method for single batch-processing machine with non-identical job sizes // *Journal of System Simulation*, 2009, vol. 21, № 9, pp. 2687–2695.
13. *Koehler F. Khuller S.* Optimal Batch Schedules for Parrallel Machines // *Algorithms and Data Structures: 13th International Symposium*. Berlin: Springer, 2013, pp. 475–486.
14. *Monch L., Balasubramanian H., Fowler J. W., Pfund M. E.* Heuristic scheduling of jobs on parallel batch machines with incompatible job families and unequal ready times // *Computers & Operations Research*, 2005, № 32, pp. 2731–2750.
15. *Dang Th.-T., Frankovic B., Budinska I.* Using heuristic search for solving single machine batch processing problems // *Computing and Informatics*, 2006, vol. 25, pp. 405–420.
16. *Kohn R., Rose O., Laroque Ch.* Study on multi-objective optimization for parallel batch machine scheduling using variable neighbourhood search // *Proceedings of the 2013 Winter Simulation Conference*. 8th-11th December 2013, pp. 3654-3670.
17. *Li Sh., Cheng T.C.E., Ng C.T., Yuan J.* Single-machine batch scheduling with job processing time compatibility // *Theoretical Computer Science*, 2015, vol. 583, pp. 57-66. DOI: 10.1016/j.tcs.2015.03.043.
18. *Jin M., Liu X., Luo W.* Single-Machine Parallel-Batch Scheduling with Nonidentical Job Sizes and Rejection // *Mathematics*, 2020, vol. 8, DOI: 10.3390/math8020258.
19. *Surjandari I., Rachman A., Purdianta, Dhini A.* The batch scheduling model for dynamic multi-item, multi-level production in an assembly job shop with parallel machines // *International Journal of Technology*, 2015, № 1, pp. 84-96. DOI: 10.14716/ijtech.v6i1.783.
20. *Joglekar G.* Using Simulation for Scheduling and Rescheduling of Batch Processes // *Processes*, 2017, № 5, Article ID 66. DOI: 10.3390/pr5040066.
21. *Ковалев М. Я.* Модели и методы календарного планирования. Курс лекций. Минск: БГУ, 2004.
22. *Кротов К.В.* Обоснование методов построения комплексных расписаний обработки партий данных при условии оперативного формировании комплектов из результатов // *Вестник Воронежского Государственного Университета. Серия: Системный анализ и информационные технологии*, 2018, № 4, с. 58–72.
23. *Кротов К.В.* Использование аппарата генетических алгоритмов при формировании решений по составам партий данных в двухуровневой задаче

построения комплексных расписаний их обработки // Автоматизированные технологии и производства. Международный научно-технический журнал, 2017, № 2(16). с. 23–34.

24. *Кротов К.В.* Построение комплексных расписаний обработки пакетов данных в конвейерной системе при задании ограничений на длительность интервалов времени ее функционирования // Труды учебных заведений связи, 2020, т. 6, № 3. с. 75–89.

Скатков Александр Владимирович – д-р техн. наук, профессор, профессор, кафедра «Информационные технологии и компьютерные системы», Севастопольский государственный университет. Область научных интересов: аналитическое и имитационное моделирование систем неоднородной структуры, исследование операций, системотехнический анализ и проектирование вычислительных комплексов, систем и сетей. Число научных публикаций – 300. Vm1945@mail.ru. ул. Университетская, 33, Севастополь. 299053. ORCID 0000-0002-5678-9587

Кротов Кирилл Викторович — канд. техн. наук, доцент, доцент, кафедра «Информационные системы», Севастопольский государственный университет. Область научных интересов: моделирование процессов в вычислительных системах, оптимальное планирование процессов в вычислительных системах, иерархические игры. Число научных публикаций – 110. krotov_k1@mail.ru. ул. Университетская, д. 33, Севастополь, 299053. ORCID 0000-0002-9670-6141

K. Krotov, A. Scatkov

CONSTRUCTION OF COMPLEX SCHEDULES FOR THE EXECUTION OF TASK PACKAGES WHEN FORMING SETS IN THE SPECIFIED DIRECTIVE TERMS

Krotov K., Scatkov A. Construction of Complex Schedules for Execution of Task Packages at Forming Sets in Specified Directive Terms.

Abstract. The current state with the solution of the problem complex planning of the execution of task packets in multistage system is characterized by the absence of universal methods of forming decisions on the composition of packets, the presence of restrictions on the dimension of the problem and the impossibility of guaranteed obtaining effective solutions for various values of its input parameters, as well the impossibility of registration the condition of the formation of sets from the results. The solution of the task of planning the execution of task packets in multistage systems with the formation of sets of results within the specified deadlines has been realized by the authors. To solve the planning problem, the generalized function of the system was decomposed into a set of hierarchically interrelated subfunctions. The use of decomposition made it possible to use a hierarchical approach for planning the execution of task packets in multistage systems, which involves defining solutions based on the composition of packets at the top level of the hierarchy and scheduling the execution of packages at the bottom level of the hierarchy. The theory of hierarchical games is used to optimize solutions for the compositions of task packets and schedules for their execution, which is a system of criteria at the decision-making levels. Evaluation of the effectiveness of decisions by the composition of packets at the top level of the hierarchy is ensured by the distribution of the results of task execution by packets in accordance with the formed schedule. To evaluate the effectiveness of decisions on the composition of packets, method for ordering the identifiers of the types of sets with registration of the deadlines and a method for distributing the results of the tasks performed by packets has been formulated, which calculates the moments of completion of the formation of sets and delays with their formation relative to the specified deadlines. The studies of planning the process of the executing task packages in multistage systems have been carried out, provided that the sets are formed within specified deadlines. On their basis, conclusions, regarding the dependence of the planning efficiency from the input parameters of the problem, were formulated.

Keywords: Multi-Stage System, Building Packages, Schedule, Result Sets, Directive DeadLines, Hierarchical Game

References

1. Krotov K.V. [A complex method for determining effective solutions for the composition of data batches and schedules of their processing in conveyor systems]. *ZHurnal «Vychislitel'nye tekhnologii»*, *Izd-vo Instituta vychislitel'nyh tekhnologij SO RAN—Computational technologies, Publishing house Of the Institute of computational technologies SB RAS*, Vol. 23, № 3, 2018. pp. 58-76 (In Russ).
2. Mendez C.A. et al. State-of-the-art review of optimization methods for short-term scheduling of batch processes // *Computers and Chemical Engineering*, 2006, vol. 30, pp. 913–946.
3. Adonyi R., Romero J., Puigjaner L., Friedler F. Incorporating heat integration in batch process scheduling // *Applied Thermal Engineering*, 2003, vol. 23, pp. 1743–1762.
4. Agha M. Integrated management of energy and production: scheduling of batch process and Combined Heat& Power plant // University of Toulouse. National Polytechnic Institute of Toulouse (FRANCE), 2009, 255 p.

5. Zeballos L.J., Henning G.P. A CP approach to the scheduling of resource-constrained multiproduct continuous facilities // *Latin American Applied Research*, 2006, № 36, pp. 205-212.
6. Díaz-Ramírez J., Huertas J.I. A continuous time model for a short-term multiproduct batch process scheduling // *Ingeniería e Investigación*, 2018, vol. 38, №1, pp.96-104. DOI: 10.15446/ing.investig.v38n1.66425
7. Ning Ch., You F. Batch Process Scheduling under Uncertainty using Data-Driven Multistage Adaptive Robust Optimization // *Chemical engineering transactions*, 2017, Vol. 61, pp. 1567-1572. DOI: 10.3303/CET1761259
8. Chaudhry I.A., Elbadawi I. A-Q., Usman M., Chugtai M. T. Minimising Total Flowtime in a No-Wait Flow Shop (NWFS) using Genetic Algorithms. // *Ingeniería e Investigación*, 2018, vol. 38, № 3, pp.68-79. DOI: 10.15446/ing.investig.v38n3.75281
9. Ogun B., Cigdem A.-U. Mathematical Models for a Batch Scheduling Problem to Minimize Earliness and Tardiness. // *Journal of Industrial Engineering and Management. JIEM*, 2018, # 11(3), p. 390-405. DOI: 10.3926/jiem.2541.
10. Li X. L., Wang Y. Scheduling Batch Processing Machine Using Max–Min Ant System Algorithm Improved by a Local Search Method. // *Mathematical Problems in Engineering*. 2018, Volume 2018, Article ID: 3124182, 10 pages. DOI: 10.1155/2018/3124182.
11. Tan Y., Huang W., Sun Y., Yue Y. Comparative Study of Different Approaches to Solve Batch Process Scheduling and Optimisation Problems. // *Proceedings of the 18th International Conference on Automation & Computing*. Loughborough University, Leicestershire. UK. 2012, pp. 424–444.
12. Cheng B.-Y., Chen H.-P., Wang S.-S. Improved ant colony optimization method for single batch-processing machine with non-identical job sizes // *Journal of System Simulation*, 2009, vol. 21, # 9, pp. 2687–2695.
13. Koehler F. Khuller S. Optimal Batch Schedules for Parallel Machines // *Algorithms and Data Structures: 13th International Symposium*. Berlin: Springer, 2013, pp. 475–486.
14. Monch L., Balasubramanian H., Fowler J. W., Pfund M. E. Heuristic scheduling of jobs on parallel batch machines with incompatible job families and unequal ready times // *Computers & Operations Research*, 2005, # 32, pp. 2731–2750.
15. Dang Th.-T., Frankovic B., Budinska I. Using heuristic search for solving single machine batch processing problems // *Computing and Informatics*, 2006, Vol. 25, pp. 405–420.
16. Kohn R., Rose O., Laroque Ch. Study on multi-objective optimization for parallel batch machine scheduling using variable neighbourhood search // *Proceedings of the 2013 Winter Simulation Conference*. 8th-11th December 2013, pp. 3654-3670.
17. Li Sh., Cheng T.C.E., Ng C.T, Yuan J. Single-machine batch scheduling with job processing time compatibility // *Theoretical Computer Science*, 2015, Vol. 583, pp. 57-66. DOI: 10.1016/j.tcs.2015.03.043.
18. Jin M., Liu X., Luo W. Single-Machine Parallel-Batch Scheduling with Nonidentical Job Sizes and Rejection // *Mathematics*, 2020, Vol. 8, Article ID:258. DOI: 10.3390/math8020258.

19. Surjandari I., Rachman A., Purdianta, Dhini A. The batch scheduling model for dynamic multi-item, multi-level production in an assembly job shop with parallel machines // International Journal of Technology, 2015, №1, pp. 84-96. DOI: 10.14716/ijtech.v6i1.783.
20. Joglekar G. Using Simulation for Scheduling and Rescheduling of Batch Processes // Processes, 2017, № 5, Article ID 66. DOI: 10.3390/pr5040066.
21. Kovalev M.M. [Models and methods of calendar planning: a course of lectures]. *Minsk: Izdatelstvo BGU – Publishing house of BSU*. 2004. 63 p. (In Russ).
22. Krotov K. V. [Justification of methods for constructing complex schedules for processing data batches under the condition of rapid formation of sets of results]. *Vestnik Voronezhskogo Gosudarstvennogo Universiteta. Seriya: Sistemyj analiz i informacionnye tekhnologii – Bulletin of The Voronezh State University. Series: System analysis and information technologies*. №4. 4, 2018, pp. 58–72 (In Russ).
23. Krotov K. V. [The use of the apparatus of genetic algorithms in the formation of decisions on the composition of data batches in the two-level task of constructing complex schedules for their processing]. *Avtomatizirovannye tekhnologii i proizvodstva. Mezhdunarodnyj nauchno-tekhnicheskij zhurnal. – Automated technologies and production. International scientific and technical journal*, 2017, №2 (16), pp. 23–34 (In Russ).
24. Krotov K. V. [Construction of complex schedules for processing data packets in a pipeline system when setting restrictions on the duration of time intervals of its operation]. *Trudy uchebnyh zavedenij svyazi. – Proceedings of educational institutions of communication*, 2020, vol. 6, No. 3. pp. 75-89 (In Russ).

Skatkov Aleksandr — Ph.D., Dr.Sci., Professor, Professor, Department of Information technologies and computer systems, Sevastopol state University. Research interests: analytical and simulation modeling of systems of heterogeneous structure, operations research, system analysis and design of computer systems, systems, and networks. Number of scientific publications – 300. Vm1945@mail.ru. 33 Universitetskaya St., Sevastopol, 299053. ORCID 0000-0002-5678-9587

Krotov Kirill — Ph.D., Associate Professor, Department of Information System, Sevastopol State University. Research interests: modeling of processes in computer systems, optimal planning of processes in computer systems, hierarchical games. Number of scientific publications – 110. krotov_k1@mail.ru. 33 Universitetskaya st., Sevastopol, 299053. ORCID 0000-0002-9670-6141

В.А. КОСТЮКОВ, М.Ю. МЕДВЕДЕВ, В.Х. ПШИХОПОВ
**МЕТОД ОПТИМИЗАЦИИ ТРАЕКТОРИИ МОБИЛЬНОГО
РОБОТА В ПОЛЕ ИСТОЧНИКОВ-РЕПЕЛЛЕРОВ**

Костюков В.А., Медведев М.Ю., Пшихопов В.Х. Метод оптимизации траектории мобильного робота в поле источников-репеллеров.

Аннотация. Рассматривается процедура корректировки траектории движения робототехнической платформы (РТП) на плоскости с целью снижения вероятности её поражения/обнаружения в поле конечного числа источников-репеллеров. Каждый из таких источников описан математической моделью некоторого фактора противодействия целостности или скрытности РТП.

Указанная процедура основана, с одной стороны, на понятии характерной вероятностной функции системы источников-репеллеров, позволяющем оценивать степень влияния этих источников на движущуюся РТП. Из этого понятия вытекает используемая здесь в качестве показателя оптимизации целевой траектории вероятность её успешного прохождения. С другой стороны, эта процедура базируется на решении локальных оптимизационных задач, позволяющих корректировать отдельные участки исходной траектории с учетом нахождения в их окрестностях конкретных источников-репеллеров с заданными параметрами. Каждый из таких источников характеризуется потенциалом, частотой воздействия, радиусом действия и параметрами спада поля.

Корректировка траектории происходит итерационно и учитывает целевое значение вероятности прохождения. Основным ограничением на вариацию исходной траектории является максимально допустимое отклонение измененной траектории от исходной. Если такого ограничения нет, то задача может потерять смысл, поскольку тогда можно выделить область, охватывающую все препятствия и источники, и обойти её по периметру. Поэтому осуществляется поиск такого локального экстремума, который соответствует допустимой кривой в смысле указанного ограничения.

Предлагаемая в настоящей работе итерационная процедура позволяет проводить поиск соответствующих локальных максимумов вероятности прохождения РТП в поле нескольких произвольно расположенных и ориентированных источников в некоторой окрестности исходной траектории.

Вначале ставится и решается задача оптимизации траектории при условии движения в поле одного источника с областью действия в виде кругового сектора, затем полученный результат распространяется на случай нескольких аналогичных источников.

Основной проблемой исследования является выбор общего вида функционала в каждой точке исходной кривой, а также его коэффициентов настройки. Показано, что выбор этих коэффициентов настройки есть адаптивная процедура, входными переменными которой являются характерные геометрические величины, описывающие текущую траекторию в поле источников.

Для устранения осцилляций, возникающих вследствие локальности предлагаемой процедуры, применяются стандартные процедуры медианного сглаживания.

Результаты моделирования показывают высокую эффективность предложенной процедуры для корректировки ранее спланированной траектории.

Ключевые слова: планирование пути, конфликтная среда, оптимизация движения, вероятность успешного прохождения, группа источников-репеллеров

1. Введение. Распространенным подходом к решению задачи учета внешних воздействий различной природы на движущуюся РТП (например, радиолокационные и прочие воздействия с целью её обнаружения) в условиях неопределенности является применение метода потенциальных полей [1-5], функций риска [6-14], и методов, используемых для обхода геометрических препятствий [15–21].

Возможности использования обобщенных функций риска в решении оптимизационных задач при наличии неопределенностей подробно рассмотрены в общем обзоре [6].

В работе [7] вводится функция риска, заданная в каждой точке области пространства, в которой расположены источники. Решается задача минимизации интеграла от функции риска при движении вдоль целевой траектории при различных условиях, среди которых можно выделить ограничения по длине пути. Для одного источника данная задача решается аналитическими методами вариационного исчисления. Для многих источников в [7] предложен алгоритм нелинейного программирования, применяемый для графа, учитывающего весовые коэффициенты ребер и соответствующие стоимости прохождения.

В работах [8, 9] показано, что в случае объекта, движущегося с переменной скоростью в окрестности одного источника, оптимальная траектория и режим изменения скорости таковы, что текущее значение суммарного «сигнала» от источника должно быть постоянно во времени. В работах [9–12] это свойство обобщено на случаи нескольких источников, подвижных и разнородных наблюдателей.

В статьях [13, 14] рассмотрены некоторые возможные стратегии поведения подвижных объектов в группе, увеличивающие эффективность проникновения через систему обороны, состоящую из нескольких источников.

Общей особенностью работ [6–14] является использование функции риска и отсутствие методов расчета вероятностей обнаружения или поражения. Например, экстремальная траектория, полученная численным методом оптимизации рисков на сетевом графе [6], требует вычисления указанной вероятности, так как знание оптимальной траектории еще не гарантирует выполнение миссии подвижным объектом/группой объектов, если нет способа рассчитать вероятность ее выполнения. Наличие методики расчета вероятности выполнения миссии позволит, в некоторых случаях, выбрать приемлемые траектории.

Одним из возможных подходов при решении задач по определению указанных вероятностей является введение аппарата характеристических вероятностных функций [22]. На основании этого

метода предложена итерационная процедура поиска траектории, минимально отклоняющейся от исходной и характеризующейся вероятностью прохождения не ниже заданного значения.

В работе [22] каждому из средств поражения/обнаружения, воздействующих на движущуюся РТП, ставится в соответствие его приближенная модель – источник полей, на основании которых может быть рассчитана вероятность успешного прохождения траектории. Под «успешным прохождением» траектории понимается: в случае источников обнаружения – факт того, что РТП не была обнаружена; в случае источника какого-либо поражающего фактора – факт того, что РТП не была поражена этим источником.

Важной проблемой является корректный учет формы той части пространства, в которой источник влияет на движущуюся в его окрестности РТП. В достаточно общем случае эту часть пространства можно описать круговым сектором, опирающимся на произвольный угол, принадлежащий интервалу от 0 до 2π . В случае независимости источника от азимутального угла эта область представляет собой круг. При решении ряда практически важных задач возможно считать радиус этого круга бесконечным, что значительно упрощает анализ, поскольку не нужно использовать функции принадлежности конечных областей влияния источников [22].

При рассмотрении указанного выше случая конечной области влияния источника необходимо дать общий алгоритм поведения объекта внутри этой области, который базируется на двух факторах. Первым фактором является стремление удалиться как можно дальше вдоль радиальной координаты от центра источника возмущений. Вторым фактором является стремление объекта как можно быстрее выйти из этой области. Основной особенностью развиваемого в настоящей статье подхода является сочетание этих двух факторов адекватно текущему положению объекта внутри области источника и его направлению движения относительно границ этой области.

При движении в поле конечного числа источников указанного типа разработан обобщенный алгоритм, базирующийся на сочетании глобального планирования пути от исходной точки до целевой с локальными отклонениями от глобального пути в окрестностях центров соответствующих источников, получаемыми на основе предлагаемой процедуры оптимизации.

Заметим, что как предлагаемый в настоящей статье, так и другие указанные выше алгоритмы обхода препятствий и уклонения от влияния источников различных воздействий могут применяться не только в отношении РТП, но также и в отношении различных

экипажных/ пилотируемых или дистанционно управляемых аппаратов. Причем наибольшую эффективность эти методы имеют в случае стационарных или даже заранее известных источников/препятствий. Тогда можно просто заранее спланировать программные траектории движения, оптимальные с точки зрения некоторого критерия, например минимизации риска или максимизации вероятности успешного прохождения траектории, и существенно облегчить исполнение оператором своих функций.

Однако в целом ряде реальных ситуаций, когда источники и препятствия могут быть существенно недетерминированными и принципиально не подлежащими какой-либо предварительной оценке по своим ключевым параметрам, применение таких методов требует использования автономной системы управления с соответствующими ограничениями по быстродействию.

2. Процедура оптимизации пути робототехнической платформы в поле нескольких источников возмущений.

Рассмотрим K источников, область действия каждого из которых $U_{S,k}$ представляет собой круговой сектор с центром в точке $O^{(k)}$, направляющим вектором средней линии $n_s^{(k)}$, радиусом $d_0^{(k)}$, углом $\Delta\alpha_s^{(k)}$, стягивающим этот сектор. Внутри каждой такой области задана характеристическая вероятностная функция (ХВФ) соответствующего источника, представляющая собой вероятность успешного прохождения достаточно малого в пределе элементарного отрезка в течение малого времени T_k ; отрезок считается малым настолько, чтобы можно было пренебречь его ориентацией относительно источника и переменностью воздействия последнего на РТП, двигающуюся по указанному отрезку. Центр отрезка находится в точке M с полярными координатами d_k, α_k , заданными в собственной системе отсчета источника. При этих условиях ХВФ есть непрерывное поле на фазовом пространстве $\mathbb{R}_2 \otimes (0, T_\delta)$, \mathbb{R}_2 есть двумерная евклидова плоскость, а $(0, T_\delta)$ есть область на дополнительной евклидовой оси \mathbb{R}_1 , T_δ имеет размерность времени. В работе [22] было предложено и обосновано следующее возможное аналитическое представление для ХВФ:

$$q_{0,S_k}(d_k, \alpha_k, T_k) = \begin{cases} \exp\left\{-\varepsilon^{(k)} T_k \exp\left[-(d_k / d_m^{(k)})^{2p_k} - \xi_{\alpha,k} \left(|\alpha_k / \alpha_m^{(k)}|\right)^{2q_k}\right]\right\}, \\ \text{при } M \in U_{S,k}; \\ 1, \text{ иначе,} \end{cases} \quad (1)$$

где $T_k \in (0, T_\delta)$, параметр T_δ определяется так, чтобы значение $q_{0,s_k}(d_k, \alpha_k, T_k)$ давало с заданной погрешностью вероятность успешного прохождения РТП вдоль указанного отрезка, причем при $T_\delta \rightarrow 0$ эта погрешность стремится к нулю. Величина $\varepsilon^{(k)}$ есть характерная частота воздействия источника на РТП; увеличение $\varepsilon^{(k)}$ снижает вероятность прохождения данного отрезка. Параметр d_k – расстояние от центра источника $O^{(k)}$ до точки M , α_k – угол отклонения вектора $O^{(k)}M$ относительно нормали $n_s^{(k)}$, причем $\alpha_k \geq 0$ при отклонении влево от нормали; p_k и q_k – положительные числа; $d_m^{(k)}$ – эффективный радиус источника, $d_m^{(k)}$ может быть не равен радиусу сектора $d_0^{(k)}$; $\alpha_m^{(k)}$ – эффективный угол отклонения от нормали. Коэффициент $\xi_{a,k} = 1$ в случае учета неравномерности углового распределения по сектору, если $\xi_{a,k} = 0$ в противном случае.

Обратим внимание, что для обозначения ХВФ у переменной q используется дополнительный нижний индекс «0», а для обозначения вероятностей прохождения конечных (не элементарных) участков траекторий этот нижний индекс отсутствует.

Каждый источник будем характеризовать набором параметров:

$$S^{(k)} \{O^{(k)}, n_s^{(k)}, d_0^{(k)}, d_m^{(k)}, \Delta\alpha_s^{(k)}, \alpha_m^{(k)}, \varepsilon^{(k)}\}, k = 1, 2, \dots, K. \quad (2)$$

Пусть РТП движется по траектории Tr , заданной совокупностью прямолинейных отрезков $A_l A_{l+1}$, $l=1, 2, \dots, N-1$; эти отрезки порождены узловыми точками A_l , $l=1, 2, \dots, N$, $N > 3$ – натуральное число.

Рассчитаем вероятность прохождения РТП по траектории Tr , которая полностью или частично принадлежит хотя бы одной из областей действия источников. В работе [22] получена общая формула расчета такой вероятности для случая, когда областью действия каждого источника является вся евклидова плоскость \mathbb{R}_2 , что равносильно $d_0^{(k)} = \infty$, $\Delta\alpha_s^{(k)} = 2\pi$ во введенных выше обозначениях.

Выражение (1) есть обобщение ХВФ, полученной в [22], на случай произвольного сектора влияния источника, который может быть как ограниченным, так и нет. В этом случае вероятность успешного прохождения траектории Tr в приближении постоянства v линейной скорости движения РТП рассчитывается по формулам:

$$\begin{aligned}
 q(Tr) &= \prod_{k=1}^K q_{S_k}(Tr) = \prod_{k=1}^K \lim_{n \rightarrow \infty} \left[\prod_{i=1}^n q_{0,S_k}(d_{k,i}, \alpha_{k,i}, \frac{|\Delta l_i|}{v}) \right] \\
 &= \exp \left[- \sum_{k=1}^K \varepsilon^{(k)} \cdot I^{(k)} \right],
 \end{aligned} \quad (3)$$

$$\begin{aligned}
 I^{(k)} &\equiv \frac{\sqrt{\pi} d_m^{(k)}}{2v} \sum_{l=1}^{N-1} \eta_l^{(k)}(A_l, A_{l+1}) \left\{ \exp \left[- \left(\frac{d_l^{(k)} \sin \varphi_l^{(k)}}{d_m^{(k)}} \right)^2 \right] \times \right. \\
 &\times \left. \left[\text{sign}(\tilde{\rho}_{2,l}^{(k)}) \text{erf}(|\tilde{\rho}_{2,l}^{(k)}|) - \text{sign}(\tilde{\rho}_{1,l}^{(k)}) \text{erf}(|\tilde{\rho}_{1,l}^{(k)}|) \right] \right\},
 \end{aligned} \quad (4)$$

$$\tilde{\rho}_{1,l}^{(k)} = \frac{d_l^{(k)} \cos \varphi_l^{(k)}}{d_m^{(k)}}, \quad \tilde{\rho}_{2,l}^{(k)} = \tilde{\rho}_{1,l}^{(k)} + \frac{\rho}{d_m^{(k)}}. \quad (5)$$

где $q_{S_k}(Tr)$ – вероятность прохождения траектории под действием источника $S^{(k)}$; $d_{(i)}^{(k)}(\rho)$ – расстояние от центра k -го источника до точки интегрирования на линейном отрезке $A_l A_{l+1}$. Здесь в формуле (3) с целью получения представления для $q_{S_k}(Tr)$ рассматривается произведение соответствующих ХВФ для всех достаточно малых интервалов $\{\Delta l_i\}_{i=1}^n$, на которые разбита траектория Tr , и далее совершается предельный переход при бесконечном измельчении таких разбиений. Функция $\eta_l^{(k)}$ определяет принадлежность хотя бы одной точки отрезка $A_l A_{l+1}$ к области k -го источника:

$$\eta_l^{(k)}(A_l, A_{l+1}) = \begin{cases} 1, & \text{если } A_l \in U_{S,k} \text{ или } A_{l+1} \in U_{S,k}; \\ 0, & \text{иначе.} \end{cases} \quad (6)$$

Величина $d_{(i)}^{(k)}(\rho)$ вычисляется следующим образом:

$$d_{(i)}^{(k)}(\rho) = \sqrt{\rho^2 + (d_l^{(k)})^2 + 2d_l^{(k)} \rho \cos \varphi_l^{(k)}}, \quad (7)$$

где $d_l^{(k)} = |O^{(k)} A_l|$; $\varphi_l^{(k)}$ – угол между направлением $O^{(k)} A_l$ и вектором $A_l A_{l+1}$. Соответствующие геометрические построения приведены на рисунке 1 для случая $K=2$. Знак приближенного равенства соответствует случаю,

когда только часть интервала $A_i A_{i+1}$ принадлежит сектору k -го источника, согласно строблирующему свойству функции $\eta_i^{(k)}(A_i, A_{i+1})$. При выборе достаточно большого числа узловых точек $N=N_{gr}$ погрешность расчета по формуле (3) будет достаточно малой, и мы будем далее предполагать $N > N_{gr}$.

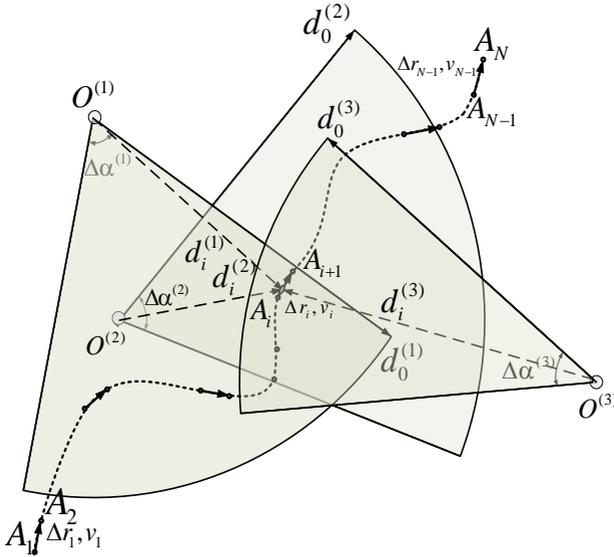


Рис. 1 Случай прохождения исходной траектории в поле действия трех источников

Поставим задачу найти максимум функционала (3) на множестве всевозможных кусочно-линейных траекторий с числом узловых точек не более N_{lim} и закрепленными концами:

$$q(Tr) \rightarrow \max,$$

при условиях: A_1 и A_N – фиксированы, $N_{gr} \leq N \leq N_{lim}$,

$$\Delta d(Tr_0, Tr_{opt}) \leq \Delta d_z, \quad (8)$$

где A_1 и A_N – начальная и конечная точки траектории Tr , $\Delta d(Tr_0, Tr_{opt})$ – отклонение оптимизированной кривой Tr_{opt} от исходной Tr_0 , которое не должно превышать заданного значения Δd_z .

В связи со сложностью решения задачи (8), преобразуем её в более простую так, чтобы отклонение её решения от соответствующего решения задачи (8) было практически приемлемым, и решение более простой задачи позволяло синтезировать более эффективную по быстродействию процедуру поиска траекторий с заданной вероятностью прохождения.

Таким образом, требуется найти такой функционал $G(Tr, p)$, чтобы решение оптимизационной задачи:

$$\begin{aligned} G(Tr, p) &\rightarrow \max, \\ &\text{при условиях:} \\ A_1 \text{ и } A_N &\text{ – фиксированы, } N_{gr} \leq N \leq N_{lim}; \\ \Delta d(Tr_0, Tr_{opt}) &\leq \Delta d_z, \end{aligned} \quad (9)$$

давало траекторию \widehat{Tr} , которая отклоняется от решения задачи (8) не более чем на заданную величину. Здесь p есть вектор параметров настройки функционала G .

Может быть дано следующее возможное решение задачи (9). Вначале для фиксированной узловой точки кусочно-линейной траектории построим некоторую новую точку так, чтобы ХВФ какого-либо из источников в новой точке имела значение большее, чем в исходной. При этом мы можем наложить ограничение на степень удаления новой точки от исходной. Затем точно так же строим аналогичные точки для всех других источников. В результате получаем некоторый набор точек, соответствующий данному узлу. Взвешенная сумма векторов, проведенных из данного узла до каждой из точек полученного набора, характеризует новую эквивалентную точку, учитывающую степень влияния каждого из источников так, что вектору большей длины соответствует и больший вклад во взвешенной сумме. Это означает, что от источника, которому соответствует вектор большей длины, нужно удалиться в первую очередь, он является более критичным, чем источник, которому соответствует вектор меньшей длины.

После получения эквивалентных точек для каждого узла указанным способом локальной оптимизации получаем некоторую новую кусочно-линейную траекторию движения РТП, которая, однако, не дает решение задачи (9), поскольку не учитывает взаимосвязь отдельных локальных участков искомой оптимальной кривой.

Этот подход лишь в ограниченном числе случаев дает решения, в достаточной степени приближенные к глобальному оптимуму вероятности прохождения. Речь идет о движении в поле источников,

удаленных друг от друга настолько, что эффект интерференции их полей проявляется слабо.

Процедура оптимизации (9) должна учитывать положение, по крайней мере, соседних точек, так как достижение локальных минимумов для отдельной точки может вступать в противоречие с достижением максимума вероятности прохождения на линейном отрезке. Аналогичный вывод справедлив и в отношении произвольного участка кусочно-ломаной, объединяющего более одного отрезка.

Локальная оптимизация может приводить к возникновению осцилляций полученной кусочно-линейной кривой. Причиной этого является независимость таких локальных процедур оптимизации. Эти осцилляции повышают длину кривой, что уменьшает вероятность ее успешного прохождения.

Хотя некоторые локальные максимумы или/и глобальный максимум исходной задачи (8) могут соответствовать и сильным отклонениям результирующих узловых точек от указанных выше эквивалентных точек, мы их рассматривать не будем, поскольку такие отклонения одновременно означают и большие смещения от исходных положений узлов, что может противоречить ограничению 2) в (8) и (9).

Высказанные соображения позволяют предложить в качестве функционала $G(Tr, p)$ в (9) специальный функционал, позволяющий:

а) учесть связанность соседних точек в плане достижения глобального максимума; б) убрать осцилляции кривой после локальных оптимизаций следующего вида:

$$G(Tr, p) = -\delta_G \sum_{i=2}^{N-1} [(x'_i - x_i)^2 + (y'_i - y_i)^2] - \delta_L \sum_{i=2}^{N-1} [(x'_i - x_{loc,i})^2 + (y'_i - y_{loc,i})^2]. \quad (10)$$

Здесь первое слагаемое есть сумма квадратов отклонений координат каждой целевой точки $A'_i(x'_i, y'_i)$ от соответствующей ей исходной $A_i(x_i, y_i)$ до оптимизации, а второе – сумма квадратов отклонений координат каждой целевой точки $A'_i(x'_i, y'_i)$ от точки $A_{loc,i}(x_{loc,i}, y_{loc,i})$, полученной в результате решения локальной оптимизационной задачи для $A_i(x_i, y_i)$. Коэффициенты $0 \leq \delta_G, \delta_L \leq 1$ характеризуют влияние факторов минимизации отклонения от исходной кривой и от кривой, полученной после локальных оптимизаций узловых точек.

Пусть точки $A_{loc,i}(x_{loc,i}, y_{loc,i})$, $i = 1, 2, \dots, N$ получены каким-либо образом, тогда рассмотрим задачу максимизации функционала (10).

Стационарная точка (10) по переменным $\{x'_i, y'_i\}$, $i = 2, 3, \dots, N-1$ получается из двух систем линейных уравнений относительно векторов

$$X' \equiv [x'_1, x'_2, \dots, x'_N]^T, Y' \equiv [y'_1, y'_2, \dots, y'_N]^T \quad (11)$$

следующим образом:

$$X'_{st} \equiv [x'_{st,1}, x'_{st,2}, \dots, x'_{st,N}]^T = A^{-1}B_X, Y'_{st} \equiv [y'_{st,1}, y'_{st,2}, \dots, y'_{st,N}]^T = A^{-1}B_Y, \quad (12)$$

$$[A] = \begin{bmatrix} -\tilde{\delta} & \delta_G & 0 & 0 & \dots & 0 & 0 & 0 \\ \delta_G & -\tilde{\delta} & \delta_G & 0 & \dots & . & 0 & 0 \\ 0 & \delta_G & -\tilde{\delta} & \delta_G & \dots & 0 & . & 0 \\ 0 & 0 & \delta_G & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & . & . & 0 & 0 \\ 0 & . & 0 & \dots & . & -\tilde{\delta} & \delta_G & \delta_G \\ 0 & 0 & . & \dots & 0 & \delta_G & -\tilde{\delta} & \delta_G \\ 0 & 0 & 0 & \dots & 0 & 0 & \delta_G & -\tilde{\delta} \end{bmatrix}, \quad (13)$$

$$B_1 = \begin{bmatrix} -\delta_L x_{loc,2} - \delta_G x_1, & -\delta_L x_{loc,3}, & \dots, & -\delta_L x_{loc,N-2}, \\ -\delta_L x_{loc,N-1} - \delta_G x_N \end{bmatrix}^T, \quad (14)$$

$$B_2 = [-\delta_L y_{loc,2} - \delta_G y_1, -\delta_L y_{loc,3}, \dots, -\delta_L y_{loc,N-2}, -\delta_L y_{loc,N-1} - \delta_G y_N]^T.$$

Можно показать, что в найденной стационарной точке, которой соответствует набор точек $A'_{st,i}(x'_{st,i}, y'_{st,i})$, $i = 2, 3, \dots, N-1$, выполняется достаточное условие локального максимума по минорам матрицы Гессе при значениях δ_G, δ_L таких, что $0 \leq \delta_G, \delta_L \leq 1$, $\delta_G + \delta_L = 1$. В этой же точке достигается и глобальный максимум:

$$A'_{opt,i}(x'_{opt,i}, y'_{opt,i}) = A'_{st,i}(x'_{st,i}, y'_{st,i}), i = 2, 3, \dots, N-1,$$

поскольку при удалении из указанного локального максимума вдоль любого направления и на любое расстояние функция $G(Tr, p)$ будет всегда монотонно убывать в силу отсутствия локальных минимумов.

3. Локальное преобразование для корректировки траектории при движении в поле одного источника. Рассмотрим задачу нахождения точек $A_{loc,i}(x_{loc,i}, y_{loc,i}), i = 1, 2, \dots, N$, которые далее называются точками локального смещения. Вначале рассмотрим движение РТП в поле действия одного источника. Будут использоваться все переменные, введенные выше для каждого k -го источника, но применительно к единственному источнику, поэтому они пишутся без индекса « k ».

Требуется сместить точку M в новое положение M' , при условии $M \in U_S$, так, чтобы в M' значение ХВФ было больше, чем в M при одном и том же T , и, во-вторых, характеризовалось большей степенью приближения к границам сектора U_S .

Пусть даны сектор $U_S \{O, n_s, d_0, \Delta\alpha_s, \varepsilon\}$ и исходная траектория Tr , пересекающая этот сектор (см. рис. 2). Часть траектории Tr , которая проходит внутри сектора, обозначим Tr_l . Также рассмотрим произвольную систему точек $M_i, i=1, 2, \dots, n$, разбивающую Tr_l на сегменты $\Delta r_i, i=1, 2, \dots, n-1$. Далее для простоты в (1) будем брать $k_a = 0$.

Найдем параметрическое преобразование:

$$M' = F(S, Tr; \rho; M), \quad (15)$$

где ρ – векторный параметр, переводящий произвольную внутреннюю для сегмента точку M , принадлежащую одновременно Tr_l , в новую точку M' таким образом, что выполнены следующие два условия.

Условие А. Если $M=M_k$ есть произвольная точка из последовательности $\{M_i\}, k \leq n$, тогда вероятность $q_s(Tr_{l,M})$ успешного прохождения той части $Tr_{l,M}$ траектории Tr_l , которая соответствует только узловым точкам подпоследовательности $\{M_i\}, i=k, k+1, \dots, n$, меньше вероятности $q_s(Tr'_{l,M})$ успешного прохождения соответствующей смещенной траектории $Tr'_{l,M}$, полученной как кусочно-ломаная кривая из узловых точек $Tr_{l,M}$ в результате применения к каждой из них преобразования F :

$$q_S(Tr'_{M'}) > q_S(Tr_{I,M}). \quad (16)$$

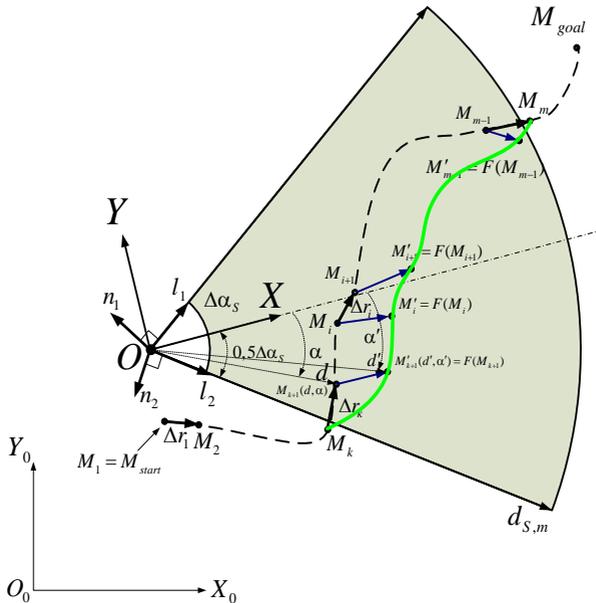


Рис. 2. К постановке задачи синтеза локальной процедуры оптимизации

Условие В. Последовательное применение преобразования F к смещенным траекториям типа $Tr'_{M'_k}$, где $M'_k = F(M_k)$, позволяет прийти за ограниченное число шагов N_{lim} к такой траектории Tr_{lim} , что вероятность успешного прохождения $q_{S,lim}$ её внутренней для сегмента S части $Tr_{I,lim}$ имеет максимально возможное значение на множестве кусочно-линейных траекторий, имеющих те же начальную и конечную точки и то же число узлов. Число $N_{it,lim}$ есть функция параметра ρ . Далее полную траекторию РТП, полученную после u -кратного применения преобразования F , то есть на u -й итерации, будем обозначать Tr_u , а её внутреннюю для сектора U_S часть как $Tr_{I,u}$.

Условие А является интегральным, подразумевающим корреляцию удаления траектории от центра опасных зон с быстрым достижением границы сегмента, за пределами которого источник не активен. Условие В гарантирует, что при надлежащем выборе

параметра ρ преобразования F удастся достичь наиболее безопасного положения траектории относительно источника.

Заметим, что достаточным условием выполнения требования А в (16) является следующее:

$$q_{0s}(d', \alpha', T') = |\delta M'_i| / v > q_{0s}(d, \alpha, T) = |\delta M_i| / v, \quad i=1, \dots, n-1, \quad (17)$$

где $\delta M_i = M_{i+1} - M_i$ – векторная длина i -го отрезка траектории Tr_i , $\delta M'_i = M'_{i+1} - M'_i$ – векторная длина отрезка, построенного на базе образов точек M_i, M_{i+1} по преобразованию (15), v – неизменный в данном приближении модуль скорости РТП.

Условие (17), накладываемое на функцию F , имеет локальный характер и означает требование смещения в более безопасную область движения, причем вероятность перемещения без поражения/обнаружения вдоль любого скорректированного участка $M'_i M'_{i+1}$ выше вероятности непоражения/необнаружения при движении вдоль исходного участка $M_i M_{i+1}$. Для исходной M и конечной M' точек использованы полярные координаты; для примера на рисунке 2 отображены точки $M_{k+1}=M(d, \alpha)$ и образ преобразования F : $M'_{k+1}(d', \alpha') = F(M_{k+1})$.

Рассмотрим теперь обобщенную методику синтеза локального преобразования для отклонения траектории в поле одного источника. Преобразование F , переводящее точку M в новое положение M' , включает следующие этапы.

1. Линейное преобразование координат точки M из базовой системы $O_0X_0Y_0$ в систему отсчета сегмента OXY , показанную на рисунке 2.

2. Операторное нелинейное преобразование $\tilde{P} = f(P)$, $P \in U_s$ сегмента на левую полуплоскость, в результате которого сектор U_s отобразится в полукруг с центром в начале координат, лежащий в левой полуплоскости. В частности, произойдут следующие геометрические отображения: а) точки сегмента, находящиеся на левом берегу разреза вдоль его средней линии, перейдут в положительную полуось OY^+ , а точки сегмента, находящиеся на правом берегу этого же разреза, перейдут в отрицательную полуось OY^- ; в) левая и правая границы сегмента перейдут в отрицательную полуось OX^- ; г) если точка M принадлежит внутренней области сегмента, то она перейдет: в случае нахождения на левой от средней линии части сегмента, то есть от оси

OX , – во второй квадрант плоскости OXY ; в случае нахождения в правой от OX части сегмента – в третий квадрант OXY (см. рис. 3). Таким образом, преобразование f совершает перевод M в некоторую точку \tilde{M} .

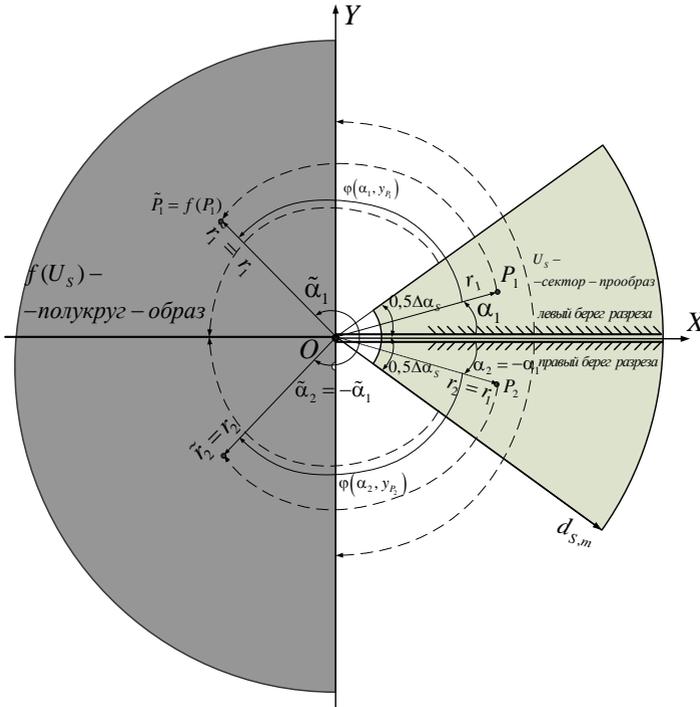


Рис. 3. Геометрия нелинейного операторного преобразования $\tilde{P}=f(P)$

3. Максимизация специального функционала $\Phi(\tilde{M}, S, Tr, \rho; P')$ на множестве точек левой полуплоскости, $P' \in OXY$ и нахождение соответствующего оптимального значения $P' \equiv \tilde{M}_{opt} = f_{\Phi}(\tilde{M}, S, Tr, \rho)$; выбор этого функционала определяется требованиями A и B , накладываемыми на искомую функцию F в (15), и свойствами операторного нелинейного преобразования f .

4. Применение обратного операторного преобразования f^{-1} к точке P' и получение искомой точки M' : $M' \equiv f^{-1}(P')$.

5. Преобразование координат M' из OXY в базовую систему $O_0X_0Y_0$.

Произведем выбор операторного нелинейного преобразования $\tilde{P}=f(P)$. Подберем такое активное преобразование, которое: а) не меняет

длины радиуса-вектора; б) левый берег разреза средней линии сектора отображает на положительную полуось OY^+ , а правый берег этого разреза – на отрицательную полуось OY^- ; в) левую и правую боковые границы сектора отображает на отрицательную полуось OX^- .

Преобразование, удовлетворяющее этому условию, имеет вид:

$$\begin{cases} x_{\tilde{P}} = \cos[\varphi(\alpha, y_P)]x_P - \sin[\varphi(\alpha, y_P)]y_P, \\ y_{\tilde{P}} = \sin[\varphi(\alpha, y_P)]x_P + \cos[\varphi(\alpha, y_P)]y_P, \end{cases} \quad (18)$$

где $\alpha = \text{sign}(y_P)\arccos(x_P/r)$ – угловое положение точки прообраза P внутри сектора U_S , $r = \sqrt{x_P^2 + y_P^2}$, а угол поворота φ равен:

$$\varphi(\alpha, y) = (\pi/2)\text{sign}(y) + (\pi/2 - \alpha_{gr})(\alpha/\alpha_{gr}). \quad (19)$$

Обратным к преобразованию (18) является:

$$x_P = r \cos \alpha, \quad y_P = r \sin \alpha, \quad (20)$$

$$\alpha = [2\tilde{\alpha} / \pi - \text{sign}(y_{\tilde{P}})]\alpha_{gr}, \quad (21)$$

α – искомый угол ориентации прообраза P точки \tilde{P} в системе координат OXY (см. рис. 3), а угол ориентации в этой же системе образа \tilde{P} есть

$$\tilde{\alpha} = \text{sign}(y_{\tilde{P}})\arccos(x_{\tilde{P}}/r); \quad r = \sqrt{x_P^2 + y_P^2} = \sqrt{x_{\tilde{P}}^2 + y_{\tilde{P}}^2} = \text{inv}.$$

Из выражений прямого (18), (19) и обратного (20), (21) преобразований можно получить следующие их свойства.

1. Если точка образа преобразования f претерпела смещение на большее расстояние от начала координат O , то и соответствующий прообраз претерпит смещение с тем же приращением модуля радиуса-вектора.
2. Если точка образа, находясь во втором/третьем квадранте, претерпевает смещение к отрицательной полуоси OX^- так, что угол между этой полуосью и радиус-вектором точки становится меньше, то точка прообраза смещается так, что угол между радиус-вектором прообраза и левой/правой границей сектора становится меньше.
3. Если точка образа, находясь во втором/ третьем квадранте, претерпевает смещение к положительной полуоси OY^+ / отрицательной полуоси OY^- так, что угол между этой полуосью

и радиус-вектором точки становится меньше, то соответствующая точка прообраза смещается в такое положение, что угол между её радиус-вектором и средней линией сектора источника становится меньше.

Рассмотрим более подробно методику синтеза и оптимизацию специального функционала $\Phi(\tilde{M}, S, Tr, \rho; P)$. Отметим, что рассмотренное операторное преобразование $\tilde{P} = f(P)$ позволяет сформулировать оптимизационную задачу так, чтобы одновременно учесть и реализовать требования A и B , накладываемые на искомый функционал F . Покажем это. Выделим четыре возможных случая движения РТП внутри сектора источника по отношению к его левой и правой линейным границам. Во-первых, РТП может двигаться по направлению к левой границе или от неё, находясь в левом полусегменте (случаи 1 и 2). Во-вторых, РТП может двигаться по направлению к правой границе или же от неё, находясь в правом полусегменте (случаи 3 и 4).

Рассмотрим случаи 1 и 3, когда РТП приближается к соответствующей границе, а скорректированная точка M' удаляется от центра O и приближается к этой границе. После преобразования $\tilde{P} = f(P)$ левая граница сектора отображается в отрицательную полуось OX . Учтем, что радиус точки \tilde{P} в системе координат OXY инвариантен относительно прямого f и обратного f^{-1} преобразований (18), (20). Поэтому для увеличения расстояния от O до M' и одновременного приближения к левой границе варьируемая точка P' функционала $\Phi(\tilde{M}, S, Tr, \rho; P)$ должна стремиться удалиться дальше вдоль отрицательного направления оси OX и одновременно приблизиться к этой оси. Возможна и смена знака вертикальной координаты, что означает выход из сектора через соответствующую его боковую границу.

Учет этих требований совместно с ограничением на степень удаления от точки M исходной траектории Tr может быть реализован в следующих функционалах для 1-го и 3-го случаев, подлежащих дальнейшей максимизации:

$$\Phi_1[\tilde{M}, S, Tr, \rho; x_p, y_p] = -\delta_1 x_p \rho - \delta_2 y_p \rho - \delta \left[(x_p - x_M)^2 + (y_p - y_M)^2 \right], \quad (22)$$

$$\Phi_3[\tilde{M}, S, Tr, \rho; x_p, y_p] = -\delta_1 x_p \rho + \delta_2 y_p \rho - \delta \left[(x_p - x_M)^2 + (y_p - y_M)^2 \right], \quad (23)$$

где $\rho = \rho(S, Tr, \tilde{M})$ – неотрицательный параметр преобразования, имеющий размерность длины и являющийся пока не определенной функцией от параметров источника S , исходной траектории Tr и точки

$\tilde{M}(x_M, y_M)$; $\delta_{1,2} = \delta_{1,2}(S, Tr, \tilde{M})$ – пока не определенные функции взвешенных факторов удаления от центра и приближения к боковой границе, зависящие от S, Tr, \tilde{M} , причем $0 \leq \delta_{1,2} \leq 1$, $\delta_1 + \delta_2 \leq 1$. Величина $\delta' = 1 - \delta_1 - \delta_2$ – взвешенный коэффициент, отражающий степень притяжения к исходной точке \tilde{M} кривой при проведении оптимизации. Координаты целевой точки функционала есть $x_{P'}$, $y_{P'}$.

Стационарные точки P'_{st} функционалов (22) и (23) имеют следующие координаты:

для случая 1

$$x_{P',st} \equiv x_{\tilde{M}_{opt}} = x_{\tilde{M}} - \frac{\delta_1 \rho}{2\delta'}, \quad y_{P',st} \equiv y_{\tilde{M}_{opt}} = y_{\tilde{M}} - \frac{\delta_2 \rho}{2\delta'}, \quad (24)$$

для случая 3:

$$x_{P',st} \equiv x_{\tilde{M}_{opt}} = x_{\tilde{M}} - \frac{\delta_1 \rho}{2\delta'}, \quad y_{P',st} \equiv y_{\tilde{M}_{opt}} = y_{\tilde{M}} + \frac{\delta_2 \rho}{2\delta'}. \quad (25)$$

Вторые производные функционала (22) принимают в стационарной точке (24) следующие значения:

$$\left(\partial^2 \Phi_{1,3} / \partial x_{P'}^2 \right) \Big|_{P'=P'_{st}} = -2\delta' < 0, \quad \left(\partial^2 \Phi_{1,3} / \partial y_{P'}^2 \right) \Big|_{P'=P'_{st}} = -2\delta' < 0,$$

$$\left(\partial^2 \Phi_{1,3} / \partial x_{P'} \partial y_{P'} \right) \Big|_{P'=P'_{st}} = 0. \quad \text{Следовательно, согласно достаточному}$$

условию существования локального максимума [23], в точках (24) и (25) функционалы (22), (23) достигают локальных максимумов, которые являются глобальными в силу единственности стационарной точки в этих двух случаях.

Таким образом, выражения (24), (25) дают координаты скорректированной точки $\tilde{M}_{opt} = P'_{st}$ в случаях 1 и 3.

Теперь рассмотрим случаи 2 и 4. В случае 2 происходит движение в левом полусекторе от левой границы. Поэтому целесообразно сделать так, чтобы после нелинейного преобразования $\tilde{P} = f(P)$ точка \tilde{P} притягивалась к образу срединной линии сектора по f , точнее, к образу левого берега разреза, произведенного вдоль этой линии. Но этот образ принадлежит, как было указано ранее, положительной полуоси OY^+ . С другой стороны, как и для случаев 1 и 3, надо стремиться отодвинуть точку \tilde{P} как можно дальше от начала координат O .

В случае 4 происходит удаление от правой границы правого сектора. Следовательно, здесь целесообразно сделать так, чтобы после нелинейного преобразования $\tilde{P} = f(P)$ точка \tilde{P} далее притягивалась к

образу по f правого берега разреза, произведенного вдоль срединной линии сектора. Но этот образ принадлежит отрицательной полуоси OY . С другой стороны, надо по-прежнему стремиться отодвинуть точку \tilde{P} как можно дальше от точки O .

Перечисленным для случаев 2 и 4 требованиям удовлетворяют следующие функционалы:

$$\Phi_2[\tilde{M}, S, Tr, \rho; x_{P'}, y_{P'}] = \delta_1 y_{P'} \rho + \delta_2 x_{P'} \rho - \delta' [(x_{P'} - x_{\tilde{M}})^2 + (y_{P'} - y_{\tilde{M}})^2]; \quad (26)$$

$$\Phi_4[\tilde{M}, S, Tr, \rho; x_{P'}, y_{P'}] = -\delta_1 y_{P'} \rho + \delta_2 x_{P'} \rho - \delta' [(x_{P'} - x_{\tilde{M}})^2 + (y_{P'} - y_{\tilde{M}})^2]. \quad (27)$$

Здесь коэффициент δ_1 отражает фактор удаления от центра, а δ_2 – фактор приближения к срединной линии со стороны левого полусектора – для случая 2, либо правого полусектора – для случая 4.

Решая задачу оптимизации функционалов (26), (27) по ранее приведенной методике, получаем точки глобальных максимумов:

для случая 2:

$$x_{P',st} \equiv x_{\tilde{M}_{opt}} = x_{\tilde{M}} + \frac{\delta_2 \rho}{2\delta'}, \quad y_{P',st} \equiv y_{\tilde{M}_{opt}} = y_{\tilde{M}} + \frac{\delta_1 \rho}{2\delta'}; \quad (28)$$

для случая 4:

$$x_{P',st} \equiv x_{\tilde{M}_{opt}} = x_{\tilde{M}} + \frac{\delta_2 \rho}{2\delta'}, \quad y_{P',st} \equiv y_{\tilde{M}_{opt}} = y_{\tilde{M}} - \frac{\delta_1 \rho}{2\delta'}; \quad (29)$$

Итоговый функционал Φ принимает вид:

$$\Phi = \{\Phi_i, \text{если } V(i), i = 1, 2, 3, 4\}, \quad (30)$$

$$V(1): M \in U_{S,l} \text{ и } (M_{prev} M \cdot n_1) \geq 0, \quad (31)$$

$$V(2): M \in U_{S,l} \text{ и } (M_{prev} M \cdot n_1) < 0,$$

$$V(3): M \in U_{S,r} \text{ и } (M_{prev} M \cdot n_1) < 0, \quad (32)$$

$$V(4): M \in U_{S,r} \text{ и } (M_{prev} M \cdot n_1) \geq 0.$$

Здесь $n_{1,2}$ – внешние нормали к левой и правой боковым границам сектора соответственно; M – точка из разбиения $\{M_i\}, i=1,2,\dots,n$ части

Tr_i исходной траектории, лежащей внутри сектора U_S на линейные участки $\Delta r_i, i=1,2,\dots,n-1; M_{prev}$ – предыдущая по отношению к M точка указанной последовательности.

Соответствующее функционалу (30) решение оптимизационной задачи дано выражением:

$$\left[x_{\tilde{M}_{opt}}, y_{\tilde{M}_{opt}} \right]^T = f_{\Phi}(\tilde{M}, S, Tr, \rho) = \left[x_{\tilde{M}}, y_{\tilde{M}} \right]^T + (\rho / 2\delta') \left[\delta_x, \delta_y \right]^T, \quad (33)$$

$$\delta_x = \begin{cases} -\delta_1, \text{ если } Y(1) \text{ или } Y(3); \\ \delta_2, \text{ иначе;} \end{cases}$$

$$\delta_y = \begin{cases} -\delta_2, \text{ если } Y(1), \\ \delta_1, \text{ если } Y(2), \\ \delta_2, \text{ если } Y(3), \\ -\delta_1, \text{ если } Y(4). \end{cases} \quad (34)$$

После нахождения точки \tilde{M}_{opt} необходимо возвратиться в область оригиналов, то есть преобразов по f , с помощью выражений (20), (21) обратного преобразования f^1 .

Указанные особенности прямого и обратного операторных отображений f и f^1 совместно с проведенным обоснованием выбора функционала (30) позволяют заключить, что свойство A параметрического преобразования (15) $M' = F(S, Tr; \rho; M)$ выполняется.

Теперь покажем, что достижение свойства B целевого преобразования F и организация итерационной процедуры коррекции исходной траектории могут быть обеспечены выбором параметра ρ – характеристической длины функционала (30). Из выражения (33) следует, что значения координат $x_{\tilde{M}}$ и $y_{\tilde{M}}$ определяются в том числе параметром ρ . Длина радиус-вектора $O\tilde{M}_{opt}$, получающегося после проведения смещения $\tilde{M} \rightarrow \tilde{M}_{opt}$ согласно соотношению (33), равна:

$$\begin{aligned} \left| \tilde{M}_{opt} \right| &= x_{\tilde{M}_{opt}}^2 + y_{\tilde{M}_{opt}}^2 \equiv \tilde{r}_{opt}^2 = \left[x_{\tilde{M}} + \frac{\delta_x \rho}{2\delta'} \right]^2 + \left[y_{\tilde{M}} + \frac{\delta_y \rho}{2\delta'} \right]^2 = \\ &= a\rho^2 + b\rho + c, \end{aligned} \quad (35)$$

где $a = \frac{\delta_x^2 + \delta_y^2}{4\delta'^2}$, $b = \frac{x_{\tilde{M}}\delta_x + y_{\tilde{M}}\delta_y}{\delta'}$, $c = x_{\tilde{M}}^2 + y_{\tilde{M}}^2 = r^2$. Здесь

использовано то, что точка \tilde{M} обязательно принадлежит второму или третьему квадранту, согласно особенности преобразования f .

Пусть поставлена задача отыскать такое значение ρ , при котором длина радиус-вектора $O\tilde{M}_{opt}$ равна заданному значению $\tilde{r}_{opt} = kr$, где число $k > 1$ в точке \tilde{M} характеризует степень нарастания радиуса при совершении преобразования Φ в \tilde{M} .

На основании (35) эта задача сводится к нахождению решения следующего уравнения:

$$a\rho^2 + b\rho + c - \tilde{r}_{opt}^2 = 0. \quad (36)$$

Решение (36) с учетом требования положительности ρ имеет вид:

$$\rho(k, \tilde{M}) = \frac{2\delta'}{\delta_1^2 + \delta_2^2} \left(-\Delta\rho + \sqrt{\Delta\rho^2 + r^2(k^2 - 1)(\delta_1^2 + \delta_2^2)} \right), \quad (37)$$

$$\Delta\rho = \begin{cases} \Delta\rho_I \equiv -(\delta_1 x_{\tilde{M}} + \delta_2 y_{\tilde{M}}), & \text{если } \mathcal{U}(1), \\ \Delta\rho_{II} \equiv \delta_2 x_{\tilde{M}} + \delta_1 y_{\tilde{M}}, & \text{если } \mathcal{U}(2), \\ \Delta\rho_{III} \equiv -\delta_1 x_{\tilde{M}} + \delta_2 y_{\tilde{M}}, & \text{если } \mathcal{U}(3), \\ \Delta\rho_{IV} \equiv \delta_2 x_{\tilde{M}} - \delta_1 y_{\tilde{M}}, & \text{если } \mathcal{U}(4). \end{cases} \quad (38)$$

Поэтому параметр ρ , выбранный по (37), (38), определяет скорость сходимости итерационной процедуры трансформации исходной траектории Tr при движении в поле одного источника; при таком выборе параметра ρ и дополнительном соблюдении условия А п. 3 обеспечивается выполнение и условия В п. 3.

Произведем выбор параметра k преобразования Φ для обеспечения выполнения условия (16). Выразим полярные координаты преобразованной точки $M'(r', \alpha')$ (согласно цепочке преобразований $M \rightarrow \tilde{M} \rightarrow P' \rightarrow M'$) через соответствующие полярные координаты исходной точки $M(r, \alpha)$.

Предположим, что исходная точка M находится в левом полусекторе источника и движется по направлению к его левой границе, то есть, что имеется случай 1. Случаи 2, 3, 4 рассматриваются аналогично.

Пусть преобразованная точка M' имеет радиальную координату $r' = kr$ в полярной системе координат источника, где r – радиальная координата исходной точки M , $k > 1$ – пока не определенный коэффициент. Определим её угловую координату α' согласно выражениям (37), (38), в которых $\Delta\rho_I = -(\delta_1 x_{\bar{M}} + \delta_2 y_{\bar{M}})$.

В силу свойств преобразования функционала Φ_I трансформированная точка $P' = \Phi_I(\bar{M})$ имеет следующую угловую координату в полярной системе источника:

$$\alpha_{p'} = \pi + \arctan\left(\frac{y_{\bar{M}} + \Delta y_{\bar{M}}}{x_{\bar{M}} + \Delta x_{\bar{M}}}\right), \quad (39)$$

где $\Delta x_{\bar{M}} = \frac{\delta_x \rho}{2\delta'}$, $\Delta y_{\bar{M}} = \frac{\delta_y \rho}{2\delta'}$ согласно выражению (33).

Применим к P' инверсное к f преобразование. Согласно (21), угол α' образа M' с учетом неравенств $y_M, y_{\bar{M}} > 0$ для рассматриваемого случая 1 может быть найден по формулам:

$$\alpha_{M'}(k, r, \alpha) = \alpha' = \left[1 + (2/\pi) \arctan\left(\frac{y_{\bar{M}} + \Delta y_{\bar{M}}}{x_{\bar{M}} + \Delta x_{\bar{M}}}\right) \right] \alpha_{gr}, \quad (40)$$

$$\begin{cases} x_{\bar{M}} = r \{ \cos[\varphi(\alpha)] \cos \alpha - \sin[\varphi(\alpha)] \sin \alpha \}, \\ y_{\bar{M}} = r \{ \sin[\varphi(\alpha)] \cos \alpha + \cos[\varphi(\alpha)] \sin \alpha \}, \end{cases} \quad (41)$$

$$\varphi(\alpha) = (\pi/2)(1 + \alpha/\alpha_{gr}) - \alpha. \quad (42)$$

Определим условия выполнения неравенств (17):

$$q_{0s}(d', \alpha', T' = |\delta M'_i|/\nu) > q_{0s}(d, \alpha, T = |\delta M_i|/\nu), \quad i=1, \dots, n-1. \quad (43)$$

Рассмотрим ХВФ с равномерным азимутальным распределением и коэффициентом $p=1$ в выражении (1). Тогда для ХВФ неравенство (17) может быть переписано в следующем виде:

$$\frac{|\delta M'_i|}{|\delta M_i|} < \exp\left[\frac{(r'^2 - r^2)}{d_m^2}\right]. \quad (44)$$

Далее, используя известные дифференциальные соотношения [24], связывающие модули малых приращений $|\delta M_i|$ и $|\delta M'_i|$ с соответствующими приращениями координат δr и $\delta \alpha$, получим:

$$\frac{|\delta M'_i|}{|\delta M_i|} \cong k \sqrt{\frac{\delta r^2 + r^2 (\partial_r \alpha' \delta r + \partial_\alpha \alpha' \delta \alpha)^2}{\delta r^2 + r^2 \delta \alpha^2}}, \quad (45)$$

где $\partial_r \alpha'$, $\partial_\alpha \alpha'$ - частные производные функции (40) по r и α , взятые в точке $M(r, \alpha)$.

В итоге неравенству (17), являющемуся достаточным условием для исходного требования A в (16), после несложных преобразований, может быть поставлено в соответствие следующее неравенство:

$$k \sqrt{\frac{\delta r^2 + r^2 (\partial_r \alpha' \delta r + \partial_\alpha \alpha' \delta \alpha)^2}{\delta r^2 + r^2 \delta \alpha^2}} < \exp\left[\left(\frac{r}{d_m}\right)^2 (k^2 - 1)\right], \quad (46)$$

Неравенство (46) имеет решение, поскольку при достаточно больших значениях k правая часть превышает левую с любым конечным коэффициентом пропорциональности. Для большинства практически важных случаев коэффициент при k в левой части (46) ограничен значением $\sqrt{3}$. Например, при $r = 2d_m$ условие (46) выполняется уже при $k=1.2$, при $r=d_m$ это условие выполняется при $k=1.4$, а при $r=0,25d_m$ минимальное значение $k=6.3$.

Поскольку максимально допустимое значение k , удовлетворяющее (46), ничем сверху не ограничено, то всегда возможен такой выбор k , что каждая следующая смещенная по F траектория будет отличаться от предыдущей не менее, чем на заданное расстояние. Например, при любом заданном минимально возможном расстоянии $\Delta r_{pr} = |MM'|$ между какой-либо исходной точкой M и её образом M' по F можно следующим выбором значения k :

$$k = \max_i \left[\max \left(k_i, \frac{\Delta r}{r_i} + 1 \right) \right], \quad (47)$$

обеспечить выполнение неравенств:

$$|M_i M_i'| \geq \Delta r_{pr}, i=1,2, \dots, n, \quad (48)$$

для каждой точки $M_i' \in Tr_{in}$ кусочно-линейной кривой на данной итерации. Здесь $k_i > 1$ есть значение параметра k для точки M_i , выбранное так, чтобы удовлетворялось неравенство (46).

Таким образом, возможность выбора параметра k , согласно соотношениям (46)-(48), обеспечивает выполнение условия A преобразования F .

Выбор преобразования $P' = f_{\Phi}(\tilde{M}, S, Tr, \rho)$ и функции $\rho(S, Tr, \tilde{M})$ с учетом (46)-(48) позволяет осуществить локальную оптимизацию каждой точки M_i по функционалу F с получением новой последовательности точек $M'_i, i=1,2,\dots,n$ так, что требования A, B целевого функционала F удовлетворяются.

Рассмотрим теперь методику выбора весовых коэффициентов δ' , δ_1, δ_2 функционала Φ . Степень влияния коэффициентов δ_1 и δ_2 факторов должна быть различной для разных внутренних точек сегмента. Действительно, для некоторой точки M_1 , которая находится достаточно близко от боковой границы сектора, коэффициент δ_2 должен иметь большее значение, чем для точки M_2 , находящейся достаточно далеко от боковых границ, но при том же радиальном удалении от центра источника, что и точка M_1 (см. рис. 4). Точно так же, для точки M_2 значение коэффициента δ_1 должно быть меньше, чем для точки M_3 , находящейся вплотную к внешней круговой границе сектора.

В первом из этих примеров большее значение коэффициента δ_2 для M_1 стимулирует в последующем РТП быстро покинуть сектор источника из положения M_1 через соответствующую близко расположенную боковую границу. Это, однако, справедливо при условии, что движение в данной точке совершается в сторону самой границы сектора. А во втором - большое значение δ_1 для M_3 стимулирует РТП быстро покинуть сектор источника из положения M_3 через соответствующую близко расположенную внешнюю границу сектора.

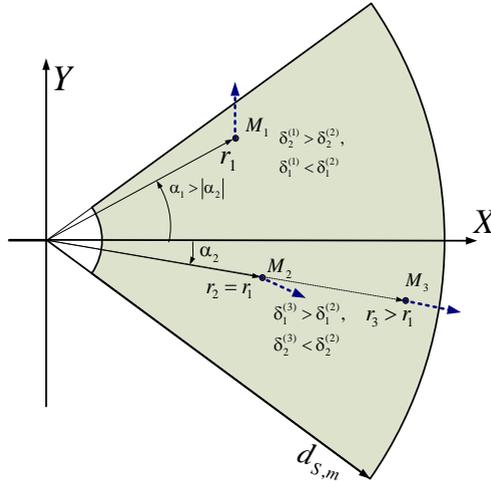


Рис. 4. Пояснение зависимости весовых коэффициентов δ_1 и δ_2 от положения исходной точки M внутри сектора источника U_S

Пусть задан коэффициент δ' , отражающий степень притяжения к исходной точке \tilde{M} кривой при оптимизации функционала (30). Введем следующее правило определения зависимости коэффициентов δ_1 и δ_2 от положения точки M :

$$\delta_1 = k_\delta / (d_{Sm} - r), \quad \delta_2 = k_\delta / (r\Delta\varphi), \quad (49)$$

$$k_\delta = (1 - \delta') / \left[(r\Delta\varphi)^{-1} + (d_{Sm} - r)^{-1} \right]. \quad (50)$$

Здесь $\Delta\varphi$ –угловое расстояние между радиусом-вектором текущей точки \tilde{M} и направляющим вектором боковой границы, к которой движется РТП. Значение $\Delta\varphi$ зависит от рассматриваемого случая:

$$\Delta\varphi = \begin{cases} \Delta\alpha/2 - |\alpha|, & \text{если } U(1) \text{ или } U(4); \\ \Delta\alpha/2 + |\alpha|, & \text{иначе,} \end{cases} \quad (51)$$

где $U(1)$ и $U(4)$ определяются согласно (31), (32).

4. Локальное преобразование для корректировки траектории движения в поле нескольких источников. Пусть имеется несколько источников возмущений с параметрами (2).

Рассмотрим участки траектории, которые целиком принадлежат секторам, минимум, двух источников (рис. 1). Введем весовые коэффициенты эффективности источников $\delta_s^{(k)}$, $k=1,2,\dots, K$, где $0 \leq \delta_s^{(k)} \leq 1$ и $\sum \delta_s^{(k)} = 1$, которые учитывают влияние, степень активности и опасность каждого из источников при планировании корректировки перемещений РТП в режиме реального времени.

Пусть для M траектории Tr применена процедура корректировки для каждого из \tilde{K} источников, в сектора которых попала точка M . В результате получены следующие точки: $M^{(k(j))'}$, $j=1,2,\dots, \tilde{K}$. Здесь $k(j)$, $j=1,2, \dots, \tilde{K}$ есть функция выделения подпоследовательности номеров источников, в которые попадает точка M . Тогда в качестве радиус-вектора O_0M' результирующей точки M' в базовой системе координат $O_0X_0Y_0$ (рис. 1), соответствующей корректировке исходной точки M с учетом всех \tilde{K} источников, примем следующий:

$$O_0M' = O_0M + \sum_{j=1}^{\tilde{K}} \delta_s^{(k(j))} \cdot MM^{(k(j))'}, \quad k = 1, 2, \dots, K, \quad (52)$$

где $MM^{(k)'}$ есть векторная длина линейного отрезка с началом в M и концом в $M^{(k)'}$.

Преобразование (52) F_k каждой точки $A_i(x_i, y_i)$, $i = 1, 2, \dots, N$ исходной кривой Tr , находящейся в поле действия нескольких источников, порождает набор точек локального смещения:

$$A_{loc,i}(x_{loc,i}, y_{loc,i}) = F_K \left[\left\{ S^{(k)} \right\}, Tr; \rho; A_i(x_i, y_i) \right], \quad (53)$$

$$i = 1, 2, \dots, N$$

5. Общая итерационная процедура достижения целевой вероятности прохождения. На основании рассмотренных в предыдущих разделах процедур может быть выработана итерационная процедура нахождения траектории, соответствующей заданной целевой вероятности прохождения.

Рассмотрим вопрос о необходимости варьирования числа узлов кусочно-линейной кривой при увеличении номера итерации. Заметим, что при неизменном значении числа узловых точек N в результате применения процедуры локальной оптимизации к кривой на текущей итерации могут найтись такие точки $A_{i,u}$, $A_{i+1,u}$, что длина отрезка $A_{i,u}A_{i+1,u}$ будет недопустимо большой, что не позволит провести

корректно поиск траектории с заданной целевой вероятностью. Например, при нахождении одной точки вне полей источников, а второй вблизи исходной кривой, но в поле действия источника, первая точка не будет подвергаться корректировке в результате применения процедуры локальной оптимизации на каждом шаге, а вторая точка – будет, поэтому для достаточно большого номера u итерации они могут удалиться на значительное расстояние. Аналогичная ситуация может иметь место для точек, соседствующих с граничными, которые закреплены. В обоих случаях могут появиться отрезки значительной длины.

По этой причине в оптимизационных задачах (8) и (9) задана возможность увеличения числа узловых точек с ростом номера итерации. Установив, например, некоторый предел $|\Delta r|_{gr}$ на максимальную длину отрезка, можно подвергать дискретизации отрезки полученной на очередной итерации кривой, которые имеют длину, большую $|\Delta r|_{gr}$.

Как уже отмечалось, первая составляющая функционала (10) способствует сокращению осцилляций кривой, возникающих вследствие применения локальной оптимизации. Степень этого сокращения регулируется коэффициентом δ_G . Однако большие значения δ_G означают малые значения коэффициента $\delta_L = 1 - \delta_G$, что уменьшает эффективность локальной оптимизации. Недостаточно большое значение δ_G может приводить к тому, что на некоторых участках будут наблюдаться осцилляции кривой, для снижения которых предлагается процедура сглаживания медианного типа, когда в окрестности радиуса R_{mean} каждой узловой точки кривой берутся соседние узловые точки, и среднее арифметическое их радиус-векторов считается новым положением данной точки.

Таким образом, алгоритм искомой итерационной процедуры следующий.

Формируется цикл $u=0, 1, 2, 3, \dots$, начиная с нулевой итерации, соответствующей начальной кривой, условием выхода из которого является превышение вероятности прохождения $q(Tr_u)$ траектории заданного значения q_{gr} ; в цикле производятся следующие операции:

1. Находятся отрезки, образованные соседними точками $A_{i,u}(x_{i,u}, y_{i,u}), i = 1, 2, \dots, N$ кривой на u -й итерации, длина которых превосходит $|\Delta r|_{gr}$, и проводится их дискретизация. В результате получается промежуточная кусочно-линейная кривая $Tr_{div,u+1}$ с точками $A_{i,u+1}^{(1)}(x_{i,u+1}^{(1)}, y_{i,u+1}^{(1)}), i = 1, 2, \dots, N(u+1)$.

2. Находится набор точек локального смещения на $u+1$ -й итерации с помощью процедуры локальной оптимизации (53):

$$A'_{i,u+1}(x'_{i,u+1}, y'_{i,u+1}) = F_K \left[\left\{ S^{(k)} \right\}, Tr_u; \rho; A'_{i,u+1}(x_{i,u+1}^{(1)}, y_{i,u+1}^{(1)}) \right], \quad (54)$$

$$i = 1, 2, \dots, N(u+1).$$

В результате получается кусочно-линейная кривая $Tr_{loc,u+1}$.

3. Находится оптимальная траектория для функционала (10) по формулам (12) – (15):

$$X_{u+1} = A^{-1}B_{X,u+1}, Y = A^{-1}B_{Y,u+1}, A'_{i,u+1}(x_{i,u+1}^{(3)}, y_{i,u+1}^{(3)}) = [X_{u+1}(i), Y_{u+1}(i)]^T, \quad (55)$$

$$i = 2, 3, \dots, N(u+1) - 1, A'_{1,u+1} = A'_{1,u+1}, A'_{N(u+1),u+1} = A'_{N(u+1),u+1},$$

где $B_{X,u+1}, B_{Y,u+1}$ есть правые части соответствующих линейных уравнений относительно координат точек $A'_{i,u+1}$, $i = 2, 3, \dots, N(u+1) - 1$ с матрицей A , не зависящей от u . В результате получаем кривую $Tr_{glob,u+1}$.

4. Производится сглаживание $Tr_{glob,u+1}$, после чего получается искомая траектория Tr_{u+1} на $u+1$ итерации с точками

$$A'_{i,u+1}(x'_{i,u+1}, y'_{i,u+1}), i = 1, 3, \dots, N(u+1).$$

Заметим, что используемая в данном алгоритме процедура локальной оптимизации узловых точек базируется на адаптивной подстройке параметров $\rho, k, \delta', \delta_1, \delta_2$ преобразования Φ для каждого источника на каждой текущей итерации, исходя из текущего расположения некоторой узловой точки траектории на этой итерации относительно сегмента источника. Для этого используются выражения (37), (47) и (49), (50).

6. Результаты моделирования. Пусть известна исходная кривая, по которой предполагается движение РТП для достижения заданной целевой точки.

Рассмотрим введенную итерационную процедуру на примере 1 движения вдоль исходной прямолинейной траектории, пересекающей сектор источника с ХФВ (1) при $k_a=0$ с параметрами $\varepsilon=3, d_m=10$, $\Delta\alpha_s = \pi/3$, $O(0;0)$; $n=[1;0]$. Параметры итерационной процедуры $\mu=0.1, \delta'=0.4$, рассмотренное число итераций $N_{iter}=50$, весовые

коэффициенты источников $\delta_s^{(k)}, k=1,2,..,K$ для простоты положены равными. Это означает, что РТП необходимо избегать попадания в зоны действия каждого из источников в одинаковой степени. Будем соотносить допустимое число итераций с требованием неперевышения некоторого критического удаления от исходной кривой. Движение РТП предполагается с постоянной скоростью $v=4$ м/с. Целевое значение вероятности прохождения везде принято равным $q_{goal}=0,9$.

Из рисунке 5а, где приведены траектории итераций 1, 5, 9, 10, 20, 30, 40, 50 для примера 1, видно, что с ростом номера u итерации вероятности прохождения образуют следующую последовательность $q_{S,u}$: $q_{S,u=1}=0.6583$; $q_{S,u=5}=0.7231$; $q_{S,u=9}=0.7885$; $q_{S,u=10}=0.8029$; $q_{S,u=20}=0.9087$; $q_{S,u=30}=0.9513$, $q_{S,u=40}=0.9667$, $q_{S,u=50}=0.9649$.

Здесь и далее исходная кусочно-линейная кривая показана синей сплошной линией, сглаженная траектория – сплошной желтой, а итерационные кривые изображены зеленым пунктиром, за исключением кривой на той итерации, на которой достигается максимальное значение вероятности успешного прохождения – она показана сплошной зеленой линией. Если на графиках не видна исходная синяя кривая, значит, она практически совпадает с желтой кривой сглаживания и визуально перекрывается ею.

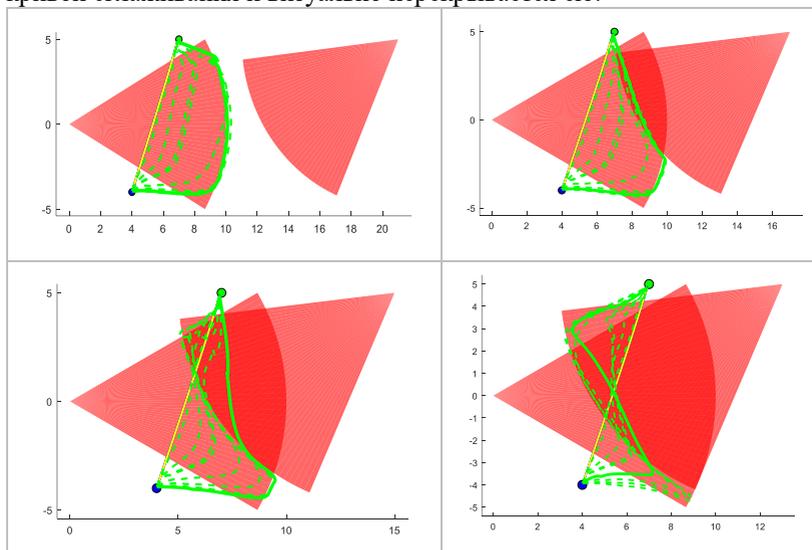


Рис. 5. Моделирование движения РТП в поле одного и двух источников: А) для одиночного источника влияния; Б) два источника, при положении второго источника $O_2(17;5)$; в) два источника, при положении второго источника $O_2(15;5)$; г) два источника, при положении второго источника $O_2(13;5)$

В примере 1 максимальное значение вероятности прохождения достигается при 40-й итерации. Дальнейшее незначительное (около 1%) уменьшение вероятности прохождения вызвано влиянием погрешностей, обусловленных шагом дискретизации и сглаживанием промежуточных кривых на каждой итерации. Целевая вероятность здесь достигается уже на 20-й итерации.

На рисунке 5б приведен пример движения РТП в поле двух источников. Источники такие же, как и в примере 1, но центр и направляющий вектор второго равен: $O_2(17;5)$; $n_2 = [-0,8; -0,6]$. Параметры итерационной процедуры те же, что и для примера 1. Вероятности прохождения образуют следующую последовательность для тех же значений u , что и в примере 1: $q_{S,u=1}=0.6587$; $q_{S,u=5}=0.7228$; $q_{S,u=9}=0.7773$; $q_{S,u=10}=0.7860$; $q_{S,u=20}=0.8420$; $q_{S,u=30}=0.8742$; $q_{S,u=40}=0.8788$; $q_{S,u=50}=0.8667$.

Снижение вероятности прохождения на 50-й итерации по сравнению с 40-й вызвано здесь усилением влияния второго источника. Здесь целевая вероятность $q_{goal}=0,9$ не достигается в пределах $N_{iter}=50$.

На рисунке 5в показан случай, когда второй источник занимает положение $O_2(15;5)$. В этом случае также не удастся достичь целевой вероятности. Соответствующие вероятности прохождения образуют следующую последовательность: $q_{S,u=1}=0.6365$; $q_{S,u=5}=0.6619$; $q_{S,u=9}=0.6852$; $q_{S,u=10}=0.6858$; $q_{S,u=20}=0.6670$; $q_{S,u=30}=0.6475$; $q_{S,u=40}=0.6976$; $q_{S,u=50}=0.7637$.

Заметим, что в окрестности $u=30$ наблюдается локальный минимум, однако далее интегральная вероятность растет и достигает значения $q_{S,u=50}=0.7637$. Наличие немонотонного поведения функции $q_{S,u}$ свидетельствует о сложном влиянии полей обоих источников на итерационные кривые РТП в условиях, когда исходная кривая проходит между этими источниками. Это напоминает попадание в область локального минимума суммарного потенциала при решении задачи огибания сложных препятствий методом потенциальных полей.

Для случая еще большего приближения второго источника к первому соответствующие результаты приведены на рисунке 5г, а последовательность вероятностей прохождения есть: $q_{S,u=1}=0.5646$; $q_{S,u=5}=0.5680$; $q_{S,u=9}=0.5831$; $q_{S,u=10}=0.5868$; $q_{S,u=20}=0.5884$; $q_{S,u=30}=0.5561$; $q_{S,u=40}=0.5414$; $q_{S,u=50}=0.5540$.

Здесь уже имеются три локальных максимума, обусловленные еще большим приближением второго источника к первому.

Когда область, в которой должен перемещаться РТП, содержит несколько произвольно ориентированных источников, растет число

итераций, необходимых для обхода локальных экстремумов и достижения глобального максимума, если последний существует.

Так, на рисунке 6 приведены результаты построения скорректированных траекторий в области с семью произвольно расположенными и ориентированными источниками на выборочных итерациях до $u=130$. Последовательность вероятностей прохождения есть: $q_{S,u=1}=0.7060$; $q_{S,u=5}=0.7437$; $q_{S,u=10}=0.7592$; $q_{S,u=20}=0.6462$; $q_{S,u=30}=0.6714$; $q_{S,u=40}=0.6931$, $q_{S,u=50}=0.6995$, $q_{S,u=80}=0.7586$, $q_{S,u=100}=0.8909$, $q_{S,u=130}=0.9719$.

Вначале происходит увеличение вероятности прохождения до $u=10$, затем РТП попадает в зону влияния одного из источников, заходя в неё с тыльной стороны, поэтому $q_{S,u}$ далее начинает уменьшаться вплоть до $u=20$, однако уже ранее 30-й итерации возобновляется рост $q_{S,u}$, в результате которого эта функция начинает превышать целевое значение вероятности $q=0.9$ при $u=130$.

На рисунке 7 представлены результаты построения скорректированных траекторий в области с девятью произвольно расположенными и ориентированными источниками на выборочных итерациях вплоть до $u=80$. Последовательность вероятностей прохождения есть: $q_{S,u=1}=0.4818$; $q_{S,u=5}=0.5947$; $q_{S,u=10}=0.6718$; $q_{S,u=20}=0.6783$; $q_{S,u=30}=0.7721$; $q_{S,u=40}=0.8518$, $q_{S,u=50}=0.9046$, $q_{S,u=80}=0.9413$. Особенность расположения источников здесь позволяет избежать осцилляционного поведения функции $q_{S,u}$, что обуславливает достижение целевой вероятности за меньшее число шагов, по сравнению с предыдущим случаем.

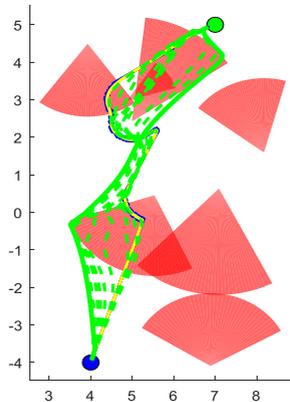


Рис. 6. Моделирование движения РТП в поле многих источников

На рисунке 8 представлен случай, когда начальная траектория проложена в области, содержащей как источники, так и препятствия (последние показаны закрашенными черным полигонами). Последовательность вероятностей прохождения есть: $q_{S,u=1}=0.5473$; $q_{S,u=5}=0.5989$; $q_{S,u=12}=0.6852$; $q_{S,u=20}=0.7798$; $q_{S,u=25}=0.7864$; $q_{S,u=30}=0.8342$; $q_{S,u=40}=0.8832$; $q_{S,u=50}=0.9570$.

В случае пересечения итерационной кривой с максимальной вероятностью прохождения одного или нескольких препятствий необходимо просто минимально отстроиться в область, безопасную от столкновения с этими препятствиями, с возможной и вынужденной потерей вероятности прохождения.

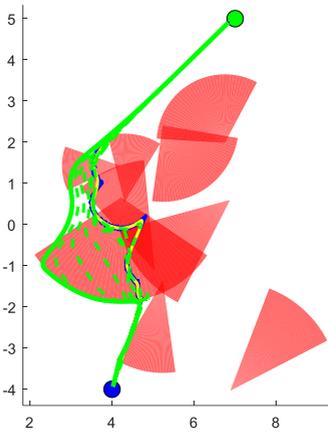


Рис. 7. Движение РТП в поле многих источников для примера 6

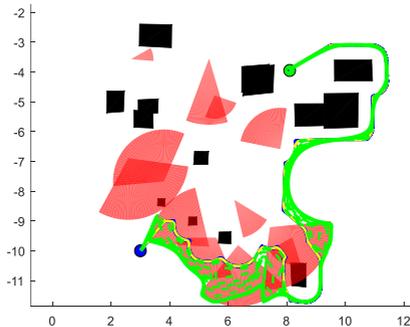


Рис. 8. Моделирование движения РТП в поле многих источников для примера 7

7. Выводы. В статье рассматривается новая процедура корректировки исходной траектории движения робототехнической платформы на плоскости с целью снижения вероятности её поражения/обнаружения в поле конечного числа источников-репеллеров, каждый из которых обладает ограниченной областью действия в виде кругового сектора. Эффективность синтезированной процедуры определяется строгим решением задачи для случая одного источника, учетом воздействия всех источников в общем случае и подтверждается результатами проведенного моделирования.

Эта процедура позволяет находить локальные максимумы указанной вероятности на множестве допустимых кривых, отклоняющихся от исходной не более чем на заданное расстояние в соответствующем функциональном пространстве.

Предложенная процедура основана, с одной стороны, на представлении о характеристической вероятностной функции одиночного источника-репеллера, позволяющем в итоге оценивать степень влияния системы таких источников на движущуюся РТП. Из этого понятия вытекает используемая в качестве критерия оптимизации целевой траектории вероятность её успешного прохождения.

С другой стороны, процедура базируется на решении локальных оптимизационных задач, позволяющих корректировать отдельные участки исходной глобальной траектории с учетом нахождения в их окрестностях конкретных источников-репеллеров с заданными параметрами. При этом локальная оптимизация узловых точек исходной траектории базируется на адаптивной подстройке ключевых параметров преобразования Φ для каждого источника на каждой текущей итерации, исходя из текущего расположения некоторой узловой точки траектории на этой итерации относительно сегмента источника.

Предложенный метод корректировки траекторий с учетом заданного распределения источников-репеллеров, основанный на указанной процедуре, позволяет эффективно организовывать соответствующее движение РТП с заданным вероятностным критерием качества/ успешности выполнения миссии.

Литература

1. Khatib O. Real-Time Obstacles Avoidance for Manipulators and Mobile Robots. *Int. Journal of Robotics Research*. 1986. vol. 5(1). pp. 90–98.
2. Montiel O., Orozco-Rosas U., Sepúlveda R. Path planning for mobile robots using Bacterial Potential Field for avoiding static and dynamic obstacles. *Expert Systems with Applications*. 2015. vol. 42 (12). pp. 5177-5191.

3. Rasekhipour Y., Khajepour A., Chen S.-K., Litkouhi B. A Potential Field-Based Model Predictive Path-Planning Controller for Autonomous Road Vehicles. *IEEE Transactions on Intelligent Transportation Systems*. 2017. vol. 18(5). pp. 1255-1267.
4. Mabrouk M.H., McInnes C.R. Solving the potential field local minimum problem using internal agent states. *Robotics and Autonomous Systems*. 2018. vol. 56(12).
5. Pshikhopov V., Medvedev M., Soloviev V. The Multicopter Control Algorithms with Unstable Modes. *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT'19)*, Paris, France. 2019. pp. 1179 – 1184.
6. Krokmal P., Zabaranin M., Uryasev S. Modeling, and optimization of risk. *Surveys in Operations Research and Management Science*. 2001. vol. 16. pp. 49–66.
7. Zabaranin M., Uryasev S., Pardalos P. Optimal Risk Path Algorithms. *Cooperative Control and Optimization*. Dordrecht: Kluwer Acad. 2002. pp. 271 – 303.
8. Галяев А.А., Маслов Е.П., Рубинович Е.Я. Об одной задаче управления движением объекта в конфликтной среде. *Известия РАН. Теория и системы управления*. 2009. vol. 3. pp. 134–140.
9. Галяев А.А., Маслов Е.П. Оптимизация законов уклонения подвижного объекта от обнаружения. *Известия РАН. Теория и системы управления*. 2010. vol. 4. pp. 43–53.
10. Абрамянц Т.Г., Маслов Е.П., Яхно В.П. Уклонение подвижного объекта от обнаружения группой наблюдателей. *Проблемы управления*. 2010. vol. 5. pp. 73–79.
11. Галяев А.А. Задача уклонения от подвижного одиночного наблюдателя на плоскости в конфликтной среде. *Автоматика. и телемеханика*. 2014. vol. 6. pp. 39–48.
12. Абрамянц Т.Г., Галяев А.А., Маслов Е.П., Рудько И.М., Яхно В.П. Уклонение подвижного объекта в конфликтной среде от обнаружения системой разнородных наблюдателей. *Проблемы управления*. 2015. vol. 2. pp. 31–37.
13. Корепанов В.О., Новиков Д.А. Задача о диффузной бомбе. *Проблемы управления*. 2011. vol. 5. pp. 66–73.
14. Корепанов В.О., Новиков Д.А. Модели стратегического поведения в задаче о диффузной бомбе. *Проблемы управления*. 2015. vol. 2. pp. 38–44.
15. Lavrenov R.O., Magid E.A., Matsuno F., Svinin M.M., Sutakorn J. Development and implementation of spline-based path planning algorithm in ROS/gazebo environment. *SPIIRAS Proceedings*. 2019. vol. 18(1), pp. 57-84.
16. Carsten J., Ferguson D., Stentz A. 3D Field D*: Improved Path Planning and Replanning in Three Dimensions. *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2006. Beijing, China, pp. 3381 – 3386.
17. Pshikhopov V., Medvedev M. Motion Planning and Control Using Bionic Approaches Based on Unstable Modes. *Path Planning for Vehicles Operating in Uncertain 2D Environments*, 2017. pp. 239 – 280.
18. Han J. An efficient approach to 3D path planning. *Information Sciences*. 2019. vol. 478, pp. 318-330.
19. Pshikhopov V., and Medvedev M. Group control of autonomous robots' motion in uncertain environment via unstable modes. *SPIIRAS Proceedings*, 2018. vol. 60(5), pp. 39-63.
20. Yu X.-Y., Fan Z.-Y., Ou L.-L., Zhu F., Guo Y.-K. Optimal Path Planning Satisfying Complex Task Requirement in Uncertain Environment // *Robotica*. 2019. vol. 37 (11), pp. 1956 – 1970.
21. Shlyakhov N.E., Vatamaniuk I.V., Ronzhin A.L. Review of the Methods and Algorithms of a Robot Swarm Aggregation. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2017. vol. 18(1). pp. 22-29.

-
22. Костюков В.А., Пшихопов В.Х., Медведев М.Ю. Оптимизация движения мобильного робота на плоскости в поле конечного числа источников-репеллеров, Труды СПИИРАН, 2020, vol. 19(1), pp. 43–78.
 23. Кудрявцев Л.Д. Курс математического анализа (в 3 томах): Т. 2. М: Дрофа. 2004. 720 с.
 24. Гельфанд И.М., Фомин С.В. Вариационное исчисление. М.: Физматлит, 1961.

Костюков Владимир Александрович – канд. техн. наук, доцент, НИИ робототехники и процессов управления, Южный федеральный университет. Область научных интересов: математическое моделирование, мобильные роботы, алгоритмы планирования пути, ветроэнергетические установки. Число научных публикаций -65. wkost-einheit@yandex.ru, <http://itpc.sfedu.ru/>; ул. Шевченко, д. 2, К-225, г. Таганрог, 347922, РФ; р.т.: +7(8634)371-694, факс: +7(8634)681-894.

Медведев Михаил Юрьевич – д-р техн. наук, доцент, ведущий научный сотрудник, НИИ робототехники и процессов управления, Южный федеральный университет. Область научных интересов: адаптивное и робастное управление мобильными роботами, оценивание возмущений, методы анализа и синтеза систем автоматического управления. Число научных публикаций-240+. medvmihal@sfedu.ru, <http://itpc.sfedu.ru> ул. Шевченко, д. 2, К-225, г. Таганрог, 347922, РФ; р.т.: +7(8634)371-694, факс: +7(8634)681-894.

Пшихопов Вячеслав Хасанович – д-р техн. наук, профессор, директор, НИИ робототехники и процессов управления, Южный федеральный университет. Область научных интересов: управление мобильными роботами в неопределенных средах; оптимальное управление роботами; анализ и синтез систем группового управления; интеллектуальное управление и планирование в робототехнике. Автор более 260 публикаций. Число научных публикаций-250. pshichop@ Rambler.ru, <http://itpc.sfedu.ru/>; ул. Шевченко, д. 2, К-225, г. Таганрог, 347922, РФ; р.т.: +7(8634)371-694, факс: +7(8634)681-894. Поддержка исследований. Работа выполнена при финансировании Российского научного фонда (грант 18-19-00621, выполняемый в ОАО «НКБ робототехники и систем управления»).

V. KOSTJUKOV, M. MEDVEDEV, V. PSHIKHOPOV

**METHOD FOR OPTIMIZING OF MOBILE
ROBOT TRAJECTORY IN REPELLER SOURCES FIELD**

Kostjukov V., Medvedev M., Pshikhovov V. Method for Optimizing of Mobile Robot Trajectory in Repeller Sources Field.

Abstract. The article discusses the procedure for correcting the trajectory of a robotic platform (RTP) on a plane in order to reduce the probability of its defeat/detection in the field of a finite number of repeller sources. Each of these sources is described by a mathematical model of some factor of counteraction to the RTP.

This procedure is based, on the one hand, on the concept of a characteristic probability function of a system of repeller sources, which allows us to assess the degree of influence of these sources on the moving RTP. From this concept follows the probability of its successful completion used here as a criterion for optimizing the target trajectory. On the other hand, this procedure is based on solving local optimization problems that allow you to correct individual sections of the initial trajectory, taking into account the location of specific repeller sources with specified parameters in their vicinity. Each of these sources is characterized by the potential, frequency of impact, radius of action, and parameters of the field decay.

The trajectory is adjusted iteratively and takes into account the target value of the probability of passing. The main restriction on the variation of the original trajectory is the maximum allowable deviation of the changed trajectory from the original one. If there is no such restriction, then the task may lose its meaning, because then you can select an area that covers all obstacles and sources, and bypass it around the perimeter. Therefore, we search for a local extremum that corresponds to an acceptable curve in the sense of the specified restriction.

The iterative procedure proposed in this paper allows us to search for the corresponding local maxima of the probability of RTP passage in the field of several randomly located and oriented sources, in some neighborhood of the initial trajectory.

First, the problem of trajectory optimization is set and solved under the condition of movement in the field of single source with the scope in the form of a circular sector, then the result is extended to the case of several similar sources.

The main problem of the study is the choice of the General form of the functional at each point of the initial curve, as well as its adjustment coefficients. It is shown that the selection of these coefficients is an adaptive procedure, the input variables of which are characteristic geometric values describing the current trajectory in the source field.

Standard median smoothing procedures are used to eliminate oscillations that occur as a result of the locality of the proposed procedure.

The simulation results show the high efficiency of the proposed procedure for correcting the previously planned trajectory.

Keywords: Path Planning, Conflict Environment, Traffic Optimization, Probability of Successful Passage, Group of Repeller Sources

Vladimir Kostjukov – Dr. Sci., associated professor, Research and Development Institute of Robotics and Control Systems, Southern Federal University. Research interests: mathematical modeling, mobile robots, path planning algorithms, wind power plants construction. Number of publications: 65. wkost-einheit@yandex.ru, <http://rirpc.sfedu.ru/>; 2, Shevchenko Street, Office 225, Taganrog, Russian Federation, 347922; phone: +7(8634)371-694, fax: +7(8634)681-894.

Mikhail Medvedev – Dr. Sci., full professor, Research and Development Institute of Robotics and Control Systems, Southern Federal University. Research interests: adaptive and robust control of mobile robots, estimation, analysis and synthesis attitudes of control systems. Number

of publications: 240. medvmihal@sfedu.ru, <http://rirpc.sfedu.ru/>; 2, Shevchenko Street, Office 225, Taganrog, Russian Federation, 347922; phone +7(8634)371-694, fax +7(8634)681-894.

Pshikhov Viacheslav – Dr. Sci., full professor, head of Research and Development Institute of Robotics and Control Systems, Southern Federal University. Research interests: control of mobile robots in uncertain environment, optimal control of robots, analysis and synthesis of collective robots control, intelligent control and planning in the robotics. The number of publications is 250. pshichop@rambler.ru, <http://rirpc.sfedu.ru/>; Shevchenko Street, 2, K-225, Taganrog, 347922, Russian Federation; phone +7(8634)371-694, fax +7(8634)681-894.

Acknowledgment. The study is supported by the Russian Science Foundation, grant 18-19-00621, executed at Joint stock Company “Scientific-Design bureau of Robotics and Control Systems”.

References

1. Khatib O. Real-Time Obstacles Avoidance for Manipulators and Mobile Robots. *Int. Journal of Robotics Research*. 1986. vol. 5(1), pp. 90–98.
2. Montiel O., Orozco-Rosas U., Sepúlveda R. Path planning for mobile robots using Bacterial Potential Field for avoiding static and dynamic obstacles. *Expert Systems with Applications*. 2015. vol. 42 (12), pp. 5177-5191.
3. Rasekhipour Y., Khajepour A., Chen S.-K., Litkouhi B. A Potential Field-Based Model Predictive Path-Planning Controller for Autonomous Road Vehicles. *IEEE Transactions on Intelligent Transportation Systems*. 2017. vol. 18(5), pp. 1255-1267.
4. Mabrouk M.H., McInnes C.R. Solving the potential field local minimum problem using internal agent states. 2008. *Robotics and Autonomous Systems*. vol. 56(12), pp. 1050 – 1060.
5. Pshikhov V., Medvedev M., Soloviev V. The Multicopter Control Algorithms with Unstable Modes. *6th International Conference on Control, Decision and Information Technologies (CoDIT'19)*, Paris, France. 2019. pp. 1179 – 1184.
6. Krokhmal P., Zabarankin M., Uryasev S. Modeling and optimization of risk. *Surveys in Operations Research and Management Science*. 2001. vol. 16, pp. 49–66.
7. Zabarankin M., Uryasev S., Pardalos P. Optimal Risk Path Algorithms. *Cooperative Control and Optimizatoin*. Dordrecht: Kluwer Acad., pp. 271 – 303.
8. Galyaev A.A., Maslov E.P., Rubinovich E.Ya. *Ob odnoj zadache upravleniya dvizheniem ob"ekta v konfliktnoj srede* [On a problem of controlling the movement of an object in a conflict environment]. *Izvestiya RAN. Teoriya i sistemy upravleniya* [Journal of Computer and Systems Sciences International]. 2009. vol. 3, pp. 134–140. (In Russ.).
9. Galyaev A.A., Maslov E.P. *Optimizaciya zakonov ukloneniya podvizhnogo ob"ekta ot obnaruzheniya* [Optimization of the laws of mobile object evasion from detection]. *Izvestiya RAN. Teoriya i sistemy upravleniya* [Journal of Computer and Systems Sciences International]. 2010. vol. 4, pp. 43–53. (In Russ.).
10. Abramyan T.G., Maslov E.P., Yakhno V.P. *Optimizaciya zakonov ukloneniya podvizhnogo ob"ekta ot obnaruzheniya* [Evasion of a moving object from detection by a group of observers]. *Problemy upravleniya* [Control Sciences]. 2010. vol. 5, pp. 73–79. (In Russ.).
11. Galyaev A.A. *Zadacha ukloneniya ot podvizhnogo odinochnogo nablyudatelya na ploskosti v konfliktnoj srede* [Evasion on plane from a single mobile observer in the conflict environment]. *Avtomatika i telemekhanika* [Automation and Remote Control]. 2014. vol. 76(6), pp. 1017–1024.

12. Abramyan T.G., Galyaev A.A., Maslov E.P., Rud'ko I.M., Yakhno V.P. *Uklonenie podvizhnogo ob"ekta v konfliktnoj srede ot obnaruzheniya sistemoy raznorodnykh nablyudatelej* [Evasion of a moving object in a conflict environment from detection by the system of heterogeneous observers]. *Problemy upravleniya* [Control Sciences]. 2015. vol. 2, pp. 31–37. (In Russ.).
13. Korepanov V.O., Novikov D.A. *Zadacha o diffuznoj bombe* [The problem of diffuse the bomb]. *Problemy upravleniya* [Control Sciences]. 2011. vol. 5, pp. 66–73. (In Russ.).
14. Korepanov V.O., Novikov D.A. *Modeli strategicheskogo povedeniya v zadache o diffuznoj bombe* [Models of strategic behavior in the diffuse bomb problem]. *Problemy upravleniya* [Control Sciences]. 2015. vol. 2, pp. 38–44. (In Russ.).
15. Lavrenov R.O., Magid E.A., Matsuno F., Svinin M.M., Sutakorn J. Development and implementation of spline-based path planning algorithm in ROS/gazebo environment. *SPIIRAS Proceedings*. vol. 18(1), pp. 57-84. (2019).
16. Carsten J., Ferguson D., Stentz A. 3D Field D*: Improved Path Planning and Replanning in Three Dimensions. *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2006. Beijing, China, pp. 3381 – 3386.
17. Pshikhopov V., Medvedev M. Motion Planning and Control Using Bionic Approaches Based on Unstable Modes. *Path Planning for Vehicles Operating in Uncertain 2D Environments*, 2017. pp. 239 – 280.
18. Han J. An efficient approach to 3D path planning. *Information Sciences*. 2019. vol. 478, pp. 318-330.
19. Pshikhopov V., Medvedev M. Group control of autonomous robots motion in uncertain environment via unstable modes. *SPIIRAS Proceedings*. 2018. vol. 60(5), pp. 39-63.
20. Yu X.-Y., Fan Z.-Y., Ou L.-L., Zhu F., Guo Y.-K. Optimal Path Planning Satisfying Complex Task Requirement in Uncertain Environment. *Robotica*. 2019. vol. 37 (11), pp. 1956 – 1970.
21. Shlyakhov N.E., Vatamaniuk I.V., Ronzhin A.L. Review of the Methods and Algorithms of a Robot Swarm Aggregation. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2017. vol. 18(1). pp. 22-29.
22. Kostjukov V. Pshikhopov V., Medvedev M. Optimization of mobile robot movement on a plane with finite number of repeller sources. *SPIIRAS Proceedings*. 2020, vol. 19(1), pp. 43-78.
23. Kudrjavcev L.D. *Kurs matematicheskogo analiza* [A course of mathematical analysis]. Moscow: Drofa. vol. 2. 2004. 720 p.
24. Gel'fand I.M., Fomin S.V. *Variacionnoe ischislenie* [Calculus of variations]. Moscow: Fizmatlit, 1961.

А.С. ГВОЗДАРЕВ, П.Е. ПАТРАЛОВ
**ВЕРОЯТНОСТНЫЙ АНАЛИЗ ОБОБЩЕННОЙ
СТАТИСТИЧЕСКОЙ МОДЕЛИ МНОГОЛУЧЕВОГО КАНАЛА
SIMO СИСТЕМЫ С ЗАМИРАНИЯМИ И
КОРРЕЛИРОВАННЫМИ ЗАТЕНЕНИЯМИ**

Гвоздарев А.С., Патралов П.Е. Вероятностный анализ обобщённой статистической модели многолучевого канала SIMO системы с замираниями и коррелированными затенениями

Аннотация. Рассмотрена задача анализа характеристик процесса передачи информации многоэлементными системами связи в условиях многолучевого канала распространения сигнала. Для обобщения эффектов распространения была выбрана модель канала k - μ с коррелированными затенениями, а в качестве используемой технологии организации многоэлементной системы – SIMO система, использующая дифференциально-взвешенное комбинирование сигнала на приёмной стороне. Для описания характеристик процесса передачи информации использовался подход на основе статистик высшего порядка эргодической пропускной способности. В рамках исследования были получены аналитические выражения для статистик произвольного порядка для рассматриваемой модели канала. Проведён анализ поведения первых четырёх статистик (эргодической пропускной способности, величины надёжности, коэффициентов асимметрии и эксцесса) в зависимости от характеристик канала (количества многопутевых кластеров распространения, доли мощности, приходящейся на доминантные компоненты, степени затенения доминантного компонент и коэффициента корреляции затенений). В рамках исследования были рассмотрены 4 ситуации поведения исследуемой модели канала, существенно различающиеся по своим свойствам. Отмечено, что в отличие от пропускной способности, статистики высшего порядка оказываются существенно более чувствительными к параметрам канала и, как следствие, являются более значимыми индикаторами флуктуации скорости передачи информации в канале связи. Обнаружено наличие ярко выраженного экстремума (минимума) зависимости надёжности эргодической пропускной способности от среднего соотношения сигнал/шум, что с практической точки зрения важно учитывать при предъявлении требований к величине отношения сигнал/шум в канале, обеспечивающей желаемое качество функционирования системы связи.

Ключевые слова: беспроводной канал, замирания, затенение, модель k - μ с коррелированными затенениями, статистики высших порядков эргодической пропускной способности, SIMO, оптимальная весовая обработка, отношение сигнал/шум

1. Введение. В настоящее время технологии связи с многоэлементными антенными устройствами [1] (SIMO (Single-Input-Multiple-Output), MISO (Multiple-Input-Single-Output), MIMO (Multiple-Input-Multiple-Output)) играют одну из ключевых ролей, обеспечивая скорость передачи и надёжность, требуемые современными стандартами связи [2-3], такими как LTE-A (Long Term Evolution-Advanced), 5G, WiGig и др. В то же время рост потребительских ожиданий приводит к повышению требований, предъявляемых к обработке сигналов, что, по сути, требует пересмотра существующих

моделей канала [4]. Это обусловлено тем, что в основе процедуры обнаружения и/или обработки сигналов на приемной стороне лежит информация об используемой вероятностной модели канала [5-6], при этом жесткие требования, предъявляемые к ней и усложняющие её, определяются необходимостью учёта тонких эффектов распространения сигнала.

Общая проблема с существующими моделями каналов заключается в том, что они обычно хорошо согласуются с реальными экспериментальными данными в окрестности максимума/ожидаемого значения распределения мгновенных флуктуаций коэффициентов передачи канала или в хвостах распределения, но обычно не в обоих случаях [4], [7]. Следовательно, требуемое повышение надежности приводит к необходимости лучшей подстройки модели под экспериментальные данные и, как следствие, более надёжного прогноза флуктуаций канала и возможности более качественной компенсации возникающих из-за них искажений.

Аналитическое моделирование каналов связи является хорошо разработанной и достаточно обширной областью современных телекоммуникаций [6-7]. Среди наиболее известных моделей следует упомянуть так называемые обобщенные модели [6-11], которые учитывают эффекты ослабления, замирания (быстрого и медленного) и затенения сигнала. Кроме того, в ряде частных случаев они вырождаются в менее общие классические модели [4, 5, 7] (включая Рэля, Райса, Накагами, Хойта, Вейбулла и др.). Одной из наиболее распространённых среди обобщённых на сегодняшний день является модель k - μ с затенениями [11].

С другой стороны, современные аналитические подходы, используемые для описания многоантенных систем, в значительной степени опираются на теорию случайных матриц, поэтому усложнение модели канала обычно приводит к невозможности замкнутого аналитического описания общей модели системы, следовательно, к существенному усложнению процедуры оценки качества и прогнозированию эффективности её функционирования. Однако существует несколько частных случаев, когда применение обобщенных моделей к многоантенным системам приводит к замкнутому аналитическому решению [5]. Одним из важных с практической точки зрения сценариев является SIMO система, использующая на приёмной стороне дифференциально-взвешенную обработку сигналов [5] (MRC, Maximum Ratio Combining). В этом случае использование для описания канала модели k - μ с затенениями позволяет преодолеть проблему

аналитической разрешимости, учтя большее (по сравнению с классическими моделями) количество эффектов распространения.

Немаловажным фактором, влияющим на возможность более точной подстройки используемой модели под реальные наблюдения (как следствие, позволяющим учесть большее число эффектов в канале связи), является использование разнообразных показателей качества, описывающих функционирования системы связи. В частности, в последнее время всё возрастающий интерес привлекает математический аппарат статистик высших порядков пропускной способности [12-18], позволяющий оценить не только её ожидаемую величину (в зависимости от различных факторов), но и характеристики её флуктуации вследствие мгновенных флуктуаций канала.

Целью настоящей работы является исследование влияния параметров многолучевого SIMO канала k -м с коррелированными затенениями на статистику высшего порядка эргодической пропускной беспроводной системы связи.

2. Модель канала k -м с затенением для случая SISO системы.

Как отмечалось, в качестве рассматриваемой модели многолучевого канала с замираниями была выбрана одна из наиболее общих на данный момент моделей – k -м канал с затенением [11], в рамках которой мощность сигнала представляется в виде совокупности мощностей синфазной и квадратурной компонент флуктуирующего сигнала:

$$W = \sum_{i=1}^n (X_i + \xi p_i)^2 + (Y_i + \xi q_i)^2,$$

где X_i, Y_i взаимно независимые реализации центрированных (то есть $E\{X_i\} = E\{Y_i\} = 0$, где $E\{\cdot\}$ – оператор математического усреднения)

Гауссовских процессов с $E\{(X_i)^2\} = E\{(Y_i)^2\} = \sigma^2$, описывающие случайные флуктуации мгновенных реализаций распространяющегося сигнала; p_i, q_i – детерминированные составляющие; ξ – случайный коэффициент, отвечающий за затенение детерминированных компонент, описывающийся распределением Накагами с параметром масштаба m и единичной интенсивностью затенений (то есть $E\{\xi^2\} = 1$)

С физической точки зрения данная модель рассматривает распространение электромагнитной волны в виде совокупности из n

кластеров, в которых выделяются доминантные компоненты $\xi p_i + j \xi q_i$ с мощностью $p_i^2 + q_i^2$ для каждого i -ого кластера (ассоциируемые с зеркальными переотражениями), и рассеянные/многопутевые компоненты $X_i + j Y_i$ (ассоциируемые с диффузным рассеянием) с суммарной мощностью $2\sigma^2$.

В рамках данной модели плотность распределения вероятности мгновенного соотношения сигнал/шум $\gamma \square E\{\gamma\}W/E\{W\}$ определяется как [11]:

$$f_\gamma(\gamma) = \frac{\mu^\mu m^m (1+\kappa)^\mu}{\Gamma(\mu) \bar{\gamma}^m (\mu\kappa + m)^m} \left(\frac{\gamma}{\bar{\gamma}}\right)^{\mu-1} e^{-\mu(1+\kappa)\frac{\gamma}{\bar{\gamma}}} {}_1F_1\left(m, \mu; \frac{\mu^2 \kappa (1+\kappa) \gamma}{\mu\kappa + m \bar{\gamma}}\right), \quad (1)$$

где $\bar{\gamma} = E\{\gamma\}$ – среднее отношение сигнал/шум (ОСШ) $\Gamma(\cdot)$ – гамма-функция Эйлера [19], ${}_1F_1(\cdot, \cdot; \cdot)$ – вырожденная гипергеометрическая функция [19], μ – обобщённое количество путей распространения (кластеров многолучевого канала), κ – отношение суммарной мощности основных лучей в каждом из кластеров к суммарной мощности всех остальных лучей, m – параметр затенения основных лучей в каждом кластере. При этом предполагается, что времена задержек многопутевых волн в рамках каждого кластера сравнимы по величине и существенно меньше, чем времена запаздываний между кластерами [11].

Параметры модели могут принимать произвольные положительные значения, однако, проведённое ранее исследование показало, что в ряде задач (например, в задачах энергетического обнаружения сигнала в когнитивных радиосетях) проявляется невысокая чувствительность характеристик, усреднённых с плотностью распределения вероятности (1), к точности оценки параметров κ – μ модели в рассматриваемых на практике диапазонах изменений значений параметров, позволяет трактовать их как целочисленные без существенного искажения оценки качества функционирования системы связи.

Нужно также отметить, что в ряде предельных случаев модель κ – μ канала с затенением переходит в часто используемые и хорошо известные упрощённые модели [11]: одностороннюю Гауссовскую, Рэлеевскую, Райсовскую, Накагами- m , κ – μ модель и Райсовскую с затенением.

3. Модификация модели k - μ с коррелированными затенениями для случая SIMO. Основным преимуществом модели k - μ канала с затенением является возможность учёта большего (по сравнению с классическими моделями) числа факторов, описывающих распространение сигнала. Однако значительно возросшая сложность математического описания существенно затрудняет непосредственный её перенос на случай систем с разнесённым приёмом/передачей (SIMO, MISO, MIMO) [5].

Данная задача может быть решена с использованием результатов, полученных в [14] и [15]. В частности, в [14] рассматривается SISO система с агрегацией сигнала из нескольких частотных поддиапазонов (каждый из которых моделируется каналом k - μ с затенениями), с неэквивалентными условиями распространения в каждом из поддиапазонов и наличием корреляции процессов затенения доминантных компонент, при этом используется дифференциально-взвешенное комбинирование сигналов на приёмной стороне с вероятностно-статистическими характеристиками, полученными в [20].

С точки же зрения математического описания, модель [14] остаётся справедливой, если заменить физическую трактовку SISO системы с несколькими частотными подканалами на SIMO систему (с M антенными элементами на приёмной стороне, при выполнении условия независимости суммируемых каналов), функционирующую в одном диапазоне частот. При этом будем считать, что до каждого приёмного антенного элемента канал связи описывается моделью k - μ с затенениями, неидентичными параметрами распространения μ_i, k_i ($i = \overline{1, M}$) и наличием корреляции затенений доминантных компонент. С практической точки зрения это эквивалентно наличию μ путей распространения с существенно различающимися условиями, что характерно для моделей городских каналов связи. Важно отметить, что MRC обработка является оптимальной при выполнении условий наличия информации о состоянии беспроводного канала и независимости трактов приёма. На практике первое требование удовлетворяется на основе оценки коэффициента передачи с использованием пилот-сигналов, а второе – за счёт организации взаимного пространственного разнесения приемных антенных элементов на расстояния, обеспечивающие условия «наиболее благоприятного» распространения [21], например, половину длины волны или более.

Предполагая, что на приемной стороне используется дифференциально-взвешенная обработка сигнала, плотность распределения вероятности мгновенных значений соотношения

сигнал/шум полученной модифицированной κ-μ модели с коррелированными затенениями может быть записана в следующей форме [14]:

$$f_\gamma(\gamma) = A \left(\frac{\eta}{\bar{\gamma}} \right)^U \gamma^{U-1} e^{-\frac{\eta\gamma}{\bar{\gamma}}} \sum_{k=0}^{\infty} D_k {}_1F_1 \left(mM + k, U; \frac{\eta}{(1+\lambda^{-1})\bar{\gamma}} \right), \quad (2)$$

где введены следующие переобозначения:

$$\gamma = \sum_{i=1}^M \gamma_i, \quad U = \sum_{i=1}^M \mu_i, \quad A = \prod_{i=1}^m \left(\frac{\lambda}{\lambda_i} \right)^m, \quad \eta = \sum_{i=1}^M \mu_i (1 + \kappa_i),$$

при этом $\bar{\gamma}$ – среднее соотношение сигнал/шум, $\lambda = \min(\lambda_1, \lambda_2, \dots, \lambda_M)$ – минимальное значение спектра собственных чисел $\{\lambda_1, \lambda_2, \dots, \lambda_M\}$ произведения матриц DC, где

$$D = \text{diag} \left(\frac{\mu_1 \kappa_1}{m}, \frac{\mu_2 \kappa_2}{m}, \dots, \frac{\mu_M \kappa_M}{m} \right), \quad [C]_{i,j} = \sqrt{\rho_{i,j}},$$

$\rho_{i,j}$ – коэффициент корреляции доминантных компонент кластеров с номерами i и j . Взвешивающие коэффициенты суммирования D_k получаются с использованием рекурсивной процедуры

$$D_k = \frac{\delta_k}{\lambda^{mM+k} \Gamma(U)} (1 + \lambda^{-1})^{-(mM+k)}, \quad \delta_{k+1} = \frac{m}{k+1} \sum_{q=1}^{k+1} \sum_{i=1}^M \left(1 - \frac{\lambda}{\lambda_i} \right)^q \delta_{k+1-i},$$

с начальным условием $\delta_0 = 1$.

4. Статистики высшего порядка эргодической пропускной способности. Как было отмечено ранее, одним из наиболее интересных и перспективных подходов к описанию многоэлементных систем связи является описание в терминах статистик эргодической пропускной способности (СПС) высшего порядка. Здесь и далее будем следовать общепринятому подходу, рассматривающему нормированную пропускную способность системы (в рамках которого она совпадает по смыслу со спектральной эффективностью), то есть функционирующей в полосе частот в 1 Гц в течение интервала времени в 1 с [5].

По определению [12], под СПС порядка n понимается:

$$\Lambda_n = \int_0^{\infty} \log_2^n (1 + \gamma) f_\gamma(\gamma) d\gamma. \quad (3)$$

Соответственно, при $n = 1$ получаем первый начальный момент, который характеризует саму эргодическую пропускную способность [5], то есть $\Lambda_1 = E \{ \log_2 (1 + \gamma) \}$.

Для описания степени разброса флуктуаций мгновенных значений пропускной способности наиболее часто используются: степень замираний η и величина надёжности пропускной способности \mathcal{R} , которые характеризуют величину потери (степень надёжности) пропускной способности при передаче 1 бита информации за счёт случайных флуктуаций мгновенного ОСШ [13]:

$$\eta = \frac{\Lambda_2}{\Lambda_1^2} - 1, \quad \mathcal{R} = 1 - \left(\frac{\Lambda_2}{\Lambda_1} - \Lambda_1 \right). \quad (4)$$

В частности, для случая нефлуктуирующего канала с аддитивным белым Гауссовым шумом (АБГШ) $\Lambda_1^2 = \Lambda_2$ и, как следствие, $\eta = 0$, что можно считать референсной точкой отсчёта. Степень надёжности оценки эргодической пропускной способности показывает вероятность того, что эргодическая пропускная способность попадает в заданный диапазон отклонений, равный величине нормированной дисперсии эргодической пропускной способности.

Так же для описания деформации распределения мгновенных значений пропускной способности используются коэффициенты асимметрии $S\check{k}$ и эксцесса $\mathcal{K}ur$:

$$S\check{k} = \frac{\Lambda_3 - \Lambda_1^3}{(\Lambda_2 - \Lambda_1^2)^{3/2}}, \quad \mathcal{K}ur = \frac{\Lambda_4 - \Lambda_1^4}{(\Lambda_2 - \Lambda_1^2)^2}. \quad (5)$$

Стоит оговориться, что, несмотря на схожую терминологию, в отличие от общего определения моментов 3 и 4 порядков в классической математической статистике, выражения (5) представляют собой строго положительные монотонно возрастающие функции значений ОСШ [12]-[14], что в свою очередь несколько затрудняет их непосредственное использование как количественных показателей, как следствие, они служат скорее для качественного сравнения различных ситуаций между собой.

5. Аналитические выражения статистик высшего порядка эргодической пропускной способности для случая SIMO системы и канала k - μ с коррелированными затенениями. Для решения задачи нахождения аналитических выражений для статистик высшего порядка эргодической пропускной способности воспользуемся подходом, предложенным в [16], в рамках которого они вычисляются через линейную комбинацию производящей функции моментов и её производной:

$$\Lambda_n = \int_0^{\infty} Z_n(p) \left\{ M_{\gamma}(-p) - \frac{\partial}{\partial p} M_{\gamma}(-p) \right\} dp \quad (6)$$

с весовым множителем $Z_n(p)$, учитывающим порядок вычисляемого момента:

$$Z_n(p) = n! e^{-p} G_{n+2, n+1}^{0, n+1} \left(\frac{1}{p} \middle| 1, 1, \dots, 1 \right),$$

где $G_{p,q}^{m,n} \left(x \middle| \begin{matrix} a_1, \dots, a_n \\ b_1, \dots, b_m \end{matrix} \right)$ – G-функция Мейера [19].

Для вычисления (6) найдём выражение для производящей функции моментов мгновенного соотношения сигнал/шум на выходе приёмной решетки, использующей дифференциально-взвешенную обработку сигнала. Принимая во внимание, что по определению $M_{\gamma}(p) = E\{e^{\gamma p}\}$ и статистическое усреднение производится с учётом плотности распределения вероятности (2), получим:

$$\begin{aligned} M_{\gamma}(p) &= A \left(\frac{\eta}{\gamma} \right)^U \sum_{k=0}^{\infty} D_k \int_0^{\infty} \gamma^{U-1} e^{-\frac{\eta \gamma}{\gamma}} e^{\gamma p} {}_1F_1 \left(mM + k, U; \frac{\eta}{(1+\lambda^{-1})} \frac{\gamma}{\gamma} \right) d\gamma = \\ &= A \left(\frac{\eta}{\gamma} \right)^U \sum_{k=0}^{\infty} D_k L \left\{ \gamma^{U-1} {}_1F_1 \left(mM + k, U; \frac{\eta}{(1+\lambda^{-1})} \frac{\gamma}{\gamma} \right); \frac{\eta}{\gamma} - p \right\}, \end{aligned} \quad (7)$$

где A – нормировочный множитель, $L\{f(x); y\}$ – преобразование Лапласа от функции $f(x)$, вычисленное в точке y . Важно отметить, что в [16] производящая функция моментов определена $M_{\gamma}(p) = E\{e^{-\gamma p}\}$, как следствие выражение (6) (используемое далее в вычислениях) и соответствующее выражение в [16] имеют противоположные знаки.

Используя [22] выражение (7) может быть преобразовано к виду:

$$M_{\gamma}(p) = A \Gamma(U) \left(\frac{\eta}{\gamma} - p \right)^{-U} \xi^{mM}(p) \left(\frac{\eta}{\gamma} \right)^U \sum_{k=0}^{\infty} D_k \xi^k(p), \quad (8)$$

где $\xi(p) = \left(\frac{\eta}{\gamma} - p \right) / \left(\frac{\eta}{\gamma(1+\lambda)} - p \right)$. Подставляя полученное

выражение (8) в (6), производя дифференцирование по переменной p и группируя слагаемые в фигурных скобках, получим:

$$\Lambda_n = A \left(\frac{\eta}{\bar{\gamma}} \right)^U \sum_{k=0}^{\infty} D_k \int_0^{\infty} \frac{Z_n(p) \xi^{m+1}(-p)}{(\eta/\bar{\gamma} + p)^{U+2}} (p + p_{0_+})(p + p_{0_-}) dp, \quad (9)$$

с учётом введённого обозначения:

$$p_{0_{\pm}} = \frac{U}{2} + \frac{\eta}{2\bar{\gamma}} \left(\frac{2+\lambda}{1+\lambda} \right) \mp \frac{1}{2} \sqrt{\left(U + \frac{\eta}{\bar{\gamma}} \left(\frac{2+\lambda}{1+\lambda} \right) \right)^2 - 4 \frac{\eta}{\bar{\gamma}} \left(\frac{m}{1+\lambda} + \frac{1}{1+\lambda} \left(U + \frac{\eta}{\bar{\gamma}} \right) \right)}.$$

Заметим, что весовой множитель $Z_n(p)$ в (9) непредставим в элементарных функциях, однако, для частных случаев выражается в виде суммы по степеням $\log p$. Например, для первых 4 порядков (существенных для рассматриваемого в работе описания) $Z_n(p)$:

$$\begin{cases} Z_1(p) = -\log p - E, & Z_2(p) = 2E \log p + \log^2 p + E^2 - \frac{\pi^2}{6}, \\ Z_3(p) = \frac{1}{2} \left((\pi^2 - 6E^2) \log p - 6E \log^2 p - 2 \log^3 p - 2(2\zeta(3) + E^3) + \pi^2 E \right), \\ Z_4(p) = \log p (8\zeta(3) + 4E^3 - 2\pi^2 E) - (\pi^2 - 6E^2) \log^2 p + \\ \quad + 4E \log^3 p + \log^4 p + 8\zeta(3)E + E^4 - \pi^2 E^2 + \frac{\pi^4}{60}, \end{cases}$$

где E – константа Эйлера-Маскерони, а $\zeta(\cdot)$ – дзета-функция Римана [19].

В общем случае интеграл в (9) не вычисляется в замкнутой форме, однако может быть рассчитан численно с использованием вычислительно эффективных алгоритмов.

6. Результаты численной апробации. Существенным для дальнейшего анализа является тот факт, что (9) зависит от одношагового коэффициента корреляции затенений доминантных компонент $\rho_{i,j}$. В рамках вероятностно-статистического описания корреляционных эффектов в канале распространения будем рассматривать классическую экспоненциальную модель корреляции:

$$\rho_{i,j} = \rho^{|i-j|}, \quad (10)$$

которая характеризуется только корреляцией между двумя любыми соседними элементами – одношаговым коэффициентом корреляции ρ . В свою очередь для ρ воспользуемся моделью Абди-Кавеха [23], обобщающей большое количество моделей пространственно-временной корреляции (например, модели изотропного рассеяния, модели Ли, модели Кларка и др.):

$$\rho = I_0 \left(\sqrt{\chi^2 - 4\pi^2 \tilde{d}^2 + 4\pi \tilde{d} \chi \sin(\varphi)} \right) / I_0(\chi), \quad (11)$$

где $I_0(\square)$ – модифицированная функция Бесселя нулевого порядка, \tilde{d} – нормированное на длину волны расстояние между антенными элементами (для эквидистантных антенных систем), χ определяется шириной пространственного спектра углов прихода (изменяется в диапазоне от 0 до ∞), φ – среднее направление прихода сигнала.

Для дальнейшего моделирования была выбрана совокупность параметров моделей (10)-(11), обеспечивающая некоторые разумные уровни ρ . Например, в случае четырёхэлементной эквидистантной приёмной решётки, ориентированной нормально к падающему фронту ($\varphi = 0$) $\chi = 3,3$ даёт значение $\rho = 0,2$, $\chi = 5,3$ обеспечивает $\rho = 0,4$, $\chi = 7$ соответствует $\rho = 0,5$ и $\chi = 9,4$ даёт $\rho = 0,6$, что соответствует ширине пространственного спектра 63° , 50° , 43° и 37° соответственно.

С точки зрения анализа поведения системы связи, существенным являются зависимости выбранных для описания статистик высшего порядка (Λ_1 , \mathcal{R} , $S\mathcal{K}$, $\mathcal{K}ur$) от параметров системы ($\bar{\gamma}$, m , ρ). При этом будем рассматривать 4 существенно различающихся по своим свойствам ситуации поведения рассматриваемой модели канала (считая, что параметры пространственных подканалов при распространении до каждого из 4 приёмных элементов идентичны):

1. $\mu = 0,5$, $\kappa = 0,1$, что соответствует малому количеству кластеров многопутевого распространения с малой долей мощности, приходящейся на доминантные компоненты;
2. $\mu = 2,5$, $\kappa = 0,1$, что соответствует большому количеству кластеров с тем же соотношением мощностей между доминантными и многопутевыми компонентами;
3. $\mu = 0,5$, $\kappa = 3$, что соответствует малому количеству кластеров с большой долей мощности, приходящейся на доминантные компоненты;
4. $\mu = 2,5$, $\kappa = 3$, что соответствует большому количеству кластеров многопутевого распространения с большой долей мощности, сконцентрированной в доминантных компонентах.

Количество слагаемых при использовании полученного выражения (9) выбиралось таким образом, чтобы гарантировать относительную ошибку усечения ряда не более 10^{-5} .

Для удобства здесь и далее на всех графиках сплошные кривые соответствуют модели № 1, длинная штрихпунктирная – модели № 4,

средняя и короткая штрихпунктирные линии моделям № 2 и № 3 соответственно.

Проведённый анализ позволил отметить, что, несмотря на схожее поведение (малый масштаб изменений) эргодической пропускной способности рассматриваемого канала связи для всех анализируемых моделей (см. рис. 1), статистики высшего порядка оказываются существенно более чувствительными. Так, в частности, переход от наименее (№ 1) к наиболее (№ 4) благоприятной модели приводит к выигрышу в Λ_1 не более чем на 0,5 бит/с/Гц в области больших ОСШ (что составляет лишь 4,6% относительного изменения), однако существенным образом сказывается на поведении \mathcal{R} (см. рис. 2), $S\bar{\kappa}$ (62% относительного изменения) и $\mathcal{K}ur$ (82% относительного изменения) (см. рис. 3 и 4, представленные в логарифмическом масштабе). На рисунках 2 и 3 изображены кривые, соответствующие $m=1$, круглыми маркерами обозначены случаи с $\rho=0,1$, квадратными – $\rho=0,5$.

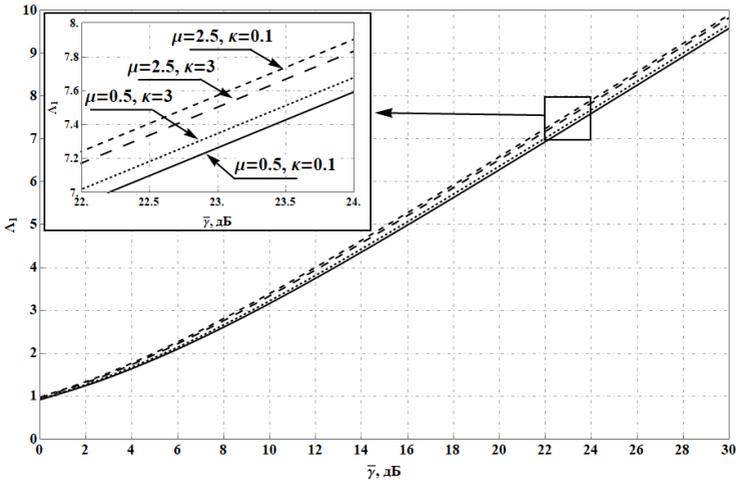


Рис. 1. Зависимость эргодической пропускной способности от $\bar{\gamma}$

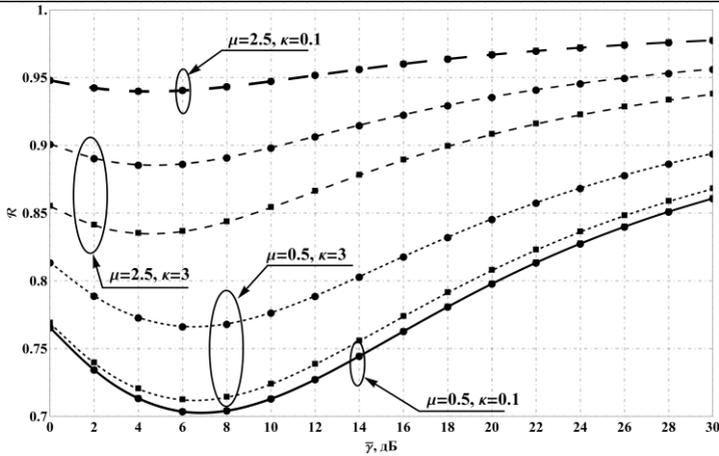


Рис. 2. Зависимость надёжности в определении эргодической пропускной способности от $\bar{\gamma}$

Стоит отметить, что определения (5) отличаются от классического для теории вероятности способа введения понятий асимметрии и эксцесса, как следствие, получаемые характеристики носят монотонно неубывающий характер и могут быть использованы скорее для качественного, нежели количественного анализа, путём сравнения скоростей роста характеристик для разных случаев.

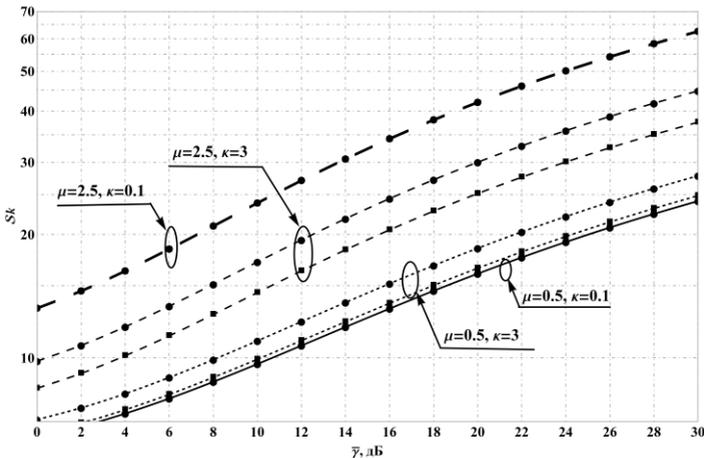
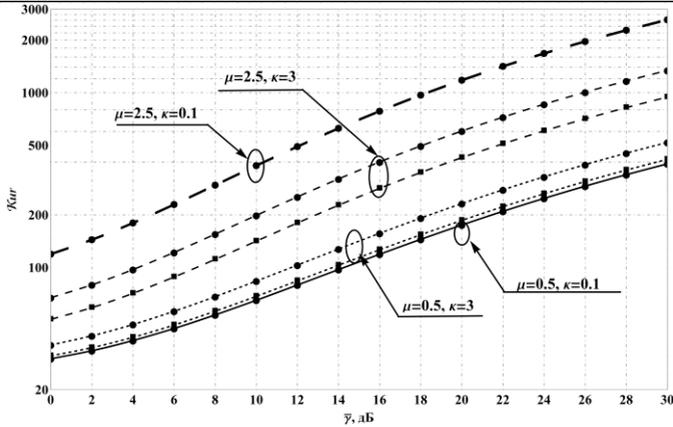


Рис. 3. Зависимость коэффициента асимметрии от $\bar{\gamma}$


 Рис. 4. Зависимость коэффициента эксцесса от $\bar{\gamma}$

Важно отметить наличие у \mathcal{R} ярко выраженных экстремумов (минимумов). Учёт СПС высшего порядка позволяет понять, что, несмотря на монотонное увеличение эргодической пропускной способности, при повышении $\bar{\gamma}$ существует диапазон ОСШ, при которых степень разброса её мгновенных значений (относительно ожидаемого) оказывается максимальной (что соответствует минимумам \mathcal{R}). Фиксируя (как это, например, сделано в [16]) \mathcal{R} на некотором уровне, например, 0,9 или 0,95, можно указать совокупность параметров ($\bar{\gamma}$, m , ρ , μ , κ) для которых не возможно обеспечить функционирование системы связи с требуемыми характеристиками. Так в частности видно, что модель канала № 1, в отличие от № 2 не способна обеспечить требуемый режим по надёжности в определении Λ_1 (вплоть до $\bar{\gamma} = 30$ дБ), несмотря на то, что с точки зрения самой Λ_1 проигрыш оказывается совершенно незначительным (не более 0,5 бит/с/Гц).

На рисунках 5–9 представлены зависимости поведения СПС высшего порядка от величины одношагового коэффициента корреляции. Результаты представлены для случая $m=0,5$, круглыми маркерами обозначены случаи с $\bar{\gamma} = 0$ дБ, квадратными – $\bar{\gamma} = 10$ дБ. Видно, что в условиях, когда основная доля мощности приходится на многопутевые компоненты (наличие сильно выраженного диффузного рассеяния, что соответствует моделям № 1 и № 2) рост ρ вплоть до существенных величин (например, до $\rho=0,8$, как на рисунках 5–9) не приводит к практически значимым изменениям.

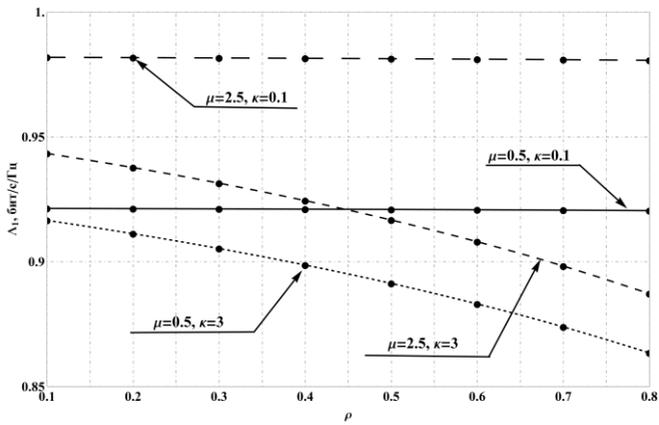


Рис. 5. Зависимость эргодической пропускной способности от ρ для случая $\bar{\gamma} = 0$ дБ

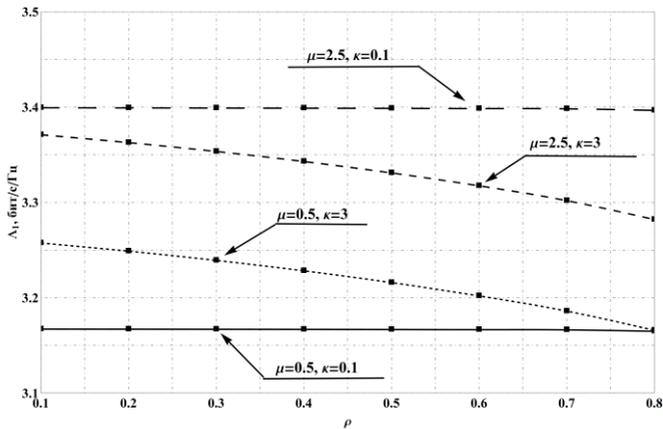


Рис. 6. Зависимость эргодической пропускной способности от ρ для случая $\bar{\gamma} = 10$ дБ

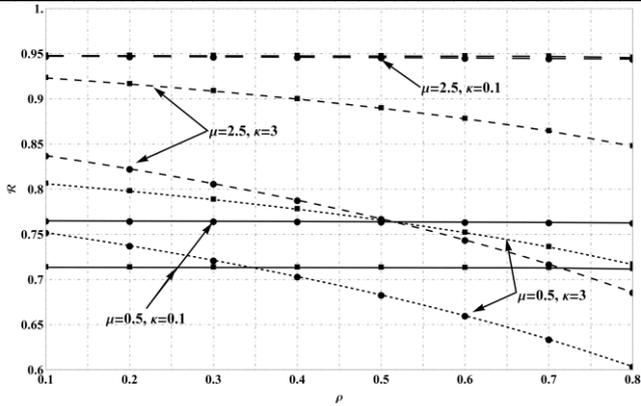


Рис. 7. Зависимость надёжности в определении эргодической пропускной способности от ρ

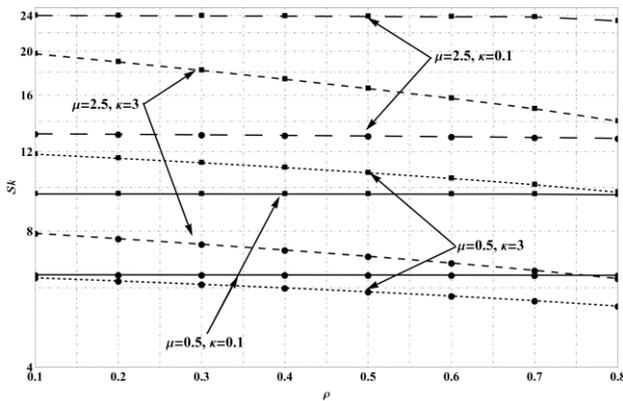


Рис. 8. Зависимость коэффициента асимметрии от ρ

При этом большее количество многопутевых кластеров обеспечивает меньшие флуктуации Λ_1 , и как следствие, большие значения \mathcal{R} (рис. 7). При наличии же ярко выраженных доминантных компонент (модели № 3 и № 4) увеличение ρ (так же как и увеличение $\bar{\gamma}$) приводит к уменьшению как \mathcal{R} , так и Sk и $\mathcal{K}ur$.

При анализе зависимости СПС высших порядков от степени затенения доминантных компонент m (см. рис. 10–12) было обнаружено, что существенные изменения наблюдаются лишь в области больших затенений (малые m). В области значений $m > 1$ влиянием изменения её величины можно пренебречь.

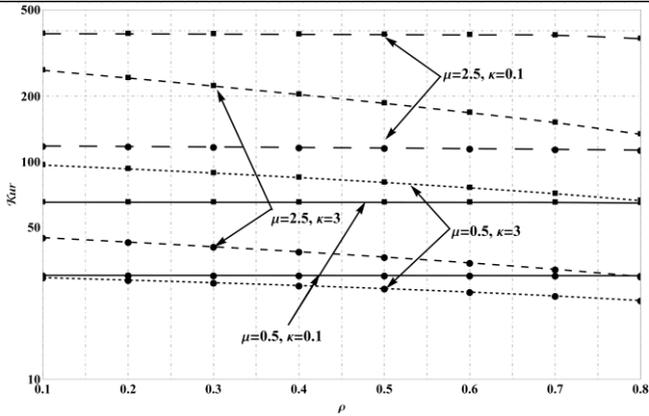


Рис. 9. Зависимость коэффициента эксцесса от ρ

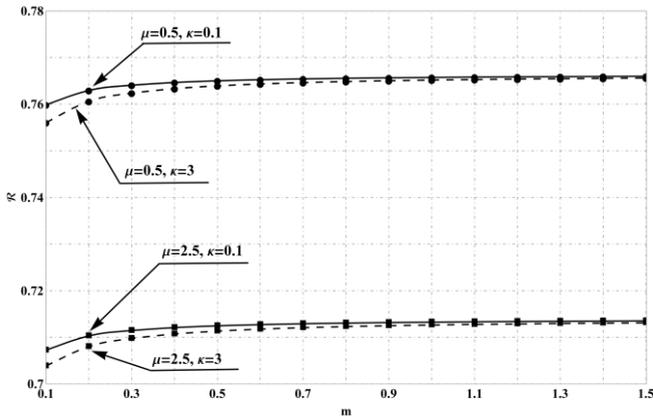


Рис. 10. Зависимость надёжности в определении эргодической пропускной способности от m при $\bar{\gamma} = 0$ дБ

С учётом вышесказанного (малый масштаб изменений) для возможности анализа эффектов влияния m на коэффициенты асимметрии и эксцесса были определены их относительные отклонения (выражаемые в процентах) при изменении одношагового коэффициента корреляции доминантных компонент:

$$\left\{ \begin{array}{l} \delta S\hat{k}(\bar{\gamma}, m) = \frac{|S\hat{k}(\bar{\gamma}, m, \rho = 0, 1) - S\hat{k}(\bar{\gamma}, m, \rho = 0, 5)|}{S\hat{k}(\bar{\gamma}, m, \rho = 0, 1)} \times 100\% \\ \delta \mathcal{K}ur(\bar{\gamma}, m) = \frac{|\mathcal{K}ur(\bar{\gamma}, m, \rho = 0, 1) - \mathcal{K}ur(\bar{\gamma}, m, \rho = 0, 5)|}{\mathcal{K}ur(\bar{\gamma}, m, \rho = 0, 1)} \times 100\% \end{array} \right. , (14)$$

При этом оказалось, что расхождение кривых для двух существенно отличающихся значений коэффициента корреляции (0,1 и 0,5) в диапазоне значений ОСШ вплоть до 30 дБ не превышает 1%, что с практической точки зрения является незначительным.

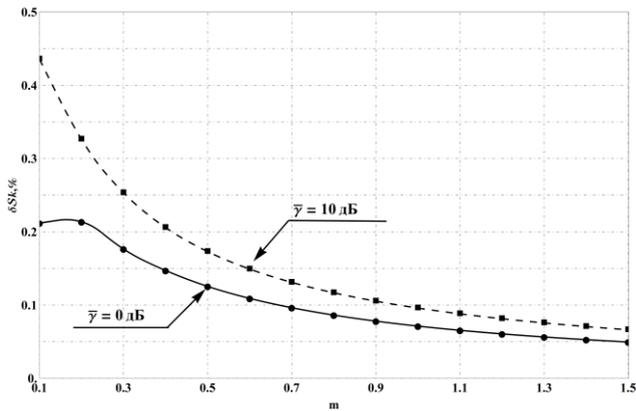


Рис.11. Зависимость относительного изменения коэффициента асимметрии от m (при $\mu=0,5, \kappa=0,1$)

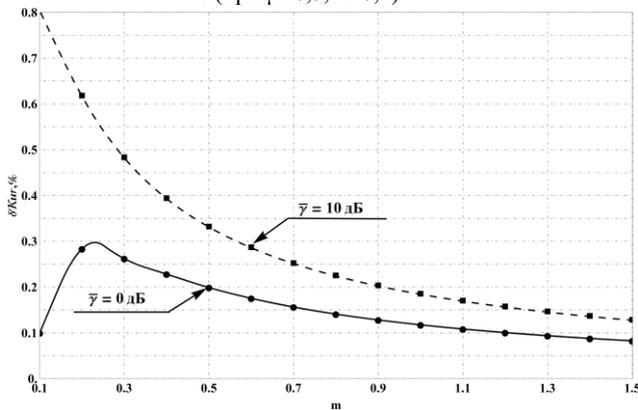


Рис.12. Зависимость относительного изменения коэффициента эксцесса от m (при $\mu=0,5, \kappa=0,1$)

Таким образом, в качестве удобного для практического использования критерия разделения степени затенения доминантных компонент на сильные и слабые (по величине m) может служить уровень $m \approx 1$.

7. Заключение. В работе, с использованием подхода на основе производящей функции моментов было получено аналитическое описание статистик высшего порядка эргодической пропускной способности канала k - μ с коррелированными затенениями для случая SIMO системы, использующей дифференциально-взвешенное комбинирование сигнала на приёмной стороне.

Проведён анализ поведения первых четырёх СПС высшего порядка (эргодической пропускной способности, величины надёжности, коэффициентов асимметрии и эксцесса) в зависимости от характеристик канала (количества многопутевых кластеров, доли мощности, приходящейся на доминантные компоненты, степени затенения доминантных компонент и коэффициента корреляции затенений). В рамках исследования были рассмотрены 4 ситуации поведения анализируемой модели канала, существенно различающиеся по своим свойствам. В результате анализа были сделаны следующие выводы:

- Несмотря на схожее поведение эргодической пропускной способности рассматриваемого канала связи для всех анализируемых моделей, статистики высшего порядка оказываются существенно более чувствительными. Как следствие являются более значимым индикатором флуктуации скорости передачи информации в канале связи.
- Зависимость надёжности эргодической пропускной способности от среднего соотношения сигнал/шум обнаруживает наличие ярко выраженного экстремума (минимума), что с практической точки зрения важно учитывать при предъявлении требований к величине ОСШ в канале, обеспечивающей некоторое желаемое качество функционирования.
- При анализе зависимости СПС высших порядков от степени затенения доминантных компонент было обнаружено, что существенные изменения наблюдаются лишь в области больших затенений (малые m). В области значений $m > 1$ влиянием изменения её величины можно пренебречь.
- При анализе зависимости относительного изменения коэффициента асимметрии и эксцесса от степени затенения оказалось, что расхождение кривых для двух существенно

отличающихся значений коэффициента корреляции (0,1 и 0,5) в диапазоне значений ОСШ вплоть до 30 дБ не превышает 1%.

Полученные результаты могут быть использованы как в рамках теоретических исследований, позволяя более полно понять степень влияния поведения сложных многопутевых каналов связи на характеристики функционирования многоэлементных антенным систем, так и в рамках практических (инженерных) приложений, давая возможность за счёт введения дополнительных ограничений на величины СПС высшего порядка предъявить новые требования к параметрам синтезируемой системы связи (включающей модель канала) для достижения желаемого качества её функционирования.

Литература

1. Massive MIMO White Paper, V. 01, Global TD-LTE Initiative, 2017-02-06.
2. 3GPP. NR; Physical layer; General description. TS 36.211 (V 16.0.0). December 2019. <https://www.3gpp.org/DynaReport/38201.html>
3. IEEE 802.22, “Part 22: Cognitive Radio Wireless Regional Area Networks (WRAN) Medium Access Control (MAC) and Physical Layer (PHY) Specifications” October 2019. https://standards.ieee.org/standard/802_22-2019.html.
4. Probability distributions relevant to radiowave propagation modelling // Recommendations ITU-R P.1057-6 (08/2019). Available at: https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.1057-6-201908-I!!PDF-E.pdf
5. Heath Jr R. W., Lozano A. Foundations of MIMO communication. Cambridge University Press, 2018. ISBN: 9780521762281
6. Kumbhani B., Kshetrimayum R. S. MIMO wireless communications over generalized fading channels. – CRC Press, 2017. ISBN 9781138033009
7. Khatib M. Advanced Trends in Wireless Communications. In-Tech, 2011. 520. DOI: 10.5772/655.
8. Yacoub M. D. The α - μ distribution: A physical fading model for the stacy distribution // IEEE Transactions on Vehicular Technology. 2007. Vol. 56 № 1. pp. 27-34. DOI: 10.1109/TVT.2006.883753.
9. Yacoub M. D. The κ - μ distribution and the η - μ distribution // IEEE Antennas and Propagation Magazine Feb. 2007. vol. 49 № 1. pp. 68–81. DOI: 10.1109/MAP.2007.370983.
10. Rasethuntsa T. R., Kumar S. An integrated performance evaluation of ED-based spectrum sensing over α - κ - μ and α - k - μ -Extreme fading channels //Transactions on Emerging Telecommunications Technologies. 2019. T. 30. №. 5. P. e3569. DOI: 10.1002/ett.3569
11. Paris J. F. Statistical Characterization of κ – μ Shadowed Fading // IEEE Transactions on Vehicular Technology. 2014. vol. 63. № 2. pp. 518–526. DOI: 10.1109/TVT.2013.2281213.
12. Fadnis C., Katiyar B. Review of Higher Order Statistics for Selection Combining Scheme in Weibull Fading Channel //2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC). IEEE, 2017.C. 648-651. DOI: 10.1109/CTCEEC.2017.8455182
13. Peppas K. P. et al. High-order statistics for the channel capacity of EGC receivers over generalized fading channels //IEEE Communications Letters. 2018. T. 22 №. 8. C. 1740–1743. DOI: 10.1109/LCOMM.2018.2846229

14. *Zhang J. et al.* On High-Order Capacity Statistics of Spectrum Aggregation Systems Over κ - μ and κ - μ Shadowed Fading Channels // *IEEE Transactions on Communications*. 2017. Т. 65 № 2. С. 935–944. DOI: 10.1109/TCOMM.2016.2637926.
15. *Sagias N.C., Lazarakis F.I., Alexandridis A.A., Dangakis K.P., Tombras G.S.* Higher order capacity statistics of diversity receivers // *Wireless Personal Communications*. vol. 56, № 4. pp. 649–668. DOI: 10.1007/s11277-009-9837-6.
16. *Yilmaz F., Tabassum H., Alouini M. S.* On the Computation of the Higher Order Statistics of the Channel Capacity for Amplify-and-Forward Multihop Transmission // *IEEE Transactions on Vehicular Technology*. 2014. С. 489–494. DOI: 10.1109/TVT.2013.2272705.
17. *Lopez-Fernandez J., Lopez-Martinez F. J.* Statistical Characterization of Second-Order Scattering Fading Channels // *IEEE Transactions on Vehicular Technology*. 2018. Т. 67. № 12. С. 11345–11353. DOI: 10.1109/TVT.2018.2877941.
18. *Gvozdarev A. S., Patralov P. E., Kanaev I. V.* An Analysis of Ergodic Capacity Higher-Order Statistics for Multiantenna Communication System in Presence of Generalized Shadowed Fading Channel // *2019 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO)*. IEEE, 2019. С. 1–4. DOI: 10.1109/SYNCHROINFO.2019.8814127.
19. *Olver F. W. J. et al.* NIST digital library of mathematical functions, release 1.0. 22. 2019. <http://dlmf.nist.gov/>
20. *Bhatnagar M. R.* On the Sum of Correlated Squared $\kappa - \mu$ Shadowed Random Variables and its Application to Performance Analysis of MRC // *IEEE Transactions on Vehicular Technology*. 2015. vol. 64, № 6. pp. 2674–2684. DOI: 10.1109/TVT.2014.2343453.
21. *Marzetta T. L.* Fundamentals of massive MIMO. – Cambridge University Press, 2016.
22. *H. Bateman and A. Erd'elyi*, Tables of integral transforms. – McGraw-Hill, 1954. <https://authors.library.caltech.edu/43489/1/Volume%201.pdf>.
23. *Abdi A., Kaveh M.* A space-time correlation model for multielement antenna systems in mobile fading channels // *IEEE Journal on Selected Areas in Communications*. – 2002. vol. 20, № 3. pp. 550–560. DOI: 10.1109/49.995514.

Гвоздарев Алексей Сергеевич — канд. физ.-мат. наук, доцент, кафедра инфокоммуникаций и радиофизики физического факультета, Ярославский государственный университет им. П.Г. Демидова. Область научных интересов: статистическая обработка сигналов в беспроводных системах связи, применение методов математической статистики в задачах обработки и передачи информации. Число научных публикаций — 60. a.gvozdarev@uniyar.ac.ru, ORCID: <https://orcid.org/0000-0001-9308-4386>; ул. Советская., д. 14, г. Ярославль, 150003, РФ; р.т.: +7(4852) 797770, факс: +7(4852)25-57-87.

Патралов Павел Евгеньевич — магистр инфокоммуникационных технологий и систем связи, кафедра инфокоммуникаций и радиофизики физического факультета, Ярославский государственный университет им. П.Г. Демидова. Область научных интересов: математическая статистика, теория вероятности, применение методов математического моделирования теории связи. Число научных публикаций — 5. p.patralov1@stud.uniya.ac.ru, ул. Советская., д. 14, г. Ярославль, 150003, РФ; р.т.: +7(4852) 797770, факс: +7 (4852)25-57-87.

A. Gvozhdarev, P. Patralov

PROBABILISTIC ANALYSIS OF GENERALISED STATISTIC MODEL FOR MULTIPATH CHANNEL OF SIMO SYSTEM WITH FADING AND CORRELATED SHADOWING

Gvozhdarev A., Patralov P. The Statistical Analysis of the Multipath Fading SIMO Channel in Presence of Correlated Shadowing.

Abstract. The paper considers the problem of analysis of the information transmission process by multi-element communication systems in presence of a multipath signal propagation channel. To generalize the propagation effects, the model of the κ - μ fading channel with correlated shadowing was assumed, and the technology used for organizing a multi-element system was the SIMO system, equipped with the maximum-ratio combiner of the signal on the receiving side. To describe the characteristics of the information transfer process, an approach based on the higher-order statistics of the ergodic capacity was used. Closed-form analytical expressions for arbitrary-order capacity higher-order statistics were obtained for the channel model under consideration. The behavior of the first four statistics (ergodic capacity, its reliability, skewness and kurtosis coefficients) is analyzed depending on the channel parameters (the number of multipath propagation clusters, the ratio of power of the dominant components to the total power of multipath waves, the degree of shadowing of the dominant components, and the shadowing correlation coefficient). Within the framework of the study, 4 distinct situations of the assumed channel model behavior were considered, which significantly differ in their properties. It is noted that, in contrast to the capacity, its higher-order statistics are significantly more sensitive to the channel parameters and, as a result, are more significant indicators of fluctuations in the information transfer rate within the communication channel. The existence of a pronounced extremum (minimum) of the reliability ergodic capacity dependence from the signal-to-noise ratio was established. It should be accounted for in practical applications, when the requirements of the signal-to-noise ratio that guarantees the desired communication link quality are set.

Keywords: Wireless Channel, Fading, Shadowing, κ - μ Model, Capacity Higher-Order Statistics, SIMO, Signal-to-Noise Ratio

Gvozhdarev Aleksey — Ph.D., Associate Professor at the Department of Infocommunications and Radiophysics (Physics Faculty) P.G. Demidov Yaroslavl State University. Research interests: statistical signal processing in wireless communication systems, application of mathematical statistics methods for information transmission and processing. The number of publications — 60. a.gvozhdarev@uniyar.ac.ru, ORCID: <https://orcid.org/0000-0001-9308-4386>; Sovetskaya str., 14, Yaroslavl, 150003, Russia; office phone: +7(4852) 797770, fax: +7 (4852)25-57-87.

Patralov Pavel — Master of Infocommunication Technologies and Communication Systems at P.G. Demidov Yaroslavl State University. Research interests: mathematical statistics, probability theory, application of mathematical modeling in wireless communications. The number of publications — 5. p.patralov1@stud.uniya.ac.ru; Sovetskaya str., 14, Yaroslavl, 150003, Russia; office phone: +7(4852) 797770, fax: +7 (4852)25-57-87.

References

1. Massive MIMO White Paper, v. 01, Global TD-LTE Initiative, 2017-02-06.
2. 3GPP. NR; Physical layer; General description. TS 36.211 (V 16.0.0). December 2019. <https://www.3gpp.org/DynaReport/38201.htm>

3. IEEE 802.22, "Part 22: Cognitive Radio Wireless Regional Area Networks (WRAN) Medium Access Control (MAC) and Physical Layer (PHY) Specifications" October 2019. https://standards.ieee.org/standard/802_22-2019.html
4. Probability distributions relevant to radiowave propagation modelling // Recommendations ITU-R P.1057-6 (08/2019). Available at: https://www.itu.int/dms_pubrec/itu-r/rec/p/R-REC-P.1057-6-201908-I!!PDF-E.pdf
5. *Heath Jr R. W., Lozano A.* Foundations of MIMO communication. Cambridge University Press, 2018. ISBN: 9780521762281.
6. *Kumbhani B., Kshetrimayum R. S.* MIMO wireless communications over generalized fading channels. – CRC Press, 2017. ISBN 9781138033009.
7. *Khatib M.* Advanced Trends in Wireless Communications. In-Tech, 2011. 520. DOI: 10.5772/655.
8. *Yacoub M. D.* The α - μ distribution: A physical fading model for the stacy distribution // IEEE Transactions on Vehicular Technology. 2007. vol. 56 № 1. pp. 27-34. DOI: 10.1109/TVT.2006.883753.
9. *Yacoub M. D.* The κ - μ distribution and the η - μ distribution // IEEE Antennas and Propagation Magazine Feb. 2007 vol. 49 №. 1. pp. 68–81. DOI: 10.1109/MAP.2007.370983.
10. *Rasethumtsa T. R., Kumar S.* An integrated performance evaluation of ED-based spectrum sensing over α - κ - μ and α - k - μ -Extreme fading channels //Transactions on Emerging Telecommunications Technologies. 2019. vol. 30. №. 5. p. e3569. DOI: 10.1002/ett.3569.
11. *Paris J. F.* Statistical Characterization of κ - μ Shadowed Fading // IEEE Transactions on Vehicular Technology. 2014. vol. 63. № 2. pp. 518–526. DOI: 10.1109/TVT.2013.2281213.
12. *Fadnis C., Katiyar B.* Review of Higher Order Statistics for Selection Combining Scheme in Weibull Fading Channel //2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC). IEEE, 2017. pp. 648–651. DOI: 10.1109/CTCEEC.2017.8455182.
13. *Peppas K. P. et al.* High-order statistics for the channel capacity of EGC receivers over generalized fading channels //IEEE Communications Letters. 2018. vol. 22. №. 8. pp. 1740–1743. DOI: 10.1109/LCOMM.2018.2846229.
14. *Zhang J. et al.* On High-Order Capacity Statistics of Spectrum Aggregation Systems Over κ - μ and κ - μ Shadowed Fading Channels //IEEE Transactions on Communications. 2017. T. 65. №. 2. c. 935–944. DOI: 10.1109/TCOMM.2016.2637926.
15. *Sagias N.C., Lazarakis F.I., Alexandridis A.A., Dangakis K.P., Tombras G.S.* Higher order capacity statistics of diversity receivers // Wireless Personal Communications. vol. 56, № 4. pp. 649–668. DOI: 10.1007/s11277-009-9837-6.
16. *Yilmaz F., Tabassum H., Alouini M. S.* On the Computation of the Higher Order Statistics of the Channel Capacity for Amplify-and-Forward Multihop Transmission // IEEE Transactions on Vehicular Technology. 2014. pp. 489-494. DOI: 10.1109/TVT.2013.2272705.
17. *Lopez-Fernandez J., Lopez-Martinez F. J.* Statistical Characterization of Second-Order Scattering Fading Channels //IEEE Transactions on Vehicular Technology. 2018. vol. 67. №. 12. pp. 11345–11353. DOI: 10.1109/TVT.2018.2877941.
18. *Gvozdev A. S., Patralov P. E., Kanaev I. V.* An Analysis of Ergodic Capacity Higher-Order Statistics for Multiantenna Communication System in Presence of Generalized Shadowed Fading Channel //2019 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). IEEE, 2019. pp. 1–4. DOI: 10.1109/SYNCHROINFO.2019.8814127.
19. *Olver F. W. J. et al.* NIST digital library of mathematical functions, release 1.0. 22. 2019. <http://dlmf.nist.gov/>

20. *Bhatnagar M. R.* On the Sum of Correlated Squared $\kappa - \mu$ Shadowed Random Variables and its Application to Performance Analysis of MRC // IEEE Transactions on Vehicular Technology. 2015. vol. 64 № 6. pp. 2674–2684. DOI: 10.1109/TVT.2014.2343453.
21. *Marzetta T. L.* Fundamentals of massive MIMO. Cambridge University Press, 2016.
22. *H. Bateman and A. Erd'elyi*, Tables of integral transforms. McGraw-Hill, 1954. <https://authors.library.caltech.edu/43489/1/Volume%201.pdf>.
23. *Abdi A., Kaveh M.* A space-time correlation model for multielement antenna systems in mobile fading channels // IEEE Journal on Selected Areas in Communications. 2002. vol. 20, № 3. pp. 550-560. DOI: 10.1109/49.995514.

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 разрешены. 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 pesoal 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 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 31.05.2021

Printed in Publishing center GUAP, 67, B. Morskaya, St. Petersburg, 190000, Russia

The journal is registered in the Russian Federal Agency for Communications and Mass-Media Supervision, certificate ПИ № ФС77-79228 dated September 25, 2020
Subscription Index П5513, Russian Post Catalog

Подписано к печати 31.05.2021. Формат 60×90 1/16. Усл. печ. л. 12,26. Заказ № 204.

Тираж 300 экз., цена свободная.

Отпечатано в Редакционно-издательском центре ГУАП, 190000, Санкт-Петербург, Б. Морская, д. 67

Журнал зарегистрирован Федеральной службой по надзору в сфере связи и массовых коммуникаций, свидетельство ПИ № ФС77-79228 от 25 сентября 2020 г.

Подписной индекс П5513 по каталогу «Почта России»

ISSN 2713-3192



9 772713 319007 >