

ISSN 2713-3192
DOI 10.15622/ia.2023.22.4
<http://ia.spcras.ru>

TOM 22 № 4

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

INFORMATICS AND AUTOMATION



СПб ФИЦ РАН

Санкт-Петербург
2023



INFORMATICS AND AUTOMATION

Volume 22 № 4, 2023

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. V. Nikulin	Prof., Ph.D., New York, United States
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
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

Editor: A. S. Lopotova

Interpreter: Ya. N. Berezina

Art editor: N. A. Dormidontova

Editorial office address

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

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

The journal is indexed in Scopus

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

© St. Petersburg Federal Research Center of the Russian Academy of Sciences, 2023

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

Том 22 № 4, 2023

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.В. Никулин	д-р наук, проф., Нью-Йорк, США
В.Ю. Осипов	д-р техн. наук, проф., Санкт-Петербург, РФ
В. Х. Пшихолопов	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
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: А. С. Лопотова

Переводчик: Я. Н. Березина

Художественный редактор: Н. А. Дормидонтова

Адрес редакции

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

e-mail: ia@spcras.ru, 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», 2023
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

Information Security

A. Sinyuk, O. Ostroumov, A. Tarasov
INFORMATION-THEORETIC REPRESENTATION OF INTERCEPTION
NETWORK CHANNEL VIRTUALIZATION 721

I. Polansky, K. Loginov
OPTIMAL NONLINEAR FILTERING OF INFORMATION IMPACT
ESTIMATES IN A STOCHASTIC MODEL OF INFORMATION WARFARE 745

M. Gidwani, A. Rao
COMPARATIVE ANALYSIS OF RUMOUR DETECTION ON SOCIAL
MEDIA USING DIFFERENT CLASSIFIERS 777

Nguyen Viet Hung, Thang Quang Loi, Nguyen Thi Huong, Tran Thi Thuy Hang,
Truong Thu Huong
AAFNDL - AN ACCURATE FAKE INFORMATION RECOGNITION MODEL
USING DEEP LEARNING FOR THE VIETNAMESE LANGUAGE 795

Digital Information Telecommunication Technologies

A. Gaiduk, N. Prokopenko, A. Bugakova
ERRORS COMPENSATION CAUSED BY TIME DELAY OF DIGITAL
SENSORS 826

I. Glukhikh, A. Prokhoshin, D. Glukhikh
COMPARISON AND RETRIEVAL OF SITUATIONS IN THE CASE-BASED
REASONING SYSTEM FOR SMART-FARM 853

Abdullah Hashem. M.A., Kumar A.V. Senthil, Qasem Ahmed A.A.,
Mosleh M.A.S.
HYBRID OPTIMIZATION BASED ON SPECTRUM AWARE
OPPORTUNISTIC ROUTING FOR COGNITIVE RADIO AD HOC
NETWORKS 880

V. Zelentsov, I. Pimanov, S. Potryasayev
INTEGRATION OF HETEROGENEOUS INFORMATION RESOURCES AND
EARTH REMOTE SENSING DATA IN MONITORING AND MANAGEMENT
OF TERRITORIAL DEVELOPMENT 906

СОДЕРЖАНИЕ

Информационная безопасность

А.Д. Синюк, О.А. Остроумов, А.А. Тарасов
ТЕОРЕТИКО-ИНФОРМАЦИОННОЕ ПРЕДСТАВЛЕНИЕ
ВИРТУАЛИЗАЦИИ СЕТЕВОГО КАНАЛА ПЕРЕХВАТА 721

И.С. Полянский, К.О. Логинов
ОПТИМАЛЬНАЯ НЕЛИНЕЙНАЯ ФИЛЬТРАЦИЯ ОЦЕНОК
ИНФОРМАЦИОННОГО ВОЗДЕЙСТВИЯ В СТОХАСТИЧЕСКОЙ
МОДЕЛИ ИНФОРМАЦИОННОГО ПРОТИВОБОРСТВА 745

М. Гидвани, А. Рао
СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВЫЯВЛЕНИЯ СЛУХОВ В СОЦИАЛЬНЫХ
СЕТЯХ С ИСПОЛЬЗОВАНИЕМ РАЗЛИЧНЫХ КЛАССИФИКАТОРОВ 777

Н.В. Хунг, Т.К. Лои, Н.Т. Хьонг, Т.Т.Т. Ханг, Т.Т. Хьонг
AAFNDL — ТОЧНАЯ МОДЕЛЬ РАСПОЗНАВАНИЯ ПОДДЕЛЬНОЙ
ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ ГЛУБОКОГО ОБУЧЕНИЯ
ВЬЕТНАМСКОГО ЯЗЫКА 795

Цифровые информационно-телекоммуникационные технологии

А.Р. Гайдук, Н.Н. Прокопенко, А.В. Бугакова
КОМПЕНСАЦИЯ ОШИБОК, ВЫЗВАННЫХ ВРЕМЕННОЙ ЗАДЕРЖКОЙ
ЦИФРОВЫХ ДАТЧИКОВ 826

И.Н. Глухих, А.С. Прохошин, Д.И. Глухих
СРАВНЕНИЕ И ОТБОР СИТУАЦИЙ В СИСТЕМАХ ВЫВОДА РЕШЕНИЙ
НА ПРЕЦЕДЕНТАХ ДЛЯ «УМНОЙ» ФЕРМЫ 853

Абдулла Хишам М.А., Кумар А.В. Сентхил, Касем Ахмед А.А.,
Мослех М.А.С.
ОПОРТУНИСТИЧЕСКАЯ МАРШРУТИЗАЦИЯ НА ОСНОВЕ
ГИБРИДНОЙ ОПТИМИЗАЦИИ С УЧЕТОМ СПЕКТРА ДЛЯ
САМООРГАНИЗУЮЩИХСЯ СЕТЕЙ КОГНИТИВНОЙ РАДИОСВЯЗИ 880

В.А. Зеленцов, И.Ю. Пиманов, С.А. Потрясаев
ИНТЕГРАЦИЯ РАЗНОРОДНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ И
ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ПРИ
МОНИТОРИНГЕ И УПРАВЛЕНИИ РАЗВИТИЕМ ТЕРРИТОРИЙ 906

А.Д. СИНЮК, О.А. ОСТРОУМОВ, А.А. ТАРАСОВ
**ТЕОРЕТИКО-ИНФОРМАЦИОННОЕ ПРЕДСТАВЛЕНИЕ
ВИРТУАЛИЗАЦИИ СЕТЕВОГО КАНАЛА ПЕРЕХВАТА**

Синюк А.Д., Остроумов О.А., Тарасов А.А. Теоретико-информационное представление виртуализации сетевого канала перехвата.

Аннотация. Сложнейшей задачей защищенных телекоммуникационных систем, использующих симметричное шифрование, в связи с необходимостью предварительной и ресурсоемкой организации секретных каналов доставки ключей сетевым корреспондентам, является управление ключами. Альтернативой выступают методы формирования ключей по открытым каналам связи. В теории информации показано, что эти методы реализуются при условии превышения информационной скорости канала корреспондентов над скоростью канала перехвата нарушителя. Актуализируется поиск методов, обеспечивающих получение информационного преимущества корреспондентов. Цель заключается в определении теоретико-информационных условий формирования виртуальной сети и канала перехвата, для которых обеспечивается лучшее у корреспондентов отношение информационных скоростей по сравнению с отношением исходных сети и канала перехвата. В работе предлагается модель передачи информации, включающая модель связности и метод передачи информации для асимптотических длин кодовых слов. Модель включает трех корреспондентов и отличается введением идеального широковещательного канала в дополнение к широковещательному каналу с ошибками. В модели введен источник «зашумляющей» информации, которая передается по каналу с ошибками, поэтому передача кодовых слов с использованием известного метода случайного кодирования производится по каналу без ошибок. Для асимптотических длин кодовых слов все действия корреспондентов по обработке и передаче информации в модели сведены в предлагаемый метод передачи информации. Использование метода корреспондентами в рамках модели передачи позволяет одновременно сформировать для них новый виртуальный широковещательный канал с информационной скоростью, как и в первоначальном канале с ошибками, а для нарушителя новый виртуальный широковещательный канал перехвата со скоростью меньшей информационной скорости первоначального канала перехвата. Теоретико-информационные условия ухудшения канала перехвата доказываются в утверждении. Практическая значимость полученных результатов заключается в возможности использования последних для оценки информационной эффективности открытого сетевого формирования ключей в предложенной модели передачи информации, а также в развитии известных научных достижений открытого ключевого согласования. Предлагаемая модель передачи может быть полезной для проведения исследований систем управления ключами и защиты информации, передаваемой по открытым каналам. Дальнейшие исследования связаны с теоретико-информационной оценкой сетевой ключевой пропускной способности, представляющей собой потенциальную теоретико-информационную скорость формирования сетевого ключа.

Ключевые слова: модель передачи информации в сети связи, сетевой корреспондент, нарушитель, источник «шумовой» последовательности, виртуальный канал перехвата, метод передачи информации для асимптотических длин кодовых слов.

1. Введение. Одной из сложных задач криптографических телекоммуникационных систем, использующих симметричные

шифры, является задача управления ключами, обусловленная предварительной организацией секретного для нарушителя канала доставки ключей корреспондентам [1]. В настоящее время задача решается с использованием защищенных каналов связи, что дорого, не оперативно и не всегда возможно. Обнаруженная компрометация ключа нарушителем приводит к существенным временным задержкам информации. В условиях функционирования защищенных сетей связи на едином ключе вероятность компрометации увеличивается с увеличением корреспондентов, что приводит к увеличению времени простоя, связанного с большим временем организации доставки нового ключа всем корреспондентам [2].

Уменьшение информационных задержек возможно с использованием методов формирования сетевого ключа по доступным открытым каналам связи. Исследования в области теории информации [3] показали, что предпосылкой к успешной реализации метода открытого ключевого согласования является выполнение условия превышения информационной скорости канала корреспондентов над скоростью канала перехвата нарушителя. Это актуализирует поиск методов, обеспечивающих получение информационного преимущества корреспондентов. Другими словами, необходимо выполнить поиск условий, при которых обеспечивается преимущество по качеству канала(ов), который используют корреспонденты для формирования общего ключа над каналом перехвата информации, утекающей к нарушителю. Поэтому исследованию модели канала перехвата (КП) в процессе открытого ключевого согласования уделяется большое внимание. Так, например, в ходе выполнения метода распределения ключей для пары корреспондентов по квантовым каналам связи [4] применение поляризованных фотонов, позволяет получить эффект, который нельзя достичь при использовании обычных средств передачи информации. Используя принцип неопределенности Гейзенберга, можно получить канал, в котором затруднен перехват нарушителем без наличия нарушений при передаче-приеме информационного фотона корреспондентами. Одной из атак нарушителя является перехват-ретрансляция поляризованного фотона, которая может быть обнаружена корреспондентами с некоторой вероятностью. В зависимости от результатов измерения нарушений корреспонденты могут отказаться или принять сеанс формирования ключа. В другом известном методе «Модели источника» [5] формирование ключа между парой корреспондентов становится возможным, если качество одного из каналов от источника к одному из корреспондентов лучше качества КП

от источника к нарушителю. Для дискретных, в том числе двоичных, каналов связи под термином «качество канала» следует понимать вероятность ошибки в канале [3]. В качестве третьего примера выступает метод формирования сетевого ключа [6], для которого возможность формирования сетевого ключа (СКЛ) обеспечивается при выполнении условия превышения числового значения вероятности ошибки КП над числовым значением суммы вероятностей ошибок в каналах связи между сетевыми корреспондентами (СК). Ужесточение требований по качеству КП связано с необходимостью формирования СКЛ по сравнению с предшествующим примером.

Цель работы заключается в определении теоретико-информационных условий формирования виртуальных сети и канала перехвата, для которых обеспечивается лучшее для корреспондентов отношение информационных скоростей по сравнению с отношением исходных сети и канала перехвата. Круг проблем, составляющих основное содержание теории информации [7], можно охарактеризовать как исследование методов кодирования для экономного представления сообщений различных источников и для надежной передачи сообщений по каналам связи с шумом основанное на статистическом описании источников сообщений, каналов связи и измерении количества информации между сообщениями по Шеннону [3]. Тогда целевые искомые теоретико-информационные условия определяются построением модели передачи информации (МПИ) в сети связи (СС), описывающей формирование информационного преимущества корреспондентов СС. Разработка МПИ определяется задачей построения модели связности (МС) в СС, включающей источники информации и каналы связи как идеальные, так и с ошибками. Разработка только МС не позволяет достигнуть цели исследования, поэтому в качестве второй задачи представляется разработка метода передачи информации (МтПИ) для асимптотических длин кодовых слов (КС), определяющего действия корреспондентов по обработке и передаче информации в МПИ. В качестве основного метода исследования используются теория информации.

В результате исследований построена новая МПИ включающая МС и МтПИ. Первая компонента включает трех СК и нарушителя (Нр). В рамках МС производится передача информации по совокупности широкополосных каналов связи [8] с ошибками и без ошибок таким образом, что в результате для корреспондентов формируются новый виртуальный широкополосный канал (ШВК) с качеством равным первоначальному каналу. Для нарушителя формируется новый виртуальный канал перехвата, значительно

уступающий своему первоначальному образу по количеству информации на своем выходе. Произведена постановка задачи. Обмен информацией основан на известных моделях случайного кодера и детерминированного декодера [6], функционирующих в новых условиях взаимосвязи с введенным источником «зашумляющей» информации. Разработан асимптотический МтПИ. Математически строго доказано утверждение об отсутствии информационных потерь в виртуальном широкополосном канале корреспондентов и их наличии в виртуальном канале перехвата. Для СС описываемой моделью двоичного ШВК показано условие по качеству каналов, выполнение которого обеспечивает достижение цели исследования.

Научная новизна полученных результатов определяется построением новой МПИ в СС, определяющей теоретико-информационные условия формирования виртуальных ШВК и КП, для которых обеспечивается информационный выигрыш для корреспондентов.

Практическая значимость результатов определяется предметом теории информации [3], сущность которого заключается в доказательстве теорем, устанавливающих предельные возможности различных методов обработки и передачи сообщений.

Практическая значимость заключается в создании условий для оценки предельного значения информационной эффективности открытого сетевого формирования ключей в предложенной МПИ СС, а также в развитии известных научных достижений открытого ключевого согласования [1, 4, 5, 6]. Кроме этого, МПИ может быть полезной для проведения исследований систем управления ключами и защиты информации, передаваемой по открытым каналам. Дальнейшие исследования связаны с теоретико-информационной оценкой сетевой ключевой пропускной способности, представляющей собой потенциальную теоретико-информационную скорость формирования сетевого ключа.

Логика проведенного исследования определяет структуру статьи. В постановке задачи предлагается МПИ в СС, которая декомпозируется на новые МС и МтПИ. Для описания передачи информационных последовательностей асимптотических длин в теории информации используется случайное кодирование [9], поэтому предлагаются известные модели случайного кодера и детерминированного декодера [6] с описанием особенностей использования. Для вышеописанных условий предлагается МтПИ определяющий порядок передачи и обработки информационных последовательностей асимптотических длин. Особенность модели нарушителя [6] заключается в наличии

дополнительного КП без ошибок в целях контроля идеального ШВК. В дальнейшем на основе анализа предложенной МПИ описывается процесс формирования виртуальных ШВК и КП. Формулируется и доказывается утверждение о формировании информационного преимущества виртуального ШВК над виртуальным КП. Произведенный анализ результатов доказательства утверждения позволяет определить требуемое условие по качеству исходных ШВК и КП. Обобщение полученных результатов, их значимость, а также направления дальнейших исследований приведены в заключении.

2. Постановка задачи. Требуется разработать модель передачи информации по открытым каналам сети связи, в которой относительно передаваемой информационной последовательности асимптотической длины качество формируемого в процессе передачи виртуального КП ухудшается относительно качества канала перехвата нарушителя до передачи информации, а качество одновременно формируемых виртуальных открытых каналов сетевых корреспондентов остается неизменным относительно качества открытых каналов СС до передачи информации. С другой точки зрения теории информации реализация этого требования определяет формирование виртуальной сети и КП, для которых обеспечивается лучшее для корреспондентов отношение информационных скоростей по сравнению с отношением исходных сети и канала перехвата.

Для этого предлагается новая МПИ в СС включающая ряд компонент, основными из которых являются МС и МтПИ. Опишем связности предлагаемой МС сети связи, показанной на рисунке 1.

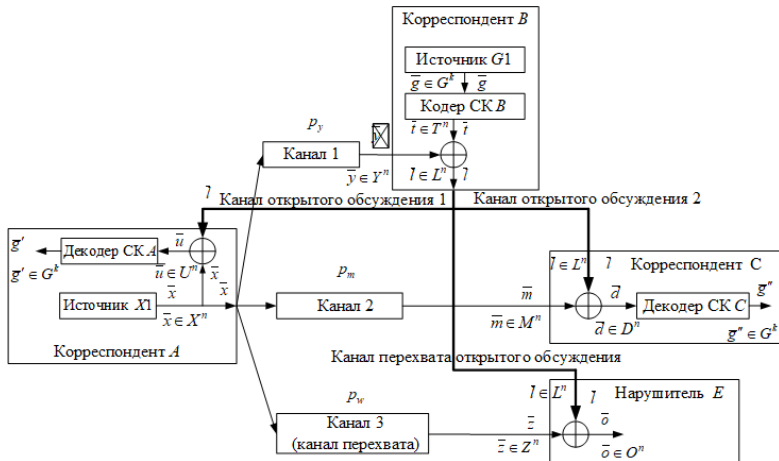


Рис. 1. Модель связности СК А, В, С и нарушителя Е

Корреспондент A имеет источник «шумовой» последовательности. Выход источника связан с первым входом сумматора по модулю 2 (СМ2) и входом ШВК (входами двух независимых составляющих каналов ШВК, описываемых моделями двоичных симметричных каналов связи (ДСК) с ошибками [3]). Канальная связность сетевых корреспондентов в СС описывается моделью двухсоставного широкополосного канала связи [10]. С выходов составляющих каналов сигналы поступают на первые входы СМ2 СК B , C . Выход ШВК контролируется нарушителем E посредством КП, описываемого моделью ДСК. Выход КП связан с первым входом СМ2 нарушителя. У корреспондента B имеется источник сообщений $G1$. Выход $G1$ связан с входом случайного кодера (СлК). Выход СлК связан со вторым входом СМ2 СК B . Выход СМ2 СК B с входом безошибочного ШВК (ШВКбо). В направлении от СК B к СК A и C имеется ШВКбо, включающий два составных канала, описываемых моделями ДСК без памяти [3], в которых отсутствуют ошибки. Назовем эти каналы – каналами открытого обсуждения (КОО) и обозначим КОО-1, КОО-2, соответственно. Выход КОО-1 связан со вторым входом СМ2, а выход СМ2 связан входом детерминированного декодера (ДД) СК A . Выход КОО-2 связан с вторым входом СМ2, а выход СМ2 связан входом ДД СК C . Выход ШВКбо контролируется Нр E посредством канала утечки информации, который описывается моделью ДСК без ошибок. Назовем его каналом перехвата открытого обсуждения (КПОО). Выход КПОО связан со вторым входом СМ2 нарушителя.

Кратко опишем элементы МС. Пусть первая компонента – источник «шумовой» последовательности $X1$ описывается моделью двоичного стационарного источника без памяти [11] и задается ансамблем $\{X, p(x)\}$. Алфавит задается множеством X и состоит символов $X = \{0, 1\}$, и в каждую единицу времени независимо выбирает i -ю букву x_i из алфавита с одинаковой вероятностью $p(x_i) = 0,5$. Источник генерирует сообщение \bar{x} , представляющее двоичную последовательность длиной n , причем $\bar{x} \in X^n$, где X^n – декартова n -я степень множества X [11].

Второй элемент МС – совокупность, состоящая из двух каналов с общим входом (выход источника $X1$, который находится у СК A) и выходами (входы сумматоров СК B и C), описывается моделью двоичного ШВК с двумя составляющими каналами связи с ошибками. Передача символов по ШВК определяется общим алфавитом X , выходными двоичными алфавитами Y и M и матрицами переходных

вероятностей $P_1 = \{p(y/x)\}$, $P_2 = \{p(m/x)\}$, $x \in X$, $y \in Y, m \in M$. Первый составляющий канал (СтК1) описывается моделью ДСК с вероятностью ошибки p_y , а второй (СтК2) – ДСК с p_m .

Третья компонента – канал перехвата, который представляет собой канал связи от выхода источника $X1$ СК A и выходом на входе СМ2 нарушителя E . Передача сигналов по КП определяется двоичными входным алфавитом X , выходным алфавитом Z и матрицей переходных вероятностей $P_3 = \{p(z/x)\}$, $x \in X, z \in Z$. КП описывается моделью ДСК с вероятностью ошибки p_w . Составляющие каналы ШВК и КП являются каналами с независимыми ошибками [3].

Четвертый элемент – источник сообщений $G1$ СК B описывается моделью двоичного стационарного источника без памяти. Алфавит источника задается множеством G и состоит из двух букв $G = \{0,1\}$. Пусть источник в каждую единицу времени независимо выбирает i -ю букву из алфавита с равной вероятностью $p(g_i) = 0,5$. Источник задается ансамблем $\{G, p(g)\}$. Источник генерирует сообщение \bar{g} , представляющее собой случайную двоичную последовательность длиной k символов источника, причем $\bar{g} \in G^k$.

Пятая составляющая МС – ШВКбо. Пусть вход ШВКбо СК B описывается множеством L^n , где L – двоичный алфавит входа. Пусть выход первого составляющего канала КОО-1 у СК A описывается множеством U^n , где U – двоичный алфавит выхода КОО-1. Пусть выход второго составляющего КОО-2 СК C описывается множеством D^n , где D – двоичный алфавит выхода второго составляющего канала КОО-2. Определим последовательности на входе ШВКбо \bar{l} , причем $\bar{l} \in L^n$, и выходе КОО-1 – \bar{u} , причем $\bar{u} \in U^n$, и выходе КОО-2 – \bar{d} , причем $\bar{d} \in D^n$.

Дополнительно у СК и Нр имеются СМ2 [3]. Выход СМ2 равен символу «1», если на его вход поступают различающие символы, иначе равен «0». Описание моделей СлК и ДД приведено ниже. В отличие от известных [6] в предлагаемой МС используется расположение СлК у СК B и ДД у СК A , поэтому введены модели ШВКбо и КПОО.

Общая постановка задачи: необходимо синтезировать и исследовать условия и порядок действий СК в процессе передачи

информации с постоянной скоростью по открытым каналам СС, описываемой совокупностью моделей ШВК и ШВКбо, обеспечивающие уменьшение информационной скорости передачи на выходе виртуального канала перехвата нарушителя, формируемого в процессе передачи.

Особенности моделей случайного кодера и детерминированного декодера. Информационная последовательность $\bar{g}, \bar{g} \in G^k$, переданная по ШВК, не может быть использована СК, т.к. может содержать ошибочно принятые символы. Это требует коррекции (отображения) принятых последовательностей в декодированные информационные последовательности: $\bar{u} \mapsto \bar{g}'$, где $\bar{g}' \in G^k$ для СК A и $\bar{d} \mapsto \bar{g}''$, где $\bar{g}'' \in G^k$ для СК C . Методология теории информации определяет, что при построении кода рассматриваются только такие коды, для которых кодер является тождественным отображением на множестве кодовых слов [5]. В таком случае коды задаются множествами кодовых слов и декодирующими отображениями, а не кодерами и декодерами, как это использовалось в [12]. Пусть задано достаточно большое n ($n \rightarrow \infty$). Необходимо выбрать код, с помощью которого производится передача информации. Для передачи информации используем модель случайного кодера, предложенного в [6] с особенностью его использования заключающуюся в том, что для формирования «виртуальных» каналов выход его связан не с входом ШВК с ошибками, как показано в [6], а с входом ШВК без ошибок у СК B и расположения ДД у СК A . Другая особенность связана с тем, что случайный кодер использует «двоичный вектор ошибок» длиной n [3] ШВК для формирования зашумленного кодового слова, а ДД у СК A и C сначала снимают «зашумление» и только потом декодируют принятые сообщения.

Кратко опишем модель СЛК. Выберем из T^n множество кодовых слов V , причем V представляет высоковероятное множество типичных последовательностей [3] и задает на T^n некоторый (n, ε_1) код со скоростью R [13] удовлетворяющей условию для $\tau, \tau > 0$:

$$R < F(X; Y; M) - \tau, \quad (1)$$

где $F(X; Y; M)$ – совместная информация (СИ) ШВК [14].

С учетом (1) мощность J множества V не превышает 2^{nR} .

В соответствии с прямой теоремой кодирования ШВК [14] для (n, ε_1) кода обеспечивается средняя вероятность ошибочного декодирования не более ε_1 , причем $\varepsilon_1 > 0$. Разделим V на M_0 непересекающихся подмножеств C_i (кодовых множеств), где $i = 1, 2, \dots, M_0$, равной мощности $D1$:

$$D1 = \frac{J}{M_0}, \quad (2)$$

где M_0 – число последовательностей $G1$:

$$M_0 = 2^k. \quad (3)$$

В соответствии теоремой о высоковероятных множествах [3] для $n \rightarrow \infty$ вероятности элементов V почти равны. Вероятность V близка к 1, т.е. для малого ρ , где $\rho > 0$, $P(V)$ удовлетворяет условию:

$$P(V) > 1 - \rho. \quad (4)$$

Тогда вероятности элементов C_i также близки друг к другу. Выражение (4) означает, что для любой \bar{t} , где $\bar{t} \in T^n$, с вероятностью:

$$p(\bar{t} \in C_i) \geq \rho, \quad (5)$$

найдется C_i , причем $\bar{t} \in C_i$. Тогда $\bar{g}_i \mapsto C_i$.

Кодирование. Пусть на выходе $G1$ появляется \bar{g}_i . Кодер с вероятностью (5) выбирает C_i . Затем с равной вероятностью выбирает кодовое слово \bar{t} , где $\bar{t} \in C_i$, и отправляет \bar{t} на вход ШВКбо. Кодирование задает распределение на T^n :

$$p(\bar{t}) = \begin{cases} \frac{P(\bar{g}_i)}{D1}, & \forall \bar{t} \in C_i; \\ 0, & \forall \bar{t} \notin \bigcup_{i=1}^{M_0} C_i, \end{cases} \quad (6)$$

где $D1$ определяется из (2), а M_0 из (3).

Особенность схем декодирования заключается в том, что ДД своим входом связаны через СМ2, в котором суммируются выход ШВКбо с выходом ШВК.

Декодирование. СК A и C известны C_i , которые выбирают в качестве соответствующих решающих областей $S_i \subseteq U^n, Q_i \subseteq D^n, i=1, \dots, M_0$. Пусть $V1$ – ансамбль решений ДД СК A . Каждая $\bar{u} \in U^n$ определяет решение $v \in V1$ по правилу:

$$v = \begin{cases} v_i, & \forall \bar{u} \in S_i; \\ v_{M_0+1}, & \forall \bar{u} \notin \bigcup_{i=1}^{M_0} S_i. \end{cases} \quad (7)$$

Решение v_i в (7) принимается с вероятностью (4). Решение v_{M_0+1} связано с отказом от декодирования. По аналогии пусть W – ансамбль решений ДД СК B . Каждая $\bar{d} \in D^n$ определяет решение $w \in W$ по правилу:

$$w = \begin{cases} w_i, & \forall \bar{d} \in Q_i, ; \\ w_{M_0+1}, & \forall \bar{d} \notin \bigcup_{i=1}^{M_0} Q_i. \end{cases} \quad (8)$$

Решение w_i в (8) принимается с вероятностью (4), которая больше $(1 - \rho)$. Решение w_{M_0+1} определяет отказ от декодирования.

Вероятность одновременного правильного декодирования \bar{u} и \bar{d} превышает величину $(1 - \rho)^2 - \varepsilon_1$.

Метод передачи информации для асимптотических длин кодовых слов. Введем символ \oplus , который обозначает поразрядное суммирование по модулю 2 двух двоичных последовательностей одинаковой длины. В общем виде метод для достаточно большого n представляет следующую последовательность действий:

1. СК A с использованием $X1$ случайно выбирает «шумовую» \bar{x} . асимптотической длины n , причем $\bar{x} \in X^n, n \rightarrow \infty$.
2. СК A подает \bar{x} на вход ШВК и передает СК B, C .

3. СК B принимает на выходе СтК1 ШВК \bar{y} длины n , причем $\bar{y} \in Y^n$. СК C принимает на выходе СтК2 – \bar{m} длины n , причем $\bar{m} \in M^n$.

4. СК B с использованием $G1$ случайно выбирает информационную последовательность \bar{g}_i асимптотической длины k , причем $k \rightarrow \infty$, $\bar{g}_i \in G^k$, $i = 1, \dots, M_0$.

5. СК B кодирует (n, ε_1) кодом со скоростью R , определенной в (1) \bar{g}_i и формирует по закону (6) кодовое слово \bar{t} длины n , причем $\bar{t} \in V$.

6. СК B поразрядно суммирует по модулю 2 в сумматоре \bar{t} и \bar{y} и получает последовательность «зашумленного» кодового слова \bar{l} длины n :

$$\bar{l} = \bar{y} \oplus \bar{t}. \quad (9)$$

7. СК B подает \bar{l} на вход ШВКбо и передает к СК A, C .

8. СК A принимает на выходе КОО-1, а СК C на выходе КОО-2 «зашумленное» кодовое слово \bar{l} .

9. СК A снимает зашумление с \bar{l} путем поразрядного суммирования \bar{l} с \bar{x} и получает двоичную принятую последовательность \bar{u} длины n , причем $\bar{u} \in U^n$:

$$\bar{u} = \bar{x} \oplus \bar{l}. \quad (10)$$

По аналогии СК C снимает зашумление с \bar{l} посредством \bar{m} и получает двоичную принятую последовательность \bar{d} длины n , причем $\bar{d} \in D^n$:

$$\bar{d} = \bar{m} \oplus \bar{l}. \quad (11)$$

10. По правилу декодирования (7) СК A с вероятностью более, чем $1 - \rho$ выносит решение о сообщении \bar{g}' длины k , причем $\bar{g}' \in G^k$. В противном случае выносит решение об отказе от декодирования. СК C по аналогии с вероятностью более $1 - \rho$ по правилу (8) выносит

решение о сообщении \bar{g} длины k , причем $\bar{g} \in G^k$. Иначе выносит решение об отказе от декодирования.

Модель нарушителя. Выше показано, что перехват информационного обмена в СС производится Нр посредством моделей КП и КПОО. Предполагается, что, как и в [6], нарушитель придерживается пассивной стратегии и может только контролировать информацию на выходах КП и КПОО. Использование Нр активной стратегии, когда он создает информацию и может обмениваться ею с СК описаны в [15]. Дополнительно предполагается, что нарушитель знает полное описание выбранного кода (n, ε_1) и источников $X1$ и $G1$, последовательность и содержание всех действий, выполняемых СК по передаче информации в СС.

Рассмотрим перехватываемую информацию Нр на основе анализа шагов вышеописанного метода передачи. На втором и третьем шагах Нр E наблюдает версию «шумовой» последовательности \bar{x} на выходе КП и принимает на выходе КП последовательность \bar{z} , причем $\bar{z} \in Z^n$. На седьмом и восьмом шагах Нр E наблюдает и принимает «зашумленное» кодовое слово \bar{t} на выходе КПОО. Зашумленная передача информации, используемая СК посредством предлагаемого метода передачи информации в СС, создает условия для формирования виртуального канала перехвата Нр [16]. Знание Нр \bar{z} не увеличивает его информацию о передаваемой СК B информационной последовательности \bar{g} , т.к. они формируются и передаются независимо разными СК. Знание нарушителем «зашумленного» кодового слова \bar{t} оставляет его информацию о переданном кодовом слове на нулевом уровне, т.к. согласно (9) \bar{t} представляет собой сумму \bar{t} и \bar{y} , причем слагаемое \bar{y} в сумме играет роль шума с максимальной собственной энтропией [17]. Это с точки зрения теории информации максимально зашумляет кодовое слово \bar{t} длины n . В [18] доказано, что при таких условиях оптимальной обработкой нарушителя E будет суммирование \bar{t} с \bar{z} для снятия случайного «зашумления» с переданного кодового слова:

$$\bar{d} = \bar{z} \oplus \bar{t}. \quad (12)$$

В результате E получает принятую последовательность \bar{d} , причем $\bar{d} \in O^n$. После выполнения СК предлагаемого МтПИ и Нр вышеописанной обработки имеющейся у него информации действий

из первоначальной модели связности формируется модель «виртуальной» связности, соединяющей выход источника сообщений GI СК B с входами декодеров СК A , C (и нарушителя E), посредством «виртуальных» ШВК и КП с «виртуальными ошибками».

3. Формирование виртуальных каналов. Исследуем, к каким преобразованиям в МС привело использование вышеописанного МтПИ. Для описания передачи в ШВК обозначим \bar{e}_y – вектор ошибок [3] (длиной n двоичных символов) в СтК1 ШВК, причем e_y – двоичный символ ошибки \bar{e}_y , который равен «1» с вероятностью $p(e_y = 1) = p_y$. Для СтК2 обозначим \bar{e}_m как вектор ошибок длиной n , причем e_m – двоичный символ ошибки \bar{e}_m , который равен «1» с вероятностью $p(e_m = 1) = p_m$. Для описания перехвата в КП по аналогии обозначим \bar{e}_w как вектор ошибок длиной n , причем e_w – двоичный символ ошибки \bar{e}_w , который равен «1» с вероятностью $p(e_w = 1) = p_w$. Определим взаимосвязь входа ШВК с его выходами и выходом КП на третьем шаге предлагаемого МтПИ:

$$\bar{y} = \bar{x} \oplus \bar{e}_y,$$

$$\bar{m} = \bar{x} \oplus \bar{e}_m, \tag{13}$$

$$\bar{z} = \bar{x} \oplus \bar{e}_w.$$

Исходя из (13), можно определить, как связаны \bar{y} и \bar{m} :

$$\bar{m} = \bar{y} \oplus \bar{e}_y \oplus \bar{e}_m. \tag{14}$$

На шестом шаге МтПИ СК B формирует \bar{l} , для которого с использованием (9) и (13) запишем:

$$\bar{l} = \bar{x} \oplus \bar{e}_y \oplus \bar{t}. \tag{15}$$

На девятом шаге СК A формирует \bar{u} , для которой на основе (10) и (15) получаем:

$$\bar{u} = \bar{t} \oplus \bar{e}_y. \quad (16)$$

СК С формирует \bar{d} , для которой из (11), (13) и (15) получаем:

$$\bar{d} = \bar{t} \oplus \bar{e}_y \oplus \bar{e}_m. \quad (17)$$

Нарушитель E формирует \bar{o} , для которой из (12), (13) и (15) получаем:

$$\bar{o} = \bar{t} \oplus \bar{e}_y \oplus \bar{e}_w. \quad (18)$$

Анализ преобразований (14), (15), (16), (17), (18) в процессе передачи информации в СС показывает, что исходные ШВК и КП в МС, представленные на рисунке 2(а), преобразуется использованием ШВКбо и КПОО к виртуальным ШВК и КП, которые показаны на рисунке 2(б). В итоге формируется виртуальная МС.

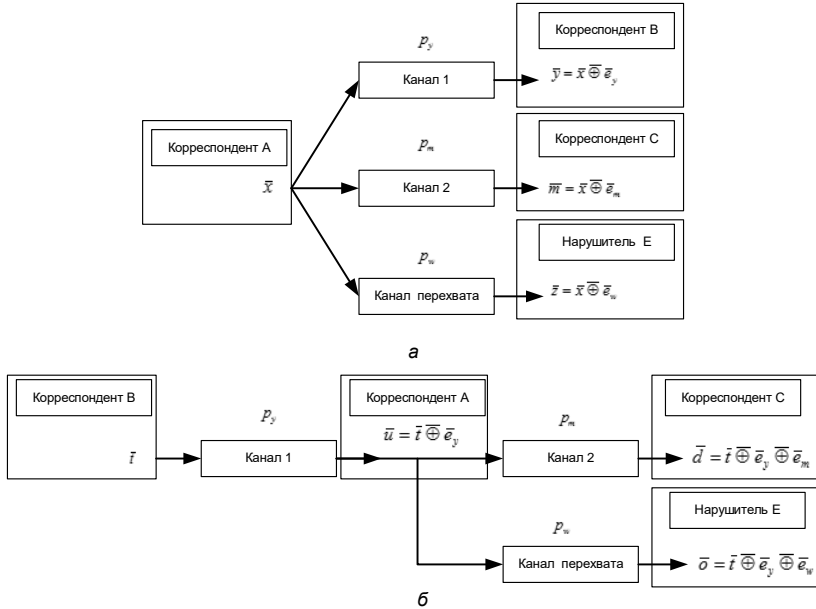


Рис. 2. Переход: а) исходных ШВК и КП; б) к «виртуальным» ШВК и КП

Анализ виртуальной модели связности на рисунке 2 показывает, что в ней, как и в первоначальной МС, имеется ШВК с ошибками, связывающий СК. Такой ШВК описывается моделью ухудшающимся широкополосным каналом (УШВК) [19]. В модели УШВК и в первоначальной модели ШВК первые составляющие ДСК каналы совпадают, а вот второй составляющий канал УШВК (между СК B и C) является ухудшенным вариантом первого составляющего канала УШВК, т.к. в его состав дополнительно входит второй составляющий канал ШВК. Ниже показано, что формирование виртуального ШВК не приводит к потере информации в процессе передачи между СК. В отношении виртуального КП можно сказать, что в нем произошли изменения по сравнению с его первоначальным вариантом. Теперь виртуальный КП (ВКП) описывается моделью УШВК, у которого первым составляющим каналом является первый составляющий канал ШВК, а второй составляющий канал является ухудшенным вариантом первоначального КП. Ниже, в Утверждении, показано, что использование СК МтПИ приводит к потере передаваемой информации между СК B и нарушителем E по сравнению с первоначальным количеством передаваемой информации между СК A и E (если дополнительно не использовать ШВКбо, как ранее показано в [6]). Это определяет целесообразность использования предлагаемого МтПИ в МС.

Утверждение. Пусть СК используют для передачи информации вышеописанный МтПИ в рамках МС с равномерным распределением символов на совмещенном входе формируемых УШВК и ВКП, тогда между информационными мерами до использования МтПИ и после его использования устанавливаются следующие соотношения:

$$\max_{\{p(t)\}} I(T; D) = \max_{\{p(x)\}} F(X; Y; M), \quad (19)$$

$$\max_{\{p(t)\}} I(T; O) \leq \max_{\{p(x)\}} I(X; Z), \quad (20)$$

причем равенство в (20) достигается при $p_y = 0$ и где $F(X; Y; M)$ – совместная информация между входом и выходами канала ШВК [14], $I(T; D)$ – взаимная информация между входом и выходом второго составляющего канала УШВК [20], $I(X; Z)$ – взаимная информация между входом и выходом первоначального КП, $I(T; O)$ – взаимная

информация между входом и выходом ВКП нарушителя.

Доказательство.

Совместная информация $F(X;Y;M)$ ШВК [14] равна:

$$F(X;Y;M) = I(Y;M), \quad (21)$$

где $I(Y;M)$ – взаимная информация между ансамблями Y и M .

Используя (14) и (21) учитывая равномерный закон распределения вероятностей на входе ШВК с ошибками можно записать для $F(X;Y;M)$:

$$F(X;Y;M) = H(Y) - H(Y/M) = 1 - h(p_y + p_m - 2p_y p_m), \quad (22)$$

где $h(q) = -q \log_2 q - (1-q) \log_2 (1-q)$ – энтропийная функция ДСК [3].

Равенство в (22) энтропии выходного ансамбля СтК1 ШВК максимально и $H(Y) = 1$, т.к. максимальна энтропия его входного ансамбля $H(X) = 1$ [3]. Условная энтропия $H(Y/M) = h(p_y + p_m - 2p_y p_m)$, т.к. она зависит от сочетания вероятностей ошибок в составляющих ШВК, обусловленных «сверткой» вероятностей в соответствии с зависимостью между \bar{y} и \bar{m} определенной в (14).

Для взаимной информации $I(T;D)$ и любых t на входе второго составляющего канала УШВК и d на его выходе в соответствии с зависимостью между \bar{y} и \bar{d} определенной в (18) можно записать:

$$I(T;D) = H(D) - H(D/T) = 1 - h(p_y + p_m - 2p_y p_m), \quad (23)$$

причем $H(D) = 1$, если $H(T) = 1$. Последнее достигается при равномерном распределении на входе второго составляющего канала УШВК [3]. Тогда (22) и (23) равны, что доказывает (19) первой части Утверждения. На основе (13) и с учетом равномерного закона распределения вероятностей на входе ШВК можно записать для

$I(X;Z)$ [3]:

$$I(X;Z) = H(Z) - H(Z/X) = 1 - h(p_w). \quad (24)$$

Равенство энтропии выходного ансамбля КП $H(Z)$ в (24) максимально, т.к. максимальна энтропия его входного ансамбля $H(X) = 1$ для ДСК канала [3]. Для оценки $I(T;O)$ и любых t на входе ВКП и символов o на его выходе с использованием (18) и на основе анализа модели нарушителя можно записать:

$$I(T;O) = H(O) - H(O/T) = 1 - h(p_y + p_w - 2p_y p_w), \quad (25)$$

причем $H(O) = 1$, если $H(T) = 1$, что достигается при равномерном распределении на входе ВКП. Сравнение (24) и (25) показывает, что $I(X;Z) \geq I(T;O)$, в силу того, что энтропийная функция канала является возрастающей функцией [3] в области определения, принадлежащей отрезку $[0;0,5]$. Это доказывает неравенство (20) во второй части Утверждения со знаком равенства в случае, если $p_y = 0$ в первом составляющем канале ШВК.

Утверждение доказано.

Анализ результатов утверждения показывает, что передаваемое СК B кодовое слово является общей информацией УШВК для СК и, следовательно, при выполнении МтПИ и условия (19) «качественная» информационная мера $F(X;Y;M)$ ШВК до выполнения МтПИ и информационная мера, характеризующая «качество» виртуального УШВК, после выполнения МтПИ равны. Это говорит о том, что использование МтПИ не ухудшает «качества» передачи общей информации сети связи. Этого нельзя сказать о виртуальном КП, формируемом после выполнения СК МтПИ и условия (20) Утверждения. «Качественная» информационная мера $I(X;Z)$ КП до выполнения МтПИ больше $I(T;O)$ – информационной меры, характеризующей «качество» «виртуального» КП после выполнения МтПИ. Это говорит о том, что использование МтПИ ухудшает «качество» формируемого ВКП. «Качество» формируемых виртуальных УШВК и КП связано с вероятностями ошибок. Анализ (22) и (23) показывает, что вероятности ошибок в ШВК и виртуального УШВК одинаковы и равны оценке $p1$:

$$p1 = p_y + p_m - 2p_y p_m . \quad (26)$$

В то же время вероятность ошибки формируемого ВКП $p2$ равна:

$$p2 = p_y + p_w - 2p_y p_w . \quad (27)$$

Анализ (27) показывает, что $p2 \geq p_w$. Это подтверждает факт того, что использование МтПИ ухудшает «качество» ВКП.

С другой стороны, взаимные информации $I(T; D)$, $I(T; O)$, $I(X; Z)$ определяется как информационная скорости соответствующих каналов связи [3]. Совместная информация $F(X; Y; M)$ является информационной скоростью ШВК [21]. В соответствии с результатами Утверждения, после виртуализации каналов информационная скорость в виртуальном УШВК остается прежней, а информационная скорость в ВКП уменьшается.

Заметим, что для того, чтобы в предлагаемой модели МПИ было возможно формирование сетевого ключа необходимо, чтобы информационная скорость в виртуальном УШВК была больше информационной скорости в ВКП [6]. Для этого необходимо выполнение условия: $p2 > p1$. Анализ (26) и (27) показывает, что это становится возможным в случае выполнения условия: $p_w > p_m$. Выполнение этого условия значительно легче для реализации, чем условие: $p_w > p_m + p_y$ для модели формирование сетевого ключа [6].

4. Заключение. В работе содержатся новые научные результаты теоретико-информационного представления виртуализации сетевого канала перехвата сети связи, включающей трех корреспондентов, связанных совокупностью ШВК с ошибками и ШВК без ошибок. Описана постановка задачи, в которой предлагается МПИ в СС, которая декомпозируется на новые МС и МтПИ. Особенностью предлагаемой МС является использование СК для передачи информации в дополнение к ШВК еще и ШВКбо. Это предопределило разработку нового МтПИ включающего 10 шагов и определяющего порядок передачи и обработки информационных последовательностей асимптотических длин. Особенностью предложенной модели нарушителя является использование им КПОО без ошибок в дополнение к КП. В целом исследуемая МПИ описывает процесс

формирования виртуальных ШВК и КП.

Методом теории информации доказано Утверждение о получении информационного преимущества виртуального ШВК над виртуальным КП, которые формируются в МПИ. Произведен анализ Утверждения как на уровне «качества» виртуальных каналов, так и на уровне информационных скоростей в формируемых каналах связи. Дальнейший анализ позволил найти условие, обеспечивающее получение информационного преимущества СК над Нр, связанное с качеством исходных ШВК и КП. Определенное условие создает предпосылки для разработки ключа с увеличенной информационной скоростью формирования ключа.

Направления дальнейших исследований связаны с теоретико-информационной оценкой в предлагаемой МПИ ключевой пропускной способности сети как оценкой потенциальной информационной скорости формирования сетевого ключа.

Полученные результаты могут быть полезны исследователям для анализа различных систем управления ключами и криптографических систем защиты информации в поиске ответов на вопросы о предельных возможностях перечисленных систем, определения, в какой мере проектируемая система уступает теоретически возможной. Кроме этого, логика вывода, используемая в теоретико-информационном исследовании подобных систем информации, может подсказать исследователям путь, на котором возможно будет найдено конструктивное решение для построения реальной системы.

Литература

1. Zhou H., El Gamal A. Network Information Theoretic Security with Omnipresent Eavesdropping // *IEEE Transactions on Information Theory*. 2021. vol. 67(12). pp. 8280–8299. DOI: 10.1109/TIT.2021.3116962.
2. Mitsugu I., Kazuo O, Junji S. Security Formalizations and their Relationships for Encryption and Key Agreement in Information-Theoretic Cryptography // *IEEE Transactions on Information Theory*. 2018. vol. 64(1). pp. 654–685. DOI: 10.1109/TIT.2017.2744650.
3. Csiszar I., Korner J. Information theory: coding theorems for discrete memoryless systems // Cambridge University Press. 2011. 523 p.
4. Chitambar E., Fortescue B., Hsieh M.-H. The Conditional Common Information in Classical and Quantum Secret Key Distillation // *IEEE Transactions on Information Theory*. 2018. vol. 64(11). pp. 7381–7394. DOI: 10.1109/TIT.2018.2851564.
5. Gohari A., Günlü O., Kramer G. Coding for Positive Rate in the Source Model Key Agreement Problem // *IEEE Transactions on Information Theory*. 2020. vol. 66(10). pp. 6303–6323. DOI: 10.1109/TIT.2020.2990750.
6. Синюк А.Д., Остроумов О.А. Теорема о ключевой пропускной способности сети связи // *Информационно-управляющие системы*. 2018. № 5(96). С. 79–87. DOI: 10.31799/1684-8853-2018-5-79-87.

7. Bloch M., Günlü O., Yener A., Oggier F., Poor H.V., Sankar L., Schaefer R.F. An Overview of Information-Theoretic Security and Privacy: Metrics, Limits and Applications // *IEEE Journal on Selected Areas in Information Theory*. 2021. vol. 2(1). pp. 5–22. DOI: 10.1109/JSAIT.2021.3062755.
8. Pappas N., Kountouris M., Ephremides A., Angelakis V. Stable Throughput Region of the Two-User Broadcast Channel // *IEEE Transactions on Communications*. 2018. vol. 66(10). pp. 4611–4621. DOI: 10.1109/TCOMM.2018.2834943.
9. Averbuch R., Merhav N. Exact Random Coding Exponents and Universal Decoders for the Asymmetric Broadcast Channel // *IEEE Transactions on Information Theory*. 2018. vol. 64(7). pp. 5070–5086. DOI: 10.1109/TIT.2018.2836668.
10. Mohapatra P., Pappas N., Lee J., Quek T.Q.S., Angelakis V. Secure Communications for the Two-User Broadcast Channel with Random Traffic // *IEEE Transactions on Information Forensics and Security*. vol. 13(9). 2018. pp. 2294–2309. DOI: 10.1109/TIFS.2018.2818076.
11. Yu L., Li H., Li W. Distortion Bounds for Source Broadcast Problems // *IEEE Transactions on Information Theory*. 2018. vol. 64. no. 9. pp. 6034–6053. DOI: 10.1109/TIT.2018.2854547.
12. Choi J. A Coding Approach with Key-Channel Randomization for Physical-Layer Authentication // *IEEE Transactions on Information Forensics and Security*. 2019. vol. 14(1). pp. 175–185. DOI: 10.1109/TIFS.2018.2847659.
13. Padakandla A., Sandeep Pradhan S. Achievable Rate Region for Three User Discrete Broadcast Channel Based on Coset Codes // *IEEE Transactions on Information Theory*. 2018. vol. 64(4). pp. 2267–2297. DOI: 10.1109/TIT.2018.2798669.
14. Остроумов О.А., Синюк А.Д. Пропускная способность широковещательного канала связи // *Вестник компьютерных и информационных технологий*. 2019. № 9(183). С. 33–42.
15. Qikun Z., Yongjiao L., Yong G., Chuanyang Z., Xiangyang L., Jun Z. Group Key Agreement Protocol Based on Privacy Protection and Attribute Authentication // *IEEE Access*. 2019. vol. 7. pp. 87085–87096. DOI: 10.1109/ACCESS.2019.2926404.
16. Ghosh S., Natarajan L. Linear Codes for Broadcasting with Noisy Side Information: Bounds and Code Constructions // *IEEE Transactions on Information Theory*. vol. 65(7). 2019. pp. 4207–4226. DOI: 10.1109/TIT.2019.2893617.
17. De Palma G. New lower bounds to the output entropy of multi-mode quantum Gaussian channels // *IEEE Transactions on Information Theory*. 2019. vol. 65. no. 9. pp. 5959–5968. DOI: 10.1109/TIT.2019.2914434.
18. Günlü O., Schaefer R.F. Controllable Key Agreement with Correlated Noise // *IEEE Journal on Selected Areas in Information Theory*. 2021. vol. 2(1). pp. 82–94. DOI: 10.1109/JSAIT.2021.3054035.
19. Zou S., Liang Y., La L., Poor H.V., Shamai S. Degraded Broadcast Channel with Secrecy Outside a Bounded Range // *IEEE Transactions on Information Theory*. 2018. vol. 64(3). pp. 2104–2120. DOI: 10.1109/TIT.2018.2791995.
20. Li C.T., Anantharam V. One-Shot Variable-Length Secret Key Agreement Approaching Mutual Information // *IEEE Transactions on Information Theory*. 2021. vol. 67(8). pp. 5509–5525. DOI: 10.1109/TIT.2021.3087963.
21. Синюк А.Д., Остроумов О.А. Информационная емкость и неопределенность дискретного широковещательного канала связи // *Вестник компьютерных и информационных технологий*. 2018. № 8(170). С. 36–45.

Синюк Александр Демьянович — д-р техн. наук, доцент, профессор кафедры, кафедра «общепрофессиональных дисциплин», Военная орденов Жукова и Ленина краснознаменная академия связи имени Маршала Советского Союза С.М. Буденного (ВАС). Область научных интересов: криптографическая защита информации,

передаваемая по открытым каналам связи. Число научных публикаций — 250. eentrop@rambler.ru; Тихорецкий проспект, 3, 194064, Санкт-Петербург, Россия; р.т.: +7(812)247-9481.

Остроумов Олег Александрович — канд. техн. наук, докторант кафедры, кафедра «защиты инфотелекоммуникационных систем специального назначения», Военная орденов Жукова и Ленина краснознаменная академия связи имени Маршала Советского Союза С.М. Буденного (ВАС). Область научных интересов: криптографическая защита информации, передаваемая по открытым каналам связи, обеспечение безопасности критически важных объектов систем связи и управления. Число научных публикаций — 134. oleg-26stav@mail.ru; Тихорецкий проспект, 3, 194064, Санкт-Петербург, Россия; р.т.: +7(812)247-9832.

Тарасов Александр Алексеевич — адъюнкт кафедры, кафедра «общепрофессиональных дисциплин», Военная орденов Жукова и Ленина краснознаменная академия связи имени Маршала Советского Союза С.М. Буденного (ВАС). Область научных интересов: криптографическая защита информации, передаваемая по открытым каналам связи. Число научных публикаций — 10. taras4912@mail.ru; Тихорецкий проспект, 3, 194064, Санкт-Петербург, Россия; р.т.: +7(812)247-9481.

A. SINYUK, O. OSTROUMOV, A. TARASOV
**INFORMATION-THEORETIC REPRESENTATION OF
INTERCEPTION NETWORK CHANNEL VIRTUALIZATION**

Sinyuk A., Ostroumov O., Tarasov A. **Information-Theoretic Representation of Interception Network Channel Virtualization.**

Abstract. The most difficult task of secure telecommunication systems using symmetric encryption, due to the need for preliminary and resource-intensive organization of secret channels for delivering keys to network correspondents, is key management. An alternative is the generating keys methods through open communication channels. In information theory, it is shown that these methods are implemented under the condition that the channel information rate of correspondents exceeds the rate of the intruder interception channel. The search for methods that provide the informational advantage of correspondents is being updated. The goal is to determine the information-theoretical conditions for the formation of a virtual network and an interception channel, for which the best ratio of information speeds for correspondents is provided compared to the ratio of the original network and interception channel. The paper proposes an information transfer model that includes a connectivity model and an information transfer method for asymptotic lengths of code words. The model includes three correspondents and is characterized by the introduction of an ideal broadcast channel in addition to an errored broadcast channel. The model introduces a source of "noisy" information, which is transmitted over the channel with errors, so the transmission of code words using the known method of random coding is carried out over the channel without errors. For asymptotic lengths of code words, all actions of correspondents in processing and transmitting information in the model are reduced to the proposed method of transmitting information. The use of the method by correspondents within the framework of the transmission model makes it possible to simultaneously form for them a new virtual broadcast channel with information rate as in the original channel with errors, and for the intruder a new virtual broadcast interception channel with a rate lower than the information rate of the initial interception channel. The information-theoretic conditions for deterioration of the interception channel are proved in the statement. The practical significance of the results obtained lies in the possibility of using the latter to assess the information efficiency of open network key formation in the proposed information transfer model, as well as in the development of well-known scientific achievements of open key agreement. The proposed transmission model can be useful for researching key management systems and protecting information transmitted over open channels. Further research is related to the information-theoretic assessment of the network key throughput, which is the potential information-theoretic speed of network key formation.

Keywords: information transmission model in a communication network, network correspondent, intruder, source of «noise» sequence, virtual interception channel, information transmission method for asymptotic lengths of code words.

References

1. Zhou H., El Gamal A. Network Information Theoretic Security with Omnipresent Eavesdropping. *IEEE Transactions on Information Theory*. 2021. vol. 67(12). pp. 8280–8299. DOI: 10.1109/TIT.2021.3116962.
2. Mitsugu I., Kazuo O, Junji S. Security Formalizations and their Relationships for Encryption and Key Agreement in Information-Theoretic Cryptography. *IEEE*

- Transactions on Information Theory. 2018. vol. 64(1). pp. 654–685. DOI: 10.1109/TIT.2017.2744650.
3. Csiszar I., Korner J. Information theory: coding theorems for discrete memoryless systems. Cambridge University Press. 2011. 523 p.
 4. Chitambar E., Fortescue B., Hsieh M.-H. The Conditional Common Information in Classical and Quantum Secret Key Distillation. IEEE Transactions on Information Theory. 2018. vol. 64(11). pp. 7381–7394. DOI: 10.1109/TIT.2018.2851564.
 5. Gohari A., Günlü O., Kramer G. Coding for Positive Rate in the Source Model Key Agreement Problem. IEEE Transactions on Information Theory. 2020. vol. 66(10). pp. 6303–6323. DOI: 10.1109/TIT.2020.2990750.
 6. Sinyuk A.D., Ostroumov O.A. [Theorem about key capacity of a communication network]. Informatsionno-upravliaiushchie sistemy – Information and Control Systems. 2018. no. 5(96). pp. 79–87. DOI: 10.31799/1684-8853-2018-5-79-87. (In Russ.).
 7. Bloch M., Günlü O., Yener A., Oggier F., Poor H.V., Sankar L., Schaefer R.F. An Overview of Information-Theoretic Security and Privacy: Metrics, Limits and Applications. IEEE Journal on Selected Areas in Information Theory. 2021. vol. 2(1). pp. 5–22. DOI: 10.1109/JSAIT.2021.3062755.
 8. Pappas N., Kountouris M., Ephremides A., Angelakis V. Stable Throughput Region of the Two-User Broadcast Channel. IEEE Transactions on Communications. 2018. vol. 66(10). pp. 4611–4621. DOI: 10.1109/TCOMM.2018.2834943.
 9. Averbuch R., Merhav N. Exact Random Coding Exponents and Universal Decoders for the Asymmetric Broadcast Channel. IEEE Transactions on Information Theory. 2018. vol. 64(7). pp. 5070–5086. DOI: 10.1109/TIT.2018.2836668.
 10. Mohapatra P., Pappas N., Lee J., Quek T.Q.S., Angelakis V. Secure Communications for the Two-User Broadcast Channel with Random Traffic. IEEE Transactions on Information Forensics and Security. vol. 13(9). 2018. pp. 2294–2309. DOI: 10.1109/TIFS.2018.2818076.
 11. Yu L., Li H., Li W. Distortion Bounds for Source Broadcast Problems. IEEE Transactions on Information Theory. 2018. vol. 64. no. 9. pp. 6034–6053. DOI: 10.1109/TIT.2018.2854547.
 12. Choi J. A Coding Approach with Key-Channel Randomization for Physical-Layer Authentication. IEEE Transactions on Information Forensics and Security. 2019. vol. 14(1). pp. 175–185. DOI: 10.1109/TIFS.2018.2847659.
 13. Padakandla A., Sandeep Pradhan S. Achievable Rate Region for Three User Discrete Broadcast Channel Based on Coset Codes. IEEE Transactions on Information Theory. 2018. vol. 64(4). pp. 2267–2297. DOI: 10.1109/TIT.2018.2798669.
 14. Ostroumov O.A., Sinyuk A.D. [Broadcast communication channel bandwidth]. Vestnik komp'yuternyh i informacionnyh tekhnologij – Bulletin of computer and information technologies. 2019. №. 9 (183). pp. 33–42. (In Russ.).
 15. Qikun Z., Yongjiao L., Yong G., Chuanyang Z., Xiangyang L., Jun Z. Group Key Agreement Protocol Based on Privacy Protection and Attribute Authentication. IEEE Access. 2019. vol. 7. pp. 87085–87096. DOI: 10.1109/ACCESS.2019.2926404.
 16. Ghosh S., Natarajan L. Linear Codes for Broadcasting with Noisy Side Information: Bounds and Code Constructions. IEEE Transactions on Information Theory. vol. 65(7). 2019. pp. 4207–4226. DOI: 10.1109/TIT.2019.2893617.
 17. De Palma G. New lower bounds to the output entropy of multi-mode quantum Gaussian channels. IEEE Transactions on Information Theory. 2019. vol. 65. no. 9. pp. 5959–5968. DOI: 10.1109/TIT.2019.2914434.
 18. Günlü O., Schaefer R.F. Controllable Key Agreement with Correlated Noise. IEEE Journal on Selected Areas in Information Theory. 2021. vol. 2(1). pp. 82–94. DOI: 10.1109/JSAIT.2021.3054035.

19. Zou S., Liang Y., La L., Poor H.V., Shamai S. Degraded Broadcast Channel with Secrecy Outside a Bounded Range. *IEEE Transactions on Information Theory*. 2018. vol. 64(3). pp. 2104–2120. DOI: 10.1109/TIT.2018.2791995.
20. Li C.T., Anantharam V. One-Shot Variable-Length Secret Key Agreement Approaching Mutual Information. *IEEE Transactions on Information Theory*. 2021. vol. 67(8). pp. 5509–5525. DOI: 10.1109/TIT.2021.3087963.
21. Sinyuk A.D., Ostroumov O.A. [Information capacity and uncertainty of a discrete broadcast communication channel]. *Vestnik komp'yuternyh i informacionnyh tekhnologij – Bulletin of computer and information technologies*. 2018. no. 8(170). pp. 36–45. (In Russ.).

Sinyuk Alexander — Ph.D., Dr.Sci., Associate Professor, Professor of the department, General professional disciplines department, Military communications academy. Research interests: cryptographic protection of information transmitted through open communication channels. The number of publications — 250. centrop@rambler.ru; 3, Tikhoretsky Av., 194064, St. Petersburg, Russia; office phone: +7(812)247-9481.

Ostroumov Oleg — Ph.D., Doctoral student of the department, Information protection department and telecommunication systems of special purpose, Military communications academy. Research interests: cryptographic protection of information transmitted through open communication channels, ensuring the security of communication and control systems critical objects. The number of publications — 134. oleg-26stav@mail.ru; 3, Tikhoretsky Av., 194064, St. Petersburg, Russia; office phone: +7(812)247-9832.

Tarasov Alexander — Adjunct of the department, General professional disciplines department, Military communications academy. Research interests: cryptographic protection of information transmitted through open communication channels. The number of publications — 10. taras4912@mail.ru; 3, Tikhoretsky Av., 194064, St. Petersburg, Russia; office phone: +7(812)247-9481.

И.С. Полянский, К.О. Логинов
**ОПТИМАЛЬНАЯ НЕЛИНЕЙНАЯ ФИЛЬТРАЦИЯ ОЦЕНОК
ИНФОРМАЦИОННОГО ВОЗДЕЙСТВИЯ В СТОХАСТИЧЕСКОЙ
МОДЕЛИ ИНФОРМАЦИОННОГО ПРОТИВОБОРСТВА**

Полянский И.С., Логинов К.О. Оптимальная нелинейная фильтрация оценок информационного воздействия в стохастической модели информационного противоборства.

Аннотация. В статье разработано вычислительно эффективное алгоритмическое решение задачи оптимальной нелинейной фильтрации оценок информационного воздействия в обобщенной стохастической модели информационного противоборства. Сформированное решение применимо при наличии разнородных правил измерения параметров модели информационного противоборства, на основании которых составляется пара систем стохастических дифференциальных уравнений. Оценка информационного воздействия в модели оптимальной нелинейной фильтрации выполняется по критерию максимального правдоподобия по определяемой эволюции апостериорной условной функции плотности вероятности на заданном интервале наблюдения. Нахождение апостериорной условной функции плотности вероятности в заданный момент времени осуществляется с учетом теоремы сложения вероятностей, как вероятность суммы двух совместных событий, функции плотности которых устанавливаются из численного решения соответствующих робастных уравнений Дункана-Мортенсена-Закаи. Для первого события полагается, что первая система стохастических дифференциальных уравнений является уравнением состояния, а вторая - уравнением наблюдения. Для второго события устанавливается их определение в обратном порядке. Решение робастного уравнения Дункана-Мортенсена-Закаи выполнено в постановке спектрального метода Галёркина при дискретизации интервала наблюдения на подынтервалы и сведении исходного решения к численному рекуррентному исследованию последовательности подзадач по так называемому Yau-Yau's алгоритму, предполагающему оценку вероятностной меры из решения прямого уравнения Колмогорова при ее последующей коррекции по наблюдению. Для выделения особенностей алгоритмической реализации составленного решения сформирован алгоритм оптимальной нелинейной фильтрации оценок информационного воздействия в обобщенной стохастической модели информационного противоборства при уточнении листинга исполняющей его функции, который представлен псевдокодом. Для выявления предпочтительности составленного алгоритмического решения по оптимальной нелинейной фильтрации оценок информационного воздействия проведена серия вычислительных экспериментов на тестовых выборках большого объема. Результат оценки информационного воздействия, получаемый по предложенному алгоритму, сравнен с определяемым решением: 1) по средневыборочным значениям из моделей наблюдения; 2) ансамблем расширенным фильтром Калмана; 3) алгоритмом фильтрации, предполагающим численное исследование уравнения Дункана-Мортенсена-Закаи. По проведенному апостериорному исследованию выделены количественные показатели, устанавливающие выигрыш составленного алгоритма и границы его применимости.

Ключевые слова: нелинейная фильтрация, оценка информационного воздействия, стохастическая модель информационного противоборства, робастное уравнение Дункана-Мортенсена-Закаи, гауссовы базисные функции, спектральный метод Галёркина.

1. Введение. Постоянное расширение аудитории социальных сетей и мессенджеров обуславливает значительный рост их влияния на жизнь человека в обществе. Это стимулирует интерес ученых к исследованию процессов информационного воздействия социальных медиа на формирование общественного мнения [1–6]. Наличие у платформ социальных медиа открытого описания способов программного взаимодействия (API – application programming interface) при одновременном развитии сторонних библиотек и расширений (numpy, requests, pandas, seaborn, networkx и пр.) высокоуровневых языков (python), автоматизирующих рутинные функции обработки, визуализации и анализа данных [7], увеличивает мотивацию подобных исследований с позиции формализации представлений при применении, совершенствовании и разработке методов математического моделирования информационного противоборства и управления в структурированном социуме.

Методологическую основу формализованного исследования социальных медиа при построении математических моделей информационного противоборства и управления составляют теории вероятностей [5, 6, 8], игр [2], графов [2, 9] и дифференциальных уравнений [3, 4, 10]. Основные подходы математических представлений по получению оценок информационного воздействия базируются на анализе активных сетевых структур [3] и дифференциальных уравнений [2]. С позиции относительной простоты и точности модели анализа реальной социальной сети в условиях неоднозначности и неполноты информации атрибутивного пространства взаимодействующих пользователей первичной предпочтительностью обладает математическая формализация, основанная на представлениях систем дифференциальных уравнений – модели информационного противоборства в структурированном социуме [2, 4, 10]. Их математическую запись, обобщенно учитывающую выделенные в [2, 4, 10] факторы распространения информации в социуме, можно определить в следующем виде [11]:

$$\frac{dx_{mk}(t)}{dt} = y_{mk}(t) \left[\alpha_{mk}(t) + \sum_{m'=1}^M \beta_{m'm}(t)x_{m'k}(t) \right] - \gamma_{mk}(t)x_{mk}(t); \tag{1}$$

$$\frac{dy_{mk}(t)}{dt} = -\frac{dx_{mk}(t)}{dt} - \gamma_{mk}(t)y_{mk}(t) + \left[\alpha_{mk}(t) + \sum_{m'=1}^M \beta_{m'm}(t)x_{m'k}(t) \right] \left[N_m - \sum_{k'=1}^K (x_{mk'}(t) + y_{mk'}(t)) \right],$$

где $m', m \in \{\overline{1, M}\}$; $k = \overline{1, K}$; K – число субъектов, в отношении которых в социуме формируются предпочтения у индивидов; M – число подгрупп, выделяемых в социуме; N_m – число индивидов в m -й подгруппе при обозначении $N_0 = \sum_{m=1}^M N_m$ общей численности социума; x_{mk} и y_{mk} – число адептов и предадептов соответственно k -го субъекта из m -й подгруппы социума; α_{mk} и γ_{mk} – интенсивности положительной и отрицательной информации, распространяемой внешними источниками в отношении k -го субъекта при учете особенности их восприятия индивидами m -й подгруппы; $\beta_{m'm}$ – интенсивность межличностной коммуникации индивидов m' -й и m -й подгрупп; $t \in [t_0, t_1]$ – момент времени ($[t_0, t_1]$ – временной интервал анализа).

Отдельно следует уточнить, что в формируемом решении на основе математических приближений вида (1) не исключается возможность комбинации моделей информационного противоборства в структурированном социуме с активными сетевыми структурами, к примеру, применяемыми для оценки измеряемых параметров α_{mk} , β_{mk} , γ_{mk} [12]. Недостаток моделей информационного противоборства в структурированном социуме [2, 4, 10] состоит в детерминированном определении их параметров (α_{mk} , β_{mk} , γ_{mk}), что не позволяет в полной мере оценить степень различия между исходными факторами (x_{mk} , y_{mk}) конкурирующих субъектов.

Для устранения указанного недостатка в работах [11, 13, 14] в развитие [2, 4, 10] предложено вводить флуктуацию в измеряемые параметры модели – интенсивности α_{mk} и γ_{mk} . При этом за методологическую основу составления стохастических математических моделей в [11, 13, 14] выбрана теория марковских процессов и процессов диффузионного типа [15, 16]. Для увеличения точности оценки x_{mk} , y_{mk} на фоне наблюдаемых α_{mk} , γ_{mk} , содержащих стохастические компоненты с параметрами отличными от нормального закона распределения [17], в [18] при введении дополнительного фиктивного стохастического дифференциального уравнения состояния разработана схема нелинейной фильтрации. Она предполагает сведение задачи анализа исходного стохастического дифференциального уравнения к численному решению уравнения Дункана–Мортенсена–Закаи [19] спектральным методом Галёркина [20].

Основной недостаток решения [18] заключается в фиктивном введении стохастического дифференциального уравнения состояния (составляется из исходного при оценке стохастических компонент наблюдаемых интенсивностей агитации методами полиспектрального

анализа [21]) в формировании модели оптимальной нелинейной фильтрации оценок информационного воздействия.

Цель настоящей статьи состоит в устранении указанного недостатка [18] при разработке вычислительно эффективного решения обобщенной задачи оптимальной нелинейной фильтрации оценок информационного воздействия в стохастических моделях информационного противоборства. Для достижения сформулированной цели в развитие результатов [18] необходимо решить следующие задачи:

1) формализовать постановку задачи оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства, не предполагающую введения фиктивного стохастического дифференциального уравнения состояния;

2) в разрабатываемой задаче оптимальной нелинейной фильтрации при определении эволюции функции плотности вероятности уточнить особенности применения робастного уравнения Дункана–Мортенсена–Закаи [22];

3) определить численную схему решения задачи оптимальной нелинейной фильтрации и при систематизированном представлении последовательности действий по получению оценок информационного воздействия в стохастической модели информационного противоборства, уточнении и конкретизации особенностей программной реализации составить соответствующий вычислительный алгоритм;

4) выполнить сравнительное апостериорное исследование разработанного алгоритмического решения при выделении особенностей по устойчивости предложенной численной схемы и точности получаемых оценок.

Для общности представлений стохастической модели информационного противоборства, в отличие от [18] установим зависимости от времени и наличие стохастической компоненты в интенсивности межличностной коммуникации $\beta_{m'm}$.

2. Постановка задачи оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства. Рассмотрим динамическую модель вида (1) для начальных условий $x_{mk}(t_0) = y_{mk}(t_0) = 0$, положив, что значения интенсивностей α_{mk} , $\beta_{m'm}$, γ_{mk} зависят от t и складываются из истинных величин $0 \leq \alpha_{mk}^0(t)$, $\beta_{m'm}^0(t)$, $\gamma_{mk}^0(t) < \infty$ и шумов наблюдения $\tilde{\alpha}_{mk}(t)$, $\tilde{\beta}_{m'm}(t)$, $\tilde{\gamma}_{mk}(t)$ с соответствующими статистическими параметрами: $\mathbb{E}[\tilde{\alpha}_{mk}] = \mathbb{E}[\tilde{\beta}_{m'm}] = \mathbb{E}[\tilde{\gamma}_{mk}] = 0$;
 $\text{cov}[\tilde{\alpha}_{mk}, \tilde{\alpha}_{m'k'}] = dt \cdot \varepsilon_{\alpha km'k'}^{\alpha}$; $\text{cov}[\tilde{\beta}_{m'm}, \tilde{\beta}_{j'j}] = dt \cdot \varepsilon_{m'mj'j}^{\beta}$;

$\text{cov} [\tilde{\gamma}_{mk}, \tilde{\gamma}_{m'k'}] = dt \cdot \varepsilon_{mkm'k'}^\gamma$ ($k, k' \in \{\overline{1, K}\}$; $m, m', j, j' \in \{\overline{1, M}\}$).
 Опираясь на представления [11, 13, 14, 18], запишем (1) в виде системы стохастических дифференциальных уравнений:

$$d\vec{Z} = \vec{A} \left(\vec{Z}, \mathbf{A}^0, \mathbf{B}^0, \mathbf{\Gamma}^0, t \right) dt + \Sigma \left(\vec{Z}, t \right) d\vec{W}, \quad (2)$$

где $\mathbf{A}^0 = (\alpha_{mk}^0)_{M \times K}$; $\mathbf{B}^0 = (\beta_{m'm}^0)_{M \times M}$; $\mathbf{\Gamma}^0 = (\gamma_{mk}^0)_{M \times K}$.

Способы определения элементов вектора переменных \vec{Z} , вектора сноса \vec{A} , диффузионной матрицы Σ и вектора шума \vec{W} , с учетом (1) и представлений [11, 13, 14, 18], уточнены в приложении.

Уравнение (2) устанавливает правило измерения численности адептов и предадептов в структурированном социуме при наблюдаемых интенсивностях α_{mk} , $\beta_{m'm}$, γ_{mk} . Потенциально соответствующие параметры α_{mk} , $\beta_{m'm}$, γ_{mk} возможно определять различными методами при обеспечении условия некоррелированности шумов наблюдения. Так, например, интенсивности α_{mk} и γ_{mk} распространения положительной и отрицательной информации о k -м кандидате в m -й подгруппе социума допустимо задавать отношением числа сообщений соответствующей тональности (положительная и отрицательная) к периоду времени наблюдения при учете их объема и вероятностных характеристик тональной оценки [18]. При этом соответствующая вероятностная оценка отдельного сообщения может получаться при применении различных алгоритмических решений. Они, например, программно-реализованы в таких библиотеках python как Dostoevsky, TextBlob, BERT, ruGPT-3 или представлены в виде самостоятельных программ [18, 23]. Их основу составляют алгоритмы машинного обучения и словарные методы [23]. Также и величина $\beta_{m'm}$ может определяться различными не взаимоувязанными способами. Частное решение по нахождению $\beta_{m'm}$ приведено в [24].

Указанное позволяет справедливо предположить о наличии второго независимого от первого правила наблюдения α'_{mk} , $\beta'_{m'm}$, γ'_{mk} соответствующих параметров интенсивностей α_{mk} , $\beta_{m'm}$, γ_{mk} в модели (1) при неопределенности в предпочтительности первого и второго правил наблюдения. При этом аналогично α_{mk} , $\beta_{m'm}$, γ_{mk} интенсивности α'_{mk} , $\beta'_{m'm}$, γ'_{mk} складываются из истинных величин α_{mk}^0 , $\beta_{m'm}^0$, γ_{mk}^0 и шумов наблюдения $\tilde{\alpha}'_{mk}(t)$, $\tilde{\beta}'_{m'm}(t)$, $\tilde{\gamma}'_{mk}(t)$ с соответствующими статистическими параметрами: $\mathbb{E}[\tilde{\alpha}'_{mk}] = \mathbb{E}[\tilde{\beta}'_{m'm}] = \mathbb{E}[\tilde{\gamma}'_{mk}] = 0$;
 $\text{cov} [\tilde{\alpha}'_{mk}, \tilde{\alpha}'_{m'k'}] = dt \cdot \varepsilon_{mkm'k'}^{\prime\alpha}$; $\text{cov} [\tilde{\beta}'_{m'm}, \tilde{\beta}'_{j'j}] = dt \cdot \varepsilon_{m'm'j'j}^{\prime\beta}$;

$\text{cov} [\tilde{\gamma}'_{mk}, \tilde{\gamma}'_{m'k'}] = dt \cdot \varepsilon'^{\gamma}_{mkm'k'}$. В этой связи, запишем вторую систему стохастических дифференциальных уравнений наблюдения за численностью адептов:

$$d\vec{Z}' = \vec{A}(\vec{Z}', \mathbf{A}^0, \mathbf{B}^0, \mathbf{\Gamma}^0, t) dt + \Sigma'(\vec{Z}', t) d\vec{V}, \quad (3)$$

где элементы вектора переменных \vec{Z}' , вектора сноса \vec{A} , диффузионной матрицы Σ' и вектора шума \vec{V} , задаются аналогичным (2) способом (см. приложение) при учете правил определения α'_{mk} , $\beta'_{m'm}$, γ'_{mk} .

Уточним, что уравнения (2), (3) будем понимать в смысле Ито [16].

Задача фильтрации состоит в получении оценок $\hat{Z} = (\hat{Z}_l)_d$ числа адептов и предадептов k -го субъекта m -й подгруппы в момент времени $t \in [t_0, t_1]$ из заданных уравнений (2), (3). Оценку \hat{Z} выполним по критерию максимального правдоподобия:

$$\hat{Z} = \arg \max_{\vec{Z} \in \Omega} [p(\vec{Z}, t)], \quad (4)$$

где $p(\vec{Z}, t)$ – апостериорная условная функция плотности вероятности распределения $\vec{Z} \in \Omega$.

Функцию $p(\vec{Z}, t)$ при $p(\vec{Z}, t) = \phi(\vec{Z}, t) / \int_{\Omega} \phi(\vec{Z}, t) d\vec{Z}$ определим по теореме сложения вероятностей [26] из:

$$\phi(\vec{Z}, t) = p_1(\vec{Z}, t) + p_2(\vec{Z}, t) - p_1(\vec{Z}, t)p_2(\vec{Z}, t), \quad (5)$$

как вероятность суммы двух совместных событий, наступающих при составлении следующих двух пар уравнений состояния и наблюдения для формирования модели оптимальной нелинейной фильтрации: 1) уравнение (2) задается состоянием, а (3) – наблюдением; 2) уравнения состояния и наблюдения определяются в обратном порядке.

В этом случае, эволюцию апостериорных условных функций плотности распределения вероятностей $p_1(\vec{Z}, t)$, $p_2(\vec{Z}, t)$ из (5) при

$p_{1,2}(\vec{Z}, t) = \phi_{1,2}(\vec{Z}, t) / \int_{\Omega} \phi_{1,2}(\vec{Z}, t) d\vec{Z}$ и заданном начальном условии $\phi_{1,2}(\vec{Z}, t_0) = p_0(\vec{Z})$ определим из решения уравнений типа Дункана–Мортенсена–Закаи (ДМЗ) [18, 19]:

$$d\phi_{1,2}(\vec{Z}, t) = \mathcal{L}_{1,2}[\phi_{1,2}(\vec{Z}, t)] dt + \phi_{1,2}(\vec{Z}, t) [\vec{A}_{2,1}(\vec{Z}, t)]^T [\mathbf{D}^{(2,1)}(t)]^{-1} d\vec{Z}_{2,1}, \quad (6)$$

где $\vec{A}_1(\vec{Z}, t) = \vec{A}(\vec{Z}, \mathbf{A}, \mathbf{B}, \mathbf{\Gamma}, t)$; $\vec{A}_2(\vec{Z}, t) = \vec{A}(\vec{Z}, \mathbf{A}', \mathbf{B}', \mathbf{\Gamma}', t)$; $\mathbf{A} = (\alpha_{mk})_{M \times K}$; $\mathbf{B} = (\beta_{m'm})_{M \times M}$; $\mathbf{\Gamma} = (\gamma_{mk})_{M \times K}$; $\mathbf{A}' = (\alpha'_{mk})_{M \times K}$; $\mathbf{B}' = (\beta'_{m'm})_{M \times M}$; $\mathbf{\Gamma}' = (\gamma'_{mk})_{M \times K}$; $\mathbf{D}^{(1)} = (D_{ll'}^{(1)})_{d \times d} = \Sigma \Sigma^T$; $\mathbf{D}^{(2)} = (D_{ll'}^{(2)})_{d \times d} = \Sigma' \Sigma'^T$; $d\vec{Z}_{2,1}$ – приращение вектора \vec{Z} , определяемое уравнениями (3), (2) соответственно; $\mathcal{L}_{1,2}[\phi]$ – диффузионные операторы Фоккера–Планка–Колмогорова [19]:

$$\mathcal{L}_{1,2}[\phi] = - \sum_{l=1}^d \frac{\partial}{\partial Z_l} [\vec{A}_{1,2} \phi] + \frac{1}{2} \sum_{l=1}^d \sum_{l'=1}^d \frac{\partial^2}{\partial Z_l \partial Z_{l'}} [D_{ll'}^{(1,2)} \phi]. \quad (7)$$

Уточним, что в заданных в выражении (6) обозначениях для определения функции ϕ_1 используются операторы и функции с соответствующим первым индексом – \mathcal{L}_1 , \vec{A}_2 , $\mathbf{D}^{(2)}$ и $d\vec{Z}_2$. Для нахождения ϕ_2 применяются компоненты из (6) с соответствующим вторым индексом. В последующих соотношениях будем придерживаться указанного правила.

Сформированные представления (4)–(6) формализуют исходную постановку задачи оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства при выполнении условий нормировки [13, 14]: $\int_{\Omega} p(\vec{Z}, t) d\vec{Z} = 1$; $\int_{\Omega} p_{1,2}(\vec{Z}, t) d\vec{Z} = 1$; $p(\vec{Z}, t) \geq 0$; $p_{1,2}(\vec{Z}, t) \geq 0$.

Следуя результатам [25, 27], для повышения точности численного решения задачи нелинейной фильтрации при снижении чувствительности

фильтра к дискретному во времени изменению траектории наблюдения для $\rho_{1,2}(\vec{Z}, t_0) = p_0(\vec{Z})$ уравнение (6) представим в робастном виде [27]:

$$\begin{aligned} \frac{\partial \rho_{1,2}}{\partial t} = & -\frac{\partial}{\partial t} \left(\vec{A}_{2,1}^T [\mathbf{D}^{(2,1)}]^{-1} \right)^T \vec{Z}_{2,1} \rho_{1,2} + \\ & + \exp \left\{ -\vec{A}_{2,1}^T [\mathbf{D}^{(2,1)}]^{-1} \vec{Z}_{2,1} \right\} \left[\mathcal{L}_{1,2} - \frac{1}{2} \vec{A}_{2,1}^T [\mathbf{D}^{(2,1)}]^{-1} \vec{A}_{2,1} \right] \times \\ & \times \left[\exp \left\{ \vec{A}_{2,1}^T [\mathbf{D}^{(2,1)}]^{-1} \vec{Z}_{2,1} \right\} \rho_{1,2} \right], \end{aligned} \quad (8)$$

при обозначении:

$$\phi_{1,2} = \exp \left\{ -\vec{A}_{2,1}^T [\mathbf{D}^{(2,1)}]^{-1} \vec{Z}_{2,1} \right\} \rho_{1,2}. \quad (9)$$

Для заданной постановки задачи (4), (5), (8), (9), учитывая результаты [11, 13, 14, 18, 25, 27], сформируем численную схему оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства.

3. Численное решение задачи оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства. Наибольшую трудоемкость и основу в сформированной задаче фильтрации (4), (5), (8), (9) составляет решение уравнения (8). Его исследование выполним в постановке спектрального метода Галёркина [20] при дискретизации интервала наблюдения $[t_0, t_1]$ на тактовые подынтервалы $t \in [\tau_n, \tau_{n+1}]$ ($\Delta t = \tau_{n+1} - \tau_n$; $n = 0, N-1$; $N = (t_1 - t_0) / \Delta t$) [28]. Следуя утверждению 1 [25] (утверждение 2.1 [27]), для заданной дискретизации $t \in [\tau_n, \tau_{n+1}]$ в формируемом численном решении сведем задачу (8), (9) к рекуррентному решению последовательности подзадач, составляющих основу Yau-Yau's алгоритма [25]:

$$\begin{aligned} \phi_{1,2}^n(\vec{Z}, \tau_{n-1}) = & \phi_{1,2}^{n-1}(\vec{Z}, \tau_{n-1}) \times \\ & \times \exp \left\{ \left[\vec{A}_{2,1}^T(\vec{Z}, \tau_{n-1}) \right]^T [\mathbf{D}^{(2,1)}(\tau_{n-1})]^{-1} d\vec{Z}_{2,1}(\tau_{n-1}) \right\}, \end{aligned} \quad (10)$$

где при $\phi_{1,2}^1(\vec{Z}, \tau_0) = \phi_{1,2}^1(\vec{Z}, t_0) = p_0(\vec{Z})$ значения $\phi_{1,2}^{n-1}(\vec{Z}, \tau_{n-1})$ определяются из решения прямого уравнения Колмогорова [25]:

$$\begin{aligned} \frac{\partial}{\partial t} \phi_{1,2}^{n-1}(\vec{Z}, t) &= \mathcal{L}_{1,2} \left[\phi_{1,2}^{n-1}(\vec{Z}, t) \right] - \\ &- \frac{1}{2} \phi_{1,2}^{n-1}(\vec{Z}, t) \left[\vec{A}_{2,1}(\vec{Z}, t) \right]^T \left[\mathbf{D}^{(2,1)}(t) \right]^{-1} \vec{A}_{2,1}(\vec{Z}, t). \end{aligned} \quad (11)$$

Учитывая результаты [13, 14, 18], решение прямого уравнения Колмогорова (11) выполним численно при введении следующих представлений.

Зададим разбиение $\Omega = \bigcup_{u=1}^U \omega^{(u)}$ набором из U симплексов $\omega^{(u)} = \left\{ \sum_{i=1}^{d+1} \zeta_i^{(u)} P_i^{(u)} : \left(\sum_{i=1}^{d+1} \zeta_i^{(u)} = 1 \right) \wedge \left(\forall i = \overline{1, d+1}, \zeta_i^{(u)} \geq 0 \right) \right\} \subset \mathbb{R}^d$ ($u = \overline{1, U}$) с $d + 1$ вершинами $P_1^{(u)}, P_2^{(u)}, \dots, P_{d+1}^{(u)}$ и барицентрическими координатами $\zeta_1^{(u)}, \dots, \zeta_{d+1}^{(u)}$ при $\omega^{(u)} \cap \omega^{(u')} = \emptyset$ ($u \neq u'; u, u' \in \{ \overline{1, U} \}$).

Обозначим $\langle \cdot, \cdot \rangle_\Omega$ скалярное произведение:

$$\langle \eta, \phi \rangle_\Omega = \int_\Omega \eta(\vec{Z}) \phi(\vec{Z}) d\vec{Z}, \quad (12)$$

для некоторых функций η и ϕ .

Зададим аппроксимацию $\tilde{\phi}_{1,2}^n(\vec{Z}, t)$ ненормированных функций $\phi_{1,2}^n(\vec{Z}, t)$ плотности вероятности:

$$\tilde{\phi}_{1,2}^n(\vec{Z}, t) = \sum_{u=1}^U \sum_{j \in \mathbb{J}_r^d} c_{uj}^{(1,2)}(t) \psi_j^{(u)}(\vec{Z}), \quad (13)$$

подстановка которых в (11) в проекционном представлении метода Галёркина сведет исходную к системе обыкновенных дифференциальных уравнений:

$$\frac{d}{dt} \vec{C}^{(1,2)}(t) = \mathbf{S}^{-1} \left(\mathbf{L}^{(1,2)}(t) - \mathbf{Q}^{(1,2)}(t) \right) \vec{C}^{(1,2)}(t); \quad (14)$$

$$\vec{C}^{(1,2)}(0) = \mathbf{S}^{-1} \vec{G}.$$

В выражениях (13), (14) приняты следующие обозначения: $\vec{G} = \text{vec}(\mathbf{G})$ при $\mathbf{G} = \left(\left\langle p_0, \psi_j^{(u)} \right\rangle_{\Omega} \right)_{U \times |\mathbb{J}_r^d|}$; $\text{vec}(\cdot)$ – операция векторизации матрицы; $\vec{C}^{(1,2)} = \text{vec}(\mathbf{C}^{(1,2)})$ – вектор искомых коэффициентов разложения, зависящих от t , при $\mathbf{C}^{(1,2)} = \left(c_{uj}^{(1,2)} \right)_{U \times |\mathbb{J}_r^d|}$; $\mathbf{L}^{(1,2)} = \left(\left\langle \psi_j^{(u)}, \mathcal{L}_{1,2} \left[\psi_{j'}^{(u')} \right] \right\rangle_{\Omega} \right)_{U_r \times U_r}$; $\mathbf{Q}^{(1,2)} = \left(\left\langle \psi_j^{(u)}, \mathbf{A}_{2,1}^T \left[\mathbf{D}^{(2,1)}(t) \right]^{-1} \mathbf{A}_{2,1} \psi_{j'}^{(u')} \right\rangle_{\Omega} \right)_{U_r \times U_r}$; $\mathbf{S} = \left(\left\langle \psi_j^{(u)}, \psi_{j'}^{(u')} \right\rangle_{\Omega} \right)_{U_r \times U_r}$; $U_r = U | \mathbb{J}_r^d |$; $u, u' \in \{\overline{1, U}\}$; \mathbb{J}_r^d – множество мультииндексов $j, j' \in \mathbb{J}_r^d$ [13, 14]:

$$\mathbb{J}_r^d = \left\{ j = (j_1, \dots, j_i, \dots, j_{d+1}) : j_i \in \mathbb{Z}_+, \sum_{i=1}^{d+1} j_i = r \right\}, \quad (15)$$

где $r \in \mathbb{N}$ – порядок аппроксимации на $\omega^{(u)}$; $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$; $\psi_j^{(u)}$ – базисная функция частичной подобласти $\omega^{(u)} \in \Omega$, которую зададим произведением:

$$\psi_j^{(u)} = \prod_{i=1}^{d+1} \varphi_{j_i}, \quad (16)$$

Гауссовых базисных функций [29]:

$$\varphi_{j_i} = \exp \left\{ - \frac{[2j_i + 1 - 2\zeta_i(r + 1)]^2}{2(d + 1)} \right\}. \quad (17)$$

Принимая во внимание заданную дискретизацию временного интервала анализа и выбранную рекуррентную схему (10) оценки

$\phi_{1,2}$, решение обыкновенного дифференциального уравнения (14) для соблюдения компромисса между точностью, вычислительной устойчивостью и сложностью выполним методом Эйлера [30]:

$$\vec{C}_n^{(1,2)} = \vec{C}_{n-1}^{(1,2)} + \Delta t \left\{ \mathbf{S}^{-1} \left[\left(\mathbf{L}_{n-1}^{(1,2)} - \mathbf{Q}_{n-1}^{(1,2)} \right) \vec{C}_{n-1}^{(1,2)} \right] \right\}, \quad (18)$$

где $\vec{C}_n^{(1,2)} \equiv \vec{C}^{(1,2)}(\tau_n)$; $\mathbf{L}_n^{(1,2)} \equiv \mathbf{L}^{(1,2)}(\tau_n)$; $\mathbf{Q}_n^{(1,2)} \equiv \mathbf{Q}^{(1,2)}(\tau_n)$.

Вычисленные выражением (17) коэффициенты $\vec{C}_n^{(1,2)}$, с учетом (10) и аппроксимации (13), нормируются по правилу:

$$\vec{C}_n^{(1,2)} = \vec{C}_n^{(1,2)} / \left(\vec{C}_n^{(1,2)} \vec{\Psi} \right), \quad (19)$$

а затем уточняются соотношением:

$$\vec{C}_n^{(1,2)} = \mathbf{S}^{-1} \left[\mathbf{Q}_n^{(1,2)} \vec{C}_n^{(1,2)} \right], \quad (20)$$

где $\mathbf{Q}'^{(1,2)} = \left(\left\langle \psi_j^{(u)}, \exp \left\{ \vec{A}_{2,1}^\Gamma [\mathbf{D}^{(2,1)}]^{-1} d\vec{Z}_{2,1} \right\} \psi_{j'}^{(u')} \right\rangle_\Omega \right)_{U_r \times U_r}$ при $\mathbf{Q}_n^{(1,2)} \equiv \mathbf{Q}'^{(1,2)}(\tau_n)$ и рассчитываемых элементов $\vec{A}_{2,1}$, $\mathbf{D}^{(2,1)}$, $d\vec{Z}_{2,1}$ в момент времени τ_n ; $\vec{\Psi} = \text{vec}(\tilde{\Psi})$ при $\tilde{\Psi} = \left(\tilde{\Psi}_{uj} \right)_{U \times |\mathbb{J}_r^d|}$ для $\tilde{\Psi}_{uj} = \int_{\omega^{(u)}} \psi_j^{(u)}(\vec{Z}) d\vec{Z}$.

Определение искомой функции плотности вероятностей $p(\vec{Z}, t)$ в момент времени τ_n осуществляется при ее аналогичной (13) аппроксимации $\tilde{p}(\vec{Z}, t) = \sum_{u=1}^U \sum_{j \in \mathbb{J}_r^d} c_{uj}(t) \psi_j^{(u)}(\vec{Z})$, где коэффициенты c_{uj} для $\vec{C} = \text{vec}(\mathbf{C})$ и $\mathbf{C} = (c_{uj})_{U \times |\mathbb{J}_r^d|}$ задаются при дополнительной нормировки (19) полученных $\vec{C}_n^{(1,2)}$ из (20) по правилу:

$$\vec{C}_n = \vec{C}_n^{(1)} + \vec{C}_n^{(2)} - \mathbf{S}^{-1} \vec{C}_n; \quad \vec{C}_n = \vec{C}_n / \left(\vec{C}_n \vec{\Psi} \right), \quad (21)$$

где $\vec{C}_n \equiv \vec{C}(\tau_n)$; $\vec{C}_n \equiv \vec{C}(\vec{C}_n, \tau_n)$ для $\vec{C}_n = \text{vec}(\tilde{C}_n)$ при $\tilde{C}_n = \left(\left\langle \tilde{\phi}_1^n, \tilde{\phi}_2^n, \psi_j^{(u)} \right\rangle_{\Omega} \right)_{U \times |\mathbb{J}_r^d|}$.

Поиск максимума (4) осуществляется методом Ньютона [31] для начального приближения $\hat{\vec{Z}}^0$, выбираемого с учетом следующего выражения: $\hat{\vec{Z}}^0 = \sum_{u=1}^U \sum_{j \in \mathbb{J}_r^d} c_{uj} \Psi_j^{(u)}$, где $\Psi_j^{(u)} = \int_{\omega^{(u)}} \vec{Z} \psi_j^{(u)}(\vec{Z}) d\vec{Z}$.

В целом, перечисленные правила вычисления эволюции функции плотности вероятности $p(\vec{Z}, \tau_n)$ и оценки $\hat{\vec{Z}}$ составляют суть численного решения задачи оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства. Исследование вопросов существования, единственности и сходимости сформированного численного решения при неизменности операторных представлений в (11) и выбранной системы базисных функций согласуется с результатами [13, 14, 18]. Задание элементов векторов и матриц в (18), (20) обеспечивается численным вычислением интегралов вида (12) с применением кубатурных формул [14] для симплексов [32]. Также для минимизации вычислительных затрат при расчете матриц \mathbf{S} , $\mathbf{Q}_n^{(1,2)}$, \tilde{C}_n из (18), (20), (21) соответственно учитывается их симметричность, а нахождение элементов матриц \mathbf{S} , \tilde{C}_n реализуется аналитически через функции ошибок (используются результаты, полученные при доказательстве лемм из [13]). Последнее определяет дополнительную предпочтительность применения в качестве аппроксимирующих полиномов Гауссовых базисных функций (16) в сравнении с другими видами специальных функций [33]. Алгоритм разбиения $\Omega = \bigcup_{u=1}^U \omega^{(u)}$ набором из U симплексов при формировании триангуляции [34] уточнен в [14]. Эффективный алгоритм формирования множества мультииндексов (15) приведен в [35].

Для систематизированного представления последовательности действий получения $\hat{\vec{Z}}$ по составленному численному решению, приведем алгоритм оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства. При формировании алгоритма используем результаты [13, 14, 18, 36].

4. Алгоритм оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства. Входными данными

алгоритма являются: 1) временной интервал анализа $[t_0, t_1]$; 2) число отсчетов дискретизации N ; 3) число K субъектов, в отношении которых формируются предпочтения у индивидов; 4) количество M подгрупп в социуме; 5) численность индивидов N_m в m -х подгруппах социума; 6) порядок аппроксимации r ; 7) разбиение $\Omega = \bigcup_{u=1}^U \omega^{(u)}$;

8) наблюдаемые матричные функции интенсивностей $\mathbf{A}, \mathbf{B}, \mathbf{\Gamma}$ и $\mathbf{A}', \mathbf{B}', \mathbf{\Gamma}'$; 9) начальное распределение функции плотности вероятности $p(\vec{Z}, t_0) = p_0(\vec{Z}) = \frac{1}{\sigma_0} \exp\left[-\frac{1}{2}(\vec{Z}^T \Sigma_0^{-1} \vec{Z})\right]$, где σ_0 – нормирующий множитель, обеспечивающий выполнение условия $\int_{\Omega} p_0(\vec{Z}) d\vec{Z} = 1$;

$$\Sigma_0 = (\Sigma_{ll'})_{d \times d}$$

Выделим основные этапы работы алгоритма, полагая неизвестными параметры $\varepsilon_{mkm'k'}^{\alpha}, \varepsilon_{m'm'j'j}^{\beta}, \varepsilon_{mkm'k'}^{\gamma}, \varepsilon_{mkm'k'}^{\prime\alpha}, \varepsilon_{m'm'j'j}^{\prime\beta}, \varepsilon_{mkm'k'}^{\prime\gamma}$.

Шаг 1. Методом полиспектрального анализа [18, 21] в отношении функций $\alpha_{mk}, \beta_{m'm}, \gamma_{mk}, \alpha'_{mk}, \beta'_{m'm}, \gamma'_{mk}$ выполнить оценку соответствующих величин интенсивностей $\hat{\alpha}_{mk}^0, \hat{\beta}_{m'm}^0, \hat{\gamma}_{mk}^0, \hat{\alpha}'_{mk}, \hat{\beta}'_{m'm}, \hat{\gamma}'_{mk}$ и усреднением $\alpha_{mk}^0 = 0,5(\hat{\alpha}_{mk}^0 + \hat{\alpha}'_{mk}^0)$, $\beta_{m'm}^0 = 0,5(\hat{\beta}_{m'm}^0 + \hat{\beta}'_{m'm}^0)$, $\gamma_{mk}^0 = 0,5(\hat{\gamma}_{mk}^0 + \hat{\gamma}'_{mk}^0)$ определить приближения для соответствующих истинных значений интенсивностей.

Шаг 2. По средневыборочным правилам [35] для $\alpha_{mk}, \beta_{m'm}, \gamma_{mk}, \alpha'_{mk}, \beta'_{m'm}, \gamma'_{mk}$ и найденных $\alpha_{mk}^0, \beta_{m'm}^0, \gamma_{mk}^0$ определить оценку соответствующих параметров $\varepsilon_{mkm'k'}^{\alpha}, \varepsilon_{m'm'j'j}^{\beta}, \varepsilon_{mkm'k'}^{\gamma}, \varepsilon_{mkm'k'}^{\prime\alpha}, \varepsilon_{m'm'j'j}^{\prime\beta}, \varepsilon_{mkm'k'}^{\prime\gamma}$ для задания ковариационных матриц шума $\mathbf{D}^{(1,2)}$.

Шаг 3. Положить $n = 0$ и установить начальные значения $\vec{Z}_n = 0, \vec{Z}_n^{(1,2)} = 0, \hat{Z}_n = 0$.

Шаг 4. При заданной $p_0(\vec{Z})$ определить $\vec{C}_n = \vec{C}_n^{(1,2)} = \mathbf{S}^{-1} \vec{G}$.

Шаг 5. Для момента времени τ_n вычислить ковариационные матрицы $\mathbf{D}_n^{(1,2)}$.

Шаг 6. По методу Эйлера (Эйлера–Маруямы [37]) выполнить экстраполяцию $\vec{Z}_{n+1}^{(1,2)}$ с учетом моделей (2), (3):

$$\vec{Z}_{n+1}^{(1)} = \vec{Z}_n^{(1)} + \Delta t \vec{A}(\vec{Z}_n^{(1)}, \mathbf{A}, \mathbf{B}, \mathbf{\Gamma}, \tau_n); \tag{22}$$

$$\vec{Z}_{n+1}^{(2)} = \vec{Z}_n^{(2)} + \Delta t \vec{A}(\vec{Z}_n^{(2)}, \mathbf{A}', \mathbf{B}', \mathbf{\Gamma}', \tau_n).$$

Шаг 7. Учитывая представления (14), (18), для τ_n рассчитать матрицы $\mathbf{L}_n^{(1,2)}$, $\mathbf{Q}_n^{(1,2)}$.

Шаг 8. Если $n > 0$, то:

а) принимая во внимание (20), рассчитать:

$$\vec{C}_n^{(1,2)} = \mathbf{S}^{-1} \left[\mathbf{Q}_n'^{(1,2)} \vec{C}_n^{(1,2)} \right],$$

где $\mathbf{Q}_n'^{(1,2)} \equiv \mathbf{Q}'^{(1,2)}(\tau_n)$,

при $\mathbf{Q}'^{(1,2)} = \left(\left\langle \psi_j^{(u)}, \exp \left\{ \bar{A}_{2,1}^T \bar{G}_n^{(2,1)} \right\} \psi_{j'}^{(u')} \right\rangle_{\Omega} \right)_{U_r \times U_r}$ для $\bar{G}_n^{(1,2)} = \left[\mathbf{D}_n^{(1,2)} \right]^{-1} \left(\bar{Z}_n^{(1,2)} - \bar{Z}_{n-1}^{(1,2)} \right)$;

б) нормировать полученные векторы $\vec{C}_n^{(1,2)}$ по правилу (19).

Шаг 9. С учетом выражения (18), вычислить $\vec{C}_{n+1}^{(1,2)}$ с последующей нормировкой результата соотношением (19).

Шаг 10. С учетом выражения (21), рассчитать \vec{C}_{n+1} .

Шаг 11. Для найденного \vec{C}_{n+1} по правилу (4) методом Ньютона [31] определить оценку \hat{Z}_{n+1} .

Шаг 12. Увеличить $n = n + 1$ и проверить условия:

а) если $n < N$, то перейти к шагу 5;

б) если $n = N$, то завершить работу алгоритма и вывести результат вычисления – $\hat{Z}_0, \hat{Z}_1, \dots, \hat{Z}_{N-1}$.

Псевдокод сформированного алгоритма оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства, с учетом указанной общей последовательности действий, приведен в виде листинга 1 – функция FILTERRATINGS.

Вспомогательные функции в приведенном листенге 1 реализуют следующие вычислительные процедуры:

1) POLYSPECT ESTIM – определение приближения для соответствующих истинных значений интенсивностей $\alpha_{mk}^0, \beta_{m'm}^0, \gamma_{mk}^0$ (шаг 1 алгоритма);

function FILTERRATINGS($t_0, t_1, N, K, M, N_1, \dots, N_M, \Omega, p_0, \mathbf{A}, \mathbf{B}, \mathbf{\Gamma}, \mathbf{A}', \mathbf{B}', \mathbf{\Gamma}'$) {

$\langle \mathbf{A}^0, \mathbf{B}^0, \mathbf{\Gamma}^0 \rangle = \text{POLYSPECTESTIM}(\mathbf{A}, \mathbf{B}, \mathbf{\Gamma}, \mathbf{A}', \mathbf{B}', \mathbf{\Gamma}')$;

$\langle \epsilon^\alpha, \epsilon^\beta, \epsilon^\gamma, \epsilon'^\alpha, \epsilon'^\beta, \epsilon'^\gamma \rangle =$

CALCVARIANCES($\mathbf{A}^0, \mathbf{B}^0, \mathbf{\Gamma}^0, \mathbf{A}, \mathbf{B}, \mathbf{\Gamma}, \mathbf{A}', \mathbf{B}', \mathbf{\Gamma}'$); $n = 0$;

$\bar{Z}_n = 0$; $\bar{Z}_n^{(1,2)} = 0$; $\hat{Z}_n = 0$;

```

 $\Delta t = (t_1 - t_0) / (N - 1); \tau_n = t_0;$ 
 $\mathbf{S} = \left( \left\langle \psi_j^{(u)}, \psi_j^{(u')} \right\rangle_{\Omega} \right)_{U_r \times U_r}; \bar{\mathbf{G}} = \left( \left\langle p_0, \psi_j^{(u)} \right\rangle_{\Omega} \right)_{U_r};$ 
 $\bar{\mathbf{C}}_n = \text{CALCSLAE}(\mathbf{S}, \bar{\mathbf{G}}); \bar{\mathbf{C}}_n = \text{NORMVECTOR}(\bar{\mathbf{C}}_n); \bar{\mathbf{C}}_n^{(1,2)} = \bar{\mathbf{C}}_n;$ 
while ( $n < N$ ) {
     $\mathbf{D}_n^{(1)} = \text{DIFFUSIONESTIM}(\epsilon^\alpha, \epsilon^\beta, \epsilon^\gamma, \bar{\mathbf{Z}}_n^{(1)});$ 
     $\mathbf{D}_n^{(2)} = \text{DIFFUSIONESTIM}(\epsilon'^\alpha, \epsilon'^\beta, \epsilon'^\gamma, \bar{\mathbf{Z}}_n^{(2)});$ 
     $\bar{\mathbf{Z}}_{n+1}^{(1)} = \bar{\mathbf{Z}}_n^{(1)} + \Delta t \bar{\mathbf{A}}(\bar{\mathbf{Z}}_n^{(1)}, \mathbf{A}, \mathbf{B}, \mathbf{\Gamma}, \tau_n);$ 
     $\bar{\mathbf{Z}}_{n+1}^{(2)} = \bar{\mathbf{Z}}_n^{(2)} + \Delta t \bar{\mathbf{A}}(\bar{\mathbf{Z}}_n^{(2)}, \mathbf{A}', \mathbf{B}', \mathbf{\Gamma}', \tau_n);$ 
     $\mathbf{L}_n^{(1,2)} = \mathbf{L}^{(1,2)}(\bar{\mathbf{A}}_{1,2}, \mathbf{D}_n^{(1,2)}, \tau_n);$ 
     $\mathbf{Q}_n^{(1,2)} = \mathbf{Q}^{(1,2)}(\bar{\mathbf{A}}_{2,1}, \mathbf{D}_n^{(2,1)}, \tau_n);$ 
if ( $n > 0$ ) {
         $\bar{\mathbf{G}}_n'^{(1,2)} = [\mathbf{D}_n^{(1,2)}]^{-1} (\bar{\mathbf{Z}}_n^{(1,2)} - \bar{\mathbf{Z}}_{n-1}^{(1,2)});$ 
         $\mathbf{Q}_n'^{(1,2)} = \mathbf{Q}'^{(1,2)}(\bar{\mathbf{A}}_{2,1}, \bar{\mathbf{G}}_n'^{(1,2)}, \tau_n);$ 
         $\bar{\mathbf{C}}_n'^{(1,2)} = \mathbf{Q}_n'^{(1,2)} \bar{\mathbf{C}}_n^{(1,2)};$ 
         $\bar{\mathbf{C}}_n^{(1,2)} = \text{CALCSLAE}(\mathbf{S}, \bar{\mathbf{C}}_n'^{(1,2)});$ 
         $\bar{\mathbf{C}}_n^{(1,2)} = \text{NORMVECTOR}(\bar{\mathbf{C}}_n^{(1,2)});$ 
         $d\bar{\mathbf{C}}_n^{(1,2)} = (\mathbf{L}_n^{(1,2)} - \mathbf{Q}_n^{(1,2)}) \bar{\mathbf{C}}_n^{(1,2)};$ 
         $d\bar{\mathbf{C}}_n^{(1,2)} = \text{CALCSLAE}(\mathbf{S}, d\bar{\mathbf{C}}_n^{(1,2)});$ 
         $\bar{\mathbf{C}}_{n+1}^{(1,2)} = \bar{\mathbf{C}}_n^{(1,2)} + \Delta t d\bar{\mathbf{C}}_n^{(1,2)};$ 
         $\bar{\mathbf{C}}_{n+1}^{(1,2)} = \text{NORMVECTOR}(\bar{\mathbf{C}}_{n+1}^{(1,2)});$ 
         $\tau_{n+1} = \tau_n + \Delta t; \bar{\mathbf{C}}_{n+1} = \bar{\mathbf{C}}(\bar{\mathbf{C}}_{n+1}^{(1,2)}, \tau_{n+1});$ 
         $\bar{\mathbf{C}}_{n+1} = \text{CALCSLAE}(\mathbf{S}, \bar{\mathbf{C}}_{n+1});$ 
         $\bar{\mathbf{C}}_{n+1} = \text{NORMVECTOR}(\bar{\mathbf{C}}_{n+1}^{(1)} + \bar{\mathbf{C}}_{n+1}^{(2)} - \bar{\mathbf{C}}_{n+1});$ 
         $\hat{\mathbf{Z}}_{n+1} = \text{ARGUMENTMAXIMUM}(\bar{\mathbf{C}}_{n+1}); n = n + 1;$ 
    }
return  $\{\hat{\mathbf{Z}}_0, \dots, \hat{\mathbf{Z}}_{N-1}\}$ 

```

Листинг 1. Оптимальная нелинейная фильтрация оценок информационного воздействия в стохастической модели информационного противоборства

- 2) CALC VARIANCES – средневывборочная оценка параметров $\epsilon_{mkm'k'}^\alpha, \epsilon_{m'mj'j'}^\beta, \epsilon_{mkm'k'}^\gamma, \epsilon_{mkm'k'}^{\prime\alpha}, \epsilon_{m'mj'j'}^{\prime\beta}, \epsilon_{mkm'k'}^{\prime\gamma}$ (шаг 2 алгоритма);
- 3) NORMVECTOR – нормировка (19);
- 4) DIFFUSIONESTIM – вычисление ковариационных матриц $\mathbf{D}_n^{(1,2)}$ (шаг 5 алгоритма), с учетом правил задания матриц $\mathbf{\Sigma}, \mathbf{\Sigma}'$, уточненных в приложении;

5) ARGUMENT MAXIMUM – максимально правдоподобная оценка \hat{Z}_{n+1} в соответствии с правилом (4) при численном нахождении оптимума методом Ньютона [31] (шаг 11 алгоритма);

6) CALC SLAE – решение системы линейных алгебраических уравнений [38].

5. Результаты вычислительного эксперимента. Для апостериорного исследования предпочтительности полученных результатов проведем серию вычислительных экспериментов при решении задачи моделирования и оценивания стохастической модели информационного противоборства вида (2), (3) для $K = 2, M = 1$ (малые значения K, M выбраны из соображений наглядности демонстрации графических результатов вычислительных экспериментов). При проведении указанных испытаний особое внимание уделим результативности сформированного алгоритма (листинг 1) при вторичном уточнении его устойчивости и вычислительной сложности. Выделенные свойства алгоритма при изменяемом уровне нестационарного негауссовского шума в наблюдаемых интенсивностях моделей (2), (3) апостериорно установим при сравнении данных моделирования точного решения $\bar{X}^0 = (x_k^0)_K, \bar{Y}^0 = (y_k^0)_K$ с оцениваемыми: 1) \tilde{Z} по средневыборочным правилам $\tilde{Z} = (\bar{Z} + \bar{Z}') / 2$ результатов наблюдений \bar{Z}, \bar{Z}' из моделей (2), (3) соответственно; 2) \tilde{Z}' из моделей (2), (3) ансамблевым расширенным фильтром Калмана [39]; 3) \hat{Z}' алгоритмом фильтрации, основу в котором составляет численное решение уравнения ДМЗ (6).

В качестве истинной модели информационного противоборства при проведении вычислительных экспериментов произвольно выбрана следующая система обыкновенных дифференциальных уравнений:

$$\frac{dy_k^0}{dt} = \gamma_k^0 (x_k^0 - y_k^0) + (\alpha_k^0 + \beta^0 x_k^0) \times \times \left[N_0 - \sum_{k'=1}^2 (x_{k'}^0 + y_{k'}^0) - y_k^0 \right]; \quad (23)$$

$$\frac{dx_k^0}{dt} = y_k^0 [\alpha_k^0 + \beta^0 x_k^0] - \gamma_k^0 x_k^0; \quad x_k^0 = 0; \quad y_k^0 = 0; \quad k = [1; 2],$$

где $\alpha_1^0(t) = 0,25 |1 + 2 \cos(15t)|$; $\alpha_2^0(t) = 0,25 |1 + 2 \sin(5t)|$; $\beta^0(t) = 0,005 |2 + 5 \cos(0,1t)|$; $\gamma_1^0(t) = 0,05 |\sin(20t)|$; $\gamma_2^0(t) = 0,1 |\cos(2t)|$.

Решение (23) при определении эволюции $\vec{X}^0 = (x_k^0)_K$, $\vec{Y}^0 = (y_k^0)_K$ осуществляется численно методом Эйлера для аналогичных (22) N , Δt .

При обеспечении условия формирования в компонентах $\alpha_k, \beta, \gamma_k$ и $\alpha'_k, \beta', \gamma'_k$ нестационарного негауссовского шума для моделей (2) и (3) соответствующие наблюдаемые интенсивности заданы следующими правилами:

1) $\alpha_k(t) = |\alpha_k^0(t) - \mathcal{N}(0, q_1)|$, $\beta(t) = |\beta^0(t) - \mathcal{N}(0, q_1)|$, $\gamma_k(t) = |\gamma_k^0(t) - \mathcal{N}(0, q_1)|$, где $\mathcal{N}(0, q_1)$ – функция генерации случайных чисел, подчиняющихся нормальному закону распределения со средним 0 и дисперсией q_1 ;

2) $\alpha'_k(t) = |\alpha_k^0(t) + \mathcal{R}(q_2^{-1}/2)|$, $\beta'(t) = |\beta^0(t) + \mathcal{R}(q_2^{-1}/2)|$, $\gamma'_k(t) = |\gamma_k^0(t) + \mathcal{R}(q_2^{-1}/2)|$, где $\mathcal{R}(q_2^{-1}/2)$ – функция генерации случайных чисел, подчиняющихся экспоненциальному закону распределения со значением показателя экспоненты равным $(2q_2)^{-1}$.

Другие входные данные алгоритма заданы следующими величинами: $t_0 = 0$; $t_1 = 4$; $N = 480$; $N_0 = 10$; $r = 33$; $q_1 = 0, 5$; $q_2 = 0, 25$.

Для компактности представлений результаты вычислительных экспериментов приведем в отношении значения суммы адептов и предадептов k -го кандидата $\xi_k(t) = x_k(t) + y_k(t)$.

На рисунке 1 представлены графики зависимости от времени элементов векторов численности адептов: 1) истинного ξ_k^0 ; 2) определяемого среднего $\tilde{\xi}_k$ из моделей (2), (3); 3) оцениваемого $\tilde{\xi}'_k$ ансамблевым расширенным фильтром Калмана [39]; 4) оцениваемого $\hat{\xi}'_k$ по алгоритму фильтрации, предполагающему численное исследование уравнения ДМЗ (6); 5) оцениваемого $\hat{\xi}_k$ по разработанному алгоритму, основанному на численном исследовании робастного уравнения ДМЗ (8).

На рисунке 2 отражены графики наблюдаемых и истинных значений интенсивностей $\alpha_k, \beta, \gamma_k, \alpha'_k, \beta', \gamma'_k$ и $\alpha_k^0, \beta^0, \gamma_k^0$ для $t \in [t_0, t_1]$, используемых при оценивании $\xi_k^0, \tilde{\xi}_k, \tilde{\xi}'_k, \hat{\xi}'_k, \hat{\xi}_k$ (рисунок 1).

На рисунке 3 приведены графики одномерного среза апостериорных условных функций плотности вероятности $p(\vec{Z}, t), p_1(\vec{Z}, t), p_2(\vec{Z}, t)$ в различные моменты времени $t \in [t_0, t_1]$. Ориентация плоскости среза выбиралась по максимальному значению дисперсии относительно оцениваемого по правилу (4) значению \vec{Z} .

Количественно сравнительную предпочтительность получаемых результатов $\tilde{\xi}_k, \tilde{\xi}'_k, \hat{\xi}'_k, \hat{\xi}_k$ по отношению к ξ_k^0 установим величиной накопленной средней квадратической ошибки,

значение которой, например, для $\tilde{\xi}_k$ определено по правилу

$$\Delta_{\tilde{\xi}} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=1}^K \left| \xi_k^0(t_n) - \tilde{\xi}_k(t_n) \right|^2} \quad [40].$$

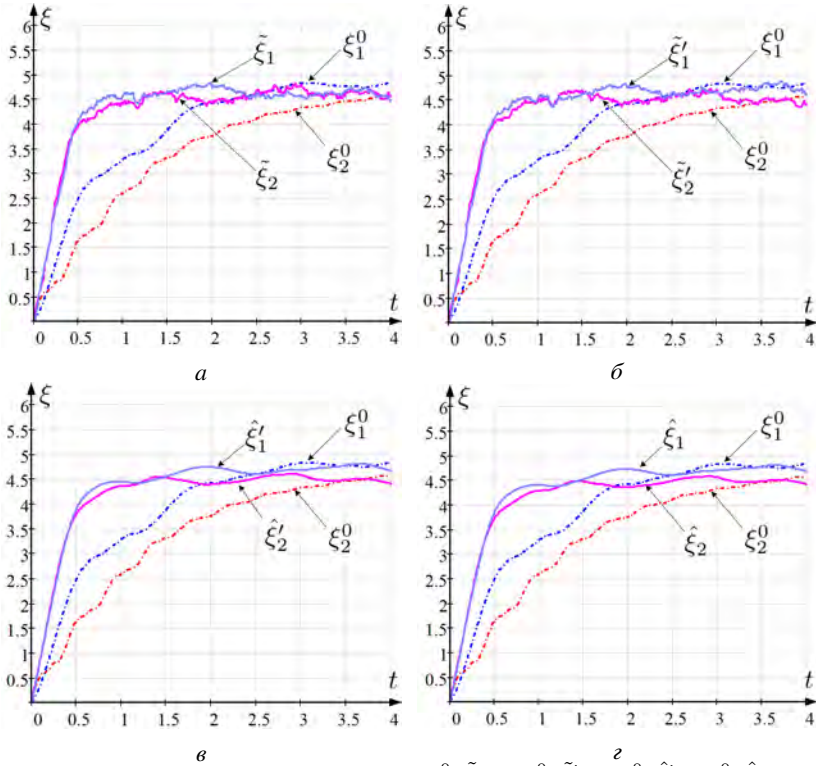


Рис. 1. Графики зависимости от t : а) $\xi_k^0, \tilde{\xi}_k$; б) $\xi_k^0, \tilde{\xi}'_k$; в) $\xi_k^0, \hat{\xi}'_k$; г) $\xi_k^0, \hat{\xi}_k$

Также сопоставление результатов выполним при определении отклонения расчетного от истинного значения для заключительного момента времени наблюдения:

$$\delta_{\tilde{\xi}} = \sum_{k=1}^K \left| \xi_k^0(t_{N-1}) - \tilde{\xi}_k(t_{N-1}) \right| \quad [40].$$

Для проведенного расчета (рисунок 1) при $q_1 = 0,5, q_2 = 0,25$ значения Δ и δ составили: $\Delta_{\tilde{\xi}} = 1,11; \Delta_{\tilde{\xi}'} = 1,076; \Delta_{\hat{\xi}'} = 0,988; \Delta_{\hat{\xi}} = 0,945; \delta_{\tilde{\xi}} = 0,366; \delta_{\tilde{\xi}'} = 0,275; \delta_{\hat{\xi}'} = 0,231; \delta_{\hat{\xi}} = 0,208.$

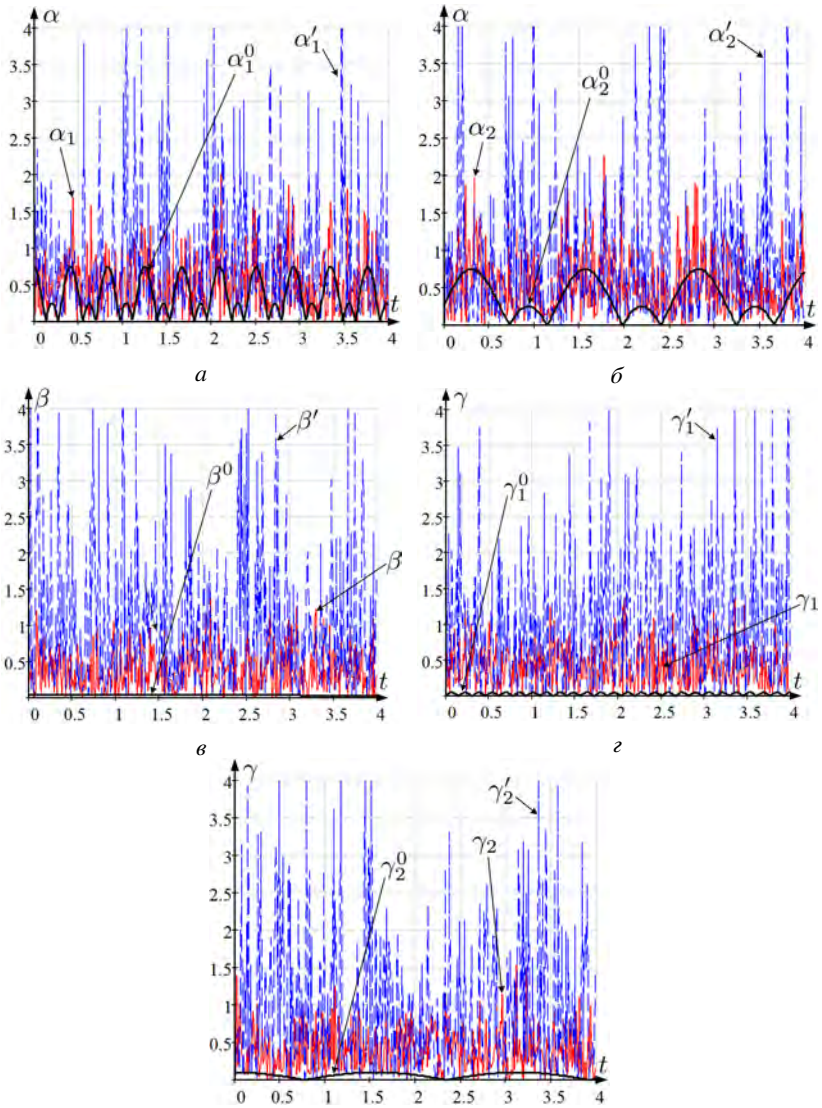


Рис. 2. Графики зависимости от t а) $\alpha_1, \alpha_1', \alpha_1^0$; б) $\alpha_2, \alpha_2', \alpha_2^0$; в) β, β', β^0 ; г) $\gamma_1, \gamma_1', \gamma_1^0$; д) $\gamma_2, \gamma_2', \gamma_2^0$

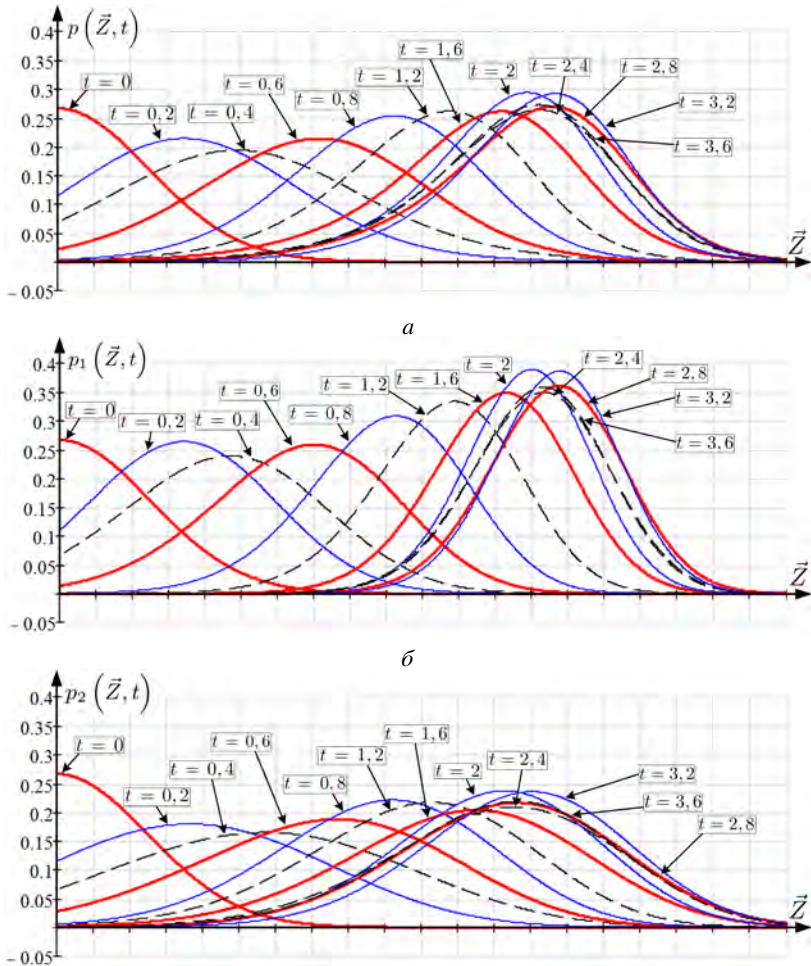
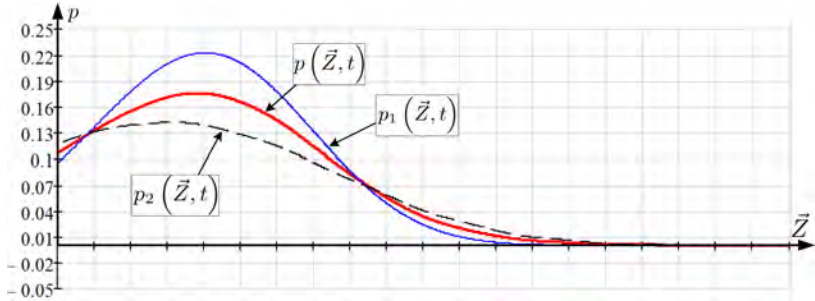


Рис. 3. Графики срезов для разных t : а) $p(\bar{Z}, t)$; б) $p_1(\bar{Z}, t)$; в) $p_2(\bar{Z}, t)$

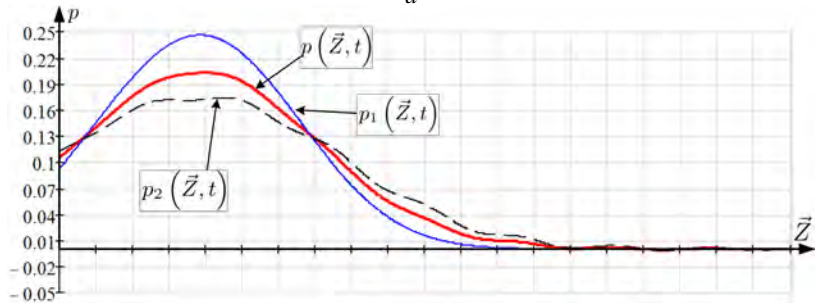
Полученные частные результаты (рисунок 1) свидетельствуют о том, что сформированный алгоритм обеспечивает более точную оценку в сравнении с известными решениями [11, 13, 14, 39]. При этом интерес вызывает исследование соотношенного выигрыша при изменении q_1, q_2 .

Очевидно, что при росте ошибок в наблюдениях $\alpha_k, \beta, \gamma_k, \alpha'_k, \beta', \gamma'_k$ вычислительная устойчивость составленного численного решения будет

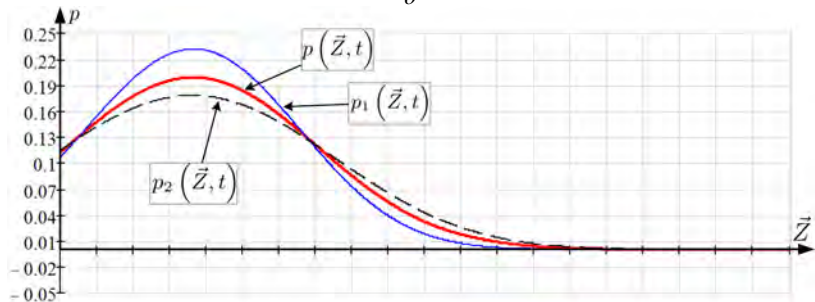
снижаться. Для ее повышения при сохранении адекватности нахождения функций $p(\vec{Z}, t)$, $p_1(\vec{Z}, t)$, $p_2(\vec{Z}, t)$ требуется увеличить число точек дискретизации N на выбранном интервале наблюдения $[t_0, t_1]$ (рисунок 4).



a



б



в

Рис. 4. Графики срезов p , p_1 , p_2 для $t = 0, 2$, $q_1 = 0, 75$; $q_2 = 0, 5q_1$, полученных при численном исследовании: а) уравнения ДМЗ при $N = 480$; б) робастного уравнения ДМЗ при $N = 480$; в) робастного уравнения ДМЗ при $N = 500$

Графическая иллюстрация отмеченной особенности приведена на рисунке 4 при представлении одномерного среза апостериорных условных функций плотности вероятности $p(\vec{Z}, t)$, $p_1(\vec{Z}, t)$, $p_2(\vec{Z}, t)$, рассчитанных при решении задачи фильтрации при применении уравнения ДМЗ (6) и робастного уравнения ДМЗ (8) (сформированный алгоритм) для $t = 0, 2$; $q_1 = 0, 75$; $q_2 = 0, 5q_1$; $N = 480$ и $N = 500$. Из представленного примера (рисунок 4) следует, что на участках быстрого изменения числа адептов и предадептов ($t \in [0; 0, 4]$, рисунок 1) при увеличении q_1 , q_2 для недостаточного числа N при численном нахождении $p(\vec{Z}, t)$, $p_1(\vec{Z}, t)$, $p_2(\vec{Z}, t)$ наблюдается неравномерная (рисунок 4, б) сходимость выбранной аппроксимации вида (13). При этом в отношении применения робастного уравнения ДМЗ (8) в сравнении с простым уравнением ДМЗ (6) эта неравномерность проявляется более выражено. Принимая во внимание указанное, выполнено сравнительное исследование точности сформированного решения при изменении q_1 , q_2 .

На рисунке 5 представлены графики зависимости усредненных по 10^6 вычислительным экспериментам $\Delta_{\tilde{\xi}}$, $\Delta_{\tilde{\xi}'}$, $\Delta_{\tilde{\xi}''}$, $\Delta_{\tilde{\xi}}$, $\delta_{\tilde{\xi}}$, $\delta_{\tilde{\xi}'}$, $\delta_{\tilde{\xi}''}$, $\delta_{\tilde{\xi}}$ от q_1 , q_2 при $q_1 = 0, 5q_2$ (указанная зависимость выбрана из соображений соизмеримости уровня шума в соответствующих наблюдаемых интенсивностях моделей (2) и (3)).

6. Заключение. Результаты проведенных вычислительных экспериментов (рисунки 1–5) подтверждают работоспособность разработанного алгоритмического решения оптимальной нелинейной фильтрации оценок информационного воздействия в стохастической модели информационного противоборства. Относительный выигрыш в сравнении со значениями, получаемыми по средневыборочным правилам наблюдения, в среднем равняется 20,7 % (рисунок 5). При этом аналогичный выигрыш применения ансамблевого расширенного фильтра Калмана составляет 7,4 %, а алгоритма нелинейной фильтрации, формируемого на основе исследования уравнения ДМЗ (6), – 17,3 %.

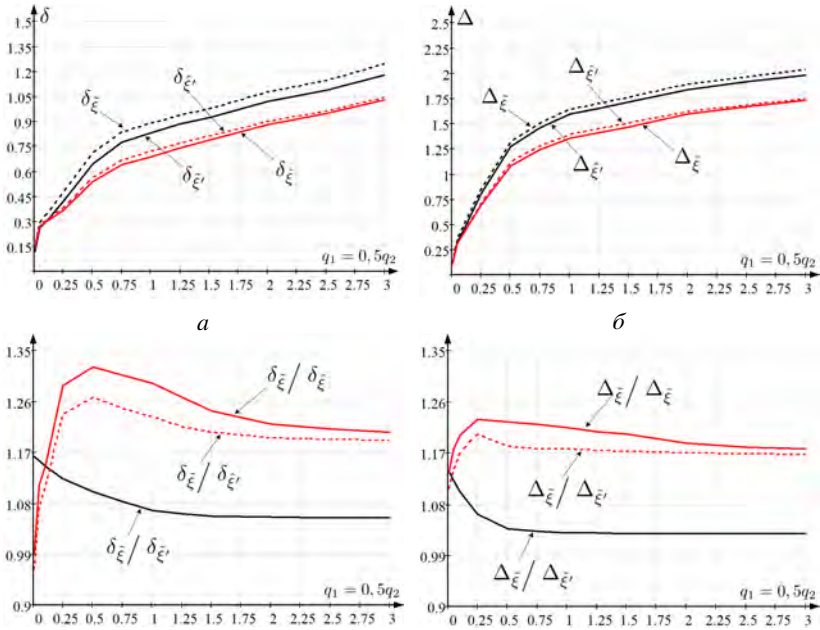


Рис. 5. Графики зависимости от q_1, q_2 при $q_1 = 0, 5q_2$: а, в) $\delta_{\bar{\xi}}, \delta_{\bar{\xi}'}, \delta_{\bar{\xi}}, \delta_{\bar{\xi}'}$; б, г) $\Delta_{\bar{\xi}}, \Delta_{\bar{\xi}'}, \Delta_{\bar{\xi}}, \Delta_{\bar{\xi}'}$

С позиции вычислительных затрат ансамблевый расширенный фильтр Калмана обладает преимуществом в сравнении со сформированным алгоритмом и способен при программной реализации на типовой ПрЭВМ обеспечить вычисления в режиме, близкому к реальному времени. По вычислительной устойчивости, определяемой равномерностью (рисунок 4(б)) сходимости выбранной аппроксимации функций плотностей вероятностей при одинаковой длительности интервала дискретизации Δt , алгоритм, который основан на исследовании уравнения ДМЗ (6), обладает преимуществом в сравнении с разработанным алгоритмом (предполагает применение робастного уравнения ДМЗ (8)). Вместе с тем, по вычислительной сложности указанные алгоритмы сопоставимы и разработанный алгоритм обеспечивает наибольшую точность оценки. Таким образом, исходя из апостериорно определенного выигрыша по результативности решения (точность оценки) при потенциальной возможности адаптивной вариации Δt в зависимости от скорости изменения числа адептов и

предадептов \vec{Z} и уровня шума q_1, q_2 наблюдений α, β, γ , разработанное алгоритмическое решение обладает наибольшей предпочтительностью. Однако его применение в сравнении с ансамблевым расширенным фильтром Калмана становится нецелесообразным (рисунок 5) при малых значениях q_1, q_2 (снижение относительного выигрыша обуславливается вычислительной погрешностью в формируемых численных решениях по кусочно-полиномиальной аппроксимации и интегрированию).

В целом, полученные результаты в развитие работ [11, 13, 14, 18], формируют вычислительно эффективные решения обобщенной задачи оптимальной нелинейной фильтрации оценок информационного воздействия в стохастических моделях информационного противоборства при выделении особенностей основных этапов алгоритмической реализации. Составленный алгоритм применим при наличии неопределенности и разнородности правил измерения соответствующих интенсивностей α, β, γ модели (1). При наличии более двух разнородных правил подобного наблюдения, сформированный алгоритм может быть обобщен, при учете основных теорем теории вероятностей [26] и модификации правила определения итоговой функции плотности распределения вероятности (5). К направлениям дальнейших исследований относится получение априорной оценки, определяющей правило адаптивного выбора Δt в составленной численной схеме.

7. Приложение. В (2) приняты следующие обозначения: $\vec{Z} \in \Omega$ при $\vec{Z} = (Z_l)_d = (\vec{z}^{(1)}, \vec{z}^{(2)}, \dots, \vec{z}^{(M)})$ определяется численностью адептов и предадептов относительно m -й подгруппы $\vec{z}^{(m)} = \left(z_k^{(m)} \right)_{2K} = (x_{m1}, y_{m1}, x_{m2}, y_{m2}, \dots, x_{mK}, y_{mK})$ [18]; $\Omega = \varepsilon^{(1)} \times \varepsilon^{(2)} \times \dots \times \varepsilon^{(M)} \subset \mathbb{R}^d$ – d -мерный ($d = 2MK$) выпуклый многогранник (правила параметризации Ω определены в [13, 14]); $\varepsilon^{(m)} \subset \mathbb{R}^{2K}$ – симплекс с $2K + 1$ вершинами [14]; $\vec{A} = (A_l)_d = (\vec{a}^{(1)}, \vec{a}^{(2)}, \dots, \vec{a}^{(M)})$ – вектор сноса при $\vec{a}^{(m)} = \left(a_k^{(m)} \right)_{2K} = \left(f_{m1}^{(1)}, f_{m1}^{(2)}, f_{m2}^{(1)}, f_{m2}^{(2)}, \dots, f_{mK}^{(1)}, f_{mK}^{(2)} \right)$ для $f_{mk}^{(2)} = \left[N_m - \sum_{k'=1}^K (x_{mk'} + y_{mk'}) - y_{mk} \right] \left[\alpha_{mk}^0 + \sum_{m'=1}^M \beta_{m'm}^0 x_{m'k} \right] + \gamma_{mk}^0 (x_{mk} - y_{mk})$ и $f_{mk}^{(1)} = y_{mk} \left[\alpha_{mk}^0 + \sum_{m'=1}^M \beta_{m'm}^0 x_{m'k} \right] - \gamma_{mk}^0 x_{mk}$; $\vec{W} = (W_l)_d = (\vec{w}^{(1)}, \vec{w}^{(2)}, \dots, \vec{w}^{(M)})$ для $\vec{w}^{(m)} = \left(w_k^{(m)} \right)_{2K} = \left(B_{m1}^{(1)}, B_{m1}^{(2)}, B_{m2}^{(1)}, B_{m2}^{(2)}, \dots, B_{mK}^{(1)}, B_{mK}^{(2)} \right)$ при определении $B_{mk}^{(1)}, B_{mk}^{(2)}$ стандартными винеровскими процессами; $\Sigma = (\Sigma_{ll'})_{d \times d} = \Sigma^\alpha + \Sigma^\beta +$

Σ^γ – матрица диффузии при:

$$\Sigma^{\alpha,\beta,\gamma} = \begin{pmatrix} \sigma_{11}^{\alpha,\beta,\gamma} & \dots & \sigma_{1M}^{\alpha,\beta,\gamma} \\ \dots & \dots & \dots \\ \sigma_{M1}^{\alpha,\beta,\gamma} & \dots & \sigma_{MM}^{\alpha,\beta,\gamma} \end{pmatrix}; \sigma_{mm'}^{\alpha,\beta,\gamma} = \begin{pmatrix} \bar{\sigma}_{11mm'}^{\alpha,\beta,\gamma} & \dots & \bar{\sigma}_{1Kmm'}^{\alpha,\beta,\gamma} \\ \dots & \dots & \dots \\ \bar{\sigma}_{K1mm'}^{\alpha,\beta,\gamma} & \dots & \bar{\sigma}_{KKmm'}^{\alpha,\beta,\gamma} \end{pmatrix};$$

$$\bar{\sigma}_{kk'mm'}^{\alpha,\beta,\gamma} = \begin{pmatrix} \sigma_{mkm'k'}^{\alpha,\beta,\gamma} \\ \bar{\sigma}_{mkm'k'}^{\alpha,\beta,\gamma} \end{pmatrix}; \sigma_{mkm'k'}^\gamma = -x_{mk} \sqrt{\varepsilon_{mkm'k'}^\gamma}; \sigma_{mkm'k'}^\alpha =$$

$$y_{mk} \sqrt{\varepsilon_{mkm'k'}^\alpha}; \bar{\sigma}_{mkm'k'}^\alpha = \left[N_m - \sum_{k''=1}^K (x_{mk''} + y_{mk''}) - y_{mk} \right] \sqrt{\varepsilon_{mkm'k'}^\alpha};$$

$$\bar{\sigma}_{mkm'k'}^\gamma = (x_{mk} - y_{mk}) \sqrt{\varepsilon_{mkm'k'}^\gamma}; \sigma_{mkm'k'}^\beta = y_{mk} \sqrt{\sum_{j=1}^M \sum_{j'=1}^M \varepsilon_{j'mjm'}^\beta x_j' x_{j'k}};$$

$$\bar{\sigma}_{mkm'k'}^\beta = \left[N_m - \sum_{k''=1}^K (x_{mk''} + y_{mk''}) - y_{mk} \right] \sqrt{\sum_{j=1}^M \sum_{j'=1}^M \varepsilon_{j'mjm'}^\beta x_j' x_{j'k}}.$$

Литература

1. Al-Oraiqat A.M., Ulichev O.S., Meleshko E.V., Al-Rawashdeh Y.S., Smirnov O.O., Polishchuk L.I. Modeling strategies for information influence dissemination in social networks. *Journal of Ambient Intelligence and Humanized Computing*. 2022. vol. 13. pp. 2463–2477. DOI: 10.1007/s12652-021-03364-w.
2. Михайлов А.П., Петров А.П.Ч. Математические модели системы «человек-общество». Москва: Физматлит, 2022. 456 с.
3. Chkhartishvili A.G., Gubanov D.A., Novikov D.A. *Social Networks: Models of information influence, control and confrontation*. Springer, 2018. 228 p.
4. Mikhailov A.P., Petrov A.P., Proncheva O.G. A model of information warfare in a society with a piecewise constant function of the destabilizing Impact. *Mathematical Models and Computer Simulations*. 2019. vol. 11. pp. 190–197. DOI: 10.1134/S2070048219020108.
5. Byzov L.G., Gubanov D.A., Kozitsin I.V., Chkhartishvili A.G. A perfect politician for social networks: an approach to analyzing ideological preferences of users. *Automation and Remote Control*. 2021. vol. 82. pp. 1614–1631. DOI: 10.1134/S0005117921090095.
6. Kozitsin I.V. Modeling Opinion Dynamics: Ranking Algorithms on Heterogeneous Populations. 4th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA). 2022. DOI: 10.1109/SUMMA57301.2022.9973437.
7. Galety M.G., Atroshi C.A., Balabantaray B.K., Mohanty S.N. *Social Network Analysis: Theory and Applications*. 2022. 232 p. DOI: 10.1002/9781119836759.
8. Chkhartishvili A.G. The Problem of Finding the Median Preference of Individuals in a Stochastic Model // *Automation and Remote Control*. 2021. vol. 82. pp. 853–862. DOI: 10.1134/S000511792105009X.
9. Губанов Д.А., Петров И.В. Информационные сообщества в социальных сетевых структурах. Ч. 2. Математические сетевые модели формирования сообществ. *Проблемы управления*. 2021. № 2. С. 18–32. DOI: 10.25728/ru.2021.1.2.
10. Михайлов А.П., Петров А.П., Маревцева Н.А., Третьякова И.В. Развитие модели распространения информации в социуме // *Математическое моделирование*. 2014. Т. 26. № 3. С. 65–74.

11. Полянский И.С., Логинов К.О., Ильин Н.И., Великих А.С. Математическая модель оценки информационного воздействия на электорат в социальных медиа при проведении выборных кампаний. Математическое моделирование. 2021. Т. 33. № 12. С. 67–81. DOI: 10.20948/mm-2021-12-05.
12. Орлов Ю.Н. Панкратов А.С. К разработке модели эволюции структуры сетевого графа. Препринты ИПМ им. М.В. Келдыша. 2021. № 24. 16 с. DOI: 10.20948/prepr-2021-24.
13. Ильинский А.С., Полянский И.С., Логинов К.О., Архипов Н.С. К вопросу численной оценки информационного воздействия на электорат при проведении выборных кампаний. Прикладная Математика и информатика: Труды факультета ВМК МГУ им. М.В. Ломоносова. – Москва: МАКС Пресс. 2021. № 68. С. 15–28.
14. Полянский И.С., Полянская И.В., Логинов К.О. Алгоритмические решения в задаче оценки информационного воздействия на электорат при проведении выборных кампаний. Физика волновых процессов и радиотехнические системы. 2021. Т. 24. № 4. С. 72–80. DOI: 10.18469/1810-3189.2021.24.4.72-80.
15. Pugachev V.S., Sinityn I.N. Stochastic differential systems analysis and filtering. Wiley: Science, 1987. 549 p.
16. Тихонов В.И. Оптимальный прием сигналов. Москва: Радио и связь, 1983. 320 с.
17. Jiang Y., Tao J., Chen X. Simulation of Non-Gaussian/Non-stationary Random Vibration. Non-Gaussian Random Vibration Fatigue Analysis and Accelerated Test. Springer, Singapore. 2022. DOI: 10.1007/978-981-16-3694-3_3.
18. Логинов К.О. Численное решение задачи фильтрации оценок информационного воздействия на электорат. Информатика и автоматизация. 2022. Т. 3(21). С. 624–652. DOI: 10.15622/ia.21.3.7.
19. Nakano Y. Kernel-based collocation methods for Zakai equations. Stochastics and Partial Differential Equations: Analysis and Computations. 2019. vol. 9. pp. 476–494. DOI: 10.1007/s40072-019-00132-y.
20. Liangliang S., Chang M.L. Galerkin spectral method for a multi-term time-fractional diffusion equation and an application to inverse source problem. Networks and Heterogeneous Media. 2022. vol. 18. no. 1. pp. 212–243. DOI: 10.3934/nhm.2023008.
21. Jiang Y., Tang Ch., Zhang X., Jiao W., Li G., Huang T. A Novel Rolling Bearing Defect Detection Method Based on Bispectrum Analysis and Cloud Model-Improved EEMD. IEEE Access. 2020. vol. 8. pp. 24323–24333. DOI: 10.1109/ACCESS.2020.2970813.
22. Chugai K.N., Kosachev I.M., Rybakov K.A. Approximate Filtering Methods in Continuous-Time Stochastic Systems. Advances in Theory and Practice of Computational Mechanics. Smart Innovation, Systems and Technologies. Springer, Singapore. 2020. vol. 173. DOI: 10.1007/978-981-15-2600-8_24.
23. Семина Т.А. Анализ тональности текста: современные подходы и существующие проблемы. Социальные и гуманитарные науки. Отечественная и зарубежная литературы. Серия 6. Языкознание. Реферативный журнал. 2020. № 4. С. 47–63.
24. Полянский И.С., Логинов К.О. К вопросу оценки интенсивности межличностной коммуникации пользователей социальной сети. XI Всероссийская научная конференция «Системный синтез и прикладная синергетика» (п. Нижний Архыз, 27 сентября – 01 октября 2022 г.). Южный федеральный университет. Ростов-на-Дону; Таганрог: Издательство Южного федерального университета, 2022. С. 132–135.
25. Dong W., Luo X., Yau S.S.-T. Solving Nonlinear Filtering Problems in Real-time by Legendre Galerkin Spectral Method. IEEE Transactions on Automatic Control. 2020. vol. 66(4). pp. 1559–1572. DOI: 10.1109/TAC.2020.3002979.
26. Borovkov A.A., Moullagaliev A. Mathematical Statistics. London: Routledge. 1998. 592 p. DOI: 10.1201/9780203749326.

27. Luo X., Yau S.S.-T. Complete real time solution of the general nonlinear filtering problem without memory. *IEEE Transactions on Automatic Control*. 2013. vol. 58. no. 10. pp. 2563–2578. DOI: 10.1109/TAC.2013.2264552.
28. Полянский И.С., Полянская И.В., Фам Т.З. Математическая модель фильтрации канонических параметров спутника-ретранслятора при орбитальном движении. *Физика волновых процессов и радиотехнические системы*. 2019. Т. 22. № 4. С. 50–57. DOI: 10.18469/1810-3189.2019.22.4.50-57.
29. Karimi N., Kazem S., Ahmadian D., Adibi H., Ballestra L.V. On a generalized Gaussian radial basis function: Analysis and applications. *Engineering Analysis with Boundary Elements*. 2020. vol. 112. pp. 46–57.
30. Epperson J.F. *An Introduction to Numerical Methods and Analysis: Solutions Manual to Accompany, Third Edition*. John Wiley and Sons, Inc. 2021. 304 p. DOI: 10.1002/9781119604570.
31. Johansson R. *Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib*. Second Edition. 2019. 700 p. DOI: 10.1007/978-1-4842-4246-9.
32. Lasserre J.B. Simple formula for integration of polynomials on a simplex. *BIT Numerical Mathematics*. 2021. vol. 61. pp. 523–533. DOI: 10.1007/s10543-020-00828-x.
33. *Encyclopedia of special functions: the Askey-Bateman project Volume 1 Univariate orthogonal polynomials* (Eds.: Mourad H. Ismail, Walter van Assche). NY: Cambridge University Press. 2020. 388 p. DOI: 10.1017/9780511979156.
34. Wagner U., Welzl E. Connectivity of Triangulation Flip Graphs in the Plane. *Discrete and Computational Geometry*. 2022. vol. 68. pp. 1227–1284. DOI: 10.1007/s00454-022-00436-2.
35. Ильинский А.С., Полянский И.С., Степанов Д.Е. О сходимости барицентрического метода в решении внутренних задач Дирихле и Неймана в R2 для уравнения Гельмгольца. *Вестник Удмуртского университета. Математика. Механика. Компьютерные науки*. 2021. Т. 31. № 1. С. 3–18. DOI: 10.35634/vm210101.
36. Polyanskiy I.S., Arkhipov N.S., Misyurin S.Yu. On solving the optimal control problem. *Automation and Remote Control*. 2019. vol. 80. pp. 66–80. DOI: 10.1134/S0005117919010065.
37. Suescun-Diaz D., Giron L.E. Valuation of Standard Call Options Using the Euler–Maryama Method with Strong Approximation. *Computational Economics*. 2022. vol. 61(4). pp. 1545–1560. DOI: 10.1007/s10614-022-10258-2.
38. Куксенко С.П. *Электромагнитная совместимость: моделирование* (Ред.: Т.Р. Газизова). Томск: В-Спектр, 2018. 188 с.
39. Oryiema R., Angwenyi D., Midenyo K. Extended Ensemble Filter for High-dimensional Nonlinear State Space Models. *Journal of Advances in Mathematics and Computer Science*. 2021. pp. 84–97. DOI: 10.9734/jamcs/2021/v36i530365.
40. Eshima N. *Statistical Data Analysis and Entropy*. Springer Singapore. 2020. 257 p. DOI: 10.1007/978-981-15-2552-0.

Полянский Иван Сергеевич — д-р физ.-мат. наук, доцент, сотрудник, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: математическое моделирование, динамические системы, дифференциальные уравнения, методы оптимизации, оптимальное управление, конформные отображения, вычислительная электродинамика, цифровая обработка сигналов. Число научных публикаций — 153. van341@mail.ru; улица Приборостроительная, 35, 302025, Орёл, Россия; р.т.: +7(953)618-7100.

Логинов Кирилл Олегович — соискатель, Академия Федеральной службы охраны Российской Федерации. Область научных интересов: математическое моделирование, методы статистического анализа данных, динамические системы, оптимальное управление, стохастические дифференциальные уравнения. Число научных публикаций — 12. kvirs@mail.ru; улица Приборостроительная, 35, 302025, Орёл, Россия; р.т.: +7(486)254-1325.

I. POLYANSKY, K. LOGINOV
**OPTIMAL NONLINEAR FILTERING OF INFORMATION IMPACT
ESTIMATES IN A STOCHASTIC MODEL OF INFORMATION
WARFARE**

Polyanskii I.S., Loginov K.O. Optimal Nonlinear Filtering of Information Impact Estimates in a Stochastic Model of Information Warfare.

Abstract. A computationally efficient algorithmic solution to the problem of optimal nonlinear filtering of information impact estimates in a generalized stochastic model of information warfare is developed in the article. The formed solution is applicable in the presence of heterogeneous rules for measuring the parameters of the information warfare model, on the basis of which a pair of systems of stochastic differential equations is formed. According to the criterion of maximum likelihood according to the determined evolution of the a posteriori conditional probability density function at a given observation interval, the evaluation of the information impact in the optimal nonlinear filtering model is performed. Taking into account the probability addition theorem, as the probability of the sum of two joint events, the density functions of which are established from the numerical solution of the corresponding robust Duncan-Mortensen-Zakai equations, finding a posteriori conditional probability density function at a given time is performed. For the first event, it is assumed that the first system of stochastic differential equations is the equation of state, and the second – is the equation of observation. For the second event, their definition is set in reverse order. The solution of the robust Duncan-Mortensen-Zakai equation is carried out in the formulation of the Galerkin spectral method when sampling the observation interval into subintervals and reducing the initial solution to a numerical recurrent study of the sequence of subtasks using the so-called Yau-Yau's algorithm, which assumes an estimate of the probability measure from the solution of the direct Kolmogorov equation with its subsequent correction by observation. To highlight the features of the algorithmic implementation of the compiled solution, an algorithm for optimal nonlinear filtering of information impact estimates in a generalized stochastic model of information confrontation when specifying the listing of the function implementing it, which is represented by a pseudocode, has been formed. To identify the preference of the compiled algorithmic solution for optimal nonlinear filtering of information impact assessments, a series of computational experiments on large-volume test samples was carried out. The result of the information impact assessment obtained by the proposed algorithm is compared with the determined solution: 1) by the average sample values from the observation models; 2) by an ensemble extended Kalman filter; 3) by a filtering algorithm involving a numerical study of the Duncan-Mortensen-Zakai equation. According to the conducted a posteriori study, quantitative indicators that establish the gain of the compiled algorithm and the limits of its applicability are highlighted.

Keywords: nonlinear filtering, information impact assessment, stochastic model of information confrontation, robust Duncan-Mortensen-Zakai equation, Gaussian basis functions, Galerkin spectral method.

References

1. Al-Oraiqat A.M., Ulichev O.S., Meleshko E.V., Al-Rawashdeh Y.S., Smirnov O.O., Polishchuk L.I. Modeling strategies for information influence dissemination in social networks. *Journal of Ambient Intelligence and Humanized Computing*. 2022. vol. 13. pp. 2463–2477. DOI: 10.1007/s12652-021-03364-w.

2. Mihajlov A.P., Petrov A.P.Ch. *Matematicheskie modeli sistemy «chelovek-obshchestvo»* [Mathematical models of the «man-society» system]. Moscow: Fizmatlit, 2022. 456 p. (In Russ.).
3. Chkhartishvili A.G., Gubanov D.A., Novikov D.A. *Social Networks: Models of information influence, control and confrontation*. Springer, 2018. 228 p.
4. Mikhailov A.P., Petrov A.P., Proncheva O.G. A model of information warfare in a society with a piecewise constant function of the destabilizing Impact. *Mathematical Models and Computer Simulations*. 2019. vol. 11. pp. 190–197. DOI: 10.1134/S2070048219020108.
5. Byzov L.G., Gubanov D.A., Kozitsin I.V., Chkhartishvili A.G. A perfect politician for social networks: an approach to analyzing ideological preferences of users. *Automation and Remote Control*. 2021. vol. 82. pp. 1614–1631. DOI: 10.1134/S0005117921090095.
6. Kozitsin I.V. Modeling Opinion Dynamics: Ranking Algorithms on Heterogeneous Populations. 4th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA). 2022. DOI: 10.1109/SUMMA57301.2022.9973437.
7. Galety M.G., Atroshi C.A., Balabantaray B.K., Mohanty S.N. *Social Network Analysis: Theory and Applications*. 2022. 232 p. DOI: 10.1002/9781119836759.
8. Chkhartishvili A.G. The Problem of Finding the Median Preference of Individuals in a Stochastic Model. *Automation and Remote Control*. 2021. vol. 82. pp. 853–862. DOI: 10.1134/S000511792105009X.
9. Gubanov D.A., Petrov I.V. Information Communities in Social Networks. Part II: Networked Models of Formation. *Control Sciences*. 2021. vol. 2. pp. 16–28. DOI: 10.25728/cs.2021.2.2.
10. Mikhailov A.P., Petrov A.P., Marevtseva N.A., Tretiakova I.V. [Development of a model of information dissemination in society]. *Matematicheskoe modelirovanie – Math modeling*. 2014. vol. 26. no. 3. pp. 65–74. (In Russ.).
11. Poljanskij I.S., Loginov K.O., Ilyin N.I., Velikikh A.S. [Mathematical model assessing the information impact on the electorate in social media during election campaigns]. *Matematicheskoe modelirovanie – Math modeling*. 2021. vol. 33. no. 12. pp. 67–81. (In Russ.).
12. Orlov Y.N., Pankratov A.S. To the evolution model of network graph structure construction. Preprinty' IPM im. M.V. Keldy'sha – IPM preprints im. M.V. Keldysh. 2021. no. 24. 16 p. DOI: 10.20948/prepr-2021-24. (In Russ.).
13. Il'inskii A.S., Polyanskii I.S., Loginov K.O., Arkhipov N.S. On the issue of numerical assessment of informational impact on the electorate during electoral campaigns. *Prikladnaya Matematika i informatika: Trudy' fakul'teta VMKMGUim. M.V. Lomonosova – Applied Mathematics and Informatics: Proceedings of the faculty of the VMK Moscow State University. M.V. Lomonosov*. Moscow: MAKS Press. 2021. no. 68. pp. 15–28. (In Russ.).
14. Polyanskii I.S., Polyanskaya I.V., Loginov K.O. [Algorithmic solutions to the problem of assessing the information impact on the electorate during election campaigns]. *Fizika volnovykh processov i radiotekhnicheskie sistemy – Physics of wave processes and radio engineering systems*. 2021. vol. 24. no. 4. pp. 72–80. (In Russ.).
15. Pugachev V.S., Sinityn I.N. *Stochastic differential systems analysis and filtering*. Wiley: Science, 1987. 549 p.
16. Tikhonov V.I. [Optimal signal reception] *Optimal'nyj priem signalov*. Moscow: Radio and communications, 1983. 320 p. (In Russ.).
17. Jiang Y., Tao J., Chen X. *Simulation of Non-Gaussian/Non-stationary Random Vibration. Non-Gaussian Random Vibration Fatigue Analysis and Accelerated Test*. Springer, Singapore. 2022. DOI: 10.1007/978-981-16-3694-3_3.

18. Loginov K.O. [Numerical solution of the problem of filtering estimates of information impact on the electorate]. *Informatika i avtomatizaciya – Informatics and automation*. 2022. vol. 3(21). pp. 624–652. DOI: 10.15622/ia.21.3.7. (In Russ.).
19. Nakano Y. Kernel-based collocation methods for Zakai equations. *Stochastics and Partial Differential Equations: Analysis and Computations*. 2019. vol. 9. pp. 476–494. DOI: 10.1007/s40072-019-00132-y.
20. Liangliang S., Chang M.L. Galerkin spectral method for a multi-term time-fractional diffusion equation and an application to inverse source problem. *Networks and Heterogeneous Media*. 2022. vol. 18. no. 1. pp. 212–243. DOI: 10.3934/nhm.2023008.
21. Jiang Y., Tang Ch., Zhang X., Jiao W., Li G., Huang T. A Novel Rolling Bearing Defect Detection Method Based on Bispectrum Analysis and Cloud Model-Improved EEMD. *IEEE Access*. 2020. vol. 8. pp. 24323–24333. DOI: 10.1109/ACCESS.2020.2970813.
22. Chugai K.N., Kosachev I.M., Rybakov K.A. Approximate Filtering Methods in Continuous-Time Stochastic Systems. *Advances in Theory and Practice of Computational Mechanics. Smart Innovation, Systems and Technologies*. Springer, Singapore. 2020. vol. 173. DOI: 10.1007/978-981-15-2600-8_24.
23. Semina T.A. [Text tonality analysis: modern approaches and existing problems]. *Social'nye i gumanitarnye nauki. Otechestvennaya i zarubezhnaya literaturny. Seriya 6. Yazykoznanie. Referativnyj zhurnal – Social and human sciences. Domestic and foreign literature. Series 6. Linguistics. Abstract journal*. 2020. no. 4. pp. 47–63. (In Russ.).
24. Polyansky I.S., Loginov K.O. [On the issue of assessing the intensity of interpersonal communication of social network users]. XI Vserossijskaya nauchnaya konferenciya «Sistemnyj sintez i prikladnaya sinergetika» [XI All-Russian Scientific Conference «System Synthesis and Applied Synergetics»]. Rostovna-Donu; Taganrog : Izdatel'stvo Yuzhnogo federal'nogo universiteta, 2022. pp. 132–135. (In Russ.).
25. Dong W., Luo X., Yau S.S.-T. Solving Nonlinear Filtering Problems in Real-time by Legendre Galerkin Spectral Method. *IEEE Transactions on Automatic Control*. 2020. vol. 66(4). pp. 1559–1572. DOI: 10.1109/TAC.2020.3002979.
26. Borovkov A.A., Moullagaliev A. *Mathematical Statistics*. London: Routledge. 1998. 592 p. DOI: 10.1201/9780203749326.
27. Luo X., Yau S.S.-T. Complete real time solution of the general nonlinear filtering problem without memory. *IEEE Transactions on Automatic Control*. 2013. vol. 58. no. 10. pp. 2563–2578. DOI: 10.1109/TAC.2013.2264552.
28. Polyanskii I.S., Polyanskaya I.V., Pham T.Z. [Mathematical model of filtering canonical parameters of a repeater satellite during orbital motion]. *Fizika volnovykh processov i radiotekhnicheskie sistemy – Physics of wave processes and radio engineering systems*. 2019. vol. 22. no. 4. pp. 50–57. (In Russ.).
29. Karimi N., Kazem S., Ahmadian D., Adibi H., Ballestra L.V. On a generalized Gaussian radial basis function: Analysis and applications. *Engineering Analysis with Boundary Elements*. 2020. vol. 112. pp. 46–57.
30. Epperson J.F. *An Introduction to Numerical Methods and Analysis: Solutions Manual to Accompany, Third Edition*. John Wiley and Sons, Inc. 2021. 304 p. DOI: 10.1002/9781119604570.
31. Johansson R. *Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib*. Second Edition. 2019. 700 p. DOI: 10.1007/978-1-4842-4246-9.
32. Lasserre J.B. Simple formula for integration of polynomials on a simplex. *BIT Numerical Mathematics*. 2021. vol. 61. pp. 523–533. DOI: 10.1007/s10543-020-00828-x.

33. Encyclopedia of special functions: the Askey-Bateman project. Volume 1: Univariate orthogonal polynomials (Eds.: Mourad H. Ismail, Walter van Assche). NY: Cambridge University Press. 2020. 388 p. DOI: 10.1017/9780511979156.
34. Wagner U., Welzl E. Connectivity of Triangulation Flip Graphs in the Plane. *Discrete and Computational Geometry*. 2022. vol. 68. pp. 1227–1284. DOI: 10.1007/s00454-022-00436-2.
35. Il'inskiia A.S., Polyanskii I.S., Stepanov D.E. [On the convergence of the barycentric method in solving internal Dirichlet and Neumann problems in R^2 for the Helmholtz equation]. *Vestn. Udmurtsk. Univ. Mat. Mekh. Komp. Nauki – Bulletin of the Udmurt University. Mathematics. Mechanics. Computer science*. 2021. vol. 31. no. 1. pp. 3–18. (In Russ.).
36. Polyanskii I.S., Arkhipov N.S., Misyurin S.Yu. On solving the optimal control problem. *Automation and Remote Control*. 2019. vol. 80. pp. 66–80. DOI: 10.1134/S0005117919010065.
37. Suescun-Díaz D., Giron L.E. Valuation of Standard Call Options Using the Euler–Maruyama Method with Strong Approximation. *Computational Economics*. 2022. vol. 61(4). pp. 1545–1560. DOI: 10.1007/s10614-022-10258-2.
38. Kuksenko S.P. [Electromagnetic compatibility: simulation] *Elektromagnitnaya sovместimost': modelirovanie*. Tomsk: B-Spectrum, 2018. 188 p. (In Russ.).
39. Oryiema R., Angwenyi D., Midenyo K. Extended Ensemble Filter for High-dimensional Nonlinear State Space Models. *Journal of Advances in Mathematics and Computer Science*. 2021. pp. 84–97. DOI: 10.9734/james/2021/v36i530365.
40. Eshima N. *Statistical Data Analysis and Entropy*. Springer Singapore. 2020. 257 p. DOI: 10.1007/978-981-15-2552-0.

Polansky Ivan — Ph.D., Dr.Sci., Associate Professor, Employee, The Academy of Federal Security Guard Service of the Russian Federation. Research interests: mathematical modeling, dynamical systems, differential equations, optimization methods, optimal control, conformal mappings, computational electrodynamics, digital signal processing. The number of publications — 153. van341@mail.ru; 35, Priborostroitel'naya St., 302025, Orel, Russia; office phone: +7(953)618-7100.

Loginov Kirill — Applicant, The Academy of Federal Security Guard Service of the Russian Federation. Research interests: mathematical modeling, methods of statistical data analysis, dynamic systems, optimal control, stochastic differential equations. The number of publications — 12. kvirs@mail.ru; 35, Priborostroitel'naya St., 302025, Orel, Russia; office phone: +7(486)254-1325.

M. GIDWANI, A. RAO
**COMPARATIVE ANALYSIS OF RUMOUR DETECTION ON
SOCIAL MEDIA USING DIFFERENT CLASSIFIERS**

Gidwani M., Rao A. Comparative Analysis of Rumour Detection on Social Media Using Different Classifiers.

Abstract. As the number of users on social media rise, information creation and circulation increase day after day on a massive basis. People can share their ideas and opinions on these platforms. A social media microblogging site such as Facebook or Twitter is the favoured medium for debating any important event, and information is shared immediately. It causes rumours to spread quickly and circulates inaccurate information, making people uneasy. Thus, it is essential to evaluate and confirm the level of veracity of such information. Because of the complexities of the text, automated detection of rumours in their early phases is challenging. This research employs various NLP techniques to extract information from tweets and then applies various machine learning models to determine whether the information is a rumour. The classification is performed using three classifiers such as SVC (Support Vector Classifier), Gradient Boosting, and Naive Bayes classifiers for five different events from the PHEME dataset. Some drawbacks include limited handling of imbalanced data, difficulty capturing complex linguistic patterns, lack of interpretability, difficulty handling large feature spaces, and insensitivity to word order and context by using the above classifiers. The stacking approach is used to overcome the above drawbacks in which the output of combined classifiers is an ensemble with LSTM. The performance of the models has been analyzed. The experimental findings reveal that the ensemble model obtained efficient outcomes compared to other classifiers, with an accuracy of 93.59%.

Keywords: Rumor detection, Machine learning, Social Media, SVC, Gradient boosting, Naive Bayes.

1. Introduction. Social networks are undisputedly the most prevalent information access environment today. It is due to massive amounts of data and massive social network connections. While researching social networks, researchers attempt to answer some issues that emerge from the rich relationships that they contain [1]. The plethora of posting methods accessible on internet communities like Twitter, Weibo, Instagram, and Facebook facilitates message dissemination. It, however, adds to the fast distribution of unverified whispers and conspiracy theories, which frequently evoke accelerated, massive, but naïve social responses. Even so, the World Economic Forum's 2013 report has identified massive digital misinformation as a significant technical and global political risk. False reports had an impact on our economies, which are not impervious to the dissemination of deceit. There are two types of rumours: new rumours and long-standing rumours. New rumours arise during breaking news, as do "long-standing rumours" that have been debated for a long time [2]. Rumours will spread and increase visibility more than previously anticipated. According to statistics, there were nearly 320 million monthly

active users on Twitter in November 2015, with 500 million tweets shared per day on average [3]. Social media platforms have grown in popularity, with users using them to share news and events and voice their opinions. The network's ease also fosters the spread of lies; For example, the fast dissemination of disinformation in a brief period can have negative things. However, determining the validity of rumours purely based on their textual substance can be challenging. As a result, people have begun to pay attention to the views expressed in relative responses. Knowing the stances that users adopt in reaction to specific rumours provides helpful insights, whereas some denying or questioning comments expose false rumours. [4, 5, 6].

Figure 1 depicts an example of a rumour that circulated during the recent COVID-2019 prompted India's Lockdown, alleging a decrease in pension payments to Indian citizens. The tale was unsubstantiated and wrong, and the Ministry of Finance of the Government of India needed to dispute it and provide clarification.

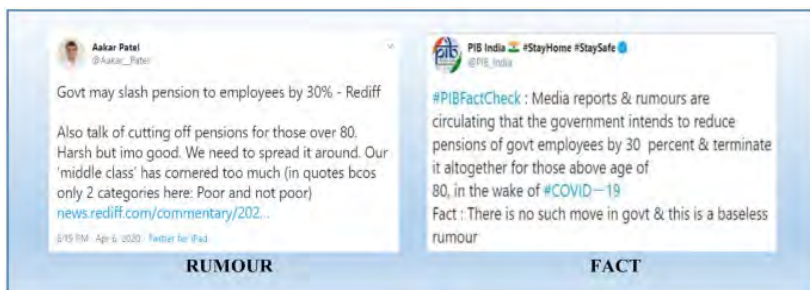


Fig. 1. Sample rumour and fact-checked in recent times

Therefore, it is crucial to be able to tell whether information elements posted on social media are real and legitimate, especially regarding current events. In this situation, building automated credibility detection systems is essential because the information seeker cannot tell true information from fake information. Nonetheless, there are several difficulties in determining the news' reliability. Nevertheless, there are several difficulties in determining the trustworthiness of news; these difficulties are caused by the limited datasets that include the aspects of the tweet that must be examined. Without technology, it might be challenging to identify accurate and reasonable postings or information. It necessitates time and work. Several researchers [7] use Twitter API to determine the creditability of the information. The limited size of tweets and casual wording is another difficulty. The phase of feature extraction and selecting

the ideal function for constructing and executing the classification devices is the most difficult problem in the rumour detection task when using noisy data. Rumour identification aims to determine whether or not an inbound social media post contains rumours. It is simply a text classification problem.

Feature engineering is a critical component of text categorization because it is required to convert data into a machine learning-friendly framework. The primary problem with text-based rumour identification is a lower discovery rate, which occurs when algorithms incorrectly categorize fake material. It has a bad result on the system's detection rate & precision. In recent years, AI approaches have produced cutting-edge outcomes for various natural language processing tasks. The PHEME dataset, which includes rumour and non-rumour tweets on five incidents – Charlie Hebdo, the German plane tragedy, the Ottawa massacre, the Sydney siege, and Ferguson – is used in this investigation.

Author contribution. In this article, we investigate the issue of automatically detecting rumours as they circulate on social media. We provide a novel approach for automatically detecting rumours on social media that combines word embeddings with machine learning. The following listing summarises the important contributions of this work:

1. To address the cross-topic problems in broadcast rumour identification, we provide a novel technique for updating word embeddings while training proceeds using word2vec.

2. The experiment takes place in five different events from the PHEME dataset having conversational structure, and it is evaluated using a variety of measures, including accuracy, precision, recall, and the AUC/ROC curve. The results show that the ensemble model outperforms better with high accuracy by comparing the four classifiers for overall events in the dataset.

The rest of the content is structured as follows: the prior effort in the area of rumour detection is described in segment 2, an overview of the proposed method is discussed in section 3, simulation results are discussed in section 4, and the paper is concluded with future research directions in section 5, which is followed by references used in this research work.

2. Related work. The amount and speed of user-generated information on social media make rumour detection imperative. The first and most essential step in locating unverified information circulating on social media is rumour detection. It serves as the foundation for all subsequent research. The outcome of future tasks can be indirectly improved by increasing this task's accuracy.

Regardless of the source's veracity and position as a verified source, social media enables the spread of information. Several studies have demonstrated automatic rumour identification in social media data.

The first technique was put out by authors [8] and begins with identifying "signal tweets," which are then sorted into several clusters, each of which stands for a rumour. Next, further similar tweets are found using the summary of each cluster. The clusters are evaluated in order of how likely they will be – rumours at the end. The whole foundation of the suggested framework is a set of user-defined regular expressions. This list must be continuously edited and updated for the model to handle new, unseen tales more effectively. Another study in this category is provided in [9]. The researcher suggested a rumour identification model based on a sequential classifier, in which the text was classified as likely a rumour or a non-rumour based on preliminary data. While this technique outperforms the earlier efforts, it is affected by the cold start problem, which means the suggested sequential classifier's effectiveness depends on the tweets' structures observed thus far. Our suggested strategy addresses this issue by categorizing each micro-post solely on its features. It is not required to use micro-post procedures.

In paper [10] the authors suggested a rumour detection approach that matches social media material to news media and employs the support vector machine (SVM) as a binary classification strategy. However, it is not suitable for large amounts of data. In study [11] the authors suggested a combined approach for rumour classification that includes deep learning and an optimized filter-wrapper classifier for naive Bayes.

Articles of news that have been purposefully written to present incorrect information are called rumours. Recently, rumour detection has gained much interest as a developing study area, particularly on social media. This field of study aims to predict whether a news item is a rumour. The work in this sector may be divided into two families: selecting news stories worth checking and assessing their authenticity. In paper [12] the authors suggested automated categorization of news stories. Their study focuses on different textual properties that can be used to discern between genuine and phoney content. Using these properties, they trained various machine learning algorithms tested on four real-world datasets using different ensemble approaches. However, it was more expensive to create, train, and deploy. In study [13] the authors focused on rumour detection using NLP and data mining techniques. Authors categorize fake news that spread on social media into two categories: rumours that have been around for a while and newly developing rumours that result from current occurrences. Authors create a four-part framework for categorizing

rumours: 1) rumour detection; 2) rumour tracking; 3) rumour stances; 4) rumour veracity.

Data imputation was used to improve the identification of fake news, according to the authors [14]. To improve performance, the authors used a novel data pre-processing method to fill in the absent value in the original dataset. Using data modelling, the authors used missing numbers for numerical and hierarchical traits. When authors selected the most common value in columns and were numerical for the average value of the column, the accuracy of multilayer perceptron (MLP) classes were shown to be improved for hierarchies.

Study [15] provided several important influencing factors, and the innovative rumour-refuting method presented in this study used NLP to analyze user microblog content before using the NLP analysis findings as extra predictor factors to identify anti-rumour spreaders. This research provides a novel and efficient decision support system for rumour countermeasures compared to prior studies. The number of pairwise similarities increases quadratically with dataset size, posing a problem for similarity-based machine learning techniques on big datasets. As a result, computing all potential pairwise similarities for big databases is virtually impossible.

3. Rumour Detection on social media by Machine Learning and Deep Learning Models. Any information, whether real or fake, has the potential to become viral on social media and touch millions of people. Even rumours spread quickly because of how quickly information is shared. As a result, it's critical to identify and suppress these rumours before they seriously affect people's lives.

The goal is to determine whether a specific micro-post on a particular bit of data is a story. It is the definition of the research issue of detecting breaking news rumours. This issue may be expressed as a binary classification problem in the manner described below: Let the word sequence $w = \{wc_1 \dots wc_N\}$ represent a micro-post w of length N . The objective is to determine if w is a rumour or non-rumour by selecting a label from the set $S = \{R, NR\}$.

Figure 2 shows the suggested rumour detection model's workflow. There are four stages in the model: 1) data collection; 2) pre-processing; 3) feature extraction; 4) training and testing of Data using a machine learning algorithm. Accuracy, precision, recall, and F1-Score measurements were employed to evaluate the effectiveness of each technique. More details on the steps are provided in the following subsections.

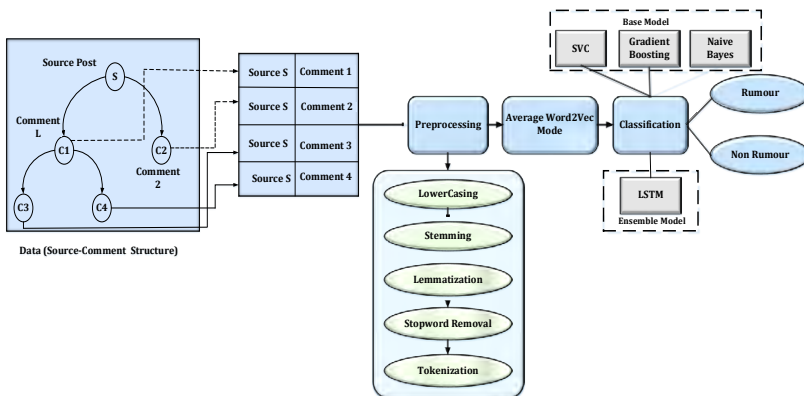


Fig. 2. Workflow of Proposed Methodology

3.1. Data Collection. The standard PHEME dataset utilized in this study for rumour identification [16] includes tweets about five significant news events that seasoned journalists have marked for the 'rumour' and 'non-rumour' classifications. Here, the objective is to simulate how a source post's comments behave in a conversational structure. The whole network was processed by taking into account the source node and their immediate comment reply node. All source posts and immediate nodes were then saved in CSV format. These are the events:

1. #charliehebdo – around noon on January 7, 2015, two gunmen attacked the Paris headquarters of the French satirical weekly newspaper Charlie Hebdo, murdering 12 persons and injuring 11. The collection comprehends 458 whispers and 1621 non-rumours.

2. #ferguson – on August 9, 2014, Darren Wilson, a white police officer, mortally shot Michael Brown Jr., an 18-year-old African American, in Ferguson, Missouri. The officer describes an encounter with Brown during which Brown assaulted him. Wilson later pursued him after he fled. Wilson fired twelve shots, six of which struck Brown from the front. Numerous demonstrations ensued. There are 284 rumours and 859 non-rumours in the dataset.

3. #germanwingscrash – on March 24, 2015, an Airbus A320-211 transporting 144 people and six crew members crashed in the French Alps, 100 kilometres (62 miles) northwest of Nice, while en transit from Barcelona-El Prat Airport in Spain to Düsseldorf Airport in Germany. The co-pilot, whom his psychiatrist had identified as having suicidal inclinations, deliberately crashed the aircraft. The collection contains 238 whispers and 231 non-rumours.

4. #ottawashooting – on October 22, 2014, many gunshots happened on Parliament Hill in Ottawa. Michael Zehaf-Bibeau mortally shot Corporal Nathan Cirillo, a Canadian soldier acting as a symbolic sentry at the Canadian National War Memorial. Zehaf-Bibeau then went inside the adjoining Centre Block of the Canadian Parliament, where MPs were conducting caucuses. After a struggle with a policeman outside, Zehaf-Bibeau rushed inside and engaged in a firefight with parliament security. Six cops shot him 31 times, and he passed away there and then. The dataset comprises 470 rumours and 420 non-rumours.

5. #sydney siege – on the 15th and 16th of December, 2014, a gunman named Man Haron Monis kidnapped ten clients and eight employees at a Lindt confectionery café in Sydney, Australia. Two individuals were murdered, and a couple more were injured during the 16-hour standoff.

One indicates a rumour, and zero indicates a non-rumour. The data cleaning technique is applied in two different processes. The deactivated ID must be eliminated once the null value has been eliminated.

3.2. Pre-processing. The process of changing data so that the machine learning program can understand is known as pre-processing. The model is fed with clean data transformed from the original raw data. Especially for social media content, data pre-processing is an important stage. These datasets must go through the following rounds of pre-processing to receive specific refinements. The crucial cleaning procedures within the pre-processing techniques applied in this study include stop-word removal, tokenization, lowercase, deleting repeated letters, auto-correct spelling, and stemming. They will assist us by removing redundant data and reducing the actual data's quantity. Then, to achieve better execution, the pre-processing entails the ordering of the processes stated in the following steps:

1. Getting rid of the unnecessary: Punctuation indicators such as commas, apostrophes, quotation marks, and question marks are deleted if they contribute no value to the performance.

2. Elimination of Stop Words: Stop words are frequently used to describe the most general terms in a language that lack substantial semantic substance in an assertion. Each tweet in the datasets has these terms removed.

3. Lowercasing is changing uppercase characters to lowercase letters in a phrase. For example, "Charlie Hebdo became well known for publishing the Muhammed cartoons two years ago" will be converted into "charlie hebdo became well known for publishing Muhammed cartoons two years ago".

4. Tokenization is the process of reducing vast quantities of substance into smaller token-sized bits. The tokens, split down into words, are counting the sentences. For example, the processed input tokenizes such as, 'charlie', 'hebdo', 'became', 'well', 'known', 'publishing', 'Muhammed', 'cartoon', 'two', 'year', 'ago'.

5. The stemming and Lemmatization process recovers the lemma form for each input word, and end users can ask any form of a base word and receive relevant results, reducing the word to its word stem.

The pre-processed tweets are fed to the feature extraction stage to extract the features to classify the rumour and non-rumour contents.

3.3. Feature Extraction using Word2Vec. A word2vec model uses a text dataset as input to generate real-valued, low-dimensional vector approximations of the words in the corpus. As a result, written information is transformed into dispersed vector representations, which may subsequently be used as input for the classifier to classify rumours and non-rumours. Word embedding is vector depictions of words.

In this research, we use the Google News Dataset [17] to train the word2vec model more efficiently using a method dubbed skip-gram. The word2vec model [18] is created using skip-gram using a text corpus. Let wc_i be a word in the corpus, and let the context of wc_i be the collection of words surrounding wc_i inside a certain window size of a phrase. By learning the word representations of each word wc_i and the words in their context, skip-gram constructs the word2vec model. The learning goal in this instance is to find meaningful representations of these phrases in the embedding space so that the model, given any other word, w_t , can predict its surrounding context words with high probabilities while the others have low probabilities. A skip-gram model's formal goal is to maximize the average log probability function shown shown in equation (1), given a list of words $w = \{wc_1... wc_N\}$ and a context window of size W .

$$\text{Average log probability function} = \frac{1}{N} \sum_{n=1}^N \sum_{-w \leq j \leq w, j \neq 0} \log p(wc_{n+j} | wc_n), \quad (1)$$

where $\log p(wc_{n+j} | wc_n)$ is approximated using negative sampling. The Word2vec probabilistic model is used to anticipate the context of the nearby words. The words were predicted using a mix of model architectures depending on the corresponding context word. The dataset contains a substantial quantity of data and is transformed into word vectors. It is passed into the classifiers.

3.4. Applying learning models. We tested the efficacy of different models using three conventional machine-learning classifiers. Three different models were tested to choose the most effective model: support vector machine (SVM), Naive Bayes (NB) and gradient boosting classifier.

3.4.1. Support Vector Classifier (SVC). SVM [19], a subset of machine learning, seeks the optimal hyperplane, which can often divide data sets into two groups. A neural network approach merely looks for hyperplane separators between classes, whereas SVM seeks the optimal hyperplane. It is said to be the best if a hyperplane has the widest margin and can categorize fresh data effectively and without mistakes – a hyperplane in the shape of a straight line for two-dimensional data. SVM is widely regarded as effective in real-world applications where data sets are usually segregated nonlinearly and provide superior results than other methods. In non-linear SVM, data is transferred to an advanced dimensional vector space by the function $\phi(x)$ and then a hyperplane can be created to serve as a class separator.

The dot product produced from the converted data, namely ϕ_x, ϕ_y , is the sole component of the SVM learning process dependent on it. Because transformation is challenging to convert, kernel $K(z_x, z_y)$ can be used instead of computing dot products. The following formula denotes this function as Kernel Trick:

$$K(z_x, z_y) = \phi_x \cdot \phi_y. \quad (2)$$

By knowing the kernel functions that will be used, the Kernel Trick can streamline identifying support vectors. In our work, we have used RBF (radial Bias Function) has been used as a kernel function.

3.4.2. Gradient Boosting Classifier. A strong classifier is gradually built from several weak learners using the additive and iterative tree-based supervised machine learning technique called "boosting" [20]. Gradient boosting creates a prediction model from a collection of poor prediction models, most commonly decision trees. When a decision tree fails to learn, the technique that results is known as gradient-enhanced trees, and it typically beats random forest. It builds the model in the same way that other boosting techniques do, but it broadens the scope by enabling the optimization of any differentiable loss function.

3.4.3. Naive Bayes Classifier. The naive Bayes technique is founded on the Bayes hypothesis [21]. It employs training data to calculate the likelihood that certain attributes belong to a certain class to forecast the unobserved data class (testing data).

The posterior probability, $P(a|z)$, can be calculated using the Bayes theorem by dividing the $P(z)$ with the product of $P(a)$ and $P(z|a)$. The naive Bayes classifier assumes that the impact of a predictor's value (z) on a certain class (a) is unrelated to the values of other predictors. The term "class conditional independence" refers to this assumption:

$$P(a|z) = \frac{P(z|a)P(a)}{P(z)}, \quad (3)$$

$$P(a|Z) = P(z_1|a) \times P(z_2|a) \times \dots \times P(z_T|a) \times P(a), \quad (4)$$

where:

$P(a|z)$ is the posterior probability of the class-given attribute.

$P(a)$ is the prior probability of class.

$P(z|a)$ is the likelihood which is the predictor given class probability.

$P(z)$ is the prior probability of the predictor.

The proposed model consists of three machine learning base models: Support Vector Classifier (SVC), Gradient Boosting, and Naive Bayes. These models are combined with an ensemble model called Long Short-Term Memory (LSTM).

3.4.4. Short-Term Long Memory (LSTM). A modified RNN network called long short-term memory (LSTM) was developed to learn long-range relationships among time-varying patterns [19]. In general, LSTM is a second-order recurrent neural network that addresses the problem of disappearing and exploding gradients by substituting memory blocks in the recurrent hidden layer for RNN simple units. In an LSTM, a memory block is a complicated multi-unit processing unit. It comprises one or more memory cells, input, output, and forget adaptive multiplicative gating units, and a self-recurrent connection with a fixed weight of 1.0. With help from adaptive multiplicative gating units, it functions as a short-term memory. Input and output gates, respectively, govern a memory cell's input and output flow during activation.

In general, an Android app features $y = (y_1, y_2, \dots, y_T)$ are transmitted as input to LSTM, which then estimates an output sequence $z_t = (z_{t_1}, z_{t_2}, \dots, z_{t_T})$ with a continuous estimation of the gates that include input gate (zg), output gate (xg), forget gate (fg), and enhancing a memory cell (ml) activations iteratively from $t = 1$ to T . It is possible to formulate the calculation of the recurrent hidden layer function at time-step T as follows:

$$y_t, hz_{t-1}, ml_{t-1} \rightarrow hz_t, ml_t, \quad (5)$$

$$zg_t = \text{sigm}\left(\omega_{yzg} y_t + \omega_{hzzg} hz_{t-1} + \omega_{mlzg} ml_{t-1} + a_{zg}\right), \quad (6)$$

$$fg_t = \text{sigm}\left(\omega_{yfg} y_t + \omega_{hzzfg} hz_{t-1} + \omega_{fgml} ml_{t-1} + a_{fg}\right), \quad (7)$$

$$ml_t = fg_t \odot ml_{t-1} + zn_t \odot \tanh\left(\omega_{mly} y_t + \omega_{mlhz} hz_{t-1} + a_{ml}\right), \quad (8)$$

$$xt_n = \text{sigm}\left(\omega_{xyt} y_t + \omega_{xht} hz_{t-1} + \omega_{xmt} ml_t + a_{xt}\right), \quad (9)$$

$$hz_t = xt_t \odot \tanh(ml_t), \quad (10)$$

where \tanh represents the element-wise hyperbolic tangent non-linear activation function with outcomes in a range of $[-1,1]$, sigm is known as element-wise sigmoid non-linear activation function with an outcome range of $[0, 1]$, \odot stands for element-wise vector multiplication, " ω terms" represents the weight matrices, " a_{in} ," " a_{fg} ," " a_{ml} ," and " a_{ot} " are bias units of input gate, forget gate, memory cell, and output gate, correspondingly and ml_t instructed LSTM network to acquire knowledge of long-term temporal dynamics in the sequence data through proactively selecting either the current input value employing input gate zg_t or forgetting earlier stated data with fg_t .

4. Result and Discussion. A developed model's performance was assessed for individual events in the PHEME dataset, considering the features from the conversation comments for the source post. Experiments were conducted on a typical computer. (Intel Core i5-6500 3.20 GHz, 8 GB RAM). Python3.11.1 with Jupyter Notebook was used to create the model. The pandas, nltk, genism for word2vec, sklearn for machine learning model, and NumPy libraries are included in the Python packages. The classifier used the percentage method (80% training and 20% testing datasets) to analyze performances.

4.1. Performance Evaluation. To accurately depict performance on the PHEME dataset, we selected Accuracy, Precision, and Recall as assessment measures for individual events. We also evaluated our model's performance using AUC. The AUC scores for the PHEME dataset's five events vary from 0.78 to 0.91 (Table 1).

Table 1. Events Wise performance evaluation

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	AUC /ROC (%)
Charlie Hebdo				
SVC	93.60	97	95	87.2
Gradient Boosting Classifier	92.33	85.0	88	84.8
Naïve Bayes	82.73	80	72	79.8
Ensemble Model	93.75	97.62	95.38	92.13
Ferguson				
SVC	93.40	98	93	89.8
Gradient Boosting Classifier	92.14	97	92	88
Naïve Bayes	77.85	81	88	75.4
Ensemble Model	93.85	98.62	95.38	92.13
Ottawa shooting				
SVC	86.92	91	84	87
Gradient Boosting Classifier	91.48	93	90	91.5
Naïve Bayes	78.76	79	77	78.8
Ensemble Model	92.41	94.02	91.44	91.98
Sydney siege				
SVC	93.50	98	94	90
Gradient Boosting Classifier	92.25	98	92	88
Naïve Bayes	78.53	83	87	75.3
Ensemble Model	94.41	98.22	94.44	91.48
German wings crash				
SVC	93.16	98	93	89.5
Gradient Boosting Classifier	92.25	97	92	88.2
Naïve Bayes	78.42	82	87	75.6
Ensemble Model	93.55	98.62	95.38	92.13

Table 1 shows the results of each unique event in PHEME investigated using different classifiers such as SVC, NB, and the gradient boosting method along with the ensemble model. The ensemble model outperforms the other classifiers for the five events, with the highest accuracy for the Sydney siege.

For the Charlie Hebdo, Ferguson Unrest events, Ottawa shooting, Sydney siege, and German wings crash, the ensemble model has an accuracy of 93.75%, 93.85%, 92.41%, 94.41%, and 93.55, respectively.

By comparing the results for four classifiers for combined overall events in the datasets in positions of performance metrics like accuracy, precision, recall, and AUC/ROC curve, the ensemble model outperformed better than other classifiers. In Table 2 the three classifiers are compared with the final dataset in terms of performance metrics, which results that the ensemble model has the highest values of accuracy is 93.59%, precision is 97.42%, recall is 94.40%, and AUC/ROC is 91.97% due to its linear function of predicting the rumour data from non-rumour data accurately.

Table 2. Performance of Different classifiers on overall events in the PHEME dataset

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	AUC /ROC (%)
SVC	93.19	96	94	91
Gradient Boosting Classifier	85.25	92	88	80.8
Naïve Bayes	77.52	80	87	77.52
Ensemble Model	93.59	97.42	94.40	91.97

5. Conclusion. In times of crisis, rumours are commonplace. Rumours grow swiftly in the virtual social environment due to the intensity and unpredictability of the issue, as well as the informational gap. The tangibility of information must thus always be called into question. This study examined a unique system for real-time rumour detection to dispel online rumours. The classification is carried out using three different supervised learning methods: SVC, Gradient Boosting, and Naïve Bayes classifiers, and are combined with an ensemble model called Long Short-Term Memory (LSTM) to overcome the drawbacks of the three classifiers. The classification takes place for each classifier by using five different events in the PHEME dataset. The performance of each classifier for five

events is calculated using several performance measures like accuracy, precision, recall, and AUC/ROC curve. The results show that the ensemble model outperformed other classifiers by comparing the performance of all events in the datasets for three different classifiers. This study may be expanded by employing various deep-learning models to increase the prediction accuracy of rumour detection. We can also increase forecast accuracy by applying efficient pre-processing processes.

References

1. Bingol H., Alatas B. Rumor Detection in Social Media using machine learning methods. 1st International Informatics and Software Engineering Conference (UBMYK). 2019. pp. 1–4.
2. He Z., Li C., Zhou F., Yang Y. Rumor detection on social media with event augmentations. Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021. pp. 2020–2024.
3. Pathak A.R., Mahajan A., Singh K., Patil A., Nair A. Analysis of techniques for rumor detection in social media. Procedia Computer Science. 2020. vol. 167. pp. 2286–2296.
4. Xuan K., Xia R. Rumor stance classification via machine learning with text, user and propagation features. International Conference on Data Mining Workshops (ICDMW). 2019. pp. 560–566.
5. Shi A., Qu Z., Jia Q., Lyu C. Rumor detection of COVID-19 pandemic on online social networks. IEEE/ACM Symposium on Edge Computing (SEC). 2020. pp. 376–381.
6. Gidwani M., Rao A. Early Identification of Rumors in Social Media: A Survey. Modern Approaches in Machine Learning & Cognitive Science: A Walkthrough. 2022. pp. 305–316. DOI: 10.1007/978-3-030-96634-8_28.
7. Hassan N.Y., Goma W.H., Khoriba G.A., Haggag M.H. Supervised learning approach for Twitter credibility detection. 13th International Conference on computer engineering and Systems (ICCES). 2018. pp. 196–201.
8. Zhao Z., Resnick P., Mei Q. Enquiring minds: Early detection of rumors in social media from enquiry posts. Proceedings of the 24th international conference on world wide web. 2015. pp. 1395–1405.
9. Zubiaga A., Liakata M., Procter R. Learning reporting dynamics during breaking news for rumour detection in social media. Computer Science. arXiv preprint arXiv:1610.07363. 2016. 20 p.
10. Kandasamy N., Murugasamy K. Detecting and filtering rumor in social media using news media event. Concurrency and Computation: Practice and Experience. 2021. vol. 33(5). DOI:10.1002/cpe.6329.
11. Kumar A., Bhatia M.P.S., Sangwan S.R. Rumour detection using deep learning and filter-wrapper feature selection in benchmark Twitter dataset. Multimedia Tools and Applications. 2022. vol. 81. no. 24. pp. 34615–34632.
12. Ahmad I., Yousaf M., Yousaf S., Ahmad M.O. Fake news detection using machine learning ensemble methods. Complexity. 2020. vol. 2020. pp. 1–11.
13. Zubiaga A., Aker A., Bontcheva K., Liakata M., Procter R. Detection and resolution of rumours in social media: A survey. ACM Computing Surveys (CSUR). 2018. vol. 51. no. 2. pp. 1–36.
14. Kotteti C.M.M., Dong X., Li N., Qian L. Fake news detection enhancement with data imputation. IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data

- Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). 2018. pp. 187–192.
15. Li Z., Zhang Q., Du X., Ma Y., Wang S. Social media rumor refutation effectiveness: Evaluation, modelling and enhancement. *Information Processing & Management*. 2021. vol. 58(1), 102420.
 16. Zubiaga A., Wong Sak Hoi G., Liakata M., Procter R. PHEME dataset of rumours and non-rumours. 2016. DOI: 10.6084/M9.FIGSHARE.4010619.V1.
 17. Karkali M., Rousseau F., Ntoulas A., Vazirgiannis M. Efficient online novelty detection in news streams. *Proceedings of International Conference on Web Information Systems Engineering*. 2013. pp. 57–71. DOI: 10.1007/978-3-642-41230-1_5.
 18. Vijayaraghavan S., Wang Y., Guo Z., Voong J., Xu W., Nasser, A., Cai J., Li L., Vuong K., Wadhwa E. Fake news detection with different models. arXiv preprint arXiv:2003.04978. 2020. DOI: 10.48550/arXiv.2003.04978.
 19. Vinothkumar S., Varadhaganapathy S., Ramalingam M., Ramkishore D., Rithik S., Tharanies K.P. Fake News Detection Using SVM Algorithm in Machine Learning. *Proceedings of International Conference on Computer Communication and Informatics (ICCCI)*. 2022. pp. 1–7. DOI: 10.1109/ICCCI54379.2022.9740886.
 20. Birunda S.S., Devi R.K. A novel score-based multi-source fake news detection using gradient boosting algorithm. *Proceedings of International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. 2021. pp. 406–414.
 21. Granik M., Mesyura V. Fake news detection using naive Bayes classifier. *Proceedings of IEEE first Ukraine conference on electrical and computer engineering (UKRCON)*. 2017. pp. 900–903.

Gidwani Manya — Research Scholar, Department of information technology, Mukesh Patel School of Technology Management and Engineering (MPSTME), Shri Vile Parle Kelavani Mandal's Narsee Monjee Institute of Management Studies (SVKM's NMIMS). Research interests: machine learning, NLP, deep learning, social media analysis. The number of publications — 7. renu.meghani@gmail.com; V.L. Mehta Road, Vile Parle, West, Maharashtra, Mumbai, India; office phone: +91(224)233-4000.

Rao Ashwini — Ph.D., Dr.Sci., Assistant professor of the department, Department of information technology, Mukesh Patel School of Technology Management and Engineering (MPSTME), Shri Vile Parle Kelavani Mandal's Narsee Monjee Institute of Management Studies (SVKM's NMIMS). Research interests: machine learning, NLP, social media analysis. The number of publications — 13. ashwini.rao@nmims.edu; V.L. Mehta Road, Vile Parle, West, Maharashtra, Mumbai, India; office phone: +91(224)233-4000.

М. ГИДВАНИ, А. РАО
**СРАВНИТЕЛЬНЫЙ АНАЛИЗ ВЫЯВЛЕНИЯ СЛУХОВ В
СОЦИАЛЬНЫХ СЕТЯХ С ИСПОЛЬЗОВАНИЕМ РАЗЛИЧНЫХ
КЛАССИФИКАТОРОВ**

Гидвани М., Рао А. Сравнительный анализ выявления слухов в социальных сетях с использованием различных классификаторов.

Аннотация. По мере того, как число пользователей социальных сетей растет, создание и распространение информации увеличиваются каждый день в огромных масштабах. Люди могут делиться своими идеями и мнениями на этих платформах. Сайт микроблогов в социальных сетях, таких как Facebook или Twitter, является предпочтительным средством обсуждения любого важного события, и информация передается немедленно. Это приводит к быстрому распространению слухов и распространению неточной информации, что вызывает у людей беспокойство. Поэтому важно оценить и подтвердить уровень достоверности такой информации. Из-за сложности текста автоматическое обнаружение слухов на ранних стадиях затруднительно. В данном исследовании используются различные методы NLP для извлечения информации из твитов, а затем применяются различные модели машинного обучения, чтобы определить, является ли информация слухом. Классификация выполняется с использованием трех классификаторов, таких как SVC (Support Vector Classifier), Gradient Boosting и классификаторы Naive Bayes для пяти различных событий из набора данных RHEME. Существуют некоторые недостатки: ограниченная обработка несбалансированных данных, трудность улавливания сложных лингвистических шаблонов, отсутствие интерпретируемости, сложности с обработкой больших пространств признаков и нечувствительность к порядку слов и контексту при использовании вышеуказанных классификаторов. Подход суммирования используется для преодоления вышеуказанных недостатков, при котором выходные данные комбинированных классификаторов представляют собой ансамбль с LSTM. Была проанализирована производительность моделей. Экспериментальные результаты показывают, что ансамблевая модель дает эффективные результаты по сравнению с другими классификаторами с точностью 93,59%.

Ключевые слова: обнаружение слухов, машинное обучение, социальные сети, SVC, градиентный бустинг, наивный байесовский классификатор.

Литература

1. Bingol H., Alatas B. Rumor Detection in Social Media using machine learning methods. 1st International Informatics and Software Engineering Conference (UBMYK). 2019. pp. 1–4.
2. He Z., Li C., Zhou F., Yang Y. Rumor detection on social media with event augmentations. Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021. pp. 2020–2024.
3. Pathak A.R., Mahajan A., Singh K., Patil A., Nair A. Analysis of techniques for rumor detection in social media. Procedia Computer Science. 2020. vol. 167. pp. 2286–2296.
4. Xuan K., Xia R. Rumor stance classification via machine learning with text, user and propagation features. International Conference on Data Mining Workshops (ICDMW). 2019. pp. 560–566.

5. Shi A., Qu Z., Jia Q., Lyu C. Rumor detection of COVID-19 pandemic on online social networks. *IEEE/ACM Symposium on Edge Computing (SEC)*. 2020. pp. 376–381.
6. Gidwani M., Rao A. Early Identification of Rumors in Social Media: A Survey. *Modern Approaches in Machine Learning & Cognitive Science: A Walkthrough*. 2022. pp. 305-316. DOI: 10.1007/978-3-030-96634-8_28.
7. Hassan N.Y., Gomaa W.H., Khoriba G.A., Haggag M.H. Supervised learning approach for Twitter credibility detection. *13th International Conference on computer engineering and Systems (ICCES)*. 2018. pp. 196–201.
8. Zhao Z., Resnick P., Mei Q. Enquiring minds: Early detection of rumors in social media from enquiry posts. *Proceedings of the 24th international conference on world wide web*. 2015. pp. 1395–1405.
9. Zubiaga A., Liakata M., Procter R. Learning reporting dynamics during breaking news for rumour detection in social media. *Computer Science*. arXiv preprint arXiv:1610.07363. 2016. 20 p.
10. Kandasamy N., Murugasamy K. Detecting and filtering rumor in social media using news media event. *Concurrency and Computation: Practice and Experience*. 2021. vol. 33(5). DOI:10.1002/cpe.6329.
11. Kumar A., Bhatia M.P.S., Sangwan S.R. Rumour detection using deep learning and filter-wrapper feature selection in benchmark Twitter dataset. *Multimedia Tools and Applications*. 2022. vol. 81. no. 24. pp. 34615–34632.
12. Ahmad I., Yousaf M., Yousaf S., Ahmad M.O. Fake news detection using machine learning ensemble methods. *Complexity*. 2020. vol. 2020. pp. 1–11.
13. Zubiaga A., Aker A., Bontcheva K., Liakata M., Procter R. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)*. 2018. vol. 51. no. 2. pp. 1–36.
14. Kotteti C.M.M., Dong X., Li N., Qian L. Fake news detection enhancement with data imputation. *IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*. 2018. pp. 187–192.
15. Li Z., Zhang Q., Du X., Ma Y., Wang S. Social media rumor refutation effectiveness: Evaluation, modelling and enhancement. *Information Processing & Management*. 2021. vol. 58(1), 102420.
16. Zubiaga A., Wong Sak Hoi G., Liakata M., Procter R. PHEME dataset of rumours and non-rumours. 2016. DOI: 10.6084/M9.FIGSHARE.4010619.V1.
17. Karkali M., Rousseau F., Ntoulas A., Vazirgiannis M. Efficient online novelty detection in news streams. *Proceedings of International Conference on Web Information Systems Engineering*. 2013. pp. 57–71. DOI: 10.1007/978-3-642-41230-1_5.
18. Vijayaraghavan S., Wang Y., Guo Z., Voong J., Xu W., Nasser, A., Cai J., Li L., Vuong K., Wadhwa E. Fake news detection with different models. arXiv preprint arXiv:2003.04978. 2020. DOI: 10.48550/arXiv.2003.04978.
19. Vinothkumar S., Varadhaganapathy S., Ramalingam M., Ramkishore D., Rithik S., Tharanies K.P. Fake News Detection Using SVM Algorithm in Machine Learning. *Proceedings of International Conference on Computer Communication and Informatics (ICCCI)*. 2022. pp. 1–7. DOI: 10.1109/ICCCI54379.2022.9740886.
20. Birunda S.S., Devi R.K. A novel score-based multi-source fake news detection using gradient boosting algorithm. *Proceedings of International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. 2021. pp. 406–414.

21. Granik M., Mesyura V. Fake news detection using naive Bayes classifier. Proceedings of IEEE first Ukraine conference on electrical and computer engineering (UKRCON). 2017. pp. 900–903.

Гидвани Манья — ученый-исследователь, кафедра информационных технологий, Школа технологического менеджмента и инженерии имени Мукеша Пателя, Институт управленческих исследований Нарси Монджи Шри Вайла Парле Келавани Мандала. Область научных интересов: машинное обучение, НЛП, глубокое обучение, анализ социальных сетей. Число научных публикаций — 7. genu.meghani@gmail.com; В.Л. Мехта-роуд, Виле Парле, Запад, Махараштра, Мумбаи, Индия; р.т.: +91(224)233-4000.

Рао Ашвини — Ph.D., Dr.Sci., доцент кафедры, кафедра информационных технологий, Школа технологического менеджмента и инженерии имени Мукеша Пателя, Институт управленческих исследований Нарси Монджи Шри Вайла Парле Келавани Мандала. Область научных интересов: машинное обучение, НЛП, анализ социальных сетей. Число научных публикаций — 13. ashwini.rao@nmims.edu; В.Л. Мехта-роуд, Виле Парле, Запад, Махараштра, Мумбаи, Индия; р.т.: +91(224)233-4000.

N.V. HUNG, T.Q. LOI, N.T. HUONG, T.T. HANG, T.T. HUONG
**AAFNDL – AN ACCURATE FAKE INFORMATION RECOGNITION
MODEL USING DEEP LEARNING FOR THE VIETNAMESE
LANGUAGE**

Nguyen Viet Hung, Thang Quang Loi, Nguyen Thi Huong, Tran Thi Thuy Hang, Truong Thu Huong. AAFNDL – An Accurate Fake Information Recognition Model Using Deep Learning for the Vietnamese Language.

Abstract. On the Internet, "fake news" is a common phenomenon that frequently disturbs society because it contains intentionally false information. The issue has been actively researched using supervised learning for automatic fake news detection. Although accuracy is increasing, it is still limited to identifying fake information through channels on social platforms. This study aims to improve the reliability of fake news detection on social networking platforms by examining news from unknown domains. Especially, information on social networks in Vietnam is difficult to detect and prevent because everyone has equal rights to use the Internet for different purposes. These individuals have access to several social media platforms. Any user can post or spread the news through online platforms. These platforms do not attempt to verify users or the content of their locations. As a result, some users try to spread fake news through these platforms to propagate against an individual, a society, an organization, or a political party. In this paper, we proposed analyzing and designing a model for fake news recognition using Deep learning (called AAFNDL). The method to do the work is: 1) first, we analyze the existing techniques such as Bidirectional Encoder Representation from Transformer (BERT); 2) we proceed to build the model for evaluation; and finally, 3) we approach some Modern techniques to apply to the model, such as the Deep Learning technique, classifier technique and so on to classify fake information. Experiments show that our method can improve by up to 8.72% compared to other methods.

Keywords: social networking, computational modeling, deep learning, feature extraction, classification algorithms, fake news, BERT, TF-IDF, PhoBERT.

1. Introduction. Nowadays, broadcasting fake news online has become standard on Social Networks [1, 2], and more information, opinions, and topics can happen worldwide [3]. Fake news has a huge impact. Detecting fake news is a critical step. Using machine learning techniques to see fake news employs three popular methods: Naive Bayes [4, 5], Neural Network [6, 7], and Support Vector Machine [8 – 10]. Normalization is essential in cleaning data before using machine learning to classify it [11].

Moreover, the analysis of fake news and information distortion detection algorithms is becoming popular [12, 13]; several methods of detecting fake news in Russia have also been proposed, such as using artificial intelligence [14] and machine learning [15].

In [16], Fake news is information that is false or misleading and is presented as news. Fake news is frequently intended to harm a person's or entity's reputation or to profit from advertising revenue. Nonetheless, the word has no fixed definition and has applied to false information. Public

figures also use it to refer to any negative news. Furthermore, disinformation is the dissemination of incorrect information with malicious intent, and it is sometimes generated and spread by hostile foreign actors, particularly during elections. Some definitions of fake news include satirical articles that are misinterpreted as genuine and articles that use deception. Figure 1 describes the process of identifying fake information.



Fig. 1. Procedures for receiving and handling fake information

We have performed the analysis and divided it into three steps; below are our implementation steps:

– **Step 01 – Received Information:** First, the information is gathered using various techniques from social networking sites like Twitter and

Facebook, as well as breaking news from CNN, BBC, or online publications. Then, this data will be classified as news content, social content, and outside knowledge.

– **Step 02 – Assessment:** Following classification, the data will be compared to standard datasets supplied by individuals or organizations to verify the correctness of the news. In the past, comparison and evaluation experts handled this work. Hence censorship groups frequently needed a lot of staff, time, and effort. However, these tasks have been mechanized by algorithms that improve comparison, contrast, and evaluation under the heavy weight of big data.

– **Step 03 – Disclosure:** Finally, the data is split and labeled as fake news, false news, and factual information.

In [17], Fake news was well-known in politics when it harmed the field. The election of Donald Trump as president has generated a lot of controversy due to false information regarding the number of votes cast in his favor. However, in the last two years [18], as the Covid-19 pandemic has become a severe problem in many nations, distance has made it easier for people to access unconventional information. For example, there is much false information about vaccines, and media campaigns to stop the spread of Covid-19 have destroyed numerous health systems. Price information also encourages people to hoard food, which contributes to inflation. The economy, health, and particularly human health have all negatives impacted by false information. We must identify and remove fake news from media outlets to combat it.

In the past [19], when looking for false news, individuals checked it manually by submitting it to professionals who would screen it; however, this requires a lot of time and money. Therefore automatic fake news search engines are now regarded as fake news, a current efficient fix. Machine learning and deep learning algorithms are a couple of them. In [20], these two AI algorithms frequently are utilized since more modern AI algorithms have been developed that better solve categorization challenges (natural text classification, voice classification, image classification, etc.) Additionally, as technology becomes more productive and affordable and as the availability of standard datasets rises, it becomes easier to assess the accuracy of false news detection models.

Although numerous datasets are available, you must use them correctly with your strategy. "Granik and colleagues Fake news detection using naive Bayes classifier" was published in 2017. However, because he used the dataset 4.9%, which is fake news, his accuracy is only 74% [20]. Compared to the entire surface of identifying fake news, it is a low number. In this case, four Kaggle datasets were used to accomplish this. These datasets are appropriate for the method.

To address readers' current needs for the most reliable information among the abundance of information on social networks, we offer a deep learning aggregation model for detecting fake news based on deep learning and machine learning algorithms with high accuracy of up to 99% in this paper. We recommend doing the following:

- We analyze the existing techniques, such as BERT;
- We proceed to build a model to evaluate to classify fake information;
- We approach some Modern techniques to apply to the model through techniques, such as Deep Learning techniques, classification techniques, etc.

This paper is organized as follows: Section 2 discusses the related work. Section 3 presents the concepts and features of identifying fake news on social networks. Section 4 describes the suggested viewport estimation technique. Section 5 contains the performance assessment. Section 6 concludes with a discussion of our conclusions and open questions.

2. Related work. Because of the increased internet use, it is much easier to spread fake news. Many people are constantly connected to the internet and social media platforms. There are no restrictions when it comes to posting notices on these platforms. Some people take advantage of these platforms and begin spreading false information about individuals or organizations. This can ruin an individual's reputation or harm a business. Fake news can also sway people's opinions about a political party. There is a need for a method to detect fake news. A new study [21] has shown that machine learning classifiers are used for various purposes, including detecting fake news. The classifiers are listed first. The classifier trainers use a data set known as the training data set. Following that, these classifiers can detect fake news automatically.

Fake news and hoaxes have been around since before the Internet. Many clickbait use flashy titles or designs to entice users to click on links to increase ad revenue. In article [22], the author examined the prevalence of fake news in light of the media advances brought about by the rise of social networking sites. In this article, the author has developed a solution that users can use to detect and filter out websites containing false and misleading information.

In [23], supervised methods have yielded encouraging results. However, they have one significant limitation: they require a reliably labeled dataset to train the model, which is frequently complex, time-consuming, expensive to obtain, or unavailable due to privacy or data access constraints. Worse, because of the dynamic nature of news, this limitation is exacerbated under the setting, as annotated information may quickly become outdated and cannot represent news articles on newly emerging events. As a result, some researchers investigate weakly supervised or unsupervised methods for detecting fake news.

Studies [24, 27] have shown that Online social media networks have developed into a powerful platform for people to access, consume, and share fake news. Additionally, this results in the widespread dissemination of fake news or purposefully false or misleading information. The models must perform better for news in unexplored fields (domains) due to domain bias, which remains a significant obstacle for practical application even though accuracy is improving. As a result, numerous reports are shared, such as [24], focusing on analyzing the various traits and varieties of fake news and suggesting an efficient solution to detect it in online social media networks. This model, however, also deals with data that falls under the purview of the Online Social Media model. On the other hand, in research [25], the author has concentrated on examining data sources, particularly those that always include pairs of false and true news about the same topic. The author also relies on that to assess the accuracy and provide a dataset for concatenation useful for cross-domain detection. By examining the connection between domain news and its news environment, the author, like the method [26], focuses on developing a framework for comprehending the historical news environment. In all earlier posts, the author has cited history and the current state of the mainstream media. The author also creates a model to identify fake news by representing perceptions through domain gates. The outcomes are also good, but due to the anti-face change, this method is still only somewhat predictive if the user consciously improves; changes the history. To accomplish this, we discover that the method [27] the authors have chosen to emphasize in the suggested research, has compared various supervised machine learning models to categorize fake news (hoax news) with reliability. Although the author has suggested the K nearest neighbor model from there to classify the sample and improve the quality of service like the advertisement the author mentioned, this is similar to the method [26].

The real issue with social media is that anyone can post or share anything, occasionally leading to issues if the shared information needs to be verified. For many recent studies, this is also a challenge before sharing. For instance, [26] shows how the author used the skill by using news from Facebook, Instagram, and other social networks. The author has improved accuracy by using the random forest to enhance the quality. The method we have extended by designing a model through the data model is rigorously testbed, and the data is analyzed, in contrast to the methods mentioned above.

Most research has concentrated on detecting fake news in a specific language. Much information, however, is disseminated not only among native English speakers but also among speakers of other languages from other cultures. It raises an important question about the applicability of current

methods for detecting fake news [23]. An extensive multilingual news database is required to train a multilingual fake news detection model. To the best of our knowledge, there are very few datasets for multilingual rumor detection. The PHEME dataset includes tweets in both German and English [29]. COVID-19 news in both English and Chinese is included in multilingual COVID-19 [18], whereas fake Covid [30] includes COVID-19 news in 40 languages. COVID-19 news is available in six languages on mm-COVID [31]. While the datasets available can assist scholars with multilingual fake news research, they could be more extensive in terms of the number of languages and data they contain.

Due to the benefits of AI algorithms, numerous researchers have used these algorithms. In 2019, in study [32] the authors used machine learning to compare three Nave Bayes algorithm classifiers, Support Vector Machine, and Logistic Regression to categorize fake news. Therein, the Nave Bayes algorithm classifier had the highest accuracy result of 83%. The HC-CB-3 approach [33], which the authors proposed in 2018, was deployed in the Facebook Messenger chatbot and verified using a real-world application, reaching 81.7% accuracy in the detection of fake news. By utilizing the binary classification method [20] in 2020, the authors could identify fake news with an accuracy of up to 93%. In the same year, in [34] the authors developed a method known as Bi-Directional Graph Convolutional Networks (Bi-GCN), which can process large amounts of data quickly and efficiently while yet maintaining accuracy to assist in the study of the propagation of rumors. In 2022, a new author suggested using the TF-IDF algorithm and a random forest classifier, but the results showed 72.8% of accuracy [35].

Additionally, numerous studies have taken an interest in recent studies developed and proposed in Vietnam, such as Bidirectional Encoder Representation from Transformer (BERT) [36]. In [36], the author uses deep learning and natural language processing to base a question on an answer. The author attempts to apply both language-specific BERT models and multilingual BERT models for the Vietnamese language, including DeepPavlov multilingual BERT and multilingual BERT refined on XQuAD (PhoBERT). The analysis in this direction, though, ends at the level of the representative model. In [37], the BERT and Hybrid fastText-BILSTM models are also improved for the rather large data set of customer reviews. However, the approach of this method clearly shows that the BERT model is superior to Deep Learning. Recently, the K-BERT model has been suggested to support language representation knowledge in specialized fields [38]. Using the knowledge graph's topic model to infer the topic for the input sentence, the author has examined and approved the model for segmenting the knowledge graph by topic. Our approach, however, relies on the BERT technique and then

normalizes the data using context analysis based on the TF-IDF evaluation model to prevent the problem of incorrect input data.

In this research, we present a deep-learning aggregation model that employs the TfidfVectorizer algorithm to train and test the data, matching and transforming the training set in practice and altering the test set to increase the accuracy in detecting fake news combined with computation economics of the current BERT scheme. Additionally, Indonesia's recent development is also making substantial progress [39]. However, the BERT model still has unresolved challenges, such as sentiment analysis, text classification, and summarization. In this report, we used the BERT and TF-IDT models to analyze and evaluate fake information. Experimental findings demonstrate our model's great effectiveness, with false news detection accuracy up to 99% higher than that of the V3MFND, TF_RFCFV, and FNED models.

3. Theory background

3.1. Definition of fake news. Defining fake news has become complicated since before 2016; it only referred to satirical and humorous news [40]. After a period of complex changes with different meanings in production, the press and the people's government are threatened [41]. Since then, fake news has become a buzzword on social networks [42]. Through the analysis of the authors [40 – 42]. We define fake news as a type of information that is inaccurate, or in other words, false with the primary information (accurate information). They can be misleading, incorrect, or intentionally created to deceive the public into attracting attention or increasing specific personal or collective interests.

In Vietnam [43 – 45], much information, including accurate and incorrect information, is transmitted throughout the country to deceive appropriate property. Many authors have proposed and built models for that information, such as [43] predictive models to transfer knowledge from one data set to another without entity or relational matching. Besides, the author in [44] carefully recommends counterfeit practices that rely on covid 19 to take advantage of it to benefit individuals and organizations. This news is never proper to reality and is given to deceive and create misunderstandings about a particular issue or event.

Unfortunately, fake news is now not only spread by word of mouth from one person to another, but through media effects, and social networks, it spreads at breakneck speed. Because it is fabricated news, it is exaggerated, so it contains thrilling, attractive, easy-to-hit emotions and the psychology of people with high "expectations".

Fake news not only wins over the curiosity of readers, but it also weakens the media. Fake news misdirects a part of society and "guides" some reporters

and press agencies - unverified information from individuals on Facebook, Zalo, etc. But there are online newspapers that still "quickly" turn into journalistic products.

Our research shows that not only in Vietnam but also in other countries, it is pretty common to identify fake information or human behavior can be classified into three main categories as follows in Figure 2:

- The group that reluctantly publishes negative information always finds bad points or distorts information.
- The group of people who need to be fully informed but rely on their limited knowledge to give false information.
- The group has no data but wants to get views and badmouths, so they are ready to spread unverified information.

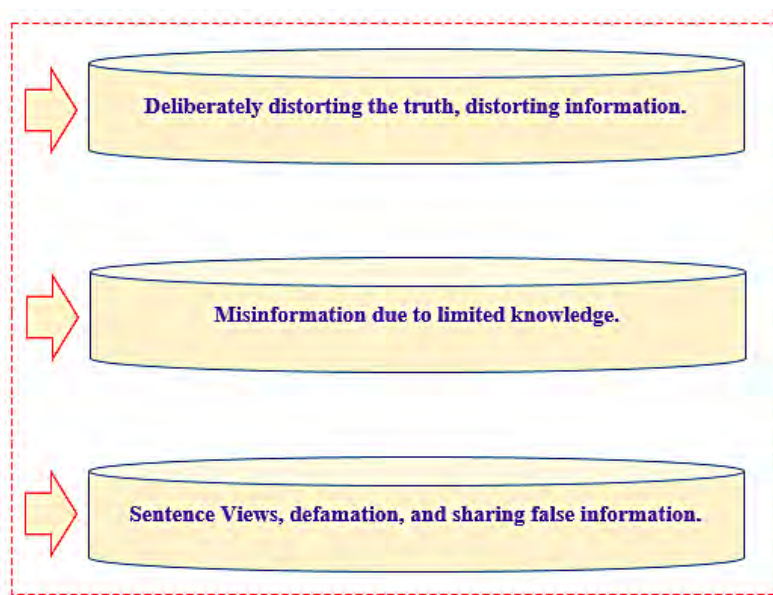


Fig. 2. Three groups of people spread fake news famous in Vietnam

3.2. Some features of recognizing fake news on social networks.

There are three fundamental characteristics to detect fake news on social networks: User, Posts, and Network.

- **User:** Fake news can be created and spread from malicious accounts on social networking sites. User features represent how those users interact with information on social media. The characteristics of social networks can

be divided into different levels: individual level and group level, in which the personal level includes relevant information such as Age, number of followers, number of posts, etc. At the group level, users will know information related to news of the posts that the user posts in the group.

– **Post:** People express their opinions or feelings through social media posts such as Feedback, sensational reactions, etc. Therefore, extracting Post features helps to find news stories and fake news through public posts. Feature Post relies on user information validation to infer authenticity from multiple aspects related to social media posts. We can extract Post information to detect fake news on social networks. These features are divided into three levels: 1) Usually, social media posts. Each post usually has characteristics such as Opinion, topic, and credibility - Post Level, 2) All relevant posts, specifically the one using “Wisdom of Crowds”, which means “Wisdom of the crowd”. For example, average confidence scores are used to assess news reliability - Group Level, and 3) Recurrent Neural Network (RNN) is used to determine when to post on social networks to attract posts that change over time. Based on the shape of this time series for different metrics of related posts (e.g., Number of Posts), mathematical features can be calculated, such as Parameters by time – Temporal Level.

– **Network:** Users use social networks to connect members with similar interests, topics, and relationships with each other – network-based extracts particular structures from users who post public posts on social networks. Network-based is built in different styles.

– **Stance Network:** Built by visible nodes for all news-related posts, the edges represent the weight of the Stance Network similarity.

– **Co-occurrence Network:** Built on user interaction by counting whether those users have posts related to the same article.

– **Friendship Network:** Indicate whether users follow or not follow related posts.

Based on the features of detecting fake news on social networks presented above. In this study, the authors use Post features to identify the information to be verified.

4. AAFNDL. In this section, we discuss some of the issues of fake news – definitions, components, types, and features of disinformation. We detail our expert model, which does some of the following work to identify forgery information printed.

4.1. Problem Formulation. Fake news has become a global issue that must be addressed immediately [46]. Defines fake news as misleading content such as conspiracy theories, rumors, clickbait, fabricated news, and satire [46]. According to reference [47], fake news is defined as misinformation

and disinformation, including false and forged information, that is spread to mislead people or fulfill propaganda.

In reality, there are several types of fake news. For example, we can take the form of a stance, satire, multi-modal, deep fake, or disinformation. There are four types of perspectives: agree, disagree, discuss, and unrelated [48]. Each concurrence is similar to the information in the fake news headline. In addition, the point of disagreement contains conflicting information.

Therefore, properly evaluating fake information to combat fake news is a big challenge. And from there, we can build a system to combat misinformation or phony information on today's social networking platforms. Most studies evaluate using English, Spanish, and Portuguese [49]. We find that English is the most commonly used language today. We recognize that the style of fake news and how it is written can also vary from country to country, so a dataset from a country that speaks that language would be a good contribution, rather than translating existing datasets into other languages.

Furthermore, we have tried more than two examples above, and the results show that our system performs reliably. With trained documents, the system always ensures high accuracy. In addition, we also have tried a few examples in addition to the training document the results are also good.

However, before processing for inclusion in the system, we have added a step before putting it into the system, that is, to process the actual data on social networking sites (the text is too long, the grammar is incorrect, or misspellings and so forth). We call this phase the "Text summarization system", but it doesn't change the meaning of the entire text. Our system will be faster due to the shorter sentence structure.

In Figure 3, we use a Text summarization that has become an essential and helpful tool for supporting and extracting textual information in today's rapidly evolving information age. Therefore, in this section, we propose a system of "Text summarization system" in three main stages as follows:

- **Analysis:** Analyze the input text to provide descriptions, including information used to search and evaluate necessary corpus units and input parameters for the summary.

- **Transformation:** The selection of extracted information is transformed to simplify and unify; as a result, corpus units have been summarized.

- **Synthetic:** From the summarized corpus, create a new text containing the primary and essential points of the original text.

Extraction plays a significant role in detecting fake information and word processing-related problems. The extraction method is built by extracting necessary textual units (sentences or paragraphs) from the original text based

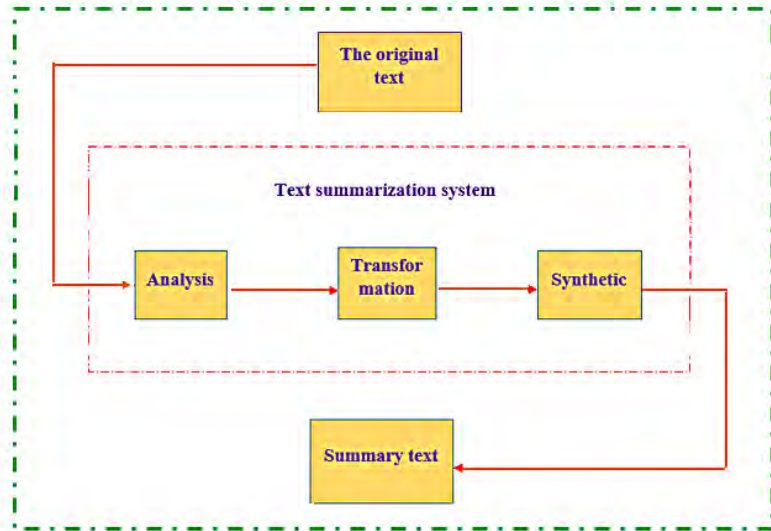


Fig. 3. Stages of the text summarization system

on analysis of words/phrases, frequencies, locations, or suggested words to determine the units' importance and extract the actual units from there as a summary. We can see it in Figure 3. Text transformation is how we use statistical and graphing algorithms to represent it. Calculate the weight of the sentence importance and select a subset of the original text to become the summary text and represent it as natural language processing.

4.2. Design and problem solve. Based on the inadequacy and explosion of information technology, we found that many studies have used tools to detect bots but have yet to detect other bots because of the constant change of the bot feature. It is hard to meet many requirements for online detection. Therefore, in this section, we analyze and build a method to detect many separate types of bots instead of one. According to [50], the authors devised an unsupervised technique for automatically clustering similar bot accounts based on a dataset and then assigning homogeneous accounts to specialized bot classifiers.

In this section, we propose a model to identify fake information. We analyze to detect phony information and use programming techniques to build an artificial information detection model. The proposed model uses neural network architecture to predict fake news in Figure 4.

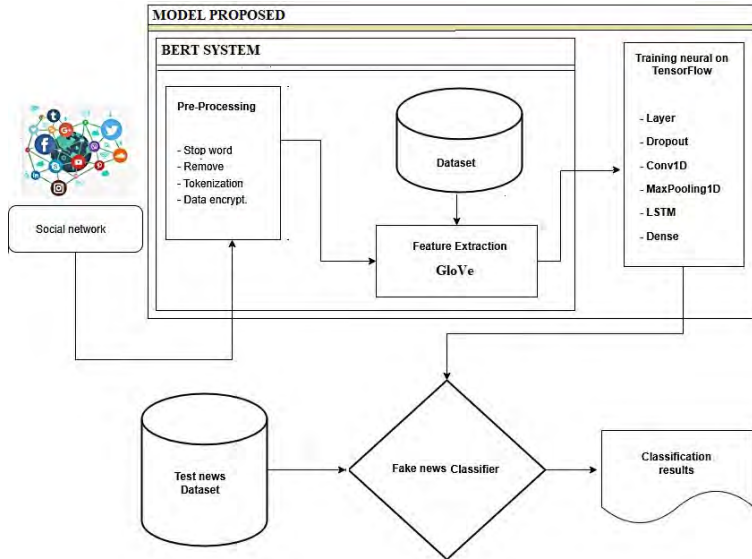


Fig. 4. AAFNDL algorithm model evaluates fake information

The model is detailed as follows:

Step 01. Data collection: The proposed approach will input the dataset from a Social network to include in the system. The data is then fed into the BERT system, a variety of techniques analyzed by many researchers such as **Word2vec** [51], and **FastText** [43]. However, in this report, we use **word2vec** for analysis in the BERT system shown in Figure 4. This system will analyze the contents of the word information based on the analyzed content, according to GloVe. We call it the benchmark license for evaluation. This is very important because it directly affects the later analysis. For example, the battery in the computer, if it is said that the battery runs fast, it is not good because the battery runs out quickly. Therefore, the BERT technique will find a vector representing each word based on a large corpus, so it cannot describe the diversity of contexts. This creation shares his direction toward the accuracy of sentences in Vietnamese. In [3], the author also shared the opinions and views of the comments. Therefore, creating a representation of each word based on the other words in the sentence yields much more meaningful results. In this step, our method will focus on data processing, and this step will be analyzed through techniques and combined with modern methods such as BERT to process the original data quickly.

In summary, at this step, we do the following two tasks:

- **First**, test data is collected from the content of articles from pages and groups of social networking sites. Collect stories with many views, comments, and shares for this data.

- **Second**, we collect real-life data sets collected from trusted websites, and the data are described in detail in the performance evaluation sections.

Moreover, we analyze using the Hidden Language Model (HLM) in this step to enable two-way learning from the text. To achieve this, we can conceal a word in a sentence and make BERT use two-way words on both sides of the sentence. To predict the hidden word, we can attempt to comprehend the words that come before and after it. By examining the two-dimensional words that come after and before the hidden text, we can quickly guess the missing word because it gives us context cues. One example set a sentence in Figure 5: "What are you doing?" We can predict and calculate the probability of this sentence.

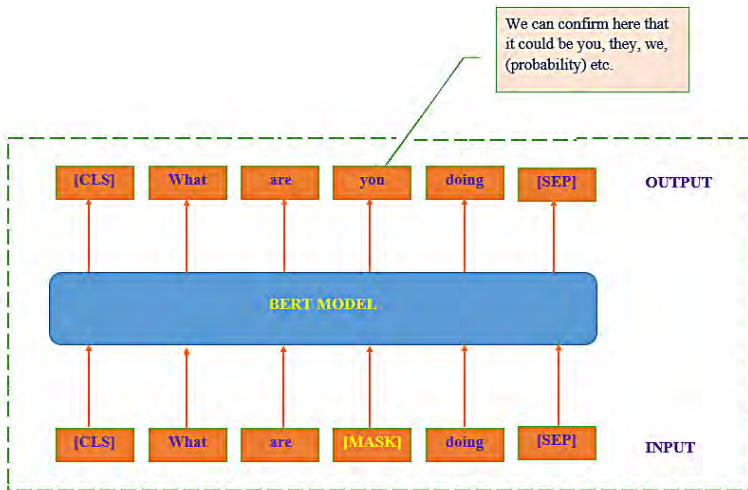


Fig. 5. BERT example uses two-way words on both sides

Step 02. Data preprocessing: The data will be analyzed and preprocessed before entering the system. We use several analytical techniques, such as word separation, unnecessary word removal, labeling, and data encryption. Furthermore, we also use the BERT [36, 39, 52, 53] feature to BERT extends the capabilities of previous methods by generating contextual

representations based on words first and then leading to a language model with richer semantics.

After collecting data from news websites and social media stories according to a particular structure, data preprocessing is performed. First, convert the data to its correct form and apply word separation measures to separate the document's content into corresponding words and phrases, remove redundant characters in Vietnamese, and keep only words mean. The result of the preprocessing stage is the index vector for each text document. The preprocessing steps are performed in Figure 6.

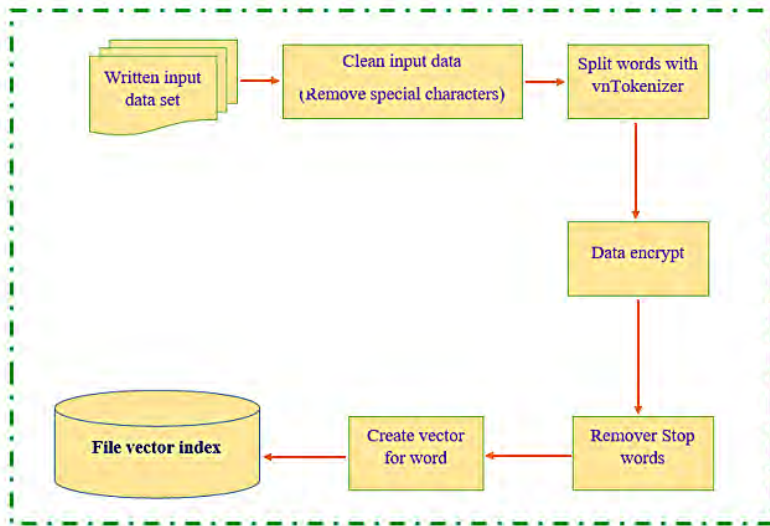


Fig. 6. Our Data Preprocessing Diagram

Step 03. Extract information: Extract the essential information from a text to create a concise version that still contains enough of the core information of the original text with the requirement to ensure grammatical and spelling correctness.

We find that the BERT model has shown superiority and responsiveness to the processing process. However, the current techniques could be more extensive in expressing the capabilities of representative vector models, especially the fine-tuning approach. The main limitation here is that language models are built based on a one-dimensional context, which limits the choice of architectural model to be used during pre-training. In OpenAI GPT [53],

for example, the authors use a left-to-right architecture, meaning the tokens depend only on the previous tokens.

Furthermore, we can see that Figure 7 is a small data collection module that performs normalization with the following specific functions:

- **Step 01:** Read news from data contextually analyzed by BERT;
- **Step 02:** Extract information, select information and remove inappropriate information;
- **Step 03:** Save information and system for proof;
- **Step 04:** Process the TF-IDF index for news data;
- **Step 05:** Build an inverse index for news for information search;
- **Step 06:** Finish.

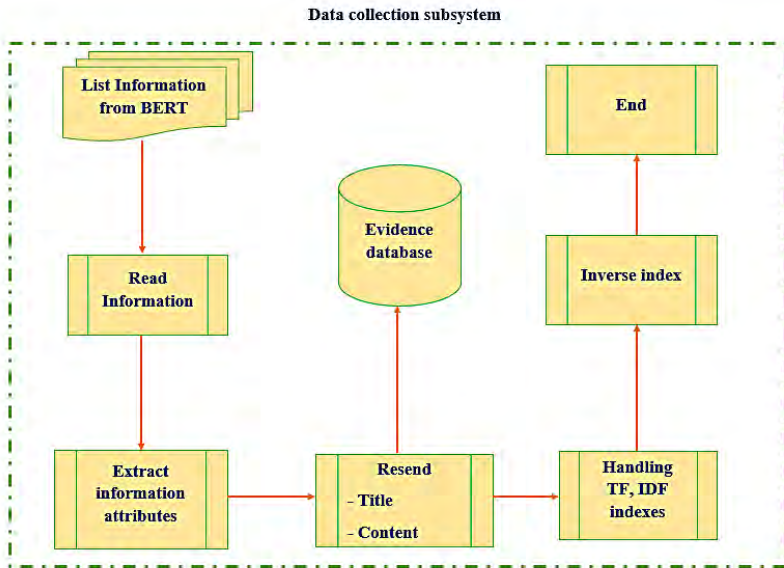


Fig. 7. Data collection subsystem

In general, this system with the idea is to build an accurate data set as evidence to deal with fraudulent and fake acts of users. In the future, we will continue to update more information, hoping that the system will automatically update and put more data into the system.

Step 04 Identify features: The study uses TF-IDF to pinpoint the characteristics of the text’s content. The most well-known statistical method for assessing the significance of a word in a text paragraph within a collection

of various text fragments is TF-IDF (Term Frequency - Inverse Document Frequency). It is frequently employed in text data mining as a weight. Text representation is transformed into vector space by TF-IDF. Therefore, we have used additional TFIDF [51, 54] features by calculating the TF by counting the keywords on the input data. However, we consider the number of keywords (for keywords that are too large, i.e., more than nine times, we consider correlation rather than using keywords). This will be more difficult because the input data processes at the basic levels. Therefore, here we have further analyzed using the following logarithmic (\log) function to calculate TF :

$$TF = \frac{1 + \log(\text{Keyword Count})}{\log(\text{Word Count})}. \quad (1)$$

This parameter Term Frequency (TF) reflects whether we use a keyword too often or too rarely. Sometimes this value does not affect a positive because we may need to measure the importance of a phrase, not just the frequency in terms of how many times it uses, but be it a preposition, pronouns, conjunctions, e.t. Therefore, to avoid that, we need the Inverse Document Frequency (IDF) index, give by:

$$IDF = \log\left(1 + \frac{\text{Total datasets}}{\text{Datasets with keyword}}\right). \quad (2)$$

The IDF formula is equivalent to the TF procedure. A linear lower function will more accurately reflect the value in situations where phrases with high IDF scores because a linear IDF function pushes a document's score too high. When phrases with high IDF scores (possibly uncommon words, misspelled terms, etc.) are in the field, the Linear IDF , like TF , can raise the document's score excessively.

Therefore, the $TF - IDF$ is a comprehensive metric, unlike the keyword density measure, which only reflects the degree of "cramming" a specific keyword into the text. Furthermore, it helps to lessen the prominence of meaningless words and phrases while elevating the importance of meaningful and uncommon terms.

Step 05 Output parameter setting: After performing the IF-IDF analysis, we conduct the analysis based on the classes trained through Layer, Dropout, Conv1D, MaxPooling1D, LSTM, and Dense layers for classification. Here we perform the assessment and divide it into two categories. If the result

is 1, it is fake news; if the result is 0, it is not. The experimental results that we define in formula 3 are the ones we tested:

$$\text{Classification results} = \begin{cases} 1 & \text{fake news} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Finally, if we want to test with our model, we can input data from the test set, and the system, through the trained data, will evaluate whether the information is fake or not.

5. Performance Evaluation. This section will detail the installation process and present some assumptions for detecting fake information through the performance evaluation process.

5.1. Experimental Settings. This section will discuss standard disinformation detection methods used in our review. In this study, we focus on processes that take only previously forged information as input. Our future work will include more advanced methods.

In particular, we evaluate three Fake News on ASUS Rog Strix G15 G513IC Laptop, with Chip: Ryzen 7- 4800H, Ram: 16 GB, and Graphics: RTX 3050 4GB methods: A comprehensive multi-domain multimodal model for identifying fake news in Vietnamese (we called V3MFND), Term Frequency - Resource Frequency combination to detect fake Vietnamese (we called TF_RFCFV), and A Deep Network for Social Media Fake News Early Detection (we called FNED). The following are the specifics of each method:

- **V3MFND [44]:** The author of this article also employs the multi-domain deep multimodal fake news detection model for Vietnamese, also known as v3MFND. Based on the evaluation of the function of each method in the multimodal model, the tests are expanded on the actual data set and demonstrate the performance of multi-domain, multimodal fake news detection for Vietnamese people.

- **TF_RFCFV [45]:** This model uses the PhoBERT [55] pre-trained language model and the Term Frequency - Resource Frequency combination to detect fake Vietnamese news on social networking sites. For word embedding and extraction for tamper detection, inverse data (TF-IDF) and convolutional neural networks (CNN) were used.

- **FNED [56]:** This method suggests a new deep neural network for early fake news detection. The technique is based on state-sensitive crowd-response feature extraction, which takes the user's text feedback and the corresponding user profile and extracts text and user features. The average aggregation mechanism, approach, and location-aware attention mechanism

highlight the importance of user feedback at specific ranking positions. To carry out feature aggregation across multiple regions using different window sizes.

In general, the methods are evaluated to be relatively stable. The advantages of every technique are demonstrated. Regarding our approach, we draw on some of the used strategies, like using tools, machine learning, and deep learning to improve our algorithms using the abovementioned methods. However, our system was created based on neural network architecture to foretell fake news.

5.2. Performance evaluation. This section uses three datasets to evaluate the methods referenced, including fake_or_real_news, news, and WELFake_Dataset in [57].

To experiment with the evaluations between the AAFNDL method and the reference methods, we divide and arrange the data sets according to 8:2, which means that 80% of the dataset uses for training and the remaining 20% uses tests. The analyzed data evaluates as follows: for the data file "fake_or_real_news" in Figure 8, "news" in Figure 9, and "WELFake_Dataset" in Figure 10. In general, our method outperforms the mentioned methods. Studies reveal that, in contrast to the FNED method, which consistently achieves results of 92%, our method achieves over 99% in Table 1.

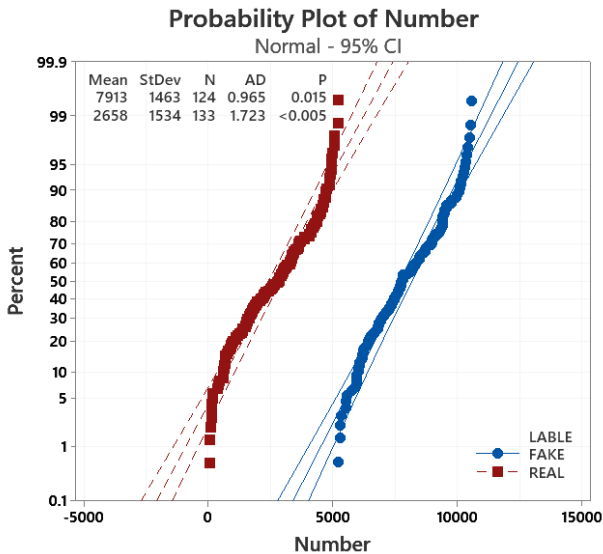


Fig. 8. Probability Plot of "fake_or_real_news" with Confidence Interval 95%

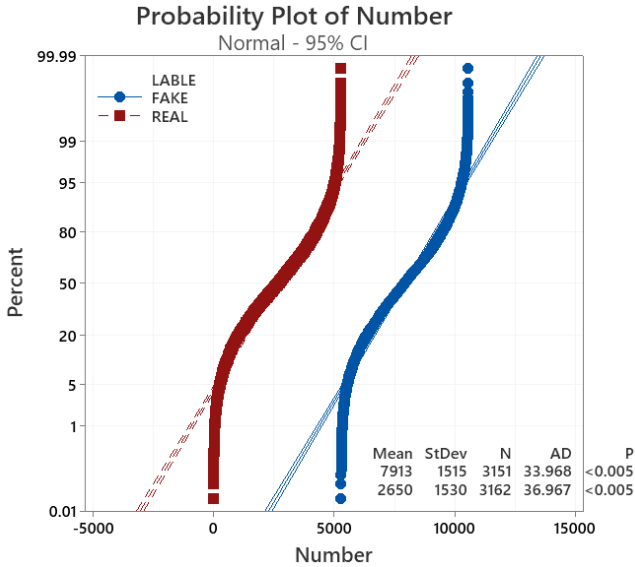


Fig. 9. Probability Plot of "news" with Confidence Interval 95%

Table 1. Findings from comparing the AAFNDL method to the referenced methods

STT	Methods	Value	Dataset names		
			fake_or_real_ news	news	WELFake_ Dataset
1	V3MFND	Accuracy	98.44	98.19	98.81
		F1-Score	92.72	92.79	95.97
2	TF_RFCFV	Accuracy	95.83	93.13	95.83
		F1-Score	82.91	82.91	90.43
3	FNED	Accuracy	90.76	90.76	91.99
		F1-Score	68.47	68.27	77.13
4	AAFNDL	Accuracy	99.22	99.48	99.70
		F1-Score	95.97	94.36	95.97

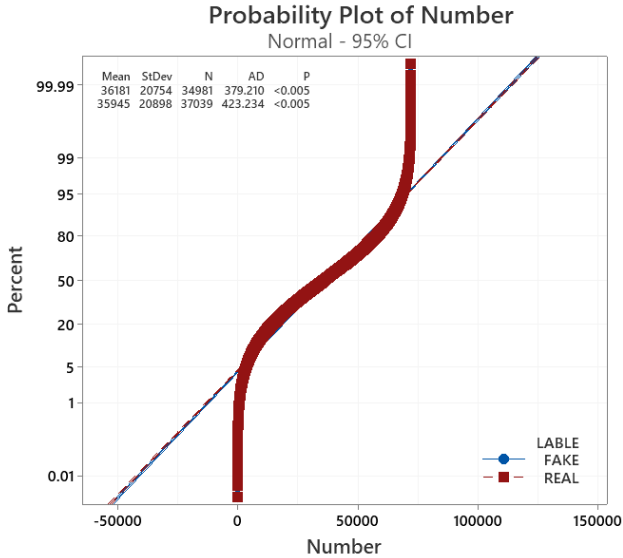


Fig. 10. Probability Plot of "WELFake_Dataset" with Confidence Interval 95%

To evaluate the performance of the proposed method with the evaluation method, we make two measurements, Accuracy, and F1-score from [58, 59], using the following formulas:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (4)$$

$$Precision = \frac{TP}{TP + FP}, \quad (5)$$

$$Recall = \frac{TP}{TP + FN}, \quad (6)$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}, \quad (7)$$

where:

- TP: The model predicts 1 while actually, it is 1;
- TN: The model predicts 0 while actually, it is 0;
- FN: The model predicts 0, but the truth is 1;
- FP: The model predicts 1, but the truth is 0.

On the one hand, for each dataset, the methods all show relatively stable results from the WELFake_Dataset to fake_or_real_news and news. For the dataset WELFake_Dataset, the techniques have the best results: V3MFND of 98.81%, TF_RFCFV of 95.83%, and FNED of the lowest at 91.99%, and our method achieves the best rate up to 99.70%. For the dataset fake_or_real_news, the methods V3MFND, TF_RFCFV, and FNED are 98.44%, 95.83%, 90.76%, and 99.22%, respectively. The remaining dataset is 98.19% with V3MFND, 93.13% with TF_RFCFV, 90.76% with FNED, and ours is 99.48%.

Moreover, the experiment also shows that the results of the F1-Score of AAFNDL are always better than those of the reference methods; the results can improve up to 3.25%, 13.06%, and 27.50% for the referenced procedures projections are V3MFND, TF_RFCFV, and FNED, respectively.

On the other hand, considering each method, the data is different. However, the experiment shows that the oscillometric methods at the set level, such as the V3MFND method (from 98.19 to 98.81) and the proposed method (from 99.22 to 99.70), fluctuate by no more than 1%, besides two of the remaining techniques ranged no more than 3% with TF_RFCFV from 93.13 to 95.83, and with FNED from 90.76 to 91.99 not more than 2%.

6. Conclusions. This paper presents the recent techniques that BERT widely uses along with the evaluation as mentioned above criteria TF-IDF. We are applying assessment techniques to analyze fake information in Vietnam and the limited problems in Vietnam and the general world. Second, we have built and analyzed an evaluation model based on the combination of two criteria of BERT and TF-IDF. Besides, we also have used Deep learning techniques to classify fake information on social networking sites like Facebook, Zalo, etc.

Also, this paper has looked at four methods for identifying false information. Our analysis also demonstrates that, when there is a method that can detect fake news up to 98.81%, the estimated performance of the current techniques is relatively good. However, experiments have shown that our method is better than the referenced methods when the proposed model and analysis technique is used. In the future, our research will enhance current approaches by enlarging the problem with more data.

References

1. Mladenova T., Valova I. Research on the ability to detect fake news in users of social networks. *International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. 2022. pp. 01–04.
2. Apuke O.D., Omar B. Fake news and covid-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*. 2021. vol. 56. p. 101475. Available at: <https://www.sciencedirect.com/science/article/pii/S0736585320301349>.
3. Nguyen H., Tan N., Quan N., Huong T., Phat N. Building a chatbot system to analyze opinions of english comments. *Informatics and Automation*. 2023. vol. 22. no. 2. pp. 289–315.
4. Yuslee N.S., Abdullah N.A.S. Fake news detection using naive bayes. *IEEE 11th International Conference on System Engineering and Technology (ICSET)*. 2021. pp. 112–117.
5. Babar A., Jagtap N., Mithari A., Shukla A., Chaudhari P. A survey on fake news detection techniques and using a blockchain based system to combat fake news. *International Journal of Computer Applications*. 2020. vol. 176. no. 27. pp. 47–53.
6. Kaliyar R.K. Fake news detection using a deep neural network. *4th International Conference on Computing Communication and Automation (ICCCA)*. 2018. pp. 1–7.
7. Sastrawan I.K., Bayupati I., Arsa D.M.S. Detection of fake news using deep learning cnn–rnn based methods. *ICT Express*. 2022. vol. 8. no. 3. pp. 396–408.
8. Vinothkumar S., Varadhaganapathy S., Ramalingam M., Ramkishore D., Rithik S., Tharanies K. Fake news detection using svm algorithm in machine learning,” in *2022 International Conference on Computer Communication and Informatics (ICCCI)*. 2022. pp. 1–7.
9. Hussain M.G., Hasan M.R., Rahman M., Protim J., Hasan S.A. Detection of bangla fake news using mnb and svm classifier. 2020. 5 p. DOI: 10.1109/iCCECE49321.2020.9231167.
10. Hussain M.G., Hasan M.R., Rahman M., Protim J., Al Hasan S. Detection of bangla fake news using mnb and svm classifier. *International Conference on Computing, Electronics Communications Engineering (iCCECE)*. 2020. pp. 81–85.
11. Aphiwongsophon S., Chongstitvatana P. Detecting fake news with machine learning method. 2018. pp. 528–531.
12. Mailjan Je.K., Kulikov A.A. [Analysis of fake news detection algorithms] *Vserossijskaja konferencija molodyh issledovatelej s mezhdunarodnym uchastiem «Social’no-gumanitarnye problemy obrazovanija i professional’noj samorealizacii «Social’nyj inzhener-2020» [All-Russian Conference of Young Researchers with International Participation “Social and Humanitarian Problems of Education and Professional Self-Realization “Social Engineer-2020”]*. 2020. pp. 204–209.
13. Vasil’kova V.V., Sadchikov D.I. [Fakes and bots as mechanisms of information distortion in social networks]. *Kazanskij social’no-gumanitarnyj vestnik – Kazan Social and Humanitarian Bulletin*. 2019. no. 2(37). pp. 24–30.
14. Tret’jakov A.O., Filatova O.G., Zhuk D.V., Gorlushkina N.N., Puchkovskaja A.A. [A method for detecting Russian-language fake news using elements of artificial intelligence]. *International Journal of Open Information Technologies*. 2018. vol. 6. no. 12. pp. 99–105.
15. Zhuk D.A., Zhuk D.V., Tret’jakov A.O. [Methods for detecting fake news in social networks using machine learning]. *Informacionnye resursy Rossii – Information resources of Russia*. 2018. no. 3(163). pp. 29–32.
16. Face news. Available at: https://en.wikipedia.org/wiki/Fake_news (accessed 10.02.2023).

17. Wang Y., Ma F., Jin Z., Yuan Y., Xun G., Jha K., Su L., Gao J. Eann: Event adversarial neural networks for multi-modal fake news detection. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. NY, USA: Association for Computing Machinery, 2018. p. 849–857. DOI: 10.1145/3219819.3219903.
18. Du J., Dou Y., Xia C., Cui L., Ma J., Yu P.S. Cross-lingual covid-19 fake news detection. International Conference on Data Mining Workshops (ICDMW). 2021. pp. 859–862. DOI: 10.1109/ICDMW53433.2021.00110.
19. Perez-Rosas V., Kleinberg B., Lefevre A., Mihalcea R. Automatic detection of fake news. Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018. pp. 3391–3401.
20. Sharma U., Saran S., Patil S. Fake news detection using machine learning algorithms. international journal of creative research thoughts – IJCRT. 2020. vol. 8(6). pp. 2320–2882.
21. Ahmed A.A.A., Aljabouh A., Donepudi P.K., Choi M.S. Detecting fake news using machine learning: A systematic literature review. Psychology and education. 2021. vol. 58(1). pp. 1932-1939.
22. Aldwairi M., Alwahedi A. Detecting fake news in social media networks. Procedia Computer Science. The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2018). 2018. vol. 141. pp. 215–222.
23. Hu L., Wei S., Zhao Z., Wu B. Deep learning for fake news detection: A comprehensive survey. AI Open. 2022. vol. 3. pp. 133–155. Available at: <https://www.sciencedirect.com/science/article/pii/S2666651022000134>.
24. Jose X., Kumar S.M., Chandran P. Characterization, classification and detection of fake news in online social media networks. 2021 IEEE Mysore Sub Section International Conference (MysuruCon). 2021. pp. 759–765.
25. Kato S., Yang L., Ikeda D. Domain bias in fake news datasets consisting of fake and real news pairs. 12th International Congress on Advanced Applied Informatics (IIAI-AAI). 2022. pp. 101–106.
26. Yu W., Ge J., Yang Z., Dong Y., Zheng Y., Dai H. Multi-domain fake news detection for history news environment perception. IEEE 17th Conference on Industrial Electronics and Applications (ICIEA). 2022. pp. 428–433.
27. Borkar T.H., Ahuja T. Comparative study of supervised learning algorithms for fake news classification. 6th International Conference on Trends in Electronics and Informatics (ICOEI). 2022. pp. 1405–1411.
28. Lu M.F., Renaldy, Ciptadi V., Nathanael R., Andaria K.S., Girsang A.S. Fake news classifier with deep learning. International Conference on Informatics Electrical and Electronics (ICIEE). 2022. pp. 1–4. DOI: 10.1109/ICIEE55596.2022.10010120.
29. Zubiaga A., Liakata M., Procter R. Exploiting context for rumour detection in social media. Social Informatics. (Eds: Ciampaglia G.L., Mashhadi A., Yasser T.). Cham: Springer International Publishing, 2017. pp. 109–123.
30. Shahi G.K., Nandini D. Fakecovid – A multilingual cross-domain fact check news dataset for COVID-19. CoRR, abs/2006.11343. 2020. 16 p. Available at: <https://arxiv.org/abs/2006.11343>.
31. Li Y., Jiang B., Shu K., Liu H. MM-COVID: A multilingual and multimodal data repository for combating COVID-19 disinformation. CoRR, abs/2011.04088. 2020. Available at: <https://arxiv.org/abs/2011.04088>.
32. Kumar V., Kumar A., Singh A.K., Pachauri A. Fake news detection using machine learning and natural language processing. International Conference on Technological Advancements and Innovations (ICTAI). 2021. pp. 547–552.

33. Della Vedova M.L., Tacchini E., Moret S., Ballarin G., DiPiero M., de Alfaro L. Automatic online fake news detection combining content and social signals. 22nd Conference of Open Innovations Association (FRUCT). 2018. pp. 272–279.
34. Bian T., Xiao X., Xu T., Zhao P., Huang W., Rong Y., Huang J. Rumor detection on social media with bi-directional graph convolutional networks. AAAI Conference on Artificial Intelligence. 2020. DOI: 10.1609/AAAI.V34I01.5393.
35. Sharma D.K., Shrivastava P., Garg S. Utilizing word embedding and linguistic features for fake news detection. 9th International Conference on Computing for Sustainable Global Development (INDIACom). 2022. pp. 844–848.
36. Trang N.T.M., Shcherbakov M. Vietnamese question answering system from multilingual BERT models to monolingual bert model. 9th International Conference System Modeling and Advancement in Research Trends (SMART). 2020. pp. 201–206.
37. Chinnalagu A., Durairaj A.K. Comparative analysis of BERT-base transformers and deep learning sentiment prediction models. 11th International Conference on System Modeling Advancement in Research Trends (SMART). 2022. pp. 874–879.
38. Min C., Ahn J., Lee T., Im D.-H. TK-BERT: Effective model of language representation using topic-based knowledge graphs. 17th International Conference on Ubiquitous Information Management and Communication (IMCOM). 2023. pp. 1–4.
39. Sebastian D., Purnomo H.D., Sembiring I. Bert for natural language processing in bahasa Indonesia. 2nd International Conference on Intelligent Cybernetics Technology Applications (ICICyTA). 2022. pp. 204–209.
40. Holbert R.L. A typology for the study of entertainment television and politics. *American Behavioral Scientist*. 2005. vol. 49. no. 3. pp. 436–453.
41. Baptista J.P., Gradim A. A working definition of fake news. *Encyclopedia*. 2022. vol. 2. no. 1. pp. 632–645.
42. Farkas J., Schou J. Fake news as a floating signifier: Hegemony, antagonism and the politics of falsehood. *Javnost-The Public*. 2018. vol. 25. no. 3. pp. 298–314.
43. Thi T.-A.N., Vuong T.-H., Le T.-H., Phan X.-H., Le T.-T., Ha Q.-T. Knowledge base completion with transfer learning using bert and fasttext. 14th International Conference on Knowledge and Systems Engineering (KSE). 2022. pp. 1–6.
44. Nguyen Thi C.-V., Vuong T.-T., Le D.-T., Ha Q.-T. v3mfnd: A deep multi-domain multimodal fake news detection model for Vietnamese. *Intelligent Information and Database Systems* (Eds.: Nguyen N.T., Tran T.K., Tukayev U., Hong T.-P., Trawinski B., Szczerbicki E.). Cham: Springer International Publishing, 2022. pp. 608–620.
45. Pham N.-D., Le T.-H., Do T.-D., Vuong T.-T., Vuong T.-H., Ha Q.-T. Vietnamese fake news detection based on hybrid transfer learning model and TF-IDF. 13th International Conference on Knowledge and Systems Engineering (KSE). 2021. pp. 1–6.
46. Shahid W., Li Y., Staples D., Amin G., Hakak S., Ghorbani A. Are you a cyborg, bot or human? – a survey on detecting fake news spreaders. *IEEE Access*, 2022. vol. 10. pp. 27069–27083.
47. Wang C.-C. Fake news and related concepts: Definitions and recent research development. *Contemporary Management Research*. 2020. vol. 16. no. 3. pp. 145–174.
48. Umer M., Imtiaz Z., Ullah S., Mehmood A., Choi G.S., On B.-W. Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access*, 2020. vol. 8. pp. 156695–156706.
49. Abonizio H.Q., de Morais J.I., Tavares G.M., Barbon Junior V. Language-independent fake news detection: English, portuguese, and spanish mutual features. *Future Internet*. 2020. vol. 12. no. 5. Available at: <https://www.mdpi.com/1999-5903/12/5/87>.

50. Sayyadiharikandeh M., Varol O., Yang K.-C., Flammini A., Menczer F. Detection of novel social bots by ensembles of specialized classifiers. CoRR, abs/2006.06867. 2020. Available at: <https://arxiv.org/abs/2006.06867>.
51. Wang R. Shi Y. Research on application of article recommendation algorithm based on word2vec and TFIDF. IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA). 2022. pp. 454–457.
52. Devlin J., Chang M.-W., Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. 2019. 16 p. DOI: 10.48550/arXiv.1810.04805.
53. Qu Y., Liu P., Song W., Liu L., Cheng M. A text generation and prediction system: Pre-training on new corpora using BERT and GPT-2. IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC). 2020. pp. 323–326.
54. Du L., Hu C. Text similarity detection method of power customer service work order based on tfidf algorithm. IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE). 2022. pp. 978–982.
55. Nguyen D.Q., Nguyen A.T. PhoBERT: Pre-trained language models for Vietnamese in Findings of the Association for Computational Linguistics: EMNLP 2020. 2020. pp. 1037–1042.
56. Liu Y., Wu Y.-F.B. FNED: A deep network for fake news early detection on social media. ACM Trans. Inf. Syst. 2020. vol. 38, no. 3. DOI: 10.1145/3386253.
57. Fake news dataset. Available at: <https://github.com/Hung1239/fake-news.git> (accessed 02.05.2023).
58. Nguyen H., Dao T.N., Pham N.S., Dang T.L., Nguyen T.D., Truong T.H. An accurate viewport estimation method for 360 video streaming using deep learning. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2022. vol. 9, no. 4. p. e2. DOI: 10.4108/eetinis.v9i4.2218.
59. Panda M., Mousa A.A.A., Hassanien A.E. Developing an efficient feature engineering and machine learning model for detecting iot-botnet cyber attacks. IEEE Access. 2021. vol. 9, pp. 91038–91052.

Nguyen Viet Hung — Ph.D., Lecturer, Faculty of information technology, East Asia University of Technology; Student, faculty of telecommunications engineering, Hanoi University of Science and Technology. Research interests: multimedia communications, network security, artificial intelligence, traffic engineering in next-generation networks, QoE/QoS guarantee for network services, green networking, applications. The number of publications — 14. hungnv@eaut.edu.vn; Ky Phu - Ky Anh, Ha Tinh, Viet Nam; office phone: +84(098)911-2079.

Thang Quang Loi — Research assistant, East Asia University of Technology. Research interests: applications, networks. The number of publications — 1. thangquangloi21@gmail.com; Xuan Long, Yen Binh, Yen Bai, Viet Nam; office phone: +84(084)6203-0902.

Nguyen Thi Huong — Research assistant, East Asia University of Technology. Research interests: applications, networks. The number of publications — 1. 20212149@eaut.edu.vn; Hoa Chinh, Chuong My, Ha Tinh, Viet Nam; office phone: +84(084)6662-1533.

Tran Thi Thuy Hang — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, database management systems, artificial intelligence, applications. The number of publications — 1. hang42c@gmail.com; Mao Khe - Dong Trieu, Quang Ninh, Viet Nam; office phone: +84(090)496-8545.

Truong Thu Huong — Ph.D., Dr.Sci., Deputy head of the department, School of electrical and electronic engineering, Hanoi University of Science and Technology. Research interests: network security, artificial intelligence, traffic engineering in next-generation networks, QoE/QoS guarantee for network services, green networking, development of the internet of things ecosystems and applications. The number of publications — 90. huong.truongthu@hust.edu.vn; 1, Dai Co Viet St., Hanoi, Viet Nam; office phone: +84(243)869-2242.

Н.В. ХУНГ, Т.К. ЛОИ, Н.Т. ХЬОНГ, Т.Т. ХАНГ, Т.Т. ХЬОНГ
**AAFNDL — ТОЧНАЯ МОДЕЛЬ РАСПОЗНАВАНИЯ
ПОДЕЛЬНОЙ ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ
ГЛУБОКОГО ОБУЧЕНИЯ ВЬЕТНАМСКОГО ЯЗЫКА**

Хунг Нгуен Вьет, Лои Тран Куанг, Хьонг Нгуен Ти, Ханг Тран Тхи Туй, Хьонг Труонг Ту. AAFNDL — точная модель распознавания поддельной информации с использованием глубокого обучения вьетнамского языка.

Аннотация. В интернете «фейковые новости» - это распространенное явление, которое часто беспокоит общество, поскольку содержит заведомо ложную информацию. Проблема активно исследовалась с использованием обучения с учителем для автоматического обнаружения фейковых новостей. Хотя точность растет, она по-прежнему ограничивается идентификацией ложной информации через каналы на социальных платформах. Это исследование направлено на повышение надежности обнаружения фейковых новостей на платформах социальных сетей путем изучения новостей с неизвестных доменов. Особенно трудно обнаружить и предотвратить распространение информации в социальных сетях во Вьетнаме, потому что все имеют равные права на использование интернета для разных целей. Эти люди имеют доступ к нескольким платформам социальных сетей. Любой пользователь может публиковать или распространять новости через онлайн-платформы. Эти платформы не пытаются проверять пользователей, их местоположение или содержимое их новостей. В результате некоторые пользователи пытаются распространять через эти платформы фейковые новости для пропаганды; против отдельного лица, общества, организации или политической партии. Мы предложили проанализировать и разработать модель распознавания фейковых новостей с использованием глубокого обучения (называемого AAFNDL). Метод выполнения работы: 1) во-первых, анализируем существующие методы, такие как представление двунаправленного кодировщика от преобразователя (BERT); 2) приступаем к построению модели для оценки; 3) подходим к применению некоторых современных методов к модели, таких как метод глубокого обучения, метод классификатора и т.д., для классификации ложной информации. Эксперименты показывают, что наш метод может улучшить результаты на 8,72% по сравнению с другими методами.

Ключевые слова: социальные сети, вычислительное моделирование, глубокое обучение, извлечение признаков, алгоритмы классификации, фейковые новости, BERT, TF-IDF, PhoBERT.

Литература

1. Mladenova T., Valova I. Research on the ability to detect fake news in users of social networks. International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). 2022. pp. 01–04.
2. Apuke O.D., Omar B. Fake news and covid-19: modelling the predictors of fake news sharing among social media users. Telematics and Informatics. 2021. vol. 56. p. 101475. Available at: <https://www.sciencedirect.com/science/article/pii/S0736585320301349>.
3. Nguyen H, Tan N., Quan N., Huong T., Phat N. Building a chatbot system to analyze opinions of english comments. Informatics and Automation. 2023. vol. 22. no. 2. pp. 289–315.

4. Yuslee N.S., Abdullah N.A.S. Fake news detection using naive bayes. IEEE 11th International Conference on System Engineering and Technology (ICSET). 2021. pp. 112–117.
5. Babar A., Jagtap N., Mithari A., Shukla A., Chaudhari P. A survey on fake news detection techniques and using a blockchain based system to combat fake news. International Journal of Computer Applications. 2020. vol. 176. no. 27. pp. 47–53.
6. Kaliyar R.K. Fake news detection using a deep neural network. 4th International Conference on Computing Communication and Automation (ICCCA). 2018. pp. 1–7.
7. Sastrawan I.K., Bayupati I., Arsa D.M.S. Detection of fake news using deep learning cnn–rnn based methods. ICT Express. 2022. vol. 8. no. 3. pp. 396–408.
8. Vinothkumar S., Varadhaganapathy S., Ramalingam M., Ramkishore D., Rithik S., Tharanies K. Fake news detection using svm algorithm in machine learning,” in 2022 International Conference on Computer Communication and Informatics (ICCCI). 2022. pp. 1–7.
9. Hussain M.G., Hasan M.R., Rahman M., Protim J., Hasan S.A. Detection of bangla fake news using mnb and svm classifier. 2020. 5 p. DOI: 10.1109/iCCECE49321.2020.9231167.
10. Hussain M.G., Hasan M.R., Rahman M., Protim J., Al Hasan S. Detection of bangla fake news using mnb and svm classifier. International Conference on Computing, Electronics Communications Engineering (iCCECE). 2020. pp. 81–85.
11. Aphiwongsophon S., Chongstitvatana P. Detecting fake news with machine learning method. 2018. pp. 528–531.
12. Маилян Э.К., Куликов А.А. Анализ алгоритмов обнаружения fake news. Всероссийская конференция молодых исследователей с международным участием «Социально-гуманитарные проблемы образования и профессиональной самореализации «Социальный инженер-2020». 2020. С. 204–209.
13. Василькова В.В., Садчиков Д.И. Фейки и боты как механизмы информационных искажений в социальных сетях. Казанский социально-гуманитарный вестник. 2019. № 2(37). С. 24–30.
14. Третьяков А.О., Филатова О.Г., Жук Д.В., Горлушкина Н.Н., Пучковская А.А. Метод определения русскоязычных фейковых новостей с использованием элементов искусственного интеллекта. International Journal of Open Information Technologies. 2018. Т. 6. № 12. С. 99–105.
15. Жук Д.А., Жук Д.В., Третьяков А.О. Методы определения поддельных новостей в социальных сетях с использованием машинного обучения. Информационные ресурсы России. 2018. № 3(163). С. 29–32.
16. Face news. Available at: https://en.wikipedia.org/wiki/Fake_news (accessed 10.02.2023).
17. Wang Y., Ma F., Jin Z., Yuan Y., Xun G., Jha K., Su L., Gao J. Eann: Event adversarial neural networks for multi-modal fake news detection. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, NY, USA: Association for Computing Machinery, 2018. p. 849–857. DOI: 10.1145/3219819.3219903.
18. Du J., Dou Y., Xia C., Cui L., Ma J., Yu P.S. Cross-lingual covid-19 fake news detection. International Conference on Data Mining Workshops (ICDMW). 2021. pp. 859–862. DOI: 10.1109/ICDMW53433.2021.00110.
19. Perez-Rosas V., Kleinberg B., Lefevre A., Mihalcea R. Automatic detection of fake news. Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018. pp. 3391–3401.

20. Sharma U., Saran S., Patil S. Fake news detection using machine learning algorithms. *international journal of creative research thoughts – IJCRT*. 2020. vol. 8(6). pp. 2320–2882.
21. Ahmed A.A.A., Aljabouh A., Donepudi P.K., Choi M.S. Detecting fake news using machine learning: A systematic literature review. *Psychology and education*. 2021. vol. 58(1). pp. 1932-1939.
22. Aldwairi M., Alwahedi A. Detecting fake news in social media networks. *Procedia Computer Science. The 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2018)*. 2018. vol. 141. pp. 215–222.
23. Hu L., Wei S., Zhao Z., Wu B. Deep learning for fake news detection: A comprehensive survey. *AI Open*. 2022. vol. 3. pp. 133–155. Available at: <https://www.sciencedirect.com/science/article/pii/S2666651022000134>.
24. Jose X., Kumar S.M., Chandran P. Characterization, classification and detection of fake news in online social media networks. *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*. 2021. pp. 759–765.
25. Kato S., Yang L., Ikeda D. Domain bias in fake news datasets consisting of fake and real news pairs. *12th International Congress on Advanced Applied Informatics (IIAI-AAI)*. 2022. pp. 101–106.
26. Yu W., Ge J., Yang Z., Dong Y., Zheng Y., Dai H. Multi-domain fake news detection for history news environment perception. *IEEE 17th Conference on Industrial Electronics and Applications (ICIEA)*. 2022. pp. 428–433.
27. Borkar T.H., Ahuja T. Comparative study of supervised learning algorithms for fake news classification. *6th International Conference on Trends in Electronics and Informatics (ICOEI)*. 2022. pp. 1405–1411.
28. Lu M.F., Renaldy, Ciptadi V., Nathanael R., Andaria K.S., Girsang A.S. Fake news classifier with deep learning. *International Conference on Informatics Electrical and Electronics (ICIEE)*. 2022. pp. 1–4. DOI: 10.1109/ICIEE55596.2022.10010120.
29. Zubiaga A., Liakata M., Procter R. Exploiting context for rumour detection in social media. *Social Informatics*. (Eds: Ciampaglia G.L., Mashhadi A., Yasser T.). Cham: Springer International Publishing, 2017. pp. 109–123.
30. Shahi G.K., Nandini D. Fakecovid – A multilingual cross-domain fact check news dataset for COVID-19. *CoRR*, abs/2006.11343. 2020. 16 p. Available at: <https://arxiv.org/abs/2006.11343>.
31. Li Y., Jiang B., Shu K., Liu H. MM-COVID: A multilingual and multimodal data repository for combating COVID-19 disinformation. *CoRR*, abs/2011.04088. 2020. Available at: <https://arxiv.org/abs/2011.04088>.
32. Kumar V., Kumar A., Singh A.K., Pachauri A. Fake news detection using machine learning and natural language processing. *International Conference on Technological Advancements and Innovations (ICTAI)*. 2021. pp. 547–552.
33. Della Vedova M.L., Tacchini E., Moret S., Ballarin G., DiPierro M., de Alfaro L. Automatic online fake news detection combining content and social signals. *22nd Conference of Open Innovations Association (FRUCT)*. 2018. pp. 272–279.
34. Bian T., Xiao X., Xu T., Zhao P., Huang W., Rong Y., Huang J. Rumor detection on social media with bi-directional graph convolutional networks. *AAAI Conference on Artificial Intelligence*. 2020. DOI: 10.1609/AAAI.V34I01.5393.
35. Sharma D.K., Shrivastava P., Garg S. Utilizing word embedding and linguistic features for fake news detection. *9th International Conference on Computing for Sustainable Global Development (INDIACom)*. 2022. pp. 844–848.

36. Trang N.T.M., Shcherbakov M. Vietnamese question answering system from multilingual BERT models to monolingual bert model. 9th International Conference System Modeling and Advancement in Research Trends (SMART). 2020. pp. 201–206.
37. Chinnalagu A., Durairaj A.K. Comparative analysis of BERT-base transformers and deep learning sentiment prediction models. 11th International Conference on System Modeling Advancement in Research Trends (SMART). 2022. pp. 874–879.
38. Min C., Ahn J., Lee T., Im D.-H. TK-BERT: Effective model of language representation using topic-based knowledge graphs. 17th International Conference on Ubiquitous Information Management and Communication (IMCOM). 2023. pp. 1–4.
39. Sebastian D., Purnomo H.D., Sembiring I. Bert for natural language processing in bahasa Indonesia. 2nd International Conference on Intelligent Cybernetics Technology Applications (ICICyTA). 2022. pp. 204–209.
40. Holbert R.L. A typology for the study of entertainment television and politics. *American Behavioral Scientist*. 2005. vol. 49. no. 3. pp. 436–453.
41. Baptista J.P., Gradim A. A working definition of fake news. *Encyclopedia*. 2022. vol. 2. no. 1. pp. 632–645.
42. Farkas J., Schou J. Fake news as a floating signifier: Hegemony, antagonism and the politics of falsehood. *Javnost-The Public*. 2018. vol. 25. no. 3. pp. 298–314.
43. Thi T.-A.N., Vuong T.-H., Le T.-H., Phan X.-H., Le T.-T., Ha Q.-T. Knowledge base completion with transfer learning using bert and fasttext. 14th International Conference on Knowledge and Systems Engineering (KSE). 2022. pp. 1–6.
44. Nguyen Thi C.-V., Vuong T.-T., Le D.-T., Ha Q.-T. v3mfnd: A deep multi-domain multimodal fake news detection model for Vietnamese. *Intelligent Information and Database Systems* (Eds.: Nguyen N.T., Tran T.K., Tukayev U., Hong T.-P., Trawinski B., Szczerbicki E.). Cham: Springer International Publishing, 2022. pp. 608–620.
45. Pham N.-D., Le T.-H., Do T.-D., Vuong T.-T., Vuong T.-H., Ha Q.-T. Vietnamese fake news detection based on hybrid transfer learning model and TF-IDF. 13th International Conference on Knowledge and Systems Engineering (KSE). 2021. pp. 1–6.
46. Shahid W., Li Y., Staples D., Amin G., Hakak S., Ghorbani A. Are you a cyborg, bot or human? – a survey on detecting fake news spreaders. *IEEE Access*, 2022. vol. 10. pp. 27069–27083.
47. Wang C.-C. Fake news and related concepts: Definitions and recent research development. *Contemporary Management Research*. 2020. vol. 16. no. 3. pp. 145–174.
48. Umer M., Imtiaz Z., Ullah S., Mehmood A., Choi G.S., On B.-W. Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access*, 2020. vol. 8. pp. 156695–156706.
49. Abonizio H.Q., de Moraes J.J., Tavares G.M., Barbon Junior V. Language-independent fake news detection: English, portuguese, and spanish mutual features. *Future Internet*. 2020. vol. 12. no. 5. Available at: <https://www.mdpi.com/1999-5903/12/5/87>.
50. Sayyadharikandeh M., Varol O., Yang K.-C., Flammini A., Menczer F. Detection of novel social bots by ensembles of specialized classifiers. *CoRR*, abs/2006.06867. 2020. Available at: <https://arxiv.org/abs/2006.06867>.
51. Wang R., Shi Y. Research on application of article recommendation algorithm based on word2vec and TFIDF. *IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*. 2022. pp. 454–457.
52. Devlin J., Chang M.-W., Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. 2019. 16 p. DOI: 10.48550/arXiv.1810.04805.

53. Qu Y., Liu P., Song W., Liu L., Cheng M. A text generation and prediction system: Pre-training on new corpora using BERT and GPT-2. IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEEC). 2020. pp. 323–326.
54. Du L., Hu C. Text similarity detection method of power customer service work order based on tfidf algorithm. IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE). 2022. pp. 978–982.
55. Nguyen D.Q., Nguyen A.T. PhoBERT: Pre-trained language models for Vietnamese in Findings of the Association for Computational Linguistics: EMNLP 2020. 2020. pp. 1037–1042.
56. Liu Y., Wu Y.-F.B. FNED: A deep network for fake news early detection on social media. ACM Trans. Inf. Syst. 2020. vol. 38. no. 3. DOI: 10.1145/3386253.
57. Fake news dataset. Available at: <https://github.com/Hung1239/fake-news.git> (accessed 02.05.2023).
58. Nguyen H., Dao T.N., Pham N.S., Dang T.L., Nguyen T.D., Truong T.H. An accurate viewport estimation method for 360 video streaming using deep learning. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems. 2022. vol. 9. no. 4. p. e2. DOI: 10.4108/eetinis.v9i4.2218.
59. Panda M., Mousa A.A.A., Hassanien A.E. Developing an efficient feature engineering and machine learning model for detecting iot-botnet cyber attacks. IEEE Access. 2021. vol. 9. pp. 91038–91052.

Хунг Нгуен Вьет — Ph.D., преподаватель, факультет информационных технологий, Восточноазиатский технологический университет; студент, факультет телекоммуникаций, Ханойский университет науки и технологий. Область научных интересов: мультимедийные коммуникации, сетевая безопасность, искусственный интеллект, управление трафиком в сетях нового поколения, гарантия QoE/QoS для сетевых услуг, экологичные сети, приложения. Число научных публикаций — 14. hungnv@eaut.edu.vn; Ки Фу - Ки Ань, Хатинь, Вьетнам; р.т.: +84(098)911-2079.

Лонг Тран Куанг — научный сотрудник, Восточноазиатский технологический университет. Область научных интересов: приложения, сети. Число научных публикаций — 1. thangquangloi21@gmail.com; Суан Лонг, Йен Бинь, Йенбай, Вьетнам; р.т.: +84(084)6203-0902.

Хьонг Нгуен Ти — научный сотрудник, Восточноазиатский технологический университет. Область научных интересов: приложения, сети. Число научных публикаций — 1. 20212149@eaut.edu.vn; Хоа Чин, Чуонг Ми, Хатинь, Вьетнам; р.т.: +84(084)6662-1533.

Ханг Тран Тхи Туь — лектор, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, системы управления базами данных, искусственный интеллект, приложения. Число научных публикаций — 1. hang42c@gmail.com; Мао Кхе - Донг Чуу, Куанг Нин, Вьетнам; р.т.: +84(090)496-8545.

Хьонг Труонг Ту — Ph.D., Dr.Sci., заместитель начальника отдела, школа электротехники и электронной инженерии, Ханойский университет науки и технологий. Область научных интересов: сетевая безопасность, искусственные интеллектуальные функции, управление трафиком в сетях следующего поколения, гарантия QoE/QoS для сети услуги, экологичные сети, развитие экосистем и приложений Интернета вещей. Число научных публикаций — 90. huong.truongthu@hust.edu.vn; улица Дай Ко Вьет, 1, Ханой, Вьетнам; р.т.: +84(243)869-2242.

А.Р. ГАЙДУК, Н.Н. ПРОКОПЕНКО, А.В. БУГАКОВА
**КОМПЕНСАЦИЯ ОШИБОК, ВЫЗВАННЫХ ВРЕМЕННОЙ
ЗАДЕРЖКОЙ ЦИФРОВЫХ ДАТЧИКОВ**

Гайдук А.Р., Прокопенко Н.Н., Бугакова А.В. Компенсация ошибок, вызванных временной задержкой цифровых датчиков.

Аннотация. Исследование посвящено повышению точности цифровых датчиков с запаздыванием по времени. Актуальность темы обусловлена широким распространением датчиков этого типа, что во многом обусловлено резким повышением требований к точности датчиков, а также расширяющимся применением цифровых технологий для обработки информации в системах управления, связи, мониторинга и многих других. Для устранения ошибок, обусловленных временной задержкой цифровых датчиков, предлагается использовать астатический быстродействующий корректор. Целесообразность применения этого корректора обосновывается свойствами дискретных динамических систем. В связи с этим сначала рассматриваются условия, при которых дискретные системы являются физически реализуемыми и имеют конечную длительность переходных процессов, поскольку в этом последнем случае они являются наиболее быстродействующими. Также показано, что для измерения полиномиального сигнала ограниченной интенсивности с нулевой ошибкой в установившемся режиме, датчик должен иметь порядок астатизма на единицу больше степени этого сигнала. На основе приведенных условий доказывается основной результат статьи – теорема, в которой устанавливаются условия существования астатического быстродействующего корректора. При включении этого корректора на выходе цифрового датчика или коррекции программного обеспечения последнего образуется модернизированный датчик, ошибка которого в установившемся режиме будет равна нулю. Это происходит вследствие того, что корректор устраняет ошибку цифрового датчика, обусловленную имевшейся в нём задержкой по времени, которая предполагается кратной периоду дискретизации. Порядок корректора как системы определяется целочисленным решением полученного в работе уравнения, которое связывает степень измеряемого полиномиального сигнала, запаздывание цифрового датчика, допустимое перерегулирование модернизированного датчика и относительный порядок искомого корректора. Это уравнение решено для случаев, когда степень измеряемого сигнала не больше единицы, перерегулирование равно часто назначаемым значениям, а задержка по времени не превышает четырёх периодов дискретизации. Порядки соответствующих модернизированных цифровых датчиков приведены в табличной форме. Это позволяет находить необходимый корректор без решения указанного уравнения во многих практических случаях. Эффективность предлагаемого подхода к повышению точности цифровых датчиков показана на численном примере. Нулевое значение ошибки модернизированного датчика подтверждается как путем компьютерного моделирования, так и численным расчетом. Полученные результаты могут использоваться при разработке высокоточных цифровых датчиков различных физических величин.

Ключевые слова: дискретная система, цифровой фильтр, цифровой датчик, порядок системы, относительный порядок системы, запаздывание, астатизм, переходный процесс, перерегулирование, конечный переходной процесс.

1. Введение. С увеличением сложности современных технических систем и ужесточением требований к их качеству, возрастает важность роли датчиков в системах различного назначения:

управления, мониторинга, диагностики и т.п. [1 – 5]. По существу, именно достоверность и точность доставляемых датчиками данных обуславливают полезность и эффективность функционирования этих систем. Эти качества датчиков зависят не только от характеристик чувствительных элементов, но и от системы предварительной обработки данных, которую целесообразно размещать ближе к считыванию данных. В связи с этим в последние годы всё более широкое распространение получают датчики, в которых чувствительный элемент объединяется со встроенной системой предварительной обработки данных. Последняя обычно представляет собой вычислитель, функции которого реализует процессор, размещаемый в непосредственной близости к чувствительному элементу датчика. Такие цифровые датчики часто называются интеллектуальными. Из-за наличия вычислителя такой датчик превращается в дискретное динамическое звено и характеризуется запаздыванием по времени. Основные проблемы, связанные с созданием интеллектуальных датчиков и измерительных систем, заключаются как в разработке эффективного, с точки зрения точности, алгоритма обработки данных, так и в выборе вычислителя. В работе [6] показано, что проблема точности интеллектуальных датчиков, тесно связана с уровнем потребления энергии компонентами системы обработки. Поэтому рекомендуется применять процессоры с минимальным потреблением энергии, что обеспечит значительно большую свободу выбора алгоритмов выполнения необходимых функций и, в частности, фильтрации.

Последнее обусловлено тем, что датчики, как правило, функционируют в условиях внешних случайных помех и воздействий, поэтому их выходные сигналы содержат случайные ошибки. Обычно эти ошибки считаются центрированными случайными гауссовскими процессами. Поэтому вычислители большинства известных интеллектуальных датчиков реализуют в основном алгоритмы цифровых фильтров различных типов. При этом современные методы цифровой фильтрации позволяют обеспечить пренебрежимо малые значения ошибок датчиков, обусловленных случайными составляющими входных сигналов [7 – 8]. Однако эти фильтры характеризуются значительной задержкой по времени, что приводит к появлению существенных ошибок датчиков [2, 6]. На рисунке 1 показаны графики входного сигнала: (а) некоторого датчика и сигнала (б) на выходе его фильтра.

Как видно ошибки, вызванные случайными помехами, по окончании переходного процесса в фильтре, практически,

отсутствуют, но из-за задержки выходного сигнала (\bar{b}) фильтра, по отношению к входному сигналу (a), возникает ошибка, часто довольно значительная, что существенно ухудшает качественные характеристики датчиков как измерительных устройств. Поэтому на практике применяются различные методы и подходы, позволяющие уменьшить влияние ошибок датчиков, обусловленных запаздыванием.

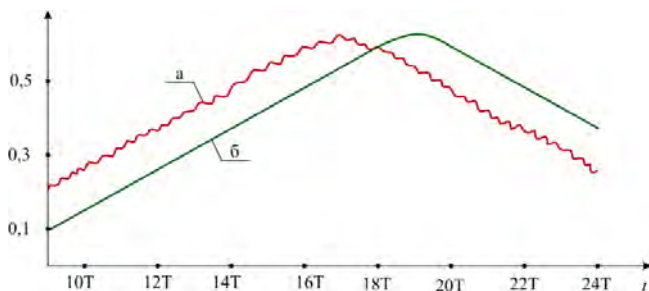


Рис. 1. Графики a -входного, b -выходного сигналов

С этой целью в ряде работ предлагаются различные алгоритмы восстановления действительного значения измеряемого сигнала. Так в [5] на основе теоремы Котельникова-Шеннона и метода регуляризации А.Н. Тихонова разработаны новый метод и алгоритм математической обработки данных, поступающих от дискретного чувствительного элемента. Вычисления реальных значений непрерывного сигнала на входе датчика осуществляется с использованием двух регуляризационных ядер и выборок соответствующего объема. При этом гарантируются точные границы ошибок датчика. С этой же целью в [9] предлагаются методы проектирования интеллектуальных, вычислительных сенсорных систем и датчиков, функционирование которых поддерживается методами машинного обучения. По мнению авторов, предлагаемые методы позволят создавать недорогие и компактные сенсорные системы, повышенной точности.

В [10] для повышения точности инерциальных датчиков, типа гироскопических акселерометров, применяется итерационный алгоритм дискретного фильтра Калмана, использующий обновление ковариационных функций, связывающих дискретные отсчеты выборок. При этом существенно снижаются вычислительные затраты, но эффективность фильтра Калмана не снижается. В непрерывных случаях для аналогичных целей используется модель непрерывного фильтра Калмана [11]. Точное измерение сигналов, несомненно, связано с безотказной работой датчиков, в том числе и

интеллектуальных. Для обеспечения устойчивой работы систем, использующих датчики, принимаются различные меры для преодоления случаев возникновения ошибок в измерительной информации, вызванных отказом датчиков. С этой целью применяются специальные наблюдатели, позволяющие оценить текущее состояние датчиков и осуществить достаточно точный прогноз изменений внешних измеряемых сигналов [12 – 15]. При этом могут использоваться адаптивные алгоритмы оценки текущего состояния датчика, разработанные на основе теории устойчивости Ляпунова [13], теории скользящих режимов, линейных матричных неравенств и интегрирования, что увеличивает скорость сходимости и позволяет свести к минимуму ошибку оценивания [14, 15].

Весьма эффективным в плане обеспечения высокой точности восстановления входного сигнала на выходе измерительной системы является эффект возникновения нулевой ошибки в следящих системах дискретного типа с конечным временем переходного процесса при действии на их входах ступенчатых воздействий [16]. Предлагается, также оценивать время запаздывания с тем, чтобы учесть его в алгоритме функционирования системы, в которой используется датчик. Для компенсации ошибок датчиков, вызванных временными задержками, предлагается применить метод линейной регрессии с минимизацией отклонений. Это позволяет обеспечить большую шумоустойчивость датчика, но нулевая ошибка, естественно, не достигается [17, 18].

В работах [19 – 21] разрабатываются специальные элементы с отрицательным групповым запаздыванием (NGD) включение которых в системы связи и передачи данных позволяет компенсировать влияние запаздывания. Эти элементы разрабатываются и изготавливаются на основе как специальных активных RC цепей, так и с применением индуктивностей и резисторов. Возникающие в этих схемах нано-, микро-, и миллисекундные эффекты NGD подтверждены численно и экспериментально. Возможность компенсации временной задержки с использованием NGD открывает широкие потенциальные возможности для улучшения качественных характеристик датчиков [19, 20]. В [21] предложена многофункциональная схема, на основе которой могут быть реализованы элементы с отрицательной групповой задержкой, включая полосовой, высокочастотный и низкочастотный фильтры, которым соответствуют различные условия включения емкости, индуктивности и сопротивления. Результаты экспериментальных исследований разработанной и изготовленной схемы с NGD подтвердили возможность компенсации запаздывания.

Однако этот подход позволяет компенсировать лишь малые запаздывания, что ограничивает его применение в реальных системах управления и системах связи.

Чаще всего известные методы компенсации ошибок интеллектуальных датчиков, вызванных запаздыванием по времени, не позволяют устранить эти ошибки полностью, даже для сигналов, определенных видов [22 – 25]. Это обуславливает необходимость разработки новых методов компенсации влияния временного запаздывания.

В данной работе для компенсации ошибок цифрового датчика, вызванных задержкой по времени, рассматривается применение астатического быстродействующего корректора (АБК), предложенного в работе [26]. Этот корректор является дискретным и реализуется также с применением цифровых элементов. Он имеет конечную длительность переходных процессов и обладает свойством астатизма [27, 28]. Это свойство динамических систем, в том числе и дискретных, достаточно часто используется для повышения качественных характеристик различных систем. В случае цифровых, интеллектуальных датчиков именно это свойство АБК позволяет обеспечить полную компенсацию его ошибок, обусловленных запаздыванием.

Обычно цифровые датчики (ЦД, рисунок 2) представляют собой совокупность чувствительного элемента (ЧЭ) и аналого-цифрового интерфейса (АЦИ) [6 – 8, 24]. Последний, в частности, предназначен для подавления (фильтрации) случайных шумов и помех, которые всегда сопровождают работу датчиков.

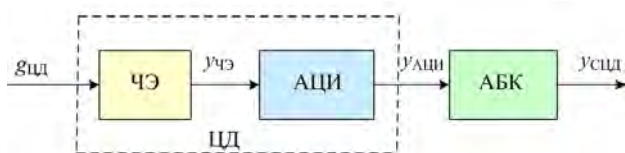


Рис. 2. Схема модернизированного цифрового датчика

Сигналы $y_{чэ}$ и $y_{аци}$ следуют с некоторым периодом T и, как показано на рисунке 3, значения выходного сигнала $y_{аци}$ в установившемся режиме не отличаются от значений $g_{цд,k}$, непрерывного сигнала $g_{цд}(t)$, $k = 0, 1, 2, \dots$. Однако, значения $y_{аци,k}$ имеют большую задержку (запаздывание) по времени t , которая может

достигать нескольких периодов, т.е. $\tau = qT$, где целое число $q \geq 1$ (на рисунке 3 $q = 2$, т.е. например, $y_{АЦЦ,14} = g_{ЦД,12}$, $y_{АЦЦ,15} = g_{ЦД,13}$ и т.д.). Эта задержка, очевидно, приводит к значительной ошибке датчика. Отметим, что в случае, представленном на рисунке 3, в интервале времени между $t = 17T$ и $t = 21T$ имеет место переходной процесс, поэтому в случае рассматриваемого датчика равенство $y_{АЦЦ,k} = g_{ЦД,k-2}$ в этом интервале, естественно, не выполняется.

Ошибка датчиков, вызванная запаздыванием, как было показано выше, отрицательно влияет на их характеристики, а также на свойства систем, в которых они используются. Поэтому возникает необходимость уменьшения влияния запаздывания на ошибку датчика, желательно полной компенсации этого влияния. Именно для полной компенсации этой ошибки предлагается дополнительно ввести в состав датчика астатический быстродействующий корректор, как показано на рисунке 2. Этот корректор, как будет показано ниже, устраняет ошибку, вызванную запаздыванием, незначительно увеличивая суммарное время запаздывания. При этом предполагается, что задержка в ЧЭ и АЦЦ кратна периоду дискретизации T [27, 28]. Датчик, схема которого показана на рисунке 2, будем называть модернизированным цифровым датчиком (upgrade digital sensor).

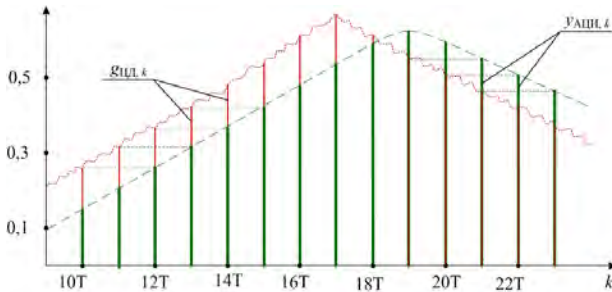


Рис. 3. Дискретные значения сигналов

Далее рассматриваются свойства модернизированных цифровых датчиков с АБК, имеющем конечную длительность переходного процесса и постоянное перерегулирование [29, 30].

2. Постановка задачи. Пусть чувствительный элемент (ЧЭ) совместно с АЦЦ цифрового датчика (рисунок 2) формируют дискретные значения $y_{АЦЦ,k} = g_{ЦД}((k-q)T)$ входного непрерывного сигнала $g_{ЦД}(t)$ (рисунок 3) с периодом T и с задержкой на q периодов.

В этом случае z -изображение выходного сигнала АЦИ $y_{\text{АЦИ},k}$ имеет вид $y_{\text{АЦИ}}(z) = z^{-q} g_{\text{ЦД}}(z)$ [29]. Значения $y_{\text{АЦИ},k}$ выходного сигнала АЦИ поступают на вход АБК, который предполагается линейным с передаточной функцией:

$$W_{\text{АБК}}(z) = y_{\text{АБК}}(z) / y_{\text{АЦИ}}(z) = D_0(z) / D(z), \quad (1)$$

где $y_{\text{АБК}}(z)$ – z -изображение выходного сигнала АБК, являющегося и выходным сигналом модернизированного цифрового датчика (МЦД, рисунок 2), т.е. $y_{\text{МЦД}}(z) = y_{\text{АБК}}(z)$; $D_0(z)$ и $D(z)$ – некоторые полиномы от z . Следовательно, в целом передаточная функция МЦД определяется выражением:

$$W_{\text{yg}}^{\text{МЦД}}(z) = \frac{y_{\text{МЦД}}(z)}{g_{\text{ЦД}}(z)} = \frac{D_0(z)}{D(z)} z^{-q}. \quad (2)$$

Величина $n_{\text{МЦД}} = \deg D(z)$ – это порядок МЦД; а величина $\mu_{\text{АБК}} = \deg D(z) - \deg D_0(z)$ – относительный порядок АБК (1) [28 – 29]. Здесь $\deg D(z)$ – обозначение степени полинома $D(z)$. Следовательно, согласно (2) относительный порядок МЦД $\mu_{\text{МЦД}} = \mu_{\text{АБК}} + q$, а порядок МЦД равен порядку АБК. Последнее обусловлено тем, что динамические свойства ЧЭ и АЦИ в данном случае проявляются в запаздывании на время $\tau = qT$ выходного сигнала АЦИ по отношению к входному сигналу $g_{\text{ЦД}}(t)$.

Заметим, что возможность физической реализации дискретных передаточных функций динамических элементов определяется их относительным порядком. Так передаточные функции типа (2) могут быть реализованы на основе цифровых элементов, если только их относительный порядок не меньше единицы [28]. В случае АБК это возможно, если только $\mu_{\text{АБК}} \geq 1$. Поэтому необходимо, чтобы относительный порядок МЦД $\mu_{\text{МЦД}} \geq q + 1$.

Пусть $\varepsilon_{\text{МЦД},k} = g_{\text{ЦД},k} - y_{\text{МЦД},k}$ – отклонение выходного сигнала МЦД от дискретных значений $g_{\text{ЦД},k} = g_{\text{ЦД}}(kT)$, $k = 0, 1, 2, \dots$, являющихся результатом дискретизации входного сигнала $g_{\text{ЦД}}(t)$

(рисунок 2). Ошибка модернизированного цифрового датчика, вызванная его входным сигналом $g_{\text{ЦД}}(t)$, равна отклонению $\varepsilon_{\text{МЦД},k}$ при $k \geq k_{\text{ПП}}$, где $k_{\text{ПП}}$ – число периодов T , в течение которых длится переходной процесс этого датчика [28, 29]. Из приведенного выражения для $\varepsilon_{\text{МЦД},k}$ и передаточной функции (2) следует, что передаточная функция МЦД по ошибке $\varepsilon_{\text{МЦД},k}$ определяется следующим выражением:

$$W_{\varepsilon g}^{\text{МЦД}}(z) = \frac{\varepsilon_{\text{МЦД}}(z)}{g_{\text{ЦД}}(z)} = 1 - W_{\text{Уг}}^{\text{МЦД}}(z) = \frac{D(z) - D_0(z)z^{-q}}{D(z)}. \quad (3)$$

Ограничимся здесь случаем полиномиальных сигналов, т.е. будем рассматривать случай, когда сигнал $g_{\text{ЦД},k}$, $k = 0, 1, 2, \dots$, является результатом дискретизации непрерывного сигнала полиномиального типа: $g_{\text{ЦД}}(t) = g_0 + g_1 t + \dots + g_r t^r$. При этом будем считать, что значение степени r приведенного полинома известно, а значения его коэффициентов g_i неизвестны, но все они являются ограниченными величинами.

Задача состоит в определении таких степеней и коэффициентов полиномов $D(z)$ и $D_0(z)$ из передаточной функции $W_{\text{АБК}}(z)$ (1), при которых ошибка датчика $\varepsilon_{\text{МЦД},k}$ будет равна нулю по окончании его переходного процесса, т.е. $\varepsilon_{\text{МЦД},k} = 0$ при всех $k_{\text{ПП}} \leq k < k_{\text{ПП}} T$, где $k_{\text{ПП}} T$ – момент времени, в который возникает новый переходной процесс цифрового датчика. Другими словами, необходимо найти такую передаточную функцию корректора, при котором будет иметь место полная компенсация ошибки модернизированного цифрового датчика, вызванной запаздыванием в исходном цифровом датчике.

3. Решение задачи. Модернизированный цифровой датчик (рисунок 2), в теоретическом плане, является дискретной системой, поэтому в данной работе используются свойства дискретных систем. Как известно, дискретная система является устойчивой и имеет конечную длительность переходного процесса, если только знаменатель её передаточной функции равен z в некоторой степени [29, 30]. Известно также, что дискретная система с такой передаточной функцией обладает наибольшим быстродействием по сравнению с дискретными системами, имеющими передаточные функции других видов. Поэтому для обеспечения устойчивости и быстродействия

МЦД, достаточно в выражениях (2) и (1) принять $D(z) = z^{n_{\text{МЦД}}}$, поскольку, как отмечалось выше, $n_{\text{МЦД}} = n_{\text{АБК}}$. Подчеркнём, что в этом случае переходной процесс АБК длится ровно $n_{\text{АБК}}T$.

Имея в виду наличие запаздывания в исходном цифровом датчике и свойства дискретных систем с конечной длительностью переходных процессов, рассмотрим некоторую дискретную систему порядка n , на вход которой поступает сигнал в виде дискретной единичной ступенчатой функции $g_k = 1(kT)$, а выходной сигнал системы описывается выражением:

$$y_k = \begin{cases} 0 \cdot 1(kT), & 0 \leq k < \mu, \\ (1 + \lambda) \cdot 1(kT), & \mu \leq k < n-1, \\ 1 \cdot 1(kT), & k \geq n. \end{cases} \quad (4)$$

Согласно (4) значения y_k равны нулю при всех $k = 0, 1, \dots, \mu-1$, следовательно, эта система имеет запаздывание по времени $\tau = \mu T$. Установившееся значение выходного сигнала рассматриваемой системы (при $k > n$) равно единице, поэтому её ошибка равна нулю, а переходный процесс имеет постоянное перерегулирование $\sigma\% = \lambda \cdot 100\%$ и конечную длительность $t_{\text{III}} = nT$, т.е. $k_{\text{III}} = n$.

Воспользовавшись свойствами Z-преобразования [29] и выражением (4), найдем передаточную функцию $W_{\text{yg}}(z) = y(z) / g(z)$ рассматриваемой дискретной системы. Как известно, z-изображение её входной функции $g_k = g(kT)$ можно обозначить $g(z) = Z\{g(kT)\}$, а z-изображение выходной функции y_k этой системы в соответствии с (4) при нулевом начальном значении запишется следующим образом $y(z) = [(1 + \lambda)z^{-\mu} - \lambda z^{-n}]g(z)$. Отсюда с учетом определения передаточных функций дискретных систем выводим:

$$W_{\text{yg}}(z) = \frac{y(z)}{g(z)} = \frac{(1 + \lambda)z^{-\mu} - \lambda z^{-n}}{1}. \quad (5)$$

Обычно передаточные функции дискретных систем принято записывать, как отношение полиномов с положительными степенями аргумента z . Поэтому, чтобы исключить отрицательные степени в выражении (5), умножим числитель и знаменатель дроби в правой части

этого равенства на z^n . В результате найдем, что искомая передаточная функция рассматриваемой дискретной системы имеет вид:

$$W_{yg}(z) = \frac{(1 + \lambda)z^{n-\mu} - \lambda}{z^n}. \quad (6)$$

Отметим, что эта система является устойчивой [28–30], а её относительный порядок, согласно приведенному выше определению, равен μ , т.е. запаздывание выходного сигнала этой системы по отношению к входному сигналу равно её относительному порядку.

Предположим, на вход системы с ПФ $W_{yg}(z)$ (6), поступает дискретный сигнал g_k , z -изображение которого является отношением полиномов от z , т.е.:

$$g(z) = Z\{g_k\} = G_0(z) / G(z), \quad (7)$$

где $G_0(z)$ и $G(z)$ – некоторые полиномы, и $\deg G_0(z) \leq \deg G(z) = n_G$.

С другой стороны, по аналогии с (3) можно заключить, что сигнал g_k воздействующий на систему с передаточной функцией $W_{yg}(z)$ вызывает ошибку $\varepsilon_{W,g}(k)$, z -изображение которой с учетом (7) имеет вид: $\varepsilon_{W,g}(z) = (1 - W_{yg}(z))G_0(z) / G(z)$. Покажем, что при $k \rightarrow \infty$ предел этой ошибки равен нулю при выполнении следующего условия:

$$W_{yg}(z) = 1 - V(z)G(z) / z^n, \quad (8)$$

где $V(z)$ – некоторый полином такой, что $\deg[V(z)G_0(z)] \leq n$. Действительно, подставляя правую часть равенства (8) в предыдущее выражение, будем иметь $\varepsilon_{W,g}(z) = V(z)G_0(z) / z^n$. Так как при выполнении условия (8) $\deg[V(z)G_0(z)] \leq n$, то в силу теоремы z -преобразования о граничных значениях [29] из полученного выражения следует $\lim \varepsilon_{W,g}(k) = 0$ при $k \rightarrow \infty$.

Как известно, z -изображение полиномиальных сигналов $g_k = g_0 + g_1kT + \dots + g_r(kT)^r$, $k = 0, 1, 2, \dots$ степени r определяется выражением:

$$g(z) = Z \{ [g_0 + g_1 kT + \dots + g_r \cdot (kT)^r] \cdot 1(kT) \} = G_0(z) / (z-1)^{r+1}. \quad (9)$$

где $G_0(z)$ – некоторый полином, коэффициенты которого однозначно определяются степенью r и коэффициентами g_i , $i = 0, 1, 2, \dots, r$. Например, если $g_k = g_0 1(kT)$, т.е. $r = 0$, то $g(z) = g_0 z / (z-1)$; если же $g_k = [g_0 + g_1 kT] \cdot 1(kT)$, то $g(z) = [g_0(z^2 - z) + g_1 Tz] / (z-1)^2$ [29].

Из выражений (8) и (9) следует, что дискретная система порядка n с ПФ $W_{yg}(z)$ будет иметь нулевую ошибку, вызванную полиномиальным сигналом степени r , если выполняется условие:

$$W_{yg}(z) = \frac{z^n - V(z)(z-1)^{r+1}}{z^n}, \quad (10)$$

где $V(z)$ – некоторый полином степени $\deg V(z) = n - r - 1$.

Отметим, что если ПФ некоторой дискретной системы удовлетворяет условию (8) или (10), то эта система называется дискретной системой с астатизмом порядка $\nu_g = r + 1$ по отношению к сигналу g_k . Важнейшее свойство астатических систем заключается в том, что коэффициенты g_i , $i = 0, 1, 2, \dots, r$ из выражения (9) не влияют на значение ошибки дискретной системы, астатической к этому сигналу. Именно по этой причине предложенный выше корректор синтезируется как астатическая система по отношению ко входному сигналу.

Из приведенных выражений (6) и (10) следует теорема, которая позволяет решить задачу синтеза АБК, обеспечивающего нулевую ошибку цифрового датчика с запаздыванием при полиномиальном входном сигнале степени r .

Теорема 1. Астатический быстродействующий корректор с относительным порядком $\mu_{\text{АБК}} = \mu_{\text{МЦД}} - q \geq 1$ и порядком астатизма $\nu_{\text{МЦД}} = r + 1$ существует, если уравнение:

$$z^{n_{\text{МЦД}}} - V(z)(z-1)^{\nu_{\text{МЦД}}} = (1 + \lambda)z^{n_{\text{МЦД}} - \mu_{\text{МЦД}}} - \lambda, \quad (11)$$

имеет целочисленное решение $n_{\text{МЦД}}$ при $\deg V(z) = n_{\text{МЦД}} - r - 1$.

Доказательство. Необходимость. Пусть астатический

быстродействующий корректор (рисунок 2) порядка $n_{\text{МЦД}}$, при котором МЦД имеет нулевую ошибку, вызванную входным полиномиальным воздействием степени $r = \nu_{\text{МЦД}} - 1$, существует. Тогда передаточная функция МЦД имеет вид (10), где $n = n_{\text{МЦД}}$ и $\deg V(z) = n_{\text{МЦД}} - r - 1$, причем $n_{\text{МЦД}}$ – это целое число, так как это порядок корректора. Если этот же МЦД имеет постоянное перерегулирование $\sigma = \lambda \cdot 100\%$ и конечную длительность переходного процесса, то согласно (6), в этом случае его передаточная функция имеет вид (5), где $n = n_{\text{МЦД}}$, $\mu = \mu_{\text{МЦД}}$, а $\nu_{\text{МЦД}} = r + 1$. Так как указанные передаточные функции описывают один и тот же МЦД и имеют равные знаменатели, то их числители тоже должны быть равны. Отсюда следует необходимость существования целочисленного решения $n_{\text{МЦД}}$ уравнения (11).

Достаточность. Пусть уравнение (11) имеет целочисленное решение $n_{\text{МЦД}}$ при $\deg V(z) = n_{\text{МЦД}} - r - 1$. Возьмём передаточную функцию МЦД в виде:

$$W_{\text{уг}}^{\text{МЦД}}(z) = [(1 + \lambda)z^{n_{\text{МЦД}} - \mu_{\text{МЦД}}} - \lambda] / z^{n_{\text{МЦД}}}. \quad (12)$$

При этом в соответствии с (6) МЦД будет иметь конечный переходной процесс длительностью $t_{\text{ПП}} = n_{\text{МЦД}}T$ и постоянное перерегулирование $\sigma = \lambda \cdot 100\%$. С другой стороны, так как указанное решение $n_{\text{МЦД}}$ уравнения (11) существует, то передаточную функцию МЦД (12) с учетом условия $\nu_{\text{МЦД}} = r + 1$ теоремы можно представить следующим образом:

$$W_{\text{уг}}^{\text{МЦД}}(z) = [z^{n_{\text{МЦД}}} - V(z)(z - 1)^{r+1}] / z^{n_{\text{МЦД}}}. \quad (13)$$

Следовательно, в соответствии с (10) ошибка МЦД с ПФ (12), обусловленная полиномиальным сигналом степени r , равна нулю. *Теорема доказана.*

В соответствии с постановкой задачи, для заданного цифрового датчика, имеющего запаздывание $\tau = qT$, необходимо найти передаточную функцию $W_{\text{АБК}}(z)$ (1) астатического быстродействующего корректора, при подключении которого к заданному датчику (рисунок 2) ошибка модернизированного датчика

$\mathcal{E}_{\text{МЦД},k}$, вызванная полиномиальным сигналом g_k степени r , будет равна нулю по окончании переходного процесса. Однако с практической точки зрения, целесообразнее находить сначала, передаточную функцию $W_{\text{yg}}^{\text{МЦД}}(z)$ модернизированного датчика по заданным к нему требованиям, а затем, используя вытекающее из (1) и (2) выражение:

$$W_{\text{АБК}}(z) = z^q W_{\text{yg}}^{\text{МЦД}}(z), \quad (14)$$

находить передаточную функцию $W_{\text{АБК}}(z)$ необходимого корректора.

Итак, пусть необходимо синтезировать МЦД, имеющий нулевую ошибку и перерегулирование $\sigma\%$ при входном полиномиальном сигнале g_k степени r . С этой целью вычисляются значения $\nu_{\text{МЦД}} = r + 1$, $\lambda = \sigma\% / 100\%$ и $\mu_{\text{МЦД}} = 1 + q$, а затем решается уравнение (11), что даёт соответствующие $n_{\text{МЦД}}$ и полином $V(z)$. При этом, если $\nu_{\text{МЦД}} = 1$, то по (11) легко установить, что при любом значении $n_{\text{МЦД}} \geq q + 2$ полином $V(z) = z^{n_{\text{МЦД}}-1} + z^{n_{\text{МЦД}}-2} + \dots + z^2 + z - \lambda$. Это позволяет при $\nu_{\text{МЦД}} = 1$ найти соответствующую передаточную функцию СЦД, не решая уравнения (11), а затем по формуле (14) найти передаточную функцию необходимого АБК. Покажем, порядок действий в этом частном случае.

Пусть запаздывание исходного цифрового датчика $\tau_{\text{ЦД}} = \tilde{q}T$, на его вход поступает сигнал $g_{\text{ЦД}}(t) = g_0 l(t)$, т.е. $r = 0$, а необходимый порядок астатизма $\nu_{\text{МЦД}} = 0 + 1 = 1$. Тогда при $\tilde{n}_{\text{МЦД}} = \tilde{q} + 2$ необходимое значение $\tilde{\mu}_{\text{МЦД}} = \tilde{q} + 1$, полином $\tilde{V}(z) = z^{\tilde{q}+1} + z^{\tilde{q}} + \dots + z - \lambda$, а передаточная функция соответствующего МЦД записывается по (10):

$$W_{\text{yg}}^{\text{МЦД}}(z) = \frac{z^{\tilde{q}+2} - (z^{\tilde{q}+1} + z^{\tilde{q}} + \dots + z - \lambda)(z - 1)}{z^{\tilde{q}+2}}. \quad (15)$$

С помощью равенства $(z^n - z) = (z^{n-1} + z^{n-2} + \dots + z^2 + z)(z - 1)$, справедливого при любом $n \geq 2$, передаточная функция (15) приводится к виду:

$$W_{\text{yg}}^{\text{МЦД}}(z) = \frac{(1 + \lambda)z - \lambda}{z^{\tilde{q}+2}}. \quad (16)$$

Полученная передаточная функция МЦД (16) соответствует (6) и (12) при $n = \tilde{n}_{\text{МЦД}} = \tilde{q} + 2$, $\mu = \tilde{\mu}_{\text{МЦД}} = \tilde{q} + 1$, поскольку в данном случае $n - \mu = \tilde{n}_{\text{МЦД}} - \tilde{\mu}_{\text{МЦД}} = 1$. Чтобы убедиться в том, что полученный МЦД обеспечивает требуемые свойства, найдем его реакцию на $g_{\text{ЦД}}(t) = g_0 1(t)$.

Модернизированному цифровому датчику с передаточной функцией (16) соответствует разностное уравнение $y_k = (1 + \lambda)g_{\text{ЦД},k-\tilde{q}-1} - \lambda g_{\text{ЦД},k-\tilde{q}-2}$. При $g_{\text{ЦД},k} = g_0 1(kT)$ и $k = 0, 1, 2, \dots$, получим:

$$y_0 = 0; y_1 = 0; \dots; y_{\tilde{q}} = (1 + \lambda)g_{\text{ЦД},-1} - \lambda g_{\text{ЦД},-2} = 0;$$

$$y_{\tilde{q}+1} = (1 + \lambda)g_{\text{ЦД},0} - \lambda g_{\text{ЦД},-1} = (1 + \lambda)g_0;$$

$$y_{\tilde{q}+2} = (1 + \lambda)g_{\text{ЦД},1} - \lambda g_{\text{ЦД},0} = (1 + \lambda)g_0 - \lambda g_0 = g_0;$$

$$y_{\tilde{q}+3} = (1 + \lambda)g_{\text{ЦД},2} - \lambda g_{\text{ЦД},1} = (1 + \lambda)g_0 - \lambda g_0 = g_0.$$

Как видно, МЦД действительно имеет запаздывание $\tau = (\tilde{q} + 1)T$, обусловленное исходным ЦД и подключенным после него АБК, перерегулирование равно $\sigma = \lambda \cdot 100\%$ и конечный переходной процесс длительностью $(\tilde{q} + 2)T$, а его ошибка, очевидно, равна нулю.

Передаточная функция АБК в этом случае, определенная по формуле (14), имеет вид $W_{\text{АБК}}(z) = [(1 + \lambda)z - \lambda] / z^2$. Как видно, относительный порядок этого АБК $\mu_{\text{АБК}} = 1$, т.е. он может быть реализован на цифровых вычислительных элементах. Соответствующий алгоритм АБК согласно (1) имеет вид $y_{\text{АБК},k} = (1 + \lambda)y_{\text{АЦД},k-1} - \lambda y_{\text{АЦД},k-2}$, $k = 0, 1, 2, \dots$.

Задача синтеза АБК существенно осложняется в тех случаях, когда необходимо обеспечить второй или более высокий порядок астатизма при малом перерегулировании, т.е. в тех случаях, когда $\nu_{\text{МЦД}} = 2$ и более при $\sigma \in [0 \div 40\%]$.

В 70-х годах прошлого века В.А.Никольский нашел (без использования уравнения (11)) полиномы $V(z) = V_N(z)$ из выражения (10) для $\nu_{\text{МЦД}} = 2$, при которых переходные процессы, вызванные

ступенчатым сигналом, имеют постоянное перерегулирование $\sigma\%$ равное 10%, 15%, 20%, 25%, 33% [31]. Существенно, что на практике именно значения 5%, 10%, 15% и 20% стремятся придать проектируемым системам управления. Однако В.А.Никольский рассматривал дискретные системы только с относительным порядком $\mu=1$, т.е. без учета запаздывания. В тоже время относительные порядки $\mu_{\text{мцд}}$ цифровых датчиков, из-за влияния запаздывания по времени, должны иметь бóльшие значения относительного порядка [6 – 8].

В соответствии с теоремой 1 для построения модернизированного цифрового датчика в общем случае достаточно по заданным значениям r , $\sigma\%$ и q вычислить $\nu_{\text{мцд}} = r + 1$, $\lambda = \sigma\% / 100\%$, $\mu_{\text{мцд}} = 1 + q$, а затем решить уравнение (11) относительно степени $n_{\text{мцд}}$. С этой целью его целесообразно представить таким образом:

$$z^{n_{\text{мцд}}} - (1 + \lambda)z^{n_{\text{мцд}} - \mu_{\text{мцд}}} + \lambda = V(z)(z - 1)^{\nu_{\text{мцд}}}. \quad (17)$$

В таблице 1 приведены порядки $n_{\text{мцд}}$ датчика при $\nu_{\text{мцд}} = 2$, $\mu_{\text{мцд}} = 1 + q$ и некоторых значениях $\sigma\%$, найденные путем решения в MATLAB уравнения (17) при $q \leq 3$.

Замечание. Если $\sigma = 15\%$, то в уравнениях (11) и (17) необходимо полагать величину $\lambda = 1/7$, а если $\sigma = 33\%$, то брать $\lambda = 100/3$.

Таблица 1. Порядок датчика $n_{\text{мцд}}$

Порядок астатизма $\nu_{\text{мцд}} = 2$				
$\sigma\%$ \ q	0	1	2	3
5 %	21	42	63	84
10 %	11	22	33	44
15 %	8	16	24	32
20 %	6	12	18	24
25 %	5	10	15	20
33 %	4	8	12	16

Из таблицы 1 следует, что порядок МЦД очень резко возрастает при уменьшении перерегулирования $\sigma\%$ и увеличении запаздывания q в исходном датчике, поэтому если МЦД должен иметь малое перерегулирование, то период T необходимо уменьшать, что требует применения более быстродействующих вычислительных средств.

Подчеркнём, что значения, приведенные во втором столбце таблицы 1 (при $q = 0$), совпадают со значениями, приведенными в [31]. Случай, когда перерегулирование $\sigma = 5\%$, в этой работе не рассматривался. Отметим также, что уравнения (11) и (17) записаны для случая полиномиальных сигналов, причем, если $v_{\text{МЦД}} \geq 3$, $0 \leq \lambda \leq 0,4$ и $\mu_{\text{МЦД}} \geq 1$, то эти уравнения не имеют решений относительно целых $n_{\text{МЦД}}$. Проблема разрешимости аналогичных уравнений при других формах входных регулярных сигналов является открытой.

Отметим, что из выражений (12), (14) и условия $\mu_{\text{МЦД}} = 1 + q$ следует, что z-изображение выходной величины АБК определяется выражением $y_{\text{АБК}}(z) = [(1 + \lambda)z^{-1} - \lambda z^{q - n_{\text{МЦД}}}]y_{\text{АЦИ}}(z)$. Этому равенству соответствует следующий алгоритм работы АБК:

$$y_{\text{АБК},k} = \begin{cases} 0, & k = 0, \\ (1 + \lambda)y_{\text{АЦИ},k-1}, & 1 \leq k < n_{\text{МЦД}} - q, \\ (1 + \lambda)y_{\text{АЦИ},k-1} - \lambda y_{\text{АЦИ},k-n_{\text{МЦД}}+q}, & n_{\text{МЦД}} - q \leq k, \\ & k = 0, 1, 2, \dots \end{cases} \quad (18)$$

Как видно, в процессе функционирования АБК на каждом k -ом периоде вызываются из памяти процессора максимум два коэффициента $1 + \lambda$ и λ ; два значения $y_{\text{АЦИ},i}$, и выполняются максимум две операции умножения и одна операция сложения с ссылкой результата в память процессора. При этом в памяти процессора необходимо иметь две ячейки для хранения коэффициентов $1 + \lambda$ и λ , а также $n_{\text{МЦД}} - q + 1$ ячеек для хранения $y_{\text{АЦИ},i}$. Очевидно, алгоритм (18) является очень простым, его выполнение требует небольших затрат машинного времени, поэтому он может быть реализован современными процессорами за период очень малой длительности. При этом реализация АБК может быть осуществлена либо в виде отдельного блока, подключаемого к выходу

цифрового датчика, или же путем небольшой коррекции алгоритма работы вычислителя этого датчика.

Подчеркнём, что предложенный подход к устранению ошибок цифровых датчиков на основе их интеллектуализации может существенно упростить реализацию систем автоматизации, создаваемых в рамках парадигм Индустрия 4 и Индустрия 5 [32]. Соответствующие системы собирают очень большие объёмы данных, поэтому интеллектуализация датчиков и устранение их ошибок существенно уменьшает объём информации, передаваемый по каналам связи информационно-управляющих подсистем.

Эффективность предложенного астатического быстродействующего корректора и метода синтеза модернизированных цифровых датчиков покажем на численном примере.

4. Пример. Найти АБК, для устранения ошибки цифрового датчика, обусловленной его запаздыванием по времени $\tau = 2T$ при $T = 0,2$ с и входными полиномиальными сигналами, степень которых не превышает единицы. Переходной процесс МЦД должен иметь перерегулирование не более 15%.

В данном случае, очевидно, $q = 2$, $r = 1$ и $\sigma \leq 15\%$, поэтому в соответствии с теоремой 1, модернизированный датчик должен иметь порядок астатизма $\nu_{\text{МЦД}} = r + 1 = 2$, а относительный порядок $\mu_{\text{МЦД}} = 1 + q = 3$. Поскольку перерегулирование $\sigma \leq 15\%$, то в соответствии с приведенным выше замечанием необходимо взять $\lambda = 1/7$. Так как $\nu_{\text{МЦД}} = 2$, то вместо решения уравнения (17) воспользуемся таблицей 1, по которой при $q = 2$ находим $n_{\text{МЦД}} = 24$. Тогда по формулам (12) и (14) получим передаточные функции МЦД и АБК:

$$\begin{aligned} W_{\text{уг}}^{\text{МЦД}}(z) &= (1 + 1/7)z^{21} - 1/7 / z^{24}, \\ W_{\text{АБК}}(z) &= (1 + 1/7)z^{21} - 1/7 / z^{22}. \end{aligned} \quad (19)$$

На рисунках 4 и 5 показаны полученные в MATLAB по передаточной функции $W_{\text{уг}}^{\text{МЦД}}(z)$ (19) графики изменения значений выходного сигнала $y_k = y_{\text{МЦД}}(kT)$ модернизированного цифрового датчика (рисунок 2), чувствительный элемент, интерфейс и найденный АБК которого работают с периодом $T = 0,2$ с. На рисунке 4 приведены значения выходного сигнала $y_k = y_{\text{МЦД}}(kT)$ и график входного сигнала

$g(t) = g_{\text{ЦД}}(t)$, который равен $1(t)$ с аддитивной случайной помехой $\psi(t)$, т.е. $g_{\text{ЦД}}(t) = 1(t) + \psi(t)$.

На рисунке 5 показаны дискретные значения $g_k = g_{\text{ЦД}}(kT) = 0,6k$ входного непрерывного сигнала $g_{\text{ЦД}}(t) = 3t$ и выходного сигнала $y_k = y_{\text{МЦД}}(kT)$ модернизированного цифрового датчика в интервале времени t от нуля до шести секунд.

По рисункам 4 и 5 можно легко заключить, что модернизированный цифровой датчик имеет запаздывание $\tau = 0,6\text{с}$ (равное $3T$), перерегулирование 15%; его переходный процесс длится ровно 4,8 секунды, т.е. $24T$, а ошибка и при ступенчатом, и при линейном сигналах, несмотря на наличие запаздывания, равна нулю.

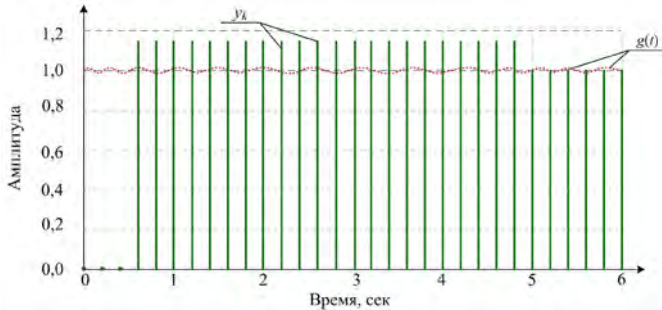


Рис. 4. Сигналы датчика при ступенчатом входном сигнале

Как видно, подключение АБК с передаточной функцией $W_{\text{АБК}}(z)$ (19) к заданному цифровому датчику обеспечивает требуемые свойства модернизированного датчика.

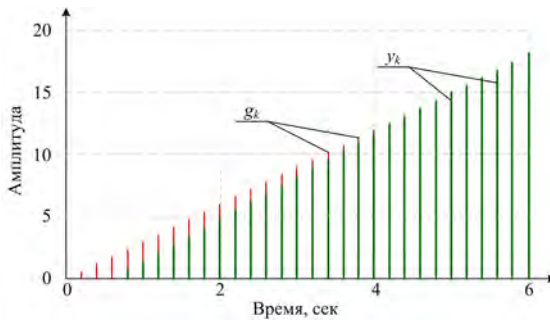


Рис. 5. Выход датчика при линейном входном сигнале

Ошибка датчика, предложенного в данной работе типа, является нулевой при произвольных, но допустимых значениях ступенчатого и линейного входного воздействия. Чтобы это показать, рассчитаем значения выходного сигнала того же датчика с ПФ $W_{yg}^{\text{МЦД}}(z)$ (19), но при действии его входе сигнала $g_{\text{ЦД}}(t) = 2t$, т.е. другой интенсивности. Передаточной функции МЦД (19) соответствует разностное уравнение:

$$y_k = (1 + 1/7)g_{\text{ЦД},k-3} - (1/7)g_{\text{ЦД},k-24}, k = 0, 1, 2, \dots, \quad (20)$$

а дискретные значения сигнала $g_{\text{ЦД}}(t) = 2t$ при $T = 0,2$ с описываются функцией $g_{\text{ЦД},k} = 0,4k$. Практически, уравнение (20) является алгоритмом работы модернизированного датчика.

Выполняя 31 шаг этого алгоритма, т.е. полагая в выражении (20) $k = 0, 1, 2, \dots, 30$, получим соответствующие значения $y_{\text{МЦД}}(kT)$. Значения k , $y_k = y_{\text{МЦД}}(kT)$, а также $g_k = g_{\text{ЦД}}(kT)$, приведены в таблице 2.

По таблице 2, как и по рисункам 4 и 5, хорошо видно, что запаздывание в датчике равно $3T$. При $k = 24$ переходной процесс заканчивается, и выходные значения модернизированного датчика становятся в точности равными значениям входного сигнала, т.е. ошибка датчика действительно является нулевой при различной интенсивности входного сигнала. Отметим, что при отсутствии синтезированного корректора ошибка рассматриваемого датчика, обусловленная запаздыванием, при $g_{\text{ЦД}}(t) = 2t$, равна $0,8$.

Подчеркнём также, что столь высокий порядок датчика (в данном случае $n_{\text{МЦД}} = 24$) обусловлен значительным запаздыванием в цифровом датчике $\tau = 2T$, а главное, условием малого перерегулирования. Например, при допустимом перерегулировании 25% и запаздывании $\tau = 2T$ порядок $n_{\text{МЦД}} = 10$, а при $\tau = T$ порядок $n_{\text{МЦД}} = 5$.

Временная длительность переходного процесса МЦД в основном определяется быстродействием применяемых в датчике вычислительных средств. Например, если период работы цифрового вычислителя датчика равен не $0,2$ с, а $0,05$ с, то длительность переходного процесса МЦД с ПФ (19) составляет всего $1,2$ с.

Таблица 2. Выходной y_k и входной g_k сигналы МЦД

k	y_k	$g_k = 0,4k$
0	$(1+1/7) \cdot 0,4 \cdot 0 - (1/7) \cdot 0,4 \cdot 0 = 0$	0
1	$(1+1/7) \cdot 0,4 \cdot 0 - (1/7) \cdot 0,4 \cdot 0 = 0$	0,4
2	$(1+1/7) \cdot 0,4 \cdot 0 - (1/7) \cdot 0,4 \cdot 0 = 0$	0,8
3	$(1+1/7) \cdot 0,4 \cdot 0 - (1/7) \cdot 0,4 \cdot 0 = 0$	1,2
4	$(1+1/7) \cdot 0,4 \cdot 1 - (1/7) \cdot 0,4 \cdot 0 = 0,45714$	1,6
5	$(1+1/7) \cdot 0,4 \cdot 2 - (1/7) \cdot 0,4 \cdot 0 = 0,91429$	2,0
6	$(1+1/7) \cdot 0,4 \cdot 3 - (1/7) \cdot 0,4 \cdot 0 = 1,37143$	2,4
⋮	⋮	⋮
22	$(1+1/7) \cdot 0,4 \cdot 19 - 0 = 8.68571$	8,8
23	$(1+1/7) \cdot 0,4 \cdot 20 - 0 = 9.14286$	9,2
24	$(1+1/7) \cdot 0,4 \cdot 21 - (1/7) \cdot 0,4 \cdot 0 = 9,60$	9,6
25	$(1+1/7) \cdot 0,4 \cdot 22 - (1/7) \cdot 0,4 \cdot 1 = 10,00$	10,0
⋮	⋮	⋮
29	$(1+1/7) \cdot 0,4 \cdot 26 - (1/7) \cdot 0,4 \cdot 5 = 11,6$	11,6
30	$(1+1/7) \cdot 0,4 \cdot 27 - (1/7) \cdot 0,4 \cdot 6 = 12,0$	12,0

5. Заключение. В работе предложен способ устранения ошибок цифровых датчиков, вызванных временной задержкой, на основе их модернизации. Это достигается либо путем включения на их выходе астатического быстродействующего корректора, либо коррекцией алгоритма работы их вычислителя. Предложенный способ позволяет обеспечить нулевую ошибку при постоянном и линейном входных сигналах, а также постоянное перерегулирование и конечный переходный процесс. Длительность переходного процесса в основном определяется быстродействием вычислителя цифрового датчика. Параметры необходимого астатического быстродействующего корректора определяются аналитически по запаздыванию исходного цифрового датчика и допустимому перерегулированию. В общем случае запаздывание интеллектуального датчика может быть равным одному или нескольким периодам дискретизации. При этом порядок модернизированного датчика резко возрастает с увеличением запаздывания и в особенности при уменьшении допустимого перерегулирования, что может привести к большой длительности переходного процесса модернизированного датчика. Однако этот эффект легко устраняется за счет повышения быстродействия применяемых процессоров.

Полученные результаты могут использоваться при разработке высокоточных цифровых датчиков различных физических величин.

Литература

1. Liu C., Liu J.-G., Kennel R. Accuracy improvement of rotational speed sensors by novel signal processing method // *Journal of Physics: Conference Series*. IOP Publishing. 2018. vol. 1065. no. 7. pp. 072013. DOI: 10.1088/1742-6596/1065/7/072013.
2. Cao M., Yang J. The Effect of the approximation method for large time delay process on the performance of IMC-PID controller // *Processing of the International Conference on Control, Power, Communication and Computing Technologies (ICCPCC'2018)*. 2018. pp. 73–77. DOI: 10.1109/ICPCC.2018.8574299.
3. Azzoni P., Caminale G., Carratù M., Iacono S.D., Fenza G., Gallo N., Liguori C., Londero E., Pietrosanto A., Rebella N. Distributed Smart Measurement Architecture for Industrial Automation // *arXiv preprint arXiv:2107.14272*. 2021. pp. 1–6. DOI: 10.48550/arXiv.2107.14272.
4. Zhang Y., Zhang S., Yin Y. Adaptive Fault Diagnosis for continuous Time-delay Repetitive System Subject to sensor Fault // *Processing of the International Conference on Advanced Mechatronic Systems (ICAMechS'2015)*. 2015. pp. 456–460. DOI: 10.1109/ICAMechS.2015.7287154.
5. Annaby M.N., Al-Abdi I.A., Abou-Dina M.S., Ghaleb A.F. Regularized sampling reconstruction of signals in the linear canonical transform domain // *Signal Processing*. 2022. vol. 198. pp. 108569. DOI: 10.1016/j.sigpro.2022.108569.
6. Butler P. Antialiasing filtering considerations for high precision SAR analog-to-digital converters // *Analog Dialogue*. 2018. vol. 52. no. 3. pp. 54–59.
7. Bakshi U.A., Bakshi L.A.V. *Electronic measurement and instrumentation* // Technical Publications. 2020. 700 p.
8. Samoylov L.K., Prokopenko N.N., Bugakova A.V. Selection of the band-pass range of the normalizing signal transducer of the sensing element in the instrumentation and control systems // *Processing of the IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT'2018)*. 2018. pp. 1–3. DOI: 10.1109/ICSICT.2018.8564937.
9. Ballard Z., Brown C., Madni A. M., Ozcan A. Machine learning and computation-enabled intelligent sensor design // *Nature Machine Intelligence*. 2021. vol. 3. pp. 556–565. DOI: 10.1038/s42256-021-00360-9.
10. Pornsarayouth S., Wongsaisuwan M. Sensor fusion of delay and non-delay signal using Kalman filter with moving covariance // *Proceedings of the IEEE International Conference on Robotics and Biomimetics*. 2009. pp. 2045–2049. DOI: 10.1109/ROBIO.2009.4913316.
11. Chachula K., Słojewski T.M., Nowak R. Multisensor data fusion for localization of pollution sources in wastewater networks // *Sensors*. 2022. vol. 22. no. 1. pp. 1–19. DOI: 10.3390/s22010387.
12. Wang W., Cai F., Cui F., Yang F. Observer-based robust reliable H_∞ control for uncertain time-delay discrete-time systems in the presence of sensor failure // *Proceedings of International Conference on Control, Automation, Robotics and Vision*. 2008. pp. 1489–1493. DOI: 10.1109/ICARCV.2008.4795744.
13. You F., Li H., Zhang Y., Guan S. A novel sensor fault diagnosis approach for time-varying delay systems with non-linear uncertainty // *Transactions of the Institute of Measurement and Control*. 2017. vol. 39. no. 7. pp. 1114–1120. DOI: 10.1177/0142331216629200.

14. Zemzemi A., Kamel M., Toumi A., Farza M. Robust integral-observer-based fault estimation for Lipschitz nonlinear systems with time-varying uncertainties // Transactions of the Institute of Measurement and Control. 2019. vol. 41. no. 7. pp. 1965–1974. DOI: 10.1177/0142331218791227.
15. Liu X., Sheng H. Active fault tolerant control of uncertain robotic system based on observer and sliding mode // IFAC-PapersOnLine. 2022. vol. 55. no. 1. pp. 598–603. DOI: 10.1016/j.ifacol.2022.04.098.
16. Stojanovic S.B., Debelkovic D.L., Antic D.S. Finite-time stability and analysis of discrete time-delay systems using discrete convolution of delayed states // Facta universitatis, Series: Automatic Control and Robotics. 2014. vol. 14. no. 32. pp. 147–158.
17. Tao R., Li X.M., Wang Y. Time delay estimation of chirp signals in the fractional Fourier domain // IEEE Transactions on Signal Processing. 2009. vol. 57. no. 3. pp. 2852–2855. DOI: 10.1109/TSP.2009.2020028.
18. Pfeiffer J., Wu X., Ayadi A. Evaluation of three different approaches for automated time delay estimation for distributed sensor systems of electric vehicles // Sensors. 2020. vol. 20. no. 2. pp. 1–18. DOI:10.3390/s20020351.
19. Chaudhary G., Jeong Y. A design of power divider with negative group-delay characteristic // IEEE Microwave Wireless Components Letters. 2015. vol. 25. no. 6. pp. 394–396. DOI: 10.1109/LMWC.2015.2421280.
20. Wan F., Miao X., Ravelo B., Yuan Q., Cheng J., Ji Q., Ge J. Design of multi-scale negative group delay circuit for sensors signal time-delay cancellation // IEEE Sensors Journal. 2019. vol. 19. no. 19. pp. 8951–8962. DOI: 10.1109/JSEN.2019.2921834.
21. Yuan A., Fang S., Wang Z., Liu H. A novel multifunctional negative group delay circuit for realizing band-pass, high-pass and low-pass // Electronics. 2021. vol. 10. no. 14. pp. 1–12. DOI: 10.3390/electronics10141742.
22. Мусаев Р.Ш., Здобнов С.А., Трофимов А.А., Цибизов П.Н. Этапы развития и основные направления совершенствования датчиков физических величин для ракетно-космической техники // Датчики и системы. 2019. № 12. С. 30–38.
23. Плясовских А.П., Разумов А.В., Саута О.И. Требования к функциям и техническим характеристикам бортовых авиационных навигационно-информационных систем // Труды СПИИРАН. 2012. № 23. С. 381–391.
24. How to Increase the Analog-to-Digital Converter Accuracy in an Application? Freescale Semiconductor, Inc., AN5250, Rev. 0, 01/2016, 20 p. Available at: <https://cache.nxp.com/docs/en/application-note/AN5250.pdf> (accessed 10.02.2022).
25. Oppenheim A.V., Schaffer R.W. Discrete-Time Signal Processing: Third edition // Pearson Higher Education. 2010. 1108 p.
26. Gaiduk A.R., Prokopenko N.N., Bugakova A.V. Accuracy Increase of Discrete Sensors with Time Delay // IEEE Sensors Journal. 2020. vol. 20. no. 19. pp. 11400–11404. DOI: 10.1109/JSEN.2020.2996079.
27. Busek J., Zitek P., Vyhldal T. Astatism analysis of time delay controllers towards effective anti-windup shemes // Proceedings of the International Conference on Process Control (PC'2019). 2019. pp. 74–79. DOI: 10.1109/PC.2019.8815283.
28. Гайдук А.Р., Плаксиенко Е.А. Анализ и аналитический синтез цифровых систем управления // СПб.: Лань, 2018. 272 с.
29. Ким Д.П. Теория автоматического управления. Линейные системы // ЮРАЙТ, 2022. 312 с.
30. Chen C.T. Linear System Theory and Design: Third edition // New York, Oxford: University Press. 1999. 334 p.
31. Никольский В.А. Об одном способе обеспечения астатизма второго порядка в дискретной системе комбинированного управления с компенсацией возмущений // Известия вузов Электромеханика. 1978. № 11. С. 1168–1173.

32. Chander B., Pal S., Di D., Buuya R. Artificial intelligence-based internet of things for industry 5.0 // Artificial intelligence-based internet of things systems. 2022. pp. 3–45. DOI: 10.1007/978-3-030-87059-1_1.

Гайдук Анатолий Романович — д-р техн. наук, профессор, ведущий научный сотрудник, кафедра «системы автоматического управления», целевая поисковая лаборатория прорывных интеллектуальных технологий группового управления робототехническими комплексами фонда перспективных исследований, Южный федеральный университет. Область научных интересов: анализ и проектирование линейных, нелинейных, непрерывных и дискретных систем управления техническими объектами, процессами, роботами и робототехническими комплексами. Число научных публикаций — 450. gaiduk_2003@mail.ru; переулок Некрасовский, 44, 347922, Таганрог, Россия; р.т.: +7(8634)371-689.

Прокопенко Николай Николаевич — д-р техн. наук, профессор, заслуженный изобретатель российской федерации, заведующий кафедрой, старший научный сотрудник, кафедра "информационные системы и радиотехника", отдел управления научных исследований, Донской государственной технической университет. Область научных интересов: элементы и устройства вычислительной техники и систем управления. Число научных публикаций — 800. prokopenko@sssu.ru; площадь Гагарина, 1, 344000, Ростов-на-Дону, Россия; р.т.: +7(928)120-1984.

Бугакова Анна Витальевна — канд. техн. наук, доцент, младший научный сотрудник, кафедра "информационные системы и радиотехника", отдел управления научных исследований, Донской государственной технической университет. Область научных интересов: элементы и устройства вычислительной техники и систем управления. Число научных публикаций — 250. annabugakova.1992@mail.ru; площадь Гагарина, 1, 344000, Ростов-на-Дону, Россия; р.т.: +7(909)416-9380.

Поддержка исследований. Исследование выполнено за счет гранта Российского научного фонда (проект № 18-79-10109-П).

A. GAIDUK, N. PROKOPENKO, A. BUGAKOVA
**ERRORS COMPENSATION CAUSED BY TIME DELAY OF
DIGITAL SENSORS**

Gaiduk A., Prokopenko N., Bugakova A. Errors Compensation Caused by Time Delay of Digital Sensors.

Abstract. The paper is devoted to improving the accuracy of digital sensors with a time lag. The relevance of the topic is due to the widespread use of sensors of this type, which is largely due to a sharp increase in the requirements for measurement accuracy. The timeliness is associated also with the extensive application of digital technologies for information processing in control systems, communications, monitoring and many others. To eliminate the errors caused by the time delay of digital sensors, it is suggested to use an astatic high-speed corrector. The applicability of this corrector is justified by the properties of discrete-time dynamical systems. In this regard, at first, the conditions are considered under which the discrete systems are physically realizable and have a finite duration of the transient since in this latter case they are the fastest. It is also shown that in order to measure a polynomial signal of limited intensity with zero error in steady-state mode, the astatic order of the sensor must be one value greater than the degree of this signal. Based on the above conditions, the main result of the article is proved – a theorem in which the conditions for the existence of the astatic high-speed corrector are established. When this corrector is switched on at the output of the digital sensor or when its software is being corrected an upgraded sensor is formed, the error of which will be zero in steady-state mode. This is due to the fact that the corrector eliminates the error of the digital sensor caused by its time delay, which is assumed to be multiple of the sampling period. The order of the corrector as a system is determined by the integer solution of the equation obtained in the work, which relates the degree of the measured polynomial signal, the time delay of the digital sensor, the permissible overshoot of the upgraded sensor and the relative order of the desired corrector. This equation is solved for the cases, where the degree of the measured signals is not greater than one, the overshoot is equal to the frequently assigned values, and the time delay does not exceed four sampling periods. The corresponding order of the upgraded sensor is given in tabular form. This makes it possible to find the required corrector without solving the shown equation in many cases. The effectiveness of the suggested approach with respect to improving the accuracy of digital sensors is shown by a numerical example. The zero error value of the upgraded sensor is confirmed both by computer simulation and numerical calculation. The results obtained can be used in the development of high-precision digital sensors of various physical quantities.

Keywords: discrete system, digital filter, digital sensor, order of system, relative order of system, time delay, astatic, transient, overshoot, finite-time transient.

References

1. Liu C., Liu J.-G., Kennel R. Accuracy improvement of rotational speed sensors by novel signal processing method. *Journal of Physics: Conference Series*. IOP Publishing. 2018. vol. 1065. no. 7. pp. 072013. DOI: 10.1088/1742-6596/1065/7/072013.
2. Cao M., Yang J. The Effect of the approximation method for large time delay process on the performance of IMC-PID controller. *Processing of the International Conference on Control, Power, Communication and Computing Technologies (ICCPCT'2018)*. 2018. pp. 73–77. DOI: 10.1109/ICCPCT.2018.8574299.

3. Azzoni P., Caminale G., Carratù M., Iacono S.D., Fenza G., Gallo N., Liguori C., Londero E., Pietrosanto A., Rebella N. Distributed Smart Measurement Architecture for Industrial Automation. arXiv preprint arXiv:2107.14272. 2021. pp. 1–6. DOI: 10.48550/arXiv.2107.14272.
4. Zhang Y., Zhang S., Yin Y. Adaptive Fault Diagnosis for continuous Time-delay Repetitive System Subject to sensor Fault. Processing of the International Conference on Advanced Mechatronic Systems (ICAMechS'2015). 2015. pp. 456–460. DOI: 10.1109/ICAMechS.2015.7287154.
5. Annaby M.N., Al-Abdi I.A., Abou-Dina M.S., Ghaleb A.F. Regularized sampling reconstruction of signals in the linear canonical transform domain. Signal Processing. 2022. vol. 198. pp. 108569. DOI: 10.1016/j.sigpro.2022.108569.
6. Butler P. Antialiasing filtering considerations for high precision SAR analog-to-digital converters. Analog Dialogue. 2018. vol. 52. no. 3. pp. 54–59.
7. Bakshi U.A., Bakshi L.A.V. Electronic measurement and instrumentation. Technical Publications. 2020. 700 p.
8. Samoylov L.K., Prokopenko N.N., Bugakova A.V. Selection of the band-pass range of the normalizing signal transducer of the sensing element in the instrumentation and control systems. Processing of the IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT'2018). 2018. pp. 1–3. DOI: 10.1109/ICSICT.2018.8564937.
9. Ballard Z., Brown C., Madni A. M., Ozcan A. Machine learning and computation-enabled intelligent sensor design. Nature Machine Intelligence. 2021. vol. 3. pp. 556–565. DOI: 10.1038/s42256-021-00360-9.
10. Pornsarayouth S., Wongsaisuwan M. Sensor fusion of delay and non-delay signal using Kalman filter with moving covariance. Proceedings of the IEEE International Conference on Robotics and Biomimetics. 2009. pp. 2045–2049. DOI: 10.1109/ROBIO.2009.4913316.
11. Chachula K., Słojewski T.M., Nowak R. Multisensor data fusion for localization of pollution sources in wastewater networks. Sensors. 2022. vol. 22. no. 1. pp. 1–19. DOI: 10.3390/s22010387.
12. Wang W., Cai F., Cui F., Yang F. Observer-based robust reliable H_∞ control for uncertain time-delay discrete-time systems in the presence of sensor failure. Proceedings of International Conference on Control, Automation, Robotics and Vision. 2008. pp. 1489–1493. DOI: 10.1109/ICARCV.2008.4795744.
13. You F., Li H., Zhang Y., Guan S. A novel sensor fault diagnosis approach for time-varying delay systems with non-linear uncertainty. Transactions of the Institute of Measurement and Control. 2017. vol. 39. no. 7. pp. 1114–1120. DOI: 10.1177/0142331216629200.
14. Zemzemi A., Kamel M., Toumi A., Farza M. Robust integral-observer-based fault estimation for Lipschitz nonlinear systems with time-varying uncertainties. Transactions of the Institute of Measurement and Control. 2019. vol. 41. no. 7. pp. 1965–1974. DOI: 10.1177/0142331218791227.
15. Liu X., Sheng H. Active fault tolerant control of uncertain robotic system based on observer and sliding mode. IFAC-PapersOnLine. 2022. vol. 55. no. 1. pp. 598–603. DOI: 10.1016/j.ifacol.2022.04.098.
16. Stojanovic S.B., Debelkovic D.L., Antic D.S. Finite-time stability and analysis of discrete time-delay systems using discrete convolution of delayed states. Facta universitatis, Series: Automatic Control and Robotics. 2014. vol. 14. no. 32. pp. 147–158.
17. Tao R., Li X.M., Wang Y. Time delay estimation of chirp signals in the fractional Fourier domain. IEEE Transactions on Signal Processing. 2009. vol. 57. no. 3. pp. 2852–2855. DOI: 10.1109/TSP.2009.2020028.

18. Pfeiffer J., Wu X., Ayadi A. Evaluation of three different approaches for automated time delay estimation for distributed sensor systems of electric vehicles. *Sensors*. 2020. vol. 20. no. 2. pp. 1–18. DOI:10.3390/s20020351.
19. Chaudhary G., Jeong Y. A design of power divider with negative group-delay characteristic. *IEEE Microwave Wireless Components Letters*. 2015. vol. 25. no. 6. pp. 394–396. DOI: 10.1109/LMWC.2015.2421280.
20. Wan F., Miao X., Ravelo B., Yuan Q., Cheng J., Ji Q., Ge J. Design of multi-scale negative group delay circuit for sensors signal time-delay cancellation. *IEEE Sensors Journal*. 2019. vol. 19. no. 19. pp. 8951–8962. DOI: 10.1109/JSEN.2019.2921834.
21. Yuan A., Fang S., Wang Z., Liu H. A novel multifunctional negative group delay circuit for realizing band-pass, high-pass and low-pass. *Electronics*. 2021. vol. 10. no. 14. pp. 1–12. DOI: 10.3390/electronics10141742.
22. Musaev R.Sh., Zdobnov S.A., Trofimov A.A., Tsbizov P.N. [Stages of development and main directions for improving sensors of physical quantities for rocket and space technology]. *Datchiki i sistemy – Sensors and systems*. 2019. no. 12. pp. 30–38. (In Russ.).
23. Plyasovskih A.P., Razumov A.V., Sauta O.I. [Requirements to functions and technical characteristics of airborne information and navigating systems]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2012. no. 23. pp. 381–391. (In Russ.).
24. How to Increase the Analog-to-Digital Converter Accuracy in an Application? *Freescale Semiconductor, Inc., AN5250, Rev. 0, 01/2016, 20 p.* Available at: <https://cache.nxp.com/docs/en/application-note/AN5250.pdf> (accessed 10.02.2022).
25. Oppenheim A.V., Schaffer R.W. *Discrete-Time Signal Processing: Third edition*. Pearson Higher Education. 2010. 1108 p.
26. Gaiduk A.R., Prokopenko N.N., Bugakova A.V. Accuracy Increase of Discrete Sensors with Time Delay. *IEEE Sensors Journal*. 2020. vol. 20. no. 19. pp. 11400–11404. DOI: 10.1109/JSEN.2020.2996079.
27. Busek J., Zitek P., Vyhliđal T. Astatism analysis of time delay controllers towards effective anti-windup shemes. *Proceedings of the International Conference on Process Control (PC'2019)*. 2019. pp. 74–79. DOI: 10.1109/PC.2019.8815283.
28. Gaiduk A.R., Plaksienko E.A. Analiz i analiticheskij sintez cifrovyyh sistem upravleniya [Analysis and analytical design of digital control systems]. *St. Petersburg: Lan Publ*, 2018. 272 p. (In Russ.).
29. Kim D.P. *Teoriya avtomaticheskogo upravleniya. Linejnye sistemy* [Theory of automatic control. Linear systems]. Publishing house URAIT, 2022. 312 p. (In Russ.).
30. Chen C.T. *Linear System Theory and Design: Third edition*. New York, Oxford: University Press, 1999. 334 p.
31. Nikolsky V.A. [On one way to ensure second-order astatism in a discrete system of combined control with disturbance compensation]. *Izvestiya vuzov. Electromechanics. – News of universities. Electromechanics*. 1978. no. 11. pp. 1168–1173. (In Russ.).
32. Chander B., Pal S., Di D., Buyya R. Artificial intelligence-based internet of things for industry 5.0. *Artificial intelligence-based internet of things systems*. 2022. pp. 3–45. DOI: 10.1007/978-3-030-87059-1_1.

Gaiduk Anatoly — Ph.D., Dr.Sci., Professor, Leading researcher, Department of automatic control systems, target search laboratory of breakthrough intelligent technologies for group control of robotic complexes of the advanced research foundation, Southern Federal University. Research interests: analysis and design of linear, nonlinear, continuous and discrete control systems for technical objects, processes, robots and robotic complexes. The number of publications — 450. gaiduk_2003@mail.ru; 44, Nekrasovsky Lane, 347922, Taganrog, Russia; office phone: +7(8634)371-689.

Prokopenko Nikolay — Ph.D., Dr.Sci., Professor, Honored inventor of the Russian Federation, Head of department, senior researcher, Department of information systems and radio engineering, department of scientific research department, Don State Technical University. Research interests: elements and devices of computer technology and control systems. The number of publications — 800. prokopenko@sssu.ru; 1, Gagarina Sq., 344000, Rostov-on-Don, Russia; office phone: +7(928)120-1984.

Bugakova Anna — Ph.D., Associate Professor, Junior researcher, Department of information systems and radio engineering, department of scientific research department, Don State Technical University. Research interests: elements and devices of computer technology and control systems. The number of publications — 250. annabugakova.1992@mail.ru; 1, Gagarina Sq., 344000, Rostov-on-Don, Russia; office phone: +7(909)416-9380.

Acknowledgements. The research has been carried out at the expense of the Grant of the Russian Science Foundation (project No. 18-79-10109-P).

И.Н. Глухих, А.С. Прохосин, Д.И. Глухих
**СРАВНЕНИЕ И ОТБОР СИТУАЦИЙ В СИСТЕМАХ ВЫВОДА
РЕШЕНИЙ НА ПРЕЦЕДЕНТАХ ДЛЯ «УМНОЙ» ФЕРМЫ**

Глухих И.Н., Прохосин А.С., Глухих Д.И. Сравнение и отбор ситуаций в системах вывода решений на прецедентах для «умной» фермы.

Аннотация. Тенденция развития умных ферм направлена на их становление полностью автономными, роботизированными предприятиями. Перспективы интеллектуализации сельскохозяйственного производства и умных ферм, в частности, сегодня связываются с развитием технологий систем, применяемых для обнаружения, распознавания сложных производственных ситуаций и поиска эффективных решений в этих ситуациях. В статье рассмотрены вопросы создания ситуационных систем поддержки принятия решения на умных фермах с применением вывода решений на основе рассуждений по прецедентам (case-based reasoning). Для разработки таких систем требуется выполнение ряда нетривиальных задач, к которым относятся, прежде всего, задачи формализации представления ситуаций и построения на этой основе способов сравнения и отбора ситуаций в базах знаний. В данном исследовании умная ферма представлена как сложный технологический объект, состоящий из взаимосвязанных компонентов, которыми являются технологические подсистемы умной фермы, производимая продукция, объекты операционного окружения, а также отношения между ними. Для реализации алгоритмов ситуационного вывода решений на основе прецедентов предложено формализованное представление ситуации в виде мультивектора, который содержит информацию о состояниях этих компонентов. Это позволило разработать ряд моделей обучаемой функции схожести между ситуациями. Проведенные эксперименты показали работоспособность предложенных моделей, на основе чего разработана ансамблевая архитектура нейросети для сравнения ситуаций и их отбора из базы знаний в процессах вывода решений. Практический интерес представляет мониторинг состояния растений по их видео-, фото- изображениям, что позволяет обнаруживать нежелательные состояния растений (болезни), которые могут служить сигналом для активизации процесса поиска решений в базе знаний.

Ключевые слова: вывод на основе прецедентов, принятие решений, нейросеть, архитектура нейросети, умная ферма.

1. Введение. Цифровизация сельского хозяйства и агропромышленного комплекса в целом с применением методов искусственного интеллекта стала одним из трендов развития отрасли в последние годы. Примерами того, как современное сельскохозяйственное предприятие становится сложным, высокотехнологическим комплексом с высоким уровнем автоматизации рабочих процессов могут служить городские фермы (контейнерные, вертикальные сити-фермы) [1 – 3], «умные» фермы и аквафермы (smart farms) [4 – 6]. Умные фермы стремятся к тому, чтобы стать полностью автономными, роботизированными предприятиями, которые при разведении сельскохозяйственных видов или пород животных не требуют участия человека [2]. Это не только существенно

облегчает данную производственную деятельность, но и снижает требования к квалификации специалистов и менеджеров, в ней участвующих. Ключевым в работе умных ферм становится организации сбора и использования данных, что позволяет не только контролировать и прогнозировать состояние производимой продукции и технологических систем, но и управлять в автоматическом режиме параметрами микроклимата, почвы, питания и т.п. для достижения желаемых результатов [5, 7]. Несмотря на широкое применение технологий мониторинга и анализа данных в работе таких высокотехнологических комплексов не исключается возникновение нештатных ситуаций, которые могут быть связаны как с болезнями растений или животных, так и с поломками оборудования, авариями. Подобные ситуации не возникают ежедневно, однако, они могут критически повлиять на работу предприятия и потому требуют оперативного, грамотного принятия решений. В то же время относительная редкость нештатных ситуаций обуславливает отсутствие, как правило, необходимого опыта и знаний у работников или менеджеров сельскохозяйственного предприятия, а специалисты-эксперты в нужной области не всегда являются доступными.

В этой связи перспективы дальнейшей интеллектуализации сельскохозяйственного производства и умных ферм, в частности, сегодня связываются с развитием технологий систем, применяемых для обнаружения, распознавания сложных производственных ситуаций и поиска эффективных решений в этих ситуациях. Так, для обнаружения и распознавания болезней растений свою эффективность уже показали искусственные нейросети, в частности, глубокие сверточные сети, которые работают в составе систем видеонаблюдения [8 – 11]. Однако, обнаруживая проблему, такие нейросети не предлагают ответов на вопросы «Что и как делать?», чтобы устранить эту проблему. На следующих этапах развития факт обнаружения и классификации болезни или любой другой проблемной ситуации требуется дополнить выводом рекомендованных или обязательных программ действий, которые нужно выполнить в этой ситуации.

Сегодня имеется опыт создания интеллектуальных систем, в том числе, в агропромышленном производстве и смежных областях, которые способны не только обнаруживать проблемные ситуации, но и предлагать для них рациональные решения. Так, в работе [12] приводится пример разработки системы, которая на основе экспертного опыта выводит рекомендации по выполнению защитных действий для источников воды. В [13] на основе анализа ситуаций и рассуждений по прецедентам предлагаются решения по ирригационной деятельности на

виноградниках, в [14] этот же подход используется для принятия решений в чрезвычайных экологических ситуациях, в статье [15] исследуются вопросы вывода решений для организации производства аквакультуры.

В этих исследованиях для вывода решений в проблемных ситуациях применяется известный в области искусственного интеллекта метод вывода решений на основе рассуждений по прецедентам (Case-based reasoning, CBR) [16]. Этот метод показал свою эффективность во многих предметных областях, когда необходимо принятие решений в условиях слабой формализации критериев и условий выбора, не позволяющих использовать строгие математические модели и методы оптимизации.

В CBR-системе используется интуитивно понятная идея о том, что, если имеется положительный опыт разрешения некоторой проблемной ситуации, то при возникновении подобной ситуации в дальнейшем мы также можем воспользоваться этот опыт для получения положительного результата. Тогда наличие достаточно большой по объему базы знаний (БЗ), где будут храниться прецеденты – пары <Ситуация-Решение>, позволит при возникновении проблемной ситуации подобрать для нее решение, сравнивая эту проблемную ситуацию с теми, которые есть в БЗ.

Важным преимуществом такого подхода является скорость вывода решений (т.к. производится выбор из готовых вариантов, а не разрабатываются новые), возможность поиска решений в условиях слабой формализации и неопределенности, а также потенциально высокое, экспертное качество принимаемых решений (за счет обеспечения такого качества решений в БЗ).

В то же время простая на первый взгляд идея CBR для реализации ее в прикладных системах требует выполнения ряда нетривиальных задач, к которым относятся, прежде всего, задачи формализации представления ситуаций и разработки на этой основе способов сравнения и отбора ситуаций в БЗ.

В данной работе предлагается решение этих задач с учетом возможностей их применения при создании ситуационных систем вывода решений, как одного из новых сервисов умной фермы [17]. Сначала в статье описывается способ математического представления сложных ситуаций, далее исследуются обучаемые нейросетевые модели схожести, на основе чего предлагается ансамблевая нейросетевая архитектура для сравнения и отбора ситуаций в процессах ситуационного вывода решений.

1.1. Методы. В CBR-системе прецедент образуется парой $\langle \text{Sit}, \text{Sol} \rangle$, где Sit является формализованным представлением известной из прошлого опыта ситуации, а Sol представляет решение, которое предлагается в данной ситуации. В зависимости от назначения CBR-системы ее решения могут содержать в себе команды для управления режимами технологического оборудования, целевые значения контролируемых параметров, программы действий персонала для устранения аварийных ситуаций, рекомендации, инструкции и др.

Если в наблюдаемой системе возникает проблемная ситуация Sit_{act} , требующая своего решения, оно ищется в БЗ путем сравнения Sit_{act} с ситуациями из базы знаний, выбора наиболее похожей ситуации Sit^* и вывода того решения Sol^* , которое входит в пару $\langle \text{Sit}^*, \text{Sol}^* \rangle$. Для определения степени похожести ситуаций некоторым образом задается функция $\text{Sim}(\text{Sit}_{\text{act}}, \text{Sit}_z)$, которая позволяет не просто зафиксировать факт похожести двух ситуаций Sit_{act} и Sit_z , но и количественно оценить степень этой похожести с тем, чтобы обеспечить обоснованный выбор на множестве ситуаций в БЗ, т.е.:

$$\text{Sit}^* = \text{argmax Sim}(\text{Sit}_{\text{act}}, \text{Sit}_z), \quad (1)$$

где z – индекс множества ситуаций в базе знаний, пробегающий значения от 1 до общего числа прецедентов, имеющих в БЗ на данный момент.

За счет этого появляется возможность вместо вывода типовых решений, основанного на распознавании класса входных ситуаций, находить конкретные и детальные решения, которые будут наибольшим образом подходить именно к данной проблемной ситуации Sit_{act} .

Условие выбора (1) может дополняться ограничениями, среди них особое место занимает ограничение на минимально допустимое значение $\text{Sim}(\cdot) \geq \text{Th}$, где Th – пороговое значение, которое ограничивает возможность применения найденного Sol^* . В случае, если $\text{Sim}(\text{Sit}_{\text{act}}, \text{Sit}^*) < \text{Th}$, найденное решение Sol^* не рекомендуется для непосредственного применения, но оно может служить основой для разработки нового решения Sol_{act} . Этот процесс называется адаптацией решений [18], по результатам которого в базу знаний (после соответствующей оценки найденного решения) добавляется новая пара $\langle \text{Sit}_{\text{act}}, \text{Sol}_{\text{act}} \rangle$.

Таким образом, в процессе эксплуатации CBR-системы обеспечивается пополнение ее базы знаний. Вопросы пополнения прецедентной БЗ, актуализации прецедентов и обслуживания объемных баз знаний составляют одно из направлений современных исследований

в данной области [18 – 20]. Но основными при разработке прецедентных систем являются задачи формализации прецедентов, в частности, представления ситуаций и связанные с ними задачи сравнения и оценки схожести ситуаций.

Наиболее очевидным способом представления ситуаций является представление их точками в многомерном пространстве параметров, которые собираются и оцениваются в наблюдаемой системе.

В этом случае оценка схожести Sim может быть определена с помощью метрик расстояния в этом пространстве, обзор которых приведен, например, в [21 – 23]. Однако, задачи представления и оценки схожести ситуаций при таком подходе усложняются с увеличением числа контролируемых параметров, при появлении наряду с количественными еще и качественных, а также неточных или неполных данных. Проблема усугубляется в сложных системах, когда в описании ситуаций для принятия решений необходимо учитывать состояния значительного числа неоднородных объектов и отношений между ними, которые могут описываться данными разной природы и разных форм представления.

В подобных условиях могут использоваться локальные метрики схожести [24], что говорит о появлении векторного критерия выбора:

$$Sit^* = \operatorname{argmax} \{Sim_1(Sit_{act}, Sit_z), \dots, Sim_k(Sit_{act}, Sit_z), \dots, Sim_N(Sit_{act}, Sit_z)\}. \quad (2)$$

Положительно то, что в (2) каждая из локальных функций схожести Sim_k может быть вычислена по-своему и на своей части данных. В то же время остается проблема дальнейшего выбора, который на практике сведется или к поиску парето-оптимальных решений, т.е. отбору нескольких несравнимых вариантов решений из БЗ, или к выбору способа агрегирования, т.е. перевода вектора в итоговую агрегированную Sim с дальнейшим выбором по (1).

В первом случае мы не избавляем пользователя (который может не обладать необходимой квалификацией и опытом) от проблемы выбора. Во втором случае с помощью экспертов необходима разработка модели компромисса между локальными метриками (например, на основе введения коэффициентов относительной важности), что может стать трудоемким. Кроме того, в сложных системах модель компромисса может оказаться не постоянной, т.е. веса относительной важности сходства по той или иной группе параметров Sim_k могут меняться в зависимости от состояний объектов и связей между ними. В

подобных случаях перспективным является использование обучаемой функции Sim и, в частности, применение нейросетевых моделей [15, 24, 25], с помощью которых функция схожести может быть представлена в виде:

$$\text{Sim} = f_m(W_m, \dots, f_1(W_1, X_{in})), \quad (3)$$

где X_{in} – вектор входных данных нейросети, который содержит в себе информацию о сравниваемых ситуациях;

W_m, \dots, W_1 – настраиваемые в процессе обучения веса межнейронных связей на слоях от 1 до m , причем для всех слоев $W \in R^{T \times V}$, T и V – размерности соответственно выходного и входного векторов на данном слое нейросети;

f_m, \dots, f_1 – функции активации на выходе нейронов на слоях от 1 до m .

При обучении на примерах похожих и непохожих пар ситуаций подбираются веса нейросетевой модели (3) так, чтобы она могла вычислять оценку схожести Sim (.) для новых, в том числе, неизвестных ей ранее ситуаций. Качество модели оценивается с помощью специальных метрик при использовании валидационных наборов данных, которые не участвовали в обучении. Подбор архитектуры (количество слоев, размерность векторов, виды активационных функций и т.п.) осуществляется экспериментально.

Далее в статье разрабатываются формализованные математические представления сложного технологического объекта умной фермы и ситуаций, которые обеспечивают возможности использования нейросетевых моделей, описываются примеры представления ситуаций, возникающие на технологическом объекте умной фермы. После этого проводятся экспериментальные исследования нейросетевых моделей и формулируются выводы для организации процесса поиска решений.

1.2. Формализация представлений объекта и ситуаций. Для формализации представления ситуации, которая может возникать в процессе работы умной фермы, далее будем использовать предложенное авторами ранее обобщенное представление сложного технологического объекта (СТО) [26]. Будем использовать абстрактный СТО, который состоит из N -го числа элементов и представляется следующим кортежем множеств:

$$\text{СТО} = \langle O, S, P, D \rangle,$$

где $O = \{ O_i \mid \forall i \in [1; N] \}$ – множество элементов сложного технологического объекта, N – число элементов сложного объекта. Под элементом понимается в данном случае любая из технологических подсистем умной фермы, собственно производимая продукция, объекты операционного окружения (при необходимости учитывать их состояние при принятии решений), а также отношения между ними. Каждый i -й элемент может находиться в некотором состоянии (например, «исправен» или «неисправен»). Кроме состояний элементы СТО могут характеризоваться дополнительными параметрами, которые ниже названы оцениваемыми;

$S = \{ S_i \mid \forall i \in [1; N] \}$ и $S_i = \{ S_{ij} \mid \forall j \in [1; N_{s_i}] \}$ – множество состояний i -го элемента СТО, N_{s_i} – число возможных состояний i -го элемента;

$P = \{ P_k \mid \forall k \in [1; N_p] \}$, $k \in \{ K_i \}$ – множество оцениваемых параметров, K_i – подмножество индексов, имеющих отношение к i -му элементу СТО, N_p – число оцениваемых параметров;

$D = \{ D_k \mid \forall k \in [1; N_p] \}$, D_k – упорядоченное множество категориальных значений k -го параметра.

Категории параметра могут задаваться лингвистическими значениями (например, параметр «Температура» может иметь область значений «Низкая», «Средняя», «Высокая»), интервалами на шкале значений оцениваемого параметра, логическими значениями (False, True) и иным способом задания перечислимых типов данных.

Каждому i -му элементу в текущий момент времени можно сопоставить вектор состояний с размерностью N_{s_i} , компоненты которого принимают значение 1 (элемент находится в j -м состоянии) или 0 (элемент не находится в данном состоянии). Аналогичным образом каждый k -й оцениваемый параметр P_k тоже представляется вектором из 0 и 1 размерность которого равна $|D_k|$. С учетом этой схожести далее для упрощения введем один универсальный оценочный вектор (состояний и параметров) S^{β} , компоненты которого могут принимать значения 1 или 0, а индекс β будет означать все множество индексов в множествах состояний и множествах категориальных значений. Набор таких векторов, которым описывается все множество элементов сложного объекта, составляет мультивектор $S^{\alpha\beta}$, где индекс α обозначает все множество индексов $r \in [1; (N + N_p)]$. Иначе говоря, α и β в $S^{\alpha\beta}$ представляют собой оси координат, где первая обозначает элементы или параметры, а вторая – их возможные состояния или значения.

Определение. Ситуацией на СТО называется такое положение дел, которое характеризуется совокупностью состояний элементов и категориальных значений оцениваемых параметров сложного объекта.

С учетом этого мультивектор $S^{\alpha\beta}$ есть формальное представление ситуации, т.е. $Sit \Leftrightarrow S^{\alpha\beta}$. Отметим, что $S^{\alpha\beta}$ не является матрицей, однако, он может быть записан в виде таблицы значений, где столбцы будут означать элементы и оцениваемые параметры, а строки соответствовать их состояниям или категориальным значениям, как это далее показано в таблице 1.

Конкатенация строк такой таблицы образует внутреннее (встроенное) представление ситуации – эмбединг ситуации, который будет использован для формирования входа нейросетевых моделей (раздел 2.1). Чтобы не вводить дополнительных обозначений, далее будем полагать, что мультивектор $S^{\alpha\beta}$ может быть представлен как в виде двумерной таблицы, так и в виде одномерного массива, получаемого конкатенацией строк этой таблицы, или с помощью иной формы представления при сохранении возможности обращения к любому из компонентов мультивектора с помощью указания значений индексов по осям α и β .

2.1. Описание объекта. Рассматриваемая умная ферма представляет собой автоматизированную систему выращивания растений на гидропонике [3]. Она осуществляет контроль управляемых параметров (температуру, влажность, освещенность, уровень CO_2 , количество и состав микроэлементов и рН в питательном растворе и др.) и регулирование режимов работы технологического оборудования для создания условий, необходимых для выращиваемых культур. За счет этого обеспечивается эффективность и качество роста растений, уменьшается зависимость от погодных условий и опыта агронома.

В целях представления и анализа ситуаций элементы СТО и оцениваемые параметры умной фермы собраны в следующие группы – компоненты комплексной ситуации: «Микроклимат», «Питание», «Освещение», «Растения». В некоторой ситуации каждая группа характеризуется состояниями входящих в нее элементов и значениями оцениваемых параметров. Например, подсистема «Микроклимат» описывается параметрами «Температура», «Влажность», «Уровень CO_2 » и элементами «Датчик температуры», «Датчик влажности», «Датчик CO_2 ». Датчики температуры, влажности, CO_2 могут быть в одном из двух состояний «Исправен» или «Неисправен», а оцениваемые параметры принимают значения из списка [«Ниже порога», «Минимум», «Оптимум», «Максимум», «Выше порога»], где первое и последнее значение являются недопустимыми.

В таблице 1 представлен пример комплексной ситуации, здесь единица в столбце соответствует тому состоянию, в котором находится данный элемент или тому значению, которое принимает данный параметр.

Таблица 1. Представление ситуации на умной ферме

Sit	Элементы и параметры	исправен	неисправен	ниже порога	минимум	оптимум	максимум	выше порога
«Микроклимат»	Температура			0	0	0	1	0
	Датчик температуры	1	0					
	Влажность			0	0	1	0	0
	Датчик влажности	1	0					
	Уровень CO ₂			0	1	0	0	0
	Датчик CO ₂	1	0					
«Полив»	Ph (кислотность раствора)			0	0	1	0	0
	pH-метр	1	0					
	ЕС по элементный (N P K+ микромакроэлементы)			0	0	0	1	0
	датчик электропроводности	1	0					
	Температура раствора			0	0	1	0	0
	Датчик температуры	1	0					
	Интенсивность полива			0	0	1	0	0
«Освещение»	Таймер полива	1	0					
	Освещение	1	0					
	Мощность освещения			0	0	1	0	0
	Периодичность освещения			0	0	0	1	0
					Низкое	Среднее	Высокое	Отсутствует
«Растения»	Увядание				1	0	0	0
	Пожелтение				0	0	1	0
	Ожоги				0	0	1	0

Таблица 2. Ситуация № 1 с неисправным таймером полива

Sit _i	Элементы и параметры							
		исправен	неисправен	ниже порога	минимум	оптимум	максимум	выше порога
«Микроклимат»	Температура			0	0	1	0	0
	Датчик температуры	1	0					
	Влажность			0	0	1	0	0
	Датчик влажности	1	0					
	Уровень CO ₂			0	0	1	0	0
	Датчик CO ₂	1	0					
«Полив»	Ph (кислотность раствора)			0	1	0	0	0
	pH-метр	1	0					
	ЕС по элементный (N P K+ микромакроэлементы)			0	1	0	0	0
	Датчик электропроводности	1	0					
	Температура раствора			0	0	1	0	0
	Датчик температуры	1	0					
	Интенсивность полива			1	0	0	0	0
	Таймер полива	0	1					
«Освещение»	Освещение	1	0					
	Мощность освещения			0	0	1	0	0
	Периодичность освещения			0	0	1	0	0
					Низкое	Среднее	Высокое	Отсутствует
«Растения»	Увядание				0	0	1	0
	Пожелтение				0	1	0	0
	Ожоги				0	1	0	0

Все вектора ситуаций в обучающем наборе данных являются результатом one-hot кодирования (вектора, соответствующие своим элементам (параметрам), принимают значение 0 или 1, при этом только один из них равен 1).

2.3. Валидационные наборы данных. Для проверки работоспособности обученных моделей дополнительно к обучающим

данным были разработаны валидационные наборы. Было подготовлено 3 валидационных набора данных (1В, 2В, 3В), позволяющих оценить способность нейросетевой модели:

- отбирать схожие ситуации;
- ранжировать схожие ситуации по степени схожести;
- работать в условиях нечеткой классификации состояний, когда компоненты вектора представления ситуаций принимают значения в интервале от 0 до 1.

Валидационные наборы содержат пары ситуаций с оцененной экспертом схожестью, отличные от пар ситуаций из обучающего набора.

Наборы 1В и 2В содержат по 15 пар ситуаций, где одна ситуация сравнивается с 15 другими разной схожести (значения схожести Sim меняются в интервале от 0,50 до 0,95).

Валидационный набор 3В содержит 10 пар ситуаций. Он имитирует возникновение нежелательной ситуации с неточным определением состояний некоторых элементов, т.е. когда компоненты вектора ситуации принимают значения в интервале $[0;1]$. Такая ситуация здесь названа нечеткой. Она иллюстрирует случай, при котором специалист, эксплуатирующий умную ферму, в силу недостатка данных (из-за неисправности оборудования) или недостатка знаний и опыта не может сделать точные выводы о значениях тех или иных параметров в текущей ситуации. Предполагается, что и в таком случае СВР-система должна уметь находить в базе знаний похожие ситуации с тем, чтобы предложить необходимое решение.

Примером нечеткой ситуации является ситуация, представленная в таблице 3. Она характеризуется неисправностью датчика регулятора кислотности раствора рН. Так как с датчика измерения рН раствора не поступает данные, то точно оценить кислотность невозможно. В силу этого используется нечеткое оценивание параметра Ph. когда при наблюдении за текущей ситуацией специалист умной фермы высказывает свои предположения относительно значения этого параметра. При этом он выставляет субъективную оценку вероятности (коэффициент уверенности) предполагаемого значения, которое принимает данный параметр в этой ситуации – в примере это значения «Ниже порога» и «Минимум» с оценками вероятности 0.4 и 0.6 соответственно. Аналогичным образом, в данной ситуации нечетко оцениваются значения параметров в группе «Растения». Когда при анализе состояния растений наблюдатель не может однозначно оценить параметры «Увядание», «Пожелтение», «Ожоги», он выставляет им значения списком [«Низкое», «Среднее», «Высокое»] с

соответствующими субъективными оценками вероятности этих значений.

Таблица 3. Пример ситуации с неточными параметрами

i_{k3}	Элементы и параметры	исправно	неисправно	ниже порога	минимум	оптимум	максимум	выше порога
«Микроклимат»	Температура			0	0	1	0	0
	Датчик температуры	1	0					
	Влажность			0	0	1	0	0
	Датчик влажности	1	0					
	Уровень CO ₂			0	0	1	0	0
	Датчик CO ₂	1	0					
«Полив»	Ph (кислотность раствора)			0.4	0.6	0	0	0
	pH-метр	0	1					
	ЕС по элементный (N P K+ микромакроэлементы)			0	0	1	0	0
	датчик электропроводности	1	0					
	Температура раствора			0	0	1	0	0
	Датчик температуры	1	0					
	Интенсивность полива			0	0	1	0	0
	Таймер полива	1	0					
«Освещение»	Освещение	1	0					
	Мощность освещения			0	0	1	0	0
	Периодичность включения освещения			0	0	1	0	0
					Низкое	Среднее	Высокое	Отсутствует
«Растения»	Увядание				0.4	0.6	0	0
	Пожелтение				0	0.6	0.4	0
	Ожоги				0	0.8	0.2	0

В нечеткой ситуации вместо one-hot кодирования векторов используется «мягкое», soft кодирование, при котором в любом из векторов могут присутствовать числа от нуля до единицы в разных позициях.

В экспериментах такая ситуация сравнивалась с 10 другими ситуациями, где состояния элементов и значения параметров были точно известны. Таким образом, проверялась гипотеза о применимости обученной нейросети для сравнения и отбора ситуаций в базе знаний, когда входная, текущая ситуация является нечеткой.

2.4. Описание метрик. Для оценки нейросетевых моделей использовались следующие метрики.

1) MAPE (Mean Absolute Percentage Error) – абсолютная средняя процентная ошибка [28] значения Sim, которая вычислялась по следующей формуле:

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| * 100\%, \quad (4)$$

где n – количество оцениваемых Sim пар ситуаций, y_t – значение Sim из валидационного файла, \hat{y}_t – модельное значение полученного Sim.

2) nDCG@k (Normalized Discounted Commulative Gain для лучших k) – метрика для оценки верности ранжирования k ситуаций [29] лучших по величине Sim. Вычислялась по следующим формулам:

$$\text{NDCG}_k = \frac{\text{DCG}_k}{\text{IDCG}_k}, \quad (5)$$

$$\text{DCG}_k = \sum_{i=1}^k \frac{2^{\hat{r}_i} - 1}{\log_2(1+i)}, \quad (6)$$

где \hat{r}_i – оцененная релевантность результата на позиции i , IDCG $_k$ – значение DCG $_k$ при ранжировании по истинным значениям (рейтингам из валидационного набора), k – количество оцениваемых ситуаций, принято $k=6$.

3) Accuracy – метрика для оценки верности определения класса [30]. В нашем случае класс пар ситуаций «похож/непохож» при пороге $Sim > 0.75$. Вычислялась по следующей формуле:

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP}, \quad (7)$$

где TP – верные положительные (похож) классификации; TN – верные отрицательные (непохож) классификации; FP – неверные положительные (похож) классификации; FN – неверные отрицательные (непохож) классификации.

2.5. Результаты. В исследовании был проведен эксперимент по анализу производительности трех нейросетевых моделей для оценки схожести ситуаций и их комбинаций. Каждая модель оперирует входными данными, получаемыми из мультивекторов двух сравниваемых ситуаций.

За основу архитектуры модели взята архитектура полносвязной нейронной сети (рисунок 1).

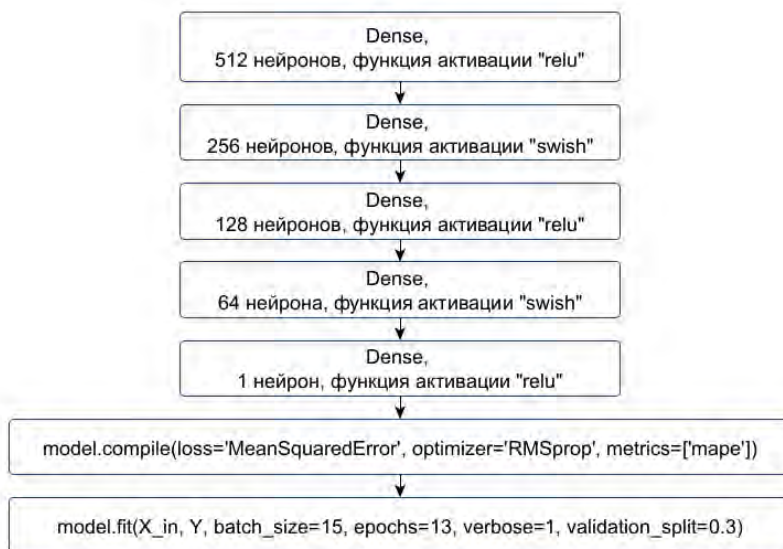


Рис. 1. Полносвязная нейронная сеть

Модель M1 на вход получает единый образ двух ситуаций X_{in} в виде конкатенации (concat) мультивекторов текущей ситуации S^{ab}_{act} и ситуации из базы знаний S^{ab}_z и преобразует его в оценку схожести этих ситуаций Sim:

$$X_{in} = \text{concat}(S^{ab}_{act}, S^{ab}_z),$$

$$M1: X_{in} \rightarrow \text{Sim} \in [0;1].$$

Этот способ формирования входного сигнала использован в одной из архитектур сиамских нейронных сетей, специально предназначенных для сравнения объектов [15].

Модель M2 на вход получает посимвольную разницу мультивекторов двух ситуаций:

$$X_{in} = |S^{ab}_{act} - S^{ab}_z|,$$

$$M2: X_{in} \rightarrow \text{Sim} \in [0;1].$$

Этот способ формирования входного сигнала был использован в работе [31]. Он позволяет сократить длину входного вектора в два раза по сравнению с тем, который используется в модели M1. Сокращение размерности входного вектора потенциально обладает положительным свойством, т.к. уменьшает количество комбинаций входных векторов, которые могут давать на выходе один и тот же результат вычислений.

Модель M3, которая на вход получает посимвольную сумму двух мультивекторов:

$$X_{in} = S_{act} + S_z,$$

$$M3: X_{in} \rightarrow \text{Sim} \in [0;1].$$

Здесь компоненты входного для модели вектора X_{in} принимают значения от 0 до 2, что интерпретируется следующим образом. Некоторый ij -й компонент входного вектора X_{in} принимает значение 2 в том случае, когда в обеих ситуациях i -й элемент находится в j -м состоянии, значение 0 – элемент не находится в этом состоянии ни в одной из ситуаций, значение 1 – i -й элемент находится в j -м состоянии в одной из сравниваемых ситуаций.

В работе [32] представлены результаты исследования предложенных моделей M1, M2, M3 для вычисления Sim. В этой статье мы приводим результаты дальнейших исследований, где в целях улучшения метрик качества рассматриваются ансамбли нейросетевых моделей.

Пример архитектуры ансамблевой модели приведен на рисунке 2, программная реализация выполнена на языке Python с использованием библиотек Keras, TensorFlow.

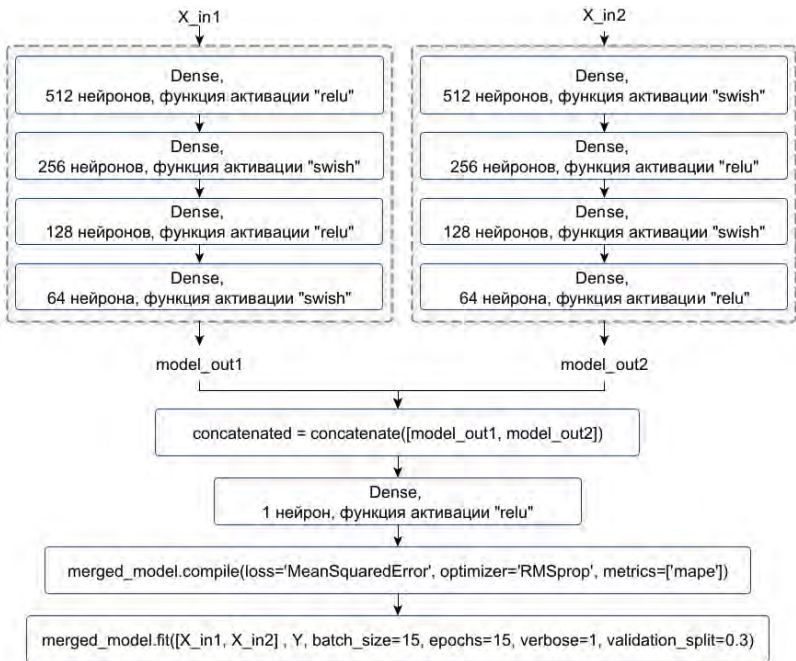


Рис. 2. Архитектура ансамблевой модели

Цикл обучения, валидации и анализа показателей производился 8 раз на одних и тех же данных для каждой модели в целях оценки стабильности показателей, что потенциально позволяет обеспечить стабильность работы на новых наборах данных.

В качестве альтернативы нейросетевым моделям рассмотрена модель агрегирования локальных метрик сходства (2) на основе аддитивной свертки с постоянными коэффициентами относительной важности, которые определяются путем экспертного опроса.

При этом формула (2) преобразуются в следующую:

$$Sim = \sum \alpha_k Sim_k, \quad (8)$$

где α_k – весовой коэффициент группы, определяемый экспертами, Sim_k – сходство k -й группы элементов СТО в двух сравниваемых ситуациях, определяемое по формуле:

$$Sim_k = \sum \beta_{ki} f_{ki},$$

где β_{ki} – коэффициент относительной важности i -го элемента (параметра) в k -й группе, принимающий значение в интервале от 0 до 1 при сумме всех коэффициентов в данной группе равной единице. Его значение, как и значение коэффициентов α , тоже определяется экспертным путем;

f_{ki} – показатель сходства i -го элемента (параметра) в k -й группе в сравниваемых ситуациях. Он принимает значение 1, когда элементы (параметры) в двух ситуациях имеют одинаковое состояние (значение) или ноль в противном случае.

Лучшие показатели метрик из восьми проведенных экспериментов с валидационными данными для отдельных моделей и их ансамблей представлены в таблице 4.

Таблица 4. Показатели метрик

№ П/П	Метрика Модель	Точная классификация состояний и оценка параметров (набор 1В, 2В)			Неточная классификация состояний и оценка параметров (набор 3В)		
		MAPE	nDCG@k	Accuracy	MAPE	nDCG@k	Accuracy
1	M1	14%	61%	92%	21%	68%	80%
2	M2	3.75%	96%	96%	12%	98%	100%
3	M3	13.40%	89%	92%	19%	78%	70%
4	Ансамбль: M2, M2	3.39%	100%	100%	9.8%	98%	100%
5	Ансамбль: M2, M2, M2	7.90%	95%	100%	12%	95%	100%
6	Четыре модели: M2, M3, M2, M2	3.48%	98.5%	100%	12%	95%	100%
7	Аддитивная свертка с постоянными коэффициентами важности	6%	92%	90%	14%	60%	90%

3. Обсуждение результатов экспериментов. Эксперименты показали, что ансамбль простых моделей может обеспечивать более качественные результаты по сравнению с отдельным их применением. На основе показателей метрик, представленных в таблице 4, можно сделать вывод о целесообразности применения в задачах сравнения ситуаций ансамбля простых моделей: M2, M2 (таблица 4). Архитектура этого ансамбля моделей показана на рисунке 2.

Данный ансамбль моделей показал лучшие (в смысле используемых метрик качества) результаты при решении задач отбора схожих ситуаций и ранжирования ситуаций по степени схожести, в том числе в условиях возникновения нечеткой ситуации.

Актуальность ранжирования возрастает с усложнением объекта (увеличением количества элементов), когда возрастает риск возникновения коллизий, т.е. случаев, при которых различие ситуаций по одним атрибутам или локальным метрикам может компенсироваться их сходством по другим атрибутам.

Аддитивная свертка при проверке на валидационном наборе, несмотря на близкие к нейросетевым моделям результаты, не показала результативность при решении задачи ранжирования, приравняв по схожести 5 ситуаций. Напротив, нейросетевая модель успешно разделила эти ситуации по уровню схожести, позволив построить ранжированную по Sim(.) последовательность, совпадающую с мнением эксперта, с помощью которого был сформирован валидационный набор.

Таким образом, эксперименты показали, что нейросетевые модели и их ансамбли, обучившись на примерах, могут быть использованы для поиска и отбора в базе знаний похожие ситуации, удовлетворяющие критериям (1)-(2). При условии наполнения БЗ прецедентами $\langle \text{Sit}, \text{Sol} \rangle$ найденная Sit^* позволяет предложить пользователю то решение Sol^* , которое ассоциировано с этой ситуацией и которое рекомендовано для таких ситуаций. Так, в рассмотренных примерах при возникновении ситуации с неисправным рН-метром (таблица 3) БЗ может содержать в себе решение, которое будет представлять собой программу действий (технологическую карту) по устранению неисправности оборудования и рекомендуемые действия для восстановления состояний растений.

4. Заключение. В статье рассмотрено использование CBR-метода с представлением ситуаций посредством векторов, которые содержат в себе информацию о состояниях элементов и значений оцениваемых параметров, применительно к задачам ситуационного вывода решений при эксплуатации умной фермы.

Умная ферма представлена как сложный технологический объект, состоящий из взаимосвязанных компонентов, которыми являются технологические подсистемы умной фермы, производимая продукция, объекты операционного окружения, а также отношения между ними. Для реализации алгоритмов ситуационного вывода решений на основе прецедентов предложено формализованное представление ситуации в виде мультивектора $S^{a\beta}$, что позволило

разработать ряд моделей обучаемой функции схожести Sim между ситуациями. Проведенные эксперименты показали работоспособность предложенных моделей, на основе чего разработана ансамблевая архитектура нейросети для сравнения ситуаций и их отбора из базы знаний в процессах вывода решений. Применение нейросетей для количественной оценки схожести ситуаций позволяет отказаться от трудоемких задач экспертного оценивания ситуаций, разработки критериев и алгоритмов их сравнения в сложных условиях, в том числе, при неточных данных для идентификации ситуации, при необходимости применения локальных метрик схожести и динамичности весов относительной важности этих метрик.

Для применения предлагаемого подхода и разработанных моделей важным становится вопрос идентификации ситуаций, что связано с классификацией состояний компонентов умной фермы по собираемым данным. Практический интерес представляет мониторинг состояния растений по их видео-, фото- изображениям [8], что позволяет обнаруживать нежелательные состояния растений (болезни), которые могут служить сигналом для активизации процесса поиска решений в базе знаний. Разработка классификаторов состояний растений и иных компонентов умной фермы для идентификации ситуаций и запуска процессов вывода решений является следующим этапом исследований авторов.

Результаты исследования могут найти применение для разработки баз знаний и систем поддержки принятия решений в сложных ситуациях, возникающих при эксплуатации умных ферм (неисправности оборудования, болезни растений, недостатки питания и др.). Можно ожидать, что применение подобных систем будет способствовать повышению оперативности и качества принимаемых решений, а также менять требования к компетенциям специалистов и обслуживающего персонала умной фермы, смещая их профиль в сторону цифровых технологий. При этом для обучения моделей и создания баз знаний потребуются и эксперты в «своей» предметной области – высококвалифицированные специалисты сельского хозяйства, чьи знания и опыт станут доступны для обеспечения работы многочисленных ферм.

Литература

1. Руткин Н.М., Лагуткин О.Ю., Лагуткина Л.Ю. Урбанизированное агропроизводство (сити-фермерство) как перспективное направление развития мирового агропроизводства и способ повышения продовольственной безопасности городов // Вестник астраханского государственного технического университета. серия: рыбное хозяйство. 2017. Т. 2017. № 4. С.95–108.

2. Сурай Н.М., Кудинова М.Г., Уварова Е.В., Жидких Е.И. Анализ развития цифровых технологий в «умных» фермах // *Инновации и инвестиции*. 2021. № 10. С. 184–188.
3. Martin M., Molin E. Environmental Assessment of an Urban Vertical Hydroponic Farming System in Sweden // *Sustainability*. 2019. vol. 11(15). no. 4124. DOI: 10.3390/su11154124.
4. Chiu M.-C., Yan W.-M., Bhat S.A., Huang N.-F. Development of smart aquaculture farm management system using IoT and AI-based surrogate models // *Journal of Agriculture and Food Research*. 2022. vol. 9. no. 100357. DOI: 10.1016/j.jafr.2022.100357.
5. Devapal D. Smart Agro Farm Solar Powered Soil and Weather Monitoring System for Farmers // *Proceedings of International Multi-conference on Computing, Communication, Electrical & Nanotechnology, I2CN-2K19*. 2020. pp. 1843–1854.
6. He L., Fu L., Fang W., Sun X., Suo R., Li G., Zhao G., Yang R., Li R. IoT-based urban agriculture container farm design and implementation for localized produce supply // *Computers and Electronics in Agriculture*. 2022. vol. 203. no. 107445. DOI: 10.1016/j.compag.2022.107445.
7. Klaina H., Gueembe I.P., Lopez-Iturri P., Campo-Bescós M.A., Azpilicueta L., Aghzout O., Alejos A.V., Falcone F. Analysis of low power wide area network wireless technologies in smart agriculture for large-scale farm monitoring and tractor communications // *Measurement*. 2022. vol. 187(5). no. 110231. DOI: 10.1016/j.measurement.2021.110231.
8. Махмудул Хасан А., Мд Ракиб Ул Ислам Р., Авинаш К. Классификация болезней листьев яблони с использованием набора данных изображений: подход многослойной сверточной нейронной сети // *Информатика и автоматизация*. 2022. Т. 21. № 4. С. 710–728. DOI: 10.15622/ia.21.4.3
9. Moreira R., Moreira L.F.R., Munhoz P.L.A., Lopes E.A., Ruas R.A.A. AgroLens: A low-cost and green-friendly Smart Farm Architecture to support real-time leaf disease diagnostics // *Internet of Things*. 2022. vol. 19. no. 100570. DOI: 10.1016/j.iot.2022.100570.
10. Hu W.-C., Chen L.-B., Huang B.-K., Lin H.-M. A Computer Vision-Based Intelligent Fish Feeding System Using Deep Learning Techniques for Aquaculture // *IEEE Sensors Journal*. 2022. vol. 22. no. 7. pp. 7185–7194. DOI: 10.1109/JSEN.2022.3151777.
11. Cho S., Kim T., Jung D.-H., Park S.H., Na Y., Ihn Y.S., Kim K.G. Plant growth information measurement based on object detection and image fusion using a smart farm robot // *Computers and Electronics in Agriculture*. 2023. vol. 207. no. 107703. DOI: 10.1016/j.compag.2023.107703.
12. Cerutti J., Abi-Zeid I., Lamontagne L., Lavoie R., Rodriguez-Pinzon M.J. A case-based reasoning tool to recommend drinking water source protection actions // *Journal of Environmental Management*. 2023. vol. 331. no. 117228. DOI: 10.1016/j.jenvman.2023.117228.
13. Zhai Z., Martínez J.F., Martínez N.L., Díaz V.H. Applying case-based reasoning and a learning-based adaptation strategy to irrigation scheduling in grape farming // *Computers and Electronics in Agriculture*. 2020. vol. 178. no. 105741. DOI: 10.1016/j.compag.2020.105741.
14. Wang D., Wan K., Ma W. Emergency decision-making model of environmental emergencies based on case-based reasoning method // *Journal of Environmental Management*. 2020. vol. 262(9). 110382. DOI: 10.1016/j.jenvman.2020.110382.
15. Mathisen B.M., Bach K., Aamodt A. Using extended siamese networks to provide decision support in aquaculture operations // *Applied Intelligence*. 2021. vol. 51(1). DOI: 10.1007/s10489-021-02251-3.

16. Aamodt A., Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches // *AI Communications*. 2001. vol. 7. pp. 39–59. DOI: 10.3233/AIC-1994-7104.
17. Скобелев П.О., Симонова Е.В., Будаев Д.В., Вошук Г.Ю., Ларюхин В.Б. Облачная интеллектуальная система SMART FARMING для управления точным земледелием // *Материалы конференции «Информационные технологии в управлении (ИТУ-2018)»* (г. Санкт-Петербург, 2–4 октября 2018 г.). Издательство: Концерн «Концерн «ЦНИИ «Электроприбор», 2018. С. 261–270.
18. Leake D., Ye X., Crandall D. Supporting Case-Based Reasoning with Neural Networks: An Illustration for Case Adaptation // *Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021)*. 2021. Available at: <https://proceedings.aaai-make.info/AAAI-MAKE-PROCEEDINGS-2021/paper1.pdf>. (accessed 26.05.2023).
19. Guo Y., Zhang B., Sun Y., Jiang K., Wu K. Machine learning based feature selection and knowledge reasoning for CBR system under big data // *Pattern Recognition*. 2021. vol. 112(6). no. 107805. DOI: 10.1016/j.patcog.2020.107805.
20. Smiti A., Elouedi Z. Dynamic maintenance case base using knowledge discovery techniques for case based reasoning systems // *Theoretical Computer Science*. 2020. vol. 817. pp 24–32. DOI: 10.1016/j.tcs.2019.06.026.
21. Liao T.W., Zhang Z., Mount C.R. Similarity measures for retrieval in case-based reasoning systems // *Applied Artificial Intelligence*. 1998. vol. 12(4). pp. 267–288. DOI: 10.1080/088395198117730.
22. Fan Z.-P., Li Y.-H., Wang X., Liu Y. Hybrid similarity measure for case retrieval in CBR and its application to emergency response towards gas explosion // *Expert Systems with Applications*. 2014. vol. 41(5). pp. 2526–2534. DOI: 10.1016/j.eswa.2013.09.051.
23. Oyelade O.N., Ezugwu A.E. A case-based reasoning framework for early detection and diagnosis of novel coronavirus // *Informatics in Medicine Unlocked*. 2020. vol. 20(6). no. 100395. DOI: 10.1016/j.imu.2020.100395.
24. Gabel T., Godehardt E. Top-down induction of similarity measures using similarity clouds. *International Conference on Case-Based Reasoning*. 2015. pp. 149–16. DOI: 10.1007/978-3-319-24586-7_11.
25. Mathisen B.M., Aamodt A., Bach K., Langseth H. Learning similarity measures from data // *Progress in Artificial Intelligence*. 2020. vol. 9. pp. 129–143. DOI: 10.1007/s13748-019-00201-2.
26. Glukhikh I., Glukhikh D. Case-Based Reasoning with an Artificial Neural Network for Decision Support in Situations at Complex Technological Objects of Urban Infrastructure // *Applied System Innovation*. 2021. vol. 4(73). 12 p. DOI: 10.3390/asi4040073.
27. Глухих И.Н., Глухих Д.И. Алгоритмы генерации обучающих множеств в системе с прецедентным выводом на основе ситуаций-примеров // *Программные продукты и системы*. 2022. Т. 35. № 4. С. 660–669.
28. Myttenaere A.D., Golden B., Grand B.L., Rossi F. Mean Absolute Percentage Error for regression models // *Neurocomputing*. 2016. vol. 192. pp. 38–48. DOI: 10.1016/j.neucom.2015.12.114.
29. Wang Y., Wang L., Li Y., He D., Liu T.-Y., Chen W. A Theoretical Analysis of NDCG Type Ranking Measures. *Computer Science*. 2013. 26 p. DOI: 10.48550/arXiv.1304.6480.
30. Taylor J.R. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, Second Edition, Paperback & Clothbound, 1997. 327 p.
31. Paulson P., Juell P. Using Reinforcement Learning for Similarity Assessment in Case-Based Systems // *IEEE Intelligent Systems*. 2003. vol. 18. no. 4. pp. 60–67. DOI: 10.1109/MIS.2003.1217629.

32. Glukhikh I., Chernysheva T., Glukhikh D. Neural Network Models for Situation Similarity Assessment in hybrid-CBR // Journal of Intelligent & Fuzzy Systems. 2023. vol. 44(15). pp. 1–14. DOI: 10.3233/JIFS-221335.

Глухих Игорь Николаевич — д-р техн. наук, профессор, заведующий кафедрой, кафедра информационных систем, Институт математики и компьютерных наук (ИМиКН); научный руководитель центра системного инжиниринга, передовая инженерная школа, Тюменский государственный университет (ТюмГУ). Область научных интересов: интеллектуальные системы поддержки принятия решений, базы знаний, вывод на основе прецедентов, нейросети в принятии решений. Число научных публикаций — 180. i.n.glukhikh@utmn.ru; улица Володарского, 6, 625003, Тюмень, Россия; р.т.: +7(932)320-0036.

Прохошин Алексей Сергеевич — магистрант, руководитель проекта, лаборатория сельскохозяйственной микологии и биологической защиты растений института экологической и сельскохозяйственной биологии (x-bio), Тюменский государственный университет (ТюмГУ). Область научных интересов: интеллектуальные системы фитосанитарного мониторинга, системы автоматизированного выращивания растений в контролируемых условиях на малообъемной гидропонике, нейросети распознавания заболеваний и идентификации вредителей. Число научных публикаций — 5. a.s.prokhoshin@utmn.ru; улица Володарского, 6, 625003, Тюмень, Россия; р.т.: +7(932)320-0230.

Глухих Дмитрий Игоревич — аспирант, ассистент, кафедра информационных систем, Тюменский государственный университет (ТюмГУ). Область научных интересов: интеллектуальные системы поддержки принятия решений, базы знаний, вывод на основе прецедентов, нейросети в принятии решений, инженерно-технические и информационные автоматизированные системы мониторинга технических систем. Число научных публикаций — 25. gluhihdmitry@gmail.com; улица Володарского, 6, 625003, Тюмень, Россия; р.т.: +7(932)320-0036.

I. GLUKHIKH, A. PROKHOSHIN, D. GLUKHIKH
**COMPARISON AND RETRIEVAL OF SITUATIONS IN THE CASE-
BASED REASONING SYSTEM FOR SMART-FARM**

Glukhikh I., Prokhoshin A., Glukhikh D. Comparison and Retrieval of Situations in the Case-Based Reasoning System for Smart-Farm.

Abstract. The trend of development of smart farms is aimed at their becoming fully autonomous, robotic enterprises. The prospects for the intellectualization of agricultural production and smart farms, in particular, today are associated with the development of technology systems used to detect, recognize complex production situations and search for effective solutions in these situations. The article presents the concept of such a decision support system on smart farms using the method of decision support based on case-based reasoning - CBR system. Its implementation requires a number of non-trivial tasks, which include, first of all, the tasks of formalizing the presentation of situations and creating methods for comparing and retrieving situations from the KB on this basis. In this study, a smart farm is presented as a complex technological object consisting of interrelated components, which are the technological subsystems of a smart farm, the products produced, the objects of the operational environment, as well as the relationships between them. To implement algorithms for situational decision-making based on precedents, a formalized representation of the situation in the form of a multivector is proposed. This allowed us to develop a number of models of the trained similarity function between situations. The conducted experiments have shown the operability of the proposed models, on the basis of which ensemble architecture of a neural network has been developed for comparing situations and selecting them from the knowledge base in decision-making processes. Of practical interest is monitoring the condition of plants by their video and photo images, which allows detecting undesirable plant conditions (diseases), which can serve as a signal to activate the process of searching for solutions in the knowledge base.

Keywords: case-based reasoning, decision making, neural network, neural network architecture, smart farm.

References

1. Rutkin N.M., Lagutkin O.Ju., Lagutkina L.Ju. [Urbanized agricultural production (city farming) as a promising direction for the development of international agricultural production and a way to ensure the food security of cities]. *Vestnik astrahanskogo gosudarstvennogo tehnikeskogo universiteta. serija: rybnoe hozjajstvo – Bulletin of the Astrakhan State Technical University. Series: fisheries.* 2017. vol. 2017. no. 4. pp. 95–108. (In Russ.).
2. Suraj N.M., Kudinova M.G., Uvarova E.V., Zhidkih E.I. [Analysis of the development of digital technologies in smart farms]. *Innovacii i investicii – Innovation and investment.* 2021. no. 10. pp.184–188. (In Russ.).
3. Martin M., Molin E. Environmental Assessment of an Urban Vertical Hydroponic Farming System in Sweden. *Sustainability.* 2019. vol. 11(15). no. 4124. DOI: 10.3390/su11154124.
4. Chiu M.-C., Yan W.-M., Bhat S.A., Huang N.-F. Development of smart aquaculture farm management system using IoT and AI-based surrogate models. *Journal of Agriculture and Food Research.* 2022. vol. 9. no. 100357. DOI: 10.1016/j.jafr.2022.100357.

5. Devapal D. Smart Agro Farm Solar Powered Soil and Weather Monitoring System for Farmers. Proceedings of International Multi-conference on Computing, Communication, Electrical & Nanotechnology, I2CN-2K19. 2020. pp. 1843–1854.
6. He L., Fu L., Fang W., Sun X., Suo R., Li G., Zhao G., Yang R., Li R. IoT-based urban agriculture container farm design and implementation for localized produce supply. *Computers and Electronics in Agriculture*. 2022. vol. 203. no. 107445. DOI: 10.1016/j.compag.2022.107445.
7. Klaina H., Guembe I.P., Lopez-Iturri P., Campo-Bescós M.A., Azpilicueta L., Aghzout O., Alejos A.V., Falcone F. Analysis of low power wide area network wireless technologies in smart agriculture for large-scale farm monitoring and tractor communications. *Measurement*. 2022. vol. 187(5). no. 110231. DOI: 10.1016/j.measurement.2021.110231.
8. Mahmudul Hasan A., Md Rakib Ul Islam R., Avinash K. [Apple Leaf Disease Classification Using Image Dataset: A Multilayer Convolutional Neural Network Approach]. *Informatika i avtomatizacija – Informatics and Automation*. 2022. vol. 21. no. 4. pp. 710–728. DOI: 10.15622/ia.21.4.3. (In Russ.).
9. Moreira R., Moreira L.F.R., Munhoz P.L.A., Lopes E.A., Ruas R.A.A. AgroLens: A low-cost and green-friendly Smart Farm Architecture to support real-time leaf disease diagnostics. *Internet of Things*. 2022. vol. 19. no. 100570. DOI: 10.1016/j.iot.2022.100570.
10. Hu W.-C., Chen L.-B., Huang B.-K., Lin H.-M. A Computer Vision-Based Intelligent Fish Feeding System Using Deep Learning Techniques for Aquaculture. *IEEE Sensors Journal*. 2022. vol. 22. no. 7. pp. 7185–7194. DOI: 10.1109/JSEN.2022.3151777.
11. Cho S., Kim T., Jung D.-H., Park S.H., Na Y., Ihn Y.S., Kim K.G. Plant growth information measurement based on object detection and image fusion using a smart farm robot. *Computers and Electronics in Agriculture*. 2023. vol. 207. no. 107703. DOI: 10.1016/j.compag.2023.107703.
12. Cerutti J., Abi-Zeid I., Lamontagne L., Lavoie R., Rodrigue-Pinzon M.J. A case-based reasoning tool to recommend drinking water source protection actions. *Journal of Environmental Management*. 2023. vol. 331. no. 117228. DOI: 10.1016/j.jenvman.2023.117228.
13. Zhai Z., Martínez J.F., Martínez N.L., Díaz V.H. Applying case-based reasoning and a learning-based adaptation strategy to irrigation scheduling in grape farming. *Computers and Electronics in Agriculture*. 2020. vol. 178. no. 105741. DOI: 10.1016/j.compag.2020.105741.
14. Wang D., Wan K., Ma W. Emergency decision-making model of environmental emergencies based on case-based reasoning method. *Journal of Environmental Management*. 2020. vol. 262(9). 110382. DOI: 10.1016/j.jenvman.2020.110382.
15. Mathisen B.M., Bach K., Aamodt A. Using extended siamese networks to provide decision support in aquaculture operations. *Applied Intelligence*. 2021. vol. 51(1). DOI: 10.1007/s10489-021-02251-3.
16. Aamodt A., Plaza E. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*. 2001. vol. 7. pp. 39–59. DOI: 10.3233/AIC-1994-7104.
17. Skobelev P.O., Simonova E.V., Budaev D.V., Voshhuk G.Ju., Larjuhin V.B. [Cloud intelligent system SMART FARMING for precision farming management]. *Materialy konferencii «Informacionnye tehnologii v upravlenii (ITU-2018)»* [Materials of the conference "Information Technologies in Management (ITU-2018)"]. St. Petersburg, Concern "Concern" Central Research Institute "Elektropribor", 2018. pp. 261–270.
18. Leake D., Ye X., Crandall D. Supporting Case-Based Reasoning with Neural Networks: An Illustration for Case Adaptation. Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021).

2021. Available at: <https://proceedings.aaai-make.info/AAAI-MAKE-PROCEEDINGS-2021/paper1.pdf> (accessed 26.05.2023).
19. Guo Y., Zhang B., Sun Y., Jiang K., Wu K. Machine learning based feature selection and knowledge reasoning for CBR system under big data. *Pattern Recognition*. 2021. vol. 112(6). no. 107805. DOI: 10.1016/j.patcog.2020.107805.
 20. Smiti A., Elouedi Z. Dynamic maintenance case base using knowledge discovery techniques for case based reasoning systems. *Theoretical Computer Science*. 2020. vol. 817. pp 24–32. DOI: 10.1016/j.tcs.2019.06.026.
 21. Liao T.W., Zhang Z., Mount C.R. Similarity measures for retrieval in case-based reasoning systems. *Applied Artificial Intelligence*. 1998. vol. 12(4). pp. 267–288. DOI: 10.1080/088395198117730.
 22. Fan Z.-P., Li Y.-H., Wang X., Liu Y. Hybrid similarity measure for case retrieval in CBR and its application to emergency response towards gas explosion. *Expert Systems with Applications*. 2014. vol. 41(5). pp. 2526–2534. DOI: 10.1016/j.eswa.2013.09.051.
 23. Oyelade O.N., Ezugwu A.E. A case-based reasoning framework for early detection and diagnosis of novel coronavirus. *Informatics in Medicine Unlocked*. 2020. vol. 20(6). no. 100395. DOI: 10.1016/j.imu.2020.100395.
 24. Gabel T., Godehardt E. Top-down induction of similarity measures using similarity clouds. *International Conference on Case-Based Reasoning*. 2015. pp. 149–16. DOI: 10.1007/978-3-319-24586-7_11.
 25. Mathisen B.M., Aamodt A., Bach K., Langseth H. Learning similarity measures from data. *Progress in Artificial Intelligence*. 2020. vol. 9. pp. 129–143. DOI: 10.1007/s13748-019-00201-2.
 26. Glukhikh I., Glukhikh D. Case-Based Reasoning with an Artificial Neural Network for Decision Support in Situations at Complex Technological Objects of Urban Infrastructure. *Applied System Innovation*. 2021. vol. 4(73). 12 p. DOI: 10.3390/asi4040073.26.
 27. Gluhih I.N., Gluhih D.I. [Algorithms for generating training sets in a system with case-based inference based on example situations]. *Programmnyye produkty i sistemy – Software&Systems*. 2022. vol. 35. no. 4. pp. 660–669. (In Russ.).
 28. Myttenaere A.D., Golden B., Grand B.L., Rossi F. Mean Absolute Percentage Error for regression models. *Neurocomputing*. 2016. vol. 192. pp. 38–48. DOI: 10.1016/j.neucom.2015.12.114.
 29. Wang Y., Wang L., Li Y., He D., Liu T.-Y., Chen W. A Theoretical Analysis of NDCG Type Ranking Measures. *Computer Science*. 2013. 26 p. DOI: 10.48550/arXiv.1304.6480.
 30. Taylor J.R. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. Second Edition. Paperback & Clothbound, 1997. 327 p.
 31. Paulson P., Juell P. Using Reinforcement Learning for Similarity Assessment in Case-Based Systems. *IEEE Intelligent Systems*. 2003. vol. 18. no. 4. pp. 60–67. DOI: 10.1109/MIS.2003.1217629.
 32. Glukhikh I., Chernysheva T., Glukhikh D. Neural Network Models for Situation Similarity Assessment in hybrid-CBR. *Journal of Intelligent & Fuzzy Systems*. 2023. vol. 44(15). pp. 1–14. DOI: 10.3233/JIFS-221335.

Glukhikh Igor — Ph.D., Dr.Sci., Professor, Head of department, Department of information systems, Institute of Mathematics and Computer Science; Scientific Director of the Center for System Engineering, Advanced Engineering School, University of Tyumen (UTMN). Research interests: intelligent decision support systems, knowledge bases, precedent-based inference, neural networks in decision making, case-based reasoning. The number of publications — 180. i.n.glukhikh@utmn.ru; 6, Volodarskogo St., 625003, Tyumen, Russia; office phone: +7(932)320-0036.

Prokoshin Alexey — Undergraduate, project leader, Laboratory of agricultural mycology and biological plant protection of the institute of ecological and agricultural biology (x-bio), University of Tyumen (UTMN). Research interests: intelligent systems for phytosanitary monitoring, systems for automated plant growing under controlled conditions in low-volume hydroponics, neural networks for disease recognition and pest identification. The number of publications — 5. a.s.prokoshin@utmn.ru; 6, Volodarskogo St., 625003, Tyumen, Russia; office phone: +7(932)320-0230.

Glukhikh Dmitry — Graduate student, assistant, Department of information systems, University of Tyumen (UTMN). Research interests: intelligent decision support systems, knowledge bases, precedent-based inference, neural networks in decision making, engineering and information automated systems for monitoring technical systems. The number of publications — 25. gluhihdmitry@gmail.com; 6, Volodarskogo St., 625003, Tyumen, Russia; office phone: +7(932)320-0036.

ABDULLAH HESHAM. M.A., KUMARA.V. SENTHIL,
QASEM AHMEDA.A., MOSLEH M.A.S.
**HYBRID OPTIMIZATION BASED ON SPECTRUM AWARE
OPPORTUNISTIC ROUTING FOR COGNITIVE RADIO AD HOC
NETWORKS**

Abdullah Hashem. M.A., Kumar A.V. Senthil, Qasem Ahmed A.A., Mosleh M.A.S. **Hybrid Optimization Based on Spectrum Aware Opportunistic Routing for Cognitive Radio Ad Hoc Networks.**

Abstract. Opportunistic routing has increased the efficiency and reliability of Cognitive Radio Ad-Hoc Networks (CRAHN). Many researchers have developed opportunistic routing models, among them the Spectrum Map-empowered Opportunistic Routing (SMOR) model, which is considered a more efficient model in this field. However, there are certain limitations in SMOR, which require attention and resolution. The issue of delay and degradation of packet delivery ratio due to non-consideration of network bandwidth and throughput are addressed in this paper. In order to resolve these issues, a hybrid optimization algorithm comprising firefly optimization and grey wolf optimization algorithms are used in the basic SMOR routing model. Thus, developed Hybrid Firefly and Grey-Wolf Optimization-based SMOR (HFGWOSMOR) routing model improves the performance by high local as well as global search optimization. Initially, the relationship between the delay and throughput is analyzed and then the cooperative multipath communication is established. The proposed routing model also computes the energy values of the received signals within the bandwidth threshold and time; hence, the performance issues found in SMOR are resolved. To evaluate its efficiency, the proposed model is compared with SMOR and other existing opportunistic routing models, which show that the proposed HFGWOSMOR performs better than other models.

Keywords: Cognitive Radio Ad Hoc Networks, Opportunistic routing, Spectrum Map-empowered Opportunistic Routing, Firefly optimization, Grey-Wolf optimization, bandwidth threshold.

1. Introduction. Cognitive radio ad hoc network (CRAHN) is a type of distributed, self-organizing, self-Configuring wireless network in which the radios in the network can adapt their transmission and reception parameters in real-time, depending on the availability of the frequency spectrum.

The cognitive radio devices can sense the presence of other radio signals in specific bands, and dynamically adjust their transmission parameters to avoid interference with other devices, and, thus, maximizing the utilization of the available spectrum. This approach leads to a more efficient use of the radio spectrum, reducing the possibility of interference and increasing the capacity of the network. This type of Network uses cognitive radio technologies to allocate network resources dynamically such as frequency, bandwidth and power.

In CRAHN, nodes (CR-users) can sense the availability of the radio spectrum and adjust their transmission parameters accordingly to avoid

interference with other users. This enables efficient utilization of the free available spectrum and improves network performance [1, 2].

CRAHNs can automatically search, monitor and use the available free spectrum to take care of the problem of the spectrum resource shortage and low utilization rate without affecting licensed users.

The cognitive radio rule has acquainted the thought with abuse spectrum holes (i.e., bands) which result from the demonstrated underutilization of the electromagnetic spectrum by present-day wireless communication and broadcasting advancements [3].

CRAHNs are often used in military, rural connectivity and emergency communication scenarios where the availability of spectrum is limited and rapidly dynamically reconfigure itself to maintain communication in the face of changing conditions [4].

The components of a Cognitive Radio Ad Hoc Network (CRAHN) are [4, 5, 6]:

- *Cognitive Radio Nodes*: The main component of the CRAHNs is the cognitive radio nodes (Secondary-User), which are equipped with the radio transceivers and the processing capabilities to monitor and adapt to the surrounding radio environment [6].

- *Spectrum Sensing*: Each node in CRAHNs infrastructure is equipped with a spectrum-sensing module to detect the presence of other users in the radio environment.

- *Decision Making*: Based on the information gathered through the sensing of spectrum, the nodes make decisions on which frequency band to use for communication, and how to allocate the available spectrum resources [7].

- *Spectrum Management*: The nodes in a CRAHN use their cognitive abilities to manage the available spectrum resources dynamically to avoid interference with other users and optimize the performance of the network.

- *Routing* the nodes in a CRAHN use routing protocols to dynamically establish and maintain communication links with other nodes in the network [8, 9].

- *Network Management*: CRAHNs use the network management techniques to monitor network performance and make adjustments to ensure optimal operation [10]. These components work together to enable dynamic, self-organizing, and efficient communication in a Cognitive Radio Ad Hoc Network [5, 6].

In cognitive radio, the secondary users (SUs) refer to a specific device that dynamically accesses and uses the underutilized portions of the radio spectrum that are licensed to primary users (PUs) such as government

agencies or licensed commercial operators. The secondary user (US) operates on a non-interfering basis with the primary users and can vacate the spectrum when the primary user requests access [7, 8].

The primary users in a cognitive radio network are the licensed or authorized users who have been assigned the use of a particular frequency band by a regulatory authority, such as a government or standardization organization [8, 9].

They have the primary rights to use the spectrum bands and are usually traditional users, such as government agencies, television and the radio broadcast stations, or mobile networks. They have priority over the secondary users in accessing the radio spectrum and can use it without interference [10].

The main aim of CR-AHNs is to increase the utilization of available spectrum by detecting and avoiding busy frequency bands, and exploiting unused ones. The nodes in CR-AHNs can also cooperate and share information with each other to make more efficient decisions about spectrum utilization. This technology results in increased network performance, efficiency and capacity, energy consumption, provides better quality of service to users, and improved overall performance [11]. On the other hand, the key idea behind CRAHNs is to allow wireless devices to sense and adapt to changes in the radio environment, such as the presence of other devices, interference, or changes in channel conditions.

In order to handle these difficulties, the opportunistic routing (OR) strategy has been connected in CRAHNs with a specific end goal to uncover the effect of the spectrum availability on the stability of the routing. Considering the predominance of the broadcast feature and the exceptionally decent variety of wireless mediums, the OR strategy has been earlier proposed in the amazingly opportunistic routing protocol (ExOR) [6, 7, 12]. Instead of firstly deciding the following hop SU and after that sending the packet to the following hop SU, a SU with the OR strategy broadcasts the packet keeping in mind the end goal to get the outcomes that all neighbors of the SU have the chance to get the packet and help with forwarding the data packets.

Contrasted with the traditional routing methodologies, the OR strategy brings the high throughput gains. Additionally, it is likewise hard to keep up the routing table for a SU because of the embodiment of dynamic spectrum access [13]. Consequently, the pre-decided end-to-end routing cannot be fitting for the CRAHN situation. Because of the way that the OR strategy does not require the earlier foundation of the routes, the OR strategy is more suitable to be utilized in the CRAHN situation with

dynamic changes of channel availability because of the dynamic behavior of PUs [9, 14].

On analysis, the SMOR routing model [10] has been found to be the most efficient OR strategy for CRAHNs. Previously some improved models of OR have been proposed [11 – 16]. However, due to the limitations of performance due to network bandwidth and throughput in SMOR, leads this paper to develop an HFGWO-SMOR routing model, which utilized hybrid Firefly, and grey-wolf optimization algorithms to improve the delay-throughput relationship analysis and improve the cooperative multipath communication.

The main challenges and issues in CRAHNs. Besides the basic challenges and issues such as (Spectrum Sensing, Spectrum Management and Allocation, Interference Management, Routing and Network Protocols, Security and Privacy), there are some of the principal issues are:

- Minimize the energy consumption of the network while ensuring the reliable data transmission; it takes into account the dynamic spectrum availability and channel conditions to make routing decisions.
- The interaction between primary users and secondary users that while achieving an optimal network performance.
- Optimization of spectrum sensing and routing in cognitive radio ad hoc networks; routing decisions to maximize network throughput while avoiding interference to primary users.
- Quality-of-service (QoS)-aware opportunistic routing in multi-channel cognitive radio ad hoc networks.

The main Contributions of this paper are:

- Deep study of the Cognitive Radio Ad Hoc Network;
- This work had made significant contributions to the understanding and development of cognitive radio systems;
- Analysis of the existing studying of the delay, and degradation of packet delivery ratio due to non-consideration of network bandwidth and throughput problems;
- Proposing a new modeled based on the “hybrid optimization model” to solve the above problems.

Structure of this research paper. The rest of the article is organized as Section 2, which presents a review of related research works. Section 3 presents the proposed system model and Section 4 explains the proposed hybrid optimization model and utilization of it in the OR strategy. Section 5 evaluates the performance of the proposed model while Section 6 makes a conclusion about this routing model.

2. Related Work. There are several research papers, which focus on the CRAHNs, these papers, serve as a starting point for understanding and

exploring the field of hybrid optimization-based spectrum-aware opportunistic routing in cognitive radio ad hoc networks.

As stated above, the interest in the CRAHN routing models has been very high recently. Many existing efforts focus on developing OR strategy with channel assignment and maximizing network throughput. In [17], a new route metric called multichannel expected any path transmission time is proposed, which exploits the channel assorted variety and resource of multiple applicant forwarders for the opportunistic routing. In light of the new metric, a distributed algorithm named channel-aware opportunistic routing is also displayed.

In study [18], an online opportunistic routing algorithm is proposed by utilizing multi-specialist support learning; introduces the concept of opportunistic spectrum access in cognitive radio networks and proposes an optimization-based approach for selecting the best available spectrum bands for communication. The proposed routing plan together addresses the connection and relay determination in light of transmission achievement probabilities. This advanced learning system effectively investigates openings in part recognizable and non-stationary conditions of CRAHNs.

In study [19], the randomization structure is summed up, which is initially proposed for the information line changing to a SNR – based interference model in multi-hop wireless networks. Further, circulated power assignment and correlation calculation are produced, which accomplishes about 100% throughput. In study [20], a Bayesian decision rule-based algorithm to take care of the throughput maximization problem ideally with steady time multifaceted nature is proposed. To organize PU transmissions, the throughput maximization problem is re-detailed by adding a constraint on the PU throughput.

In study [21], the throughput execution of the network is portrayed by utilizing a lining theoretic investigation, and throughput is additionally boosted by means of the use of the Lagrangian duality hypothesis. In study [22], by applying the convex optimization method, the shut-shape articulation for the ideal time portions is acquired to boost the sum throughput. To beat this problem, another execution metric known as the common throughput is proposed, which considers the additional constraint that all users ought to be designated with an equivalent rate paying little respect to their distances to the H-AP.

In study [11], presents a hybrid optimization algorithm for opportunistic routing in cognitive radio ad hoc networks. This algorithm uses the hybrid artificial bee colony optimization to achieve a trade-off between exploration and exploitation in the route selection process, considering spectrum availability and energy efficiency. The authors

propose a routing protocol that takes into account the variation in channel conditions and utilizes a particle swarm optimization algorithm to select the best routes based on spectrum availability and link quality.

In study [10] the authors developed the SMOR model, the Spectrum-Map-Empowered Opportunistic Routing (SMOR) model focuses on leveraging spectrum mapping techniques to enhance opportunistic routing in the cognitive radio ad hoc networks (CRAHNs). The model is designed to address the challenges posed by dynamic spectrum availability in CRAHNs; which was developed separately for both regular CRAHNs as SMOR-1 algorithm and large scale as regular CRAHNs as SMOR-2 algorithm. By incorporating spectrum mapping and opportunistic routing, the SMOR model likely aims to improve spectrum utilization, enhance overall network performance, and mitigate the effects of varying spectrum availability in CRAHNs.

SMOR-1 Algorithm (for regular CRAHNs): The SMOR-1 algorithm, specifically tailored for regular CRAHNs, aims to optimize opportunistic routing by utilizing a spectrum map. The spectrum map provides information about the availability and quality of different spectrum bands in the network. Based on this information, the SMOR-1 algorithm selects the most suitable spectrum band and path for data transmission, considering factors such as channel conditions and interference [10, 11, 26].

SMOR-2 Algorithm (for large-scale CRAHNs): The SMOR-2 algorithm, developed for large-scale CRAHNs, extends the concepts of the SMOR-1 to address the scalability issues inherent in larger networks. It aims to efficiently utilize spectrum resources while considering the challenges of topology dynamics and resource limitations in large-scale CRAHNs. The SMOR-2 algorithm may incorporate additional optimizations or techniques to handle the increased complexity and scale of the network [10].

In Stochastic geometry analysis for regular CRAHNs, the mathematical analysis for transmission delay of multi-hop communications is examined via Markov chain modelling and queuing network theory, and the SMOR-1 algorithm is proposed to exploit opportunistic selections for cooperative relay regarding link transmission qualities. For large-scale CRAHNs, the corresponding delay of opportunistic links is derived via stochastic geometry and queuing network analysis, and the SMOR-2 algorithm is proposed to fulfill geographic opportunistic routing, exhibiting cooperative diversity in such large-scale networks [10, 26].

3. System Model. Due to the challenges that face the decentralized infrastructure of Cognitive radio ad hoc network, and due to the fact that CRAHN has no infrastructure backbone, we considered the system to

involve with a finite of ($M= 10$) Primary users (PUs), every PU has its own licensed spectrum to communication in specific spectrum band $X' \subseteq X$ where X is estimated area of 500m x500 m.

PUs share an unused channel with Secondary Users (SUs) which are specified with ($N=100$) SUs when PU is in the off state, SU is able to find a PU spectrum hole to establish connection and communication with a single transmitter Tx and K receivers Rx over the time interval $[t0, T]$.

Let n denote the number of the transmitting and receiving pairs for SU and $T_n = \{1,2,...,n\}$ denotes the set of SU where the pairs of transmitting and receiving of SU i (SU i for $i \in N$) are changeable based on the PU activities; that means, the licensed spectrum of PU i should be busy during transmitting and receiving of PU i , otherwise the opportunistic Spectrum will be available for SU i . Figure 1 shows the system model utilized in this paper [25 – 31].

This system model proposes a hybrid optimization-based routing protocol for cognitive radio ad hoc networks. It combines genetic algorithms and particle swarm optimization to optimize the routing path selection while considering spectrum availability.

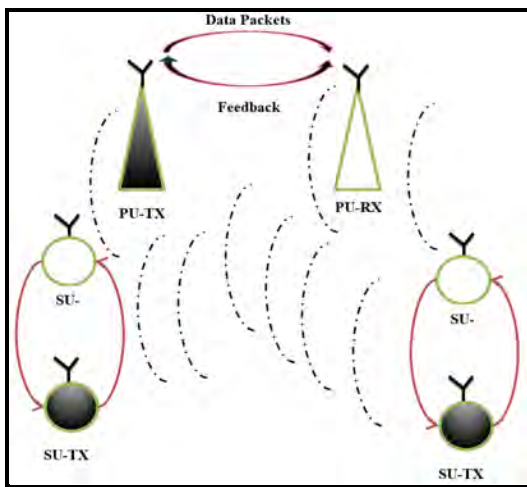


Fig. 1. System Model

The following Table 1 illustrates the main parameters in the system model, the energy of transmission in PU i is denoted by p_i^p , as well as the transmission energy in SU i is denoted by p_i^s , while it is supposed that the pairs of SU i transmitter/receiver are inside the communication range of each other.

Table 1. Simulation Parameter

No.	Simulation Parameters	
	Parameter	Parameter Value
1	Simulation Area Size	500m x500 m.
2	Simulation time	120s
3	Number of CR-Nodes (SUs)	100
4	Number of PU-Nodes	10
5	Size of frames to be scheduled	64 to 196 kb
6	Variable transmission time ranges	10–50 μ s
7	Number of channels utilized to schedule the frame transmission	35
8	Size bandwidth available per channel	2 MB/s
9	Distance between the Nodes	Random
10	Node Energy Capacity	250mAH

4. Hybrid Firefly and Gray-Wolf optimization based on SMOR Model. This proposed routing model follows the processes in SMOR; the existing SMOR model has been shown to improve and enhance the network throughput, reduce the delay and the packet loss, then, enhance the network resilience to channel variations and node failures. However, it also required careful design and optimization of the spectrum sensing, channel selection, and opportunistic routing algorithms, as well as the handoff criteria and the routing metrics.

Overall, the SMOR model is a promising approach to improve the performance and efficiency of cognitive radio ad hoc networks as explained in the previous relative work section in this paper; therefore, the proposed new approach is developed to enhance the performance, efficient opportunistic routing in CRAHN via hybrid firefly and Gray-Wolf optimization approach.

Based on these strategies, the proposed HFGWO-SMOR model has developed. The relationship between the delay and throughput is optimized using the hybrid algorithms.

The behavior of the fireflies and the Gray wolves are merged to develop this model. First, the basic concept of these two optimization models has been discussed in [25]. The flashing behaviors of fireflies are utilized to develop firefly-inspired algorithms.

Firefly Approach (FA) and Gray Wolf Optimization Approach are a metaheuristic optimization algorithm.

4.1. Firefly Approach (FA). Xin-She Yang defined the Firefly Algorithm, which is an optimization algorithm that is based on the flashing characteristics of fireflies. It was proposed by the author in 2008 as a novel optimization technique to solve complex optimization problems. The algorithm is inspired by the properties of fireflies, which use their flashing behavior to attract mates and communicate with each other [33].

The algorithm then simulates the flashing characteristics of the fireflies, where the intensity of their flashes represents the quality of the solution they represent. On the other hand, the algorithm models the behavior of fireflies, which communicate with each other through bioluminescence. The brightness of a firefly's light is proportional to its attractiveness to other fireflies, and fireflies tend to move toward the brightest light they can see [23, 24].

Firefly Approach; this algorithm uses a set of parameters, such as the light absorption coefficient and the step size, to control the movement of the fireflies. On the other hand, the firefly's movements are also influenced by the distance between the fireflies, with closer fireflies having a stronger attraction.

The Firefly algorithm has been shown to be effective in solving a wide range of optimization problems, including function optimization, parameter estimation, and machine learning. It is also known for its simplicity and fast convergence rate.

The proposed hybrid algorithm is developed by hybridizing both of these behaviors. For a maximization problem, it obtains the highest possible value of the fireflies function, the brightness and flashing can be proportional to the value of the possible objective function. In maximum optimization problems, the brightness I of a firefly at a particular location x can be chosen as $I(x) \propto f(x)$. However, the attractiveness β is relative; it should be seen in the eyes of the beholder or judged by the other fireflies.

Thus, it will vary with the distance r_{ij} between firefly i and firefly j . In addition, light intensity decreases with the distance from its source, and light is also absorbed in the media, so we should allow the attractiveness to vary with the degree of absorption [23 – 26]:

$$\beta_{i,j} = \beta_0 e^{-\gamma r_{ij}^2}, \quad (1)$$

where β_0 is the attractiveness at $r=0$. and γ is the light exhaust coefficient. The distance between two transmitters of i and receiver j is arrived using def. The movement of transmitter i as its being powered by the brighter receiver j is calculated as follows:

$$\Delta x_i = \beta_0 e^{-\gamma r_{ij}^2} (x_j^t - x_i^t) + \alpha \varepsilon_i, \quad (2)$$

where the γ second term is due to the attraction. The third term α is a randomization vector drawn from a Gaussian disposal.

The algorithm uses the following phases:

1. Initialization: Generate an initial population of fireflies with random positions and intensities.
2. Fitness Evaluation: Evaluate the fitness of each firefly based on the problem's objective function.
3. Attraction: Move each firefly towards the brightest firefly (i.e., the one with the highest intensity) in its vicinity, where the degree of attraction is based on the distance between the fireflies and their intensities.
4. Randomization: Introduce random movement to each firefly to prevent premature convergence and to explore new areas of the search space.
5. Updating: Update the positions and intensities of the fireflies based on their movements and fitness values.
6. Termination: The algorithm stops when a certain stopping criterion is met, such as a maximum number of iterations, or when the desired accuracy is achieved.

Definition: Light intensity: The light intensity of each firefly is calculated as follows:

$$I_i = f(x_i), \quad (3)$$

where I_i is the light intensity of firefly i and $f(x_i)$ is the fitness value of firefly i .

Attraction: Each firefly is attracted to other fireflies based on their light intensity and distance. The attraction of firefly i towards firefly j is calculated as follows:

$$r_{ij} = \|x_j - x_i\|, \quad (4)$$

$$\beta = \beta_0 * \exp(-\gamma * r_{ij}^2), \quad (5)$$

$$x_i = x_i + \alpha * (x_j - x_i) + \beta * \varepsilon_i, \quad (6)$$

where r_{ij} is the Euclidean distance between fireflies i and j , β_0 is the initial attractiveness, γ is the light absorption coefficient, α is the step size, and ϵ_i is a random vector with values drawn from a Gaussian distribution [27 – 32].

4.2. Gray Wolf Optimization Approach. In 2014, in studies [31, 34] the authors proposed the Gray Wolf Optimization (GWO) algorithm, inspired by the social hierarchy and hunting behavior of gray wolves in the wild.

It simulates the hunting behavior of wolves, the algorithm starts with an initial population of wolf packs where the range of each pack is from 5 to 12 wolves, each pack consisting of alpha, beta, and delta wolves. In the GWO algorithm, a population of wolves is used to search for the optimal solution to a problem. The alpha wolf is responsible for leading the hunting, while the beta and delta wolves assist the alpha wolf in the hunting process [33, 34].

The GWO algorithm optimizes a function by updating the positions of wolves, which represent the best solutions found so far. The algorithm is based on the social behavior of gray wolves, where each wolf has a specific role in the pack and works together to achieve a common goal.

Definition: The algorithm iteratively searches for the optimal solution by simulating the hunting behavior of the wolf packs. During each iteration, the alpha wolf updates its position based on its hunting experience, while the beta and delta wolves adjust their positions based on the alpha wolf's position. The alpha wolf represents the best solution found so far, the beta wolf represents the second-best solution, and the delta wolf represents the third-best solution [35].

At each iteration, the positions of the wolves are updated using the following equation [32, 33]:

$$x'_{ij} = x_{ij} + a * (2 * r1 - 1) * |A * x_{alpha} - x_{ij}|, \quad (7)$$

where x'_{ij} is the updated position of the i -th wolf in the j -th dimension, x_{ij} is the current position of the i -th wolf in the j -th dimension, a is a coefficient that decreases linearly from 2 to 0 as the number of iterations increases, $r1$ is a random number between 0 and 1, and $A * x_{alpha}$ is the position of the alpha wolf.

The Gray Wolf Optimization algorithm steps:

1. *Initialization:* The algorithm starts with an initial population of n search wolves (where CR user searches for free Spectrum holes) that are randomly distributed in the search space.

2. *Fitness evaluation*: The fitness of each search wolf (get a Spectrum-hole that belongs to PUs) is evaluated by applying the objective function of the optimization problem to its corresponding search space. The search wolf is ranked according to their fitness values, with the best (i.e., lowest) fitness values having the highest rank.

3. *Pack updating (alpha, beta, and delta) wolves*: The algorithm identifies the three best wolves in the population. The position of these wolves is then updated based on the positions of the other wolves in the population. The three search agents with the highest ranks are designated as the alpha, beta, and delta wolves, respectively.

4. *Solution update*: The algorithm updates the positions of the candidate solutions, and checks if the new solutions improve the overall fitness of the pack.

Definition: The algorithm continues to update the positions of the wolves until a stopping criterion is met, such as reaching a maximum number of iterations or a satisfactory solution [35].

Updating the position of the alpha wolf:

$$D_alpha = |C1 * X_alpha - X_i|, \quad (8)$$

$$X1 = X_alpha - A1 * D_alpha. \quad (9)$$

Updating the position of the beta wolf:

$$D_beta = |C2 * X_beta - X_i|, \quad (10)$$

$$X2 = X_beta - A2 * D_beta. \quad (11)$$

Updating the position of the delta wolf:

$$D_delta = |C3 * X_delta - X_i|, \quad (12)$$

$$X3 = X_delta - A3 * D_delta. \quad (13)$$

Updating the position of the other wolves:

$$X_i = (X1 + X2 + X3) / 3, \quad (14)$$

where X_i is the position of the i -th wolf, X_alpha , X_beta , and X_delta are the positions of the alpha, beta, and delta wolves, respectively, $C1$, $C2$, and

$C3$ are random vectors between 0 and 1, $A1$, $A2$, and $A3$ are constants that control the step size of the update.

These equations of GWO are applied to each candidate solution (wolf) in the population in each iteration of the algorithm, and the process continues until a stopping criterion is met (e.g., a maximum number of iterations or a desired level of convergence).

The steps of the HFGWO Approach:

1. Initialize values of Firefly Approach (FA) parameters: population, maximum iterations, attraction coefficient, etc....
2. Initialize Gray Wolf Optimization (GWO) parameters: population (primary-users), search agents (CR-User).
3. Generate initial fireflies.
4. Evaluate fitness and update light intensity.
5. Find the brightest firefly.
6. Update information.
7. Feed FA results to GWO.
8. GWO initializes search agents and solutions.
9. Evaluate fitness.
10. Compare with other agents to determine the best search agent.
11. Verify the result of FA.
12. Return the best grey-wolf firefly agent.

Based on this concept of HFGWO, the SMOR routing model is modified and improved. The proposed model initializes the nodes as fireflies and selects the best firefly using FA while it is cross-checked using GWO to verify the best selection. This concept is presented in the following algorithm.

Algorithm 1. HFGWO-SMOR

```

Initialize network parameters (Number of PUs, SUs, Data Rate, etc...)
Partition traffic into batches of packets
For each time slot
    Source Collect link information
    Prioritize forwarding nodes
    Select data packets for each path via HFGWO
    Initialize FA & GWO parameters: (population, maximum iterations, an
        attraction coefficient and algorithm parameters).
    Find the brightest of fireflies with a high attraction coefficient
    Change attractive level and distance
    Select the best Firefly node
    Feed FA result to GWO
    Verify the node information using GWO
    Initialize the best three solutions, the first best solution as  $\chi\alpha$ , the second
        best solution as  $\chi\beta$ , and the third best solution as  $\chi\delta$ , respectively.
    
```

```
While (k < maximum number of iterations or a desired level of
convergence)
  For i= 1:n
    Update the current position of the search agent based on the
    desired level of convergence
  End for
  Evaluate the fitness.
  Update the coefficient vector
  If there is a better solution, then update the best agents,  $x\alpha$ ,  $x\beta$  and  $x\delta$ .
  k=k+1;
  Return the best forwarding node
  Update the parameters
  Send test data
  If ACK is not received
  Initiate path-checking process
  For each relay node
    Check the packet transmission information
  Update lists
  Return packet data
End While
Transmit data
End For
```

5. Performance Evaluation. The proposed HFGWO-SMOR routing model is evaluated using MATLAB. The routing performance of this model is compared with that of SMOR [10], HABC-SOR [11], HB-SOR [12] and HFSA-SOR [13]. The simulation environment is set as in [10 – 16] and the comparisons are simulated in concepts of end-to-end delay (EED), Bit Error Rate (BER), throughput and packet delivery ratio. MATLAB simulators provide a framework for modeling the various network components and their interactions.

5.1. Delay. Simulating delays in cognitive radio ad hoc networks involves modeling the various factors that contribute to delays in the network. Delays in the network can be caused by factors such as propagation delay, queuing delay, processing delay, and transmission delay. Delay simulation in cognitive radio ad hoc networks can be represented mathematically using a queuing model; queuing models provide a framework for modeling the arrival and service processes in a network, and can be used to estimate the queuing delay and other performance metrics [29].

One commonly used queuing model for delay simulation in cognitive radio networks is the M/G/1 queuing model. In this model, packets arrive according to a Poisson process with rate λ . The queuing delay for each packet can then be calculated as:

$$D = (\rho/(\mu-\lambda)) * (1/2 + (V/2\mu)^2), \quad (15)$$

where $\rho = \lambda/\mu$ is the traffic intensity, and V is the coefficient of variation of the service time distribution. This equation assumes that the service time distribution is memoryless, which is a reasonable assumption for many communication protocols in cognitive radio networks.

$$\text{Total delay} = \text{Propagation delay} + \text{Queuing delay} + \text{Processing delay} + \text{Transmission delay}. \quad (16)$$

Propagation delay: Propagation delay is the time it takes for a signal to be traveled from the transmitter to the receiver, and is dependent on the distance between the nodes and the propagation speed of the medium. Mathematically, propagation delay can be expressed as:

$$\text{Propagation delay} = \text{distance between nodes} / \text{propagation speed of the medium}. \quad (17)$$

Queuing delay: Queuing delay is the time it takes for packets to wait in a buffer before they can be transmitted, and is dependent on the network congestion and the size of the buffer:

$$\text{Queuing delay} = \text{packet size} / \text{available bandwidth}. \quad (18)$$

Processing delay: Processing delay is the time it takes for the node to process a packet before forwarding it, and is dependent on the processing power of the node [30 – 37]. Mathematically, processing delay can be expressed as:

$$\text{Processing delay} = \text{packet size} / \text{processing power of the node}. \quad (19)$$

Transmission delay: Transmission delay is the time it takes for the packet to be transmitted over the wireless medium, and is dependent on the bandwidth of the channel and the size of the packet. Mathematically, transmission delay can be expressed as:

$$\text{Transmission delay} = \text{packet size} / \text{available bandwidth}. \quad (20)$$

5.2. Throughput: The throughput of a cognitive radio ad hoc network is affected by various factors such as the number of nodes in the

network, the data rate of the channel, the propagation delay, the processing delay, and the queuing delay [28, 29].

By using appropriate values for these parameters and applying the following equation, one can simulate the throughput of the network and analyze the performance of the network under different scenarios.

The throughput in HFGWO-SMOR Model is modeled mathematically using the following equation:

$$\text{Throughput} = \text{total number of bits received} / \text{total time}, \quad (21)$$

where the total number of bits received is the number of bits received by all the nodes in the network during a given period of time, and the total time is the time taken for all the bits to be received. The total number of bits received can be calculated as:

$$\text{Total number of bits received} = \text{number of nodes} * \text{data rate} * \text{time}, \quad (22)$$

where the number of nodes is the number of nodes in the network, data rate is the data rate of the channel, and time is the period of time for which the data rate is measured. The total time can be calculated as:

$$\text{Total time} = \text{transmission time} + \text{propagation delay} + \text{processing delay} + \text{queuing delay}. \quad (23)$$

By the way, Table 1 shows the main parameters to simulate the delays and throughput in our system model, in order to model delays in the network; one can configure the simulator to include parameters such as the distance between nodes, the buffer size, the processing power of the nodes, and the bandwidth of the channel. By adjusting these parameters, one can simulate different network scenarios and measure the resulting delays.

It is also important to consider the impact of interference in the network, as cognitive radio networks rely on the ability to detect and avoid interference. Simulating interference was done by introducing competing signals in the network, or by modeling the spectrum sensing capabilities of the nodes.

We obtained the total delay in a cognitive radio ad hoc network in our research. The resulting delay value is used to evaluate the performance of the network and to compare different network configurations and scenarios, the Figure 2 shows the EED vs. lambda comparison of SMOR, HABC-SOR, HB-SOR, HFSA-SOR and the proposed HFGWO-SMOR.

HFGWO-SMOR shows a remarkable improvement in the packet delay aspect, which leads to minimizing the delay in all levels of the offered load with an average of 4%, HFGWO-SMOR model reduced delay than other models because of the improved optimal selection of the routing paths.

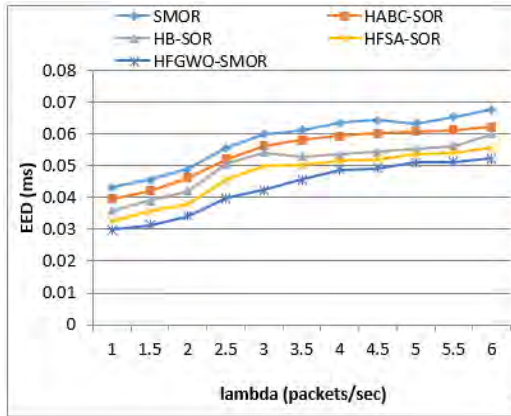


Fig. 2. End-to-end delay

Figure 3 shows the BER vs. lambda comparison of SMOR, HABC-SOR, HB-SOR, HFSA-SOR and the proposed HFGWO-SMOR. HFGWO-SMOR shows a lower error rate with a 4% decrease on average while other models have comparatively higher BER. This is because the path selection is highly reliable in the proposed model.

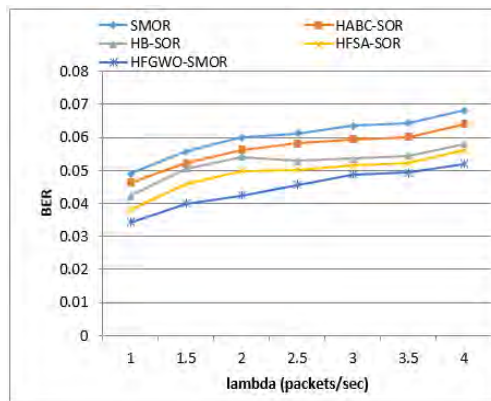


Fig. 3. BER comparison

Figure 4 shows the Throughput vs. lambda comparison of SMOR, HABC-SOR, HB-SOR, HFSA-SOR and the proposed HFGWO-SMOR. HFGWO-SMOR provides a higher throughput rate with a 3% increase on average due to a significant selection of optimal paths while other models have comparatively less throughput.

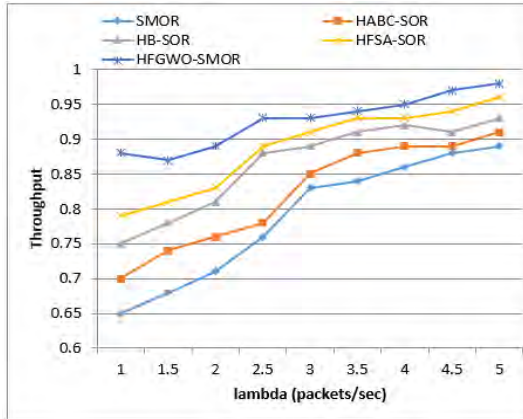


Fig. 4. Throughput comparison

Figure 5 shows the Packet delivery ratio vs. no. of nodes comparison of SMOR, HABC-SOR, HB-SOR, HFSA-SOR and the proposed HFGWO-SMOR.

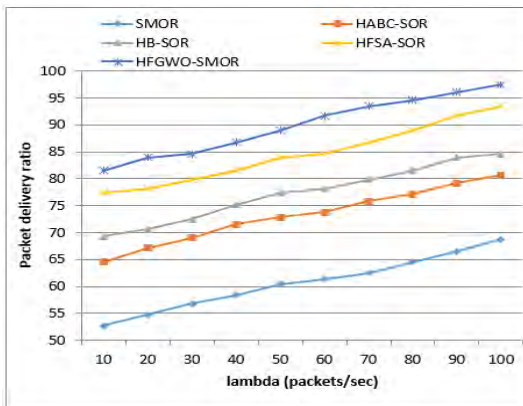


Fig. 5. Packet delivery ratio

HFGWO – SMOR provides a higher packet delivery ratio, which is almost 22% higher than the SMOR model and significantly higher than other models. From the performance evaluation results, it was found that the proposed HFGWO-SMOR model has better performance than other models in providing efficient opportunistic routing.

6. Conclusions. This paper aimed at developing an improved opportunistic routing model that can resolve the limitations of the SMOR model. This has been achieved by the HFGWO-based SMOR routing model that further improves the opportunistic routing behavior. The proposed HFGWO-SMOR model follows the process of SMOR with additional improvement achieved in the optimal selection routing paths. The experimental results also prove that the proposed model has reduced delay, less error rate, improved throughput and improved packet delivery ratio. This model provides more efficient opportunistic routing performance than the other models compared including SMOR, which is evident from the evaluation results. In the future, the feasibility of improving this model by adding viable concepts of path loss, node failures, and power consumption will be examined.

References

1. Akyildiz I.F., Lee W.Y., Chowdhury K.R. CRAHNS: Cognitive radio ad hoc networks. *AD hoc networks*. 2009. vol. 7(5). pp. 810–836.
2. Peha J.M. Approaches to spectrum sharing. *IEEE Communications magazine*. 2005. vol. 43(2). pp. 10–12.
3. Cesana M., Cuomo F., Ekici E. Routing in cognitive radio networks: Challenges and solutions. *Ad Hoc Networks*. 2011. vol. 9(3). pp. 228–248.
4. Chowdhury K.R. Communication protocols for wireless cognitive radio ad-hoc networks. Georgia Institute of Technology, 2009. 153 p.
5. Biswas S., Morris R. Opportunistic routing in multi-hop wireless networks. *ACM SIGCOMM Computer Communication Review*. 2004. vol. 34(1). pp. 69–74.
6. Biswas S., Morris R. ExOR: opportunistic multi-hop routing for wireless networks. *ACM SIGCOMM Computer Communication Review*. 2005. vol. 35(4). pp. 133–144.
7. Badarneh O.S., Salameh H.B. Opportunistic routing in cognitive radio networks: exploiting spectrum availability and rich channel diversity. *IEEE Global Telecommunications Conference (GLOBECOM 2011)*. 2011. pp. 1–5. DOI: 10.1109/GLOCOM.2011.6134241.
8. Abdullah H.M.A., Kumar A.S. A Survey on Spectrum-Map Based on Normal Opportunistic Routing Methods for Cognitive Radio Ad Hoc Networks. *International Journal of Advanced Networking and Applications*. 2015. vol. 7(3), pp. 2761–2770.
9. Lin S.-C., Chen K.C. Spectrum-map-empowered opportunistic routing for cognitive radio ad hoc networks. *IEEE Transactions on Vehicular Technology*. 2014. vol. 63(6). pp. 2848–2861.
10. Lin S.-C., Chen K.-C. Spectrum-Map-Empowered Opportunistic Routing for Cognitive Radio Ad Hoc Networks, *IEEE Transactions on Vehicular Technology*. 2014. vol. 63(6). pp. 2848–2861.
11. Abdullah H.M.A., Kumar A.S. A Hybrid Artificial Bee Colony Based Spectrum Opportunistic Routing Algorithm for Cognitive Radio Ad Hoc Networks.

- International Journal of Scientific and Engineering Research. 2016. vol. 7(6). pp. 294–303.
12. Abdullah H.M.A., Kumar A.S. HB-SOR: Hybrid Bat Spectrum Map Empowered Opportunistic Routing and Energy Reduction for Cognitive Radio Ad Hoc Networks (CRAHNs). International Journal of Scientific and Research Publications (IJSRP). 2017. vol. 7(5). pp. 284–297.
 13. Abdullah H.M.A., Kumar A.V.S. HFSA-SORA: Hybrid firefly simulated annealing based spectrum opportunistic routing algorithm for Cognitive Radio Ad hoc Networks (CRAHN). Conference on Intelligent Computing and Control (I2C2). 2017. DOI: 10.1109/I2C2.2017.8321943.
 14. Abdullah H.M.A., Kumar A.V.S. Modified SMOR Using Sparsity Aware Distributed Spectrum Map for Enhanced Opportunistic Routing in Cognitive Radio Adhoc Networks. Journal of Advanced Research in Dynamical and Control Systems. 2017. vol. 9(6). pp. 184–196.
 15. Abdullah H.M.A., Kumar, A.V.S. Vertex Search based Energy-efficient Optimal Resource Allocation in Cognitive Radio ad hoc Networks. SPIRAS Proceedings. 2018. vol. 2(57). pp. 5–25. DOI: 10.15622/sp.57.1.
 16. Abdullah H.M.A., Kumar A.V.S. Proficient Opportunistic Routing by Queuing Based Optimal Channel Selection for the Primary Users in CRAHN. ARPN Journal of Engineering and Applied Sciences. 2018. vol. 13(5). pp. 1649–1657.
 17. He S.M., Zhang D.F., Xie K., Qiao H., Zhang J. Channel aware opportunistic routing in multi-radio multi-channel wireless mesh networks. Journal of Computer Science and Technology. 2014. vol. 29(3). pp. 487–501.
 18. Barve S.S., Kulkarni P. Multi-agent reinforcement learning based opportunistic routing and channel assignment for mobile cognitive radio ad hoc network. Mobile Networks and Applications. 2014. vol. 19(6). pp. 720–730.
 19. Lee H.W., Modiano E., Le L.B. Distributed throughput maximization in wireless networks via random power allocation. IEEE transactions on mobile computing. 2012. vol. 11(4). pp. 577–590.
 20. Li S., Zheng Z., Ekici E., Shroff N. Maximizing system throughput by cooperative sensing in cognitive radio networks. IEEE/ACM Transactions on Networking (TON). 2014. vol. 22(4). pp. 1245–1256.
 21. El-Sherif A.A., Mohamed A. Decentralized throughput maximization in cognitive radio wireless mesh networks. IEEE Transactions on Mobile Computing. 2014. vol. 13(9). pp. 1967–1980.
 22. Ju H., Zhang R. Throughput maximization in wireless powered communication networks. IEEE Transactions on Wireless Communications. 2014. vol. 13(1). pp. 418–428.
 23. Ping S., Aijaz A., Holland O., Aghvami A.H. Energy and interference aware cooperative routing in cognitive radio ad-hoc networks. IEEE Wireless Communications and Networking Conference (WCNC). 2014. pp. 87–92. DOI: 10.1109/WCNC.2014.6951927.
 24. Meghanathan N., Fanuel M. A minimum channel switch routing protocol for cognitive radio ad hoc networks. 12th International Conference on Information Technology-New Generations (ITNG). 2015. pp. 280–285.
 25. Mirjalili S., Saremi S., Mirjalili S.M., Coelho L.D.S. Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization. Expert Systems with Applications. 2016. vol. 47. pp. 106–119.
 26. Poornimha J., Kumar A.V.S., Abdullah H.M.A. A New Approach to Improve Energy Consumption Time and Life Time using Energy Based Routing in WSN. Emerging Trends in Industry 4.0 (ETI 4.0). 2021. pp. 1–6. DOI: 10.1109/ETI4.051663.2021.9619412.

27. Yang Qin, Xiaoxiong Zhong, Yuanyuan Yang, Yanlin Li, Li Li. Joint channel assignment and opportunistic routing for maximizing throughput in cognitive radio networks. IEEE, 2014. DOI: 10.1109/GLOCOM.2014.7037532.
28. Dutta N., Sarma H.K.D., Polkowski Z. Cluster based routing in cognitive radio adhoc networks: Reconnoitering SINR and ETT impact on clustering. Computer Communications. 2018. vol. 115. pp. 10-20. DOI: 10.1016/j.comcom.2017.09.002.
29. Kumar A.V.S., Abdullah H.M.A., Hemashree P. An Efficient Geographical Opportunistic Routing Algorithm Using Diffusion and Sparse Approximation Models for Cognitive Radio Ad Hoc Networks. (Eds.: Smys S., Ilyyasu A.M., Bestak R., Shi F.) New Trends in Computational Vision and Bio-inspired Computing: Selected works presented at the ICCVBIC 2018. 2020. pp. 323–333. DOI: 10.1007/978-3-030-41862-5_30.
30. Abdullah H.M.A., Kumar A.V.S. Selective Cooperative Jamming Based Relay Selection and Blowfish Encryption for Enhancing Channel and Data Security in CRAHN Routing (Ed.: Elkhodr M.). Enabling Technologies and Architectures for Next-Generation Networking Capabilities. 2019. pp. 105–124. DOI: 10.4018/978-1-5225-6023-4.ch00.
31. Mirjalili S., Mirjalili S.M., Lewis A. Grey Wolf Optimizer. Advances in Engineering Software. 2014. vol. 69. pp. 46–61. DOI: 10.1016/j.advengsoft.2013.12.007.
32. Li J., Zhang L. Analytical Model for Dynamic Spectrum Decision in Cognitive Radio Ad Hoc Networks: A Stochastic Framework. Future Intelligent Information Systems. 2011. pp. 325–332.
33. Keskin R., Aliskan I. MultiObjective Optimisation-based Robust H_{∞} Controller Design Approach for a Multi-Level DC-DC Voltage Regulator. Elektronika ir Elektrotechnika. 2023. vol. 29(1). pp. 4–14.
34. Yang X.-S. Cuckoo Search and Firefly Algorithm. Springer Science and Business Media LLC. 2014. vol. 516. 366 p. DOI: 10.1007/978-3-319-02141-6.
35. Emary E., Zawbaa H.M., Grosan C. Experienced Gray Wolf Optimization Through Reinforcement Learning and Neural Networks. IEEE Transactions on Neural Networks and Learning Systems. 2018. vol. 29(3). pp. 681–694. DOI: 10.1109/TNNLS.2016.2634548.
36. Zhang Q., Li H., Liu C., Hu W. A New Extreme Learning Machine Optimized by Firefly Algorithm. Sixth International Symposium on Computational Intelligence and Design. 2013. vol. 1. pp. 133–136. DOI: 10.1109/ISCID.2013.147.
37. Raj R.N., Nayak A., Kumar M.S. QoS-aware routing protocol for Cognitive Radio Ad Hoc Networks. Ad Hoc Networks. 2021. vol. 113. DOI: 10.1016/j.adhoc.2020.102386.

Hesham Mohammed Ali Abdullah — Ph.D., Assistant professor, College of Computer Science – Jouf University, KSA. Research interests: Cognitive Radio Ad Hoc Networks, network security, wireless network. The number of publications — 10. heshammohammedali@gmail.com; King Khalid Road, 8013, Sakakah, Saudi Arabia; office phone: + 966(014)654-4444.

A.V. Senthil Kumar — Ph.D., Professor, Director, Pg & research Department of computer applications, Hindustan College of Arts and science, Bharathiar University. Research interests: data mining, fuzzy expert systems, networks, software engineering, information systems (business informatics), artificial intelligence. The number of publications — 100. avsenthilkumar@yahoo.com; 2/23, Avanashi Road, Peelamedu, Uppilpalayam, 641028, Coimbatore, India; office phone: + 91(984)301-3009.

Ammar Abdullah Qasem Ahmed — Ph.D., Assistant professor, College of Computer Science – Jouf University, KSA. Research interests: big data, cloud computing, management information systems, networking. The number of publications — 5. ammar.aqahmed@gmail.com; King Khalid Road, 8013, Sakaka, Aljouf, Saudi Arabia; office phone: + 966(552)778-831.

Mohammed Abdullatef Saeed Mosleh — Ph.D., Assistant professor, AL-Fayha College, Jubail industrial city. Research interests: cloud computing, DataBase, networks. The number of publications — 7. ma.mosleh2010@gmail.com; Jubail St., Dumat al-Jandal, Saudi Arabia; office phone: + 966(535)866-629.

Х.М. АБДУЛЛА, А. КУМАР, А.А. КАСЕМ АХМЕД, М.А. САИД МОСЛЕХ
**ОПОРТУНИСТИЧЕСКАЯ МАРШРУТИЗАЦИЯ НА ОСНОВЕ
ГИБРИДНОЙ ОПТИМИЗАЦИИ С УЧЕТОМ СПЕКТРА ДЛЯ
САМООРГАНИЗУЮЩИХСЯ СЕТЕЙ КОГНИТИВНОЙ
РАДИОСВЯЗИ**

Абдулла Х.М., Кумар А., Касем Ахмед А.А., Саид Мослех М.А. **Опportunистическая маршрутизация на основе гибридной оптимизации с учетом спектра для самоорганизующихся сетей когнитивной радиосвязи.**

Аннотация. Опportunистическая маршрутизация повысила эффективность и надежность в самоорганизующихся сетях когнитивной радиосвязи (CRAHN). Многие исследователи разработали модели опportunистической маршрутизации, в том числе модель опportunистической маршрутизации на базе карты спектра (SMOR), которая считается более эффективной моделью в этой области. Однако в SMOR существуют определенные ограничения, которые требуют внимания и устранения. В данной статье рассматривается проблема задержки и ухудшения коэффициента доставки пакетов из-за неучета пропускной способности сети. Чтобы решить эти проблемы, в базовой модели маршрутизации SMOR используется гибридный алгоритм оптимизации, состоящий из алгоритмов оптимизации Firefly и Grey Wolf. Разработанная таким образом гибридная модель маршрутизации SMOR на основе оптимизации Firefly и Grey-Wolf (HFGWOSMOR) повышает производительность за счет высокой локальной и глобальной поисковой оптимизации. Первоначально анализируется взаимосвязь между задержкой и пропускной способностью, а затем устанавливается совместная многолучевая связь. Предлагаемая модель маршрутизации также вычисляет значения энергии принимаемых сигналов в пределах порога полосы пропускания и периода времени, и, следовательно, проблемы с производительностью, обнаруженные в SMOR, решаются. Чтобы оценить её эффективность, предложенная модель сравнивается со SMOR и другими существующими моделями опportunистической маршрутизации, которые показывают, что предлагаемая модель HFGWOSMOR работает лучше, чем другие модели.

Ключевые слова: самоорганизующиеся сети когнитивной радиосвязи, опportunистическая маршрутизация, опportunистическая маршрутизация на базе карты спектра, оптимизация Firefly, оптимизация Grey-Wolf, порог пропускной способности.

Литература

1. Akyildiz I.F., Lee W.Y., Chowdhury K.R. CRAHNs: Cognitive radio ad hoc networks. *AD hoc networks*. 2009. vol. 7(5). pp. 810–836.
2. Peha J.M. Approaches to spectrum sharing. *IEEE Communications magazine*. 2005. vol. 43(2). pp. 10–12.
3. Cesana M., Cuomo F., Ekici E. Routing in cognitive radio networks: Challenges and solutions. *Ad Hoc Networks*. 2011. vol. 9(3). pp. 228–248.
4. Chowdhury K.R. Communication protocols for wireless cognitive radio ad-hoc networks. Georgia Institute of Technology, 2009. 153 p.
5. Biswas S., Morris R. Opportunistic routing in multi-hop wireless networks. *ACM SIGCOMM Computer Communication Review*. 2004. vol. 34(1). pp. 69–74.
6. Biswas S., Morris R. ExOR: opportunistic multi-hop routing for wireless networks. *ACM SIGCOMM Computer Communication Review*. 2005. vol. 35(4). pp. 133–144.

7. Badarneh O.S., Salameh H.B. Opportunistic routing in cognitive radio networks: exploiting spectrum availability and rich channel diversity. IEEE Global Telecommunications Conference (GLOBECOM 2011). 2011. pp. 1–5. DOI: 10.1109/GLOCOM.2011.6134241.
8. Abdullah H.M.A., Kumar A.S. A Survey on Spectrum-Map Based on Normal Opportunistic Routing Methods for Cognitive Radio Ad Hoc Networks. International Journal of Advanced Networking and Applications. 2015. vol. 7(3), pp. 2761–2770.
9. Lin S.C., Chen K.C. Spectrum-map-empowered opportunistic routing for cognitive radio ad hoc networks. IEEE Transactions on Vehicular Technology. 2014. vol. 63(6). pp. 2848–2861.
10. Lin S.-C., Chen K.-C. Spectrum-Map-Empowered Opportunistic Routing for Cognitive Radio Ad Hoc Networks, IEEE Transactions on Vehicular Technology. 2014. vol. 63(6). pp. 2848–2861.
11. Abdullah H.M.A., Kumar A.S. A Hybrid Artificial Bee Colony Based Spectrum Opportunistic Routing Algorithm for Cognitive Radio Ad Hoc Networks. International Journal of Scientific and Engineering Research. 2016. vol. 7(6). pp. 294–303.
12. Abdullah H.M.A., Kumar A.S. HB-SOR: Hybrid Bat Spectrum Map Empowered Opportunistic Routing and Energy Reduction for Cognitive Radio Ad Hoc Networks (CRAHNs). International Journal of Scientific and Research Publications (JSRP). 2017. vol. 7(5). pp. 284–297.
13. Abdullah H.M.A., Kumar A.V.S. HFSA-SORA: Hybrid firefly simulated annealing based spectrum opportunistic routing algorithm for Cognitive Radio Ad hoc Networks (CRAHN). Conference on Intelligent Computing and Control (I2C2). 2017. DOI: 10.1109/I2C2.2017.8321943.
14. Abdullah H.M.A., Kumar A.V.S. Modified SMOR Using Sparsity Aware Distributed Spectrum Map for Enhanced Opportunistic Routing in Cognitive Radio Adhoc Networks. Journal of Advanced Research in Dynamical and Control Systems. 2017. vol. 9(6). pp. 184–196.
15. Abdullah H.M.A., Kumar, A.V.S. Vertex Search based Energy-efficient Optimal Resource Allocation in Cognitive Radio ad hoc Networks. SPIIRAS Proceedings. 2018. vol. 2(57). pp. 5-25. DOI: 10.15622/sp.57.1.
16. Abdullah H.M.A., Kumar A.V.S. Proficient Opportunistic Routing by Queuing Based Optimal Channel Selection for the Primary Users in CRAHN. ARPN Journal of Engineering and Applied Sciences. 2018. vol. 13(5). pp. 1649–1657.
17. He S.M., Zhang D.F., Xie K., Qiao H., Zhang J. Channel aware opportunistic routing in multi-radio multi-channel wireless mesh networks. Journal of Computer Science and Technology. 2014. vol. 29(3). pp. 487–501.
18. Barve S.S., Kulkarni P. Multi-agent reinforcement learning based opportunistic routing and channel assignment for mobile cognitive radio ad hoc network. Mobile Networks and Applications. 2014. vol. 19(6). pp. 720–730.
19. Lee H.W., Modiano E., Le L.B. Distributed throughput maximization in wireless networks via random power allocation. IEEE transactions on mobile computing. 2012. vol. 11(4). pp. 577–590.
20. Li S., Zheng Z., Ekici E., Shroff N. Maximizing system throughput by cooperative sensing in cognitive radio networks. IEEE/ACM Transactions on Networking (TON). 2014. vol. 22(4). pp. 1245–1256.
21. El-Sherif A.A., Mohamed A. Decentralized throughput maximization in cognitive radio wireless mesh networks. IEEE Transactions on Mobile Computing. 2014. vol. 13(9). pp. 1967–1980.

22. Ju H., Zhang R. Throughput maximization in wireless powered communication networks. *IEEE Transactions on Wireless Communications*. 2014. vol. 13(1). pp. 418–428.
23. Ping S., Aijaz A., Holland O., Aghvami A.H. Energy and interference aware cooperative routing in cognitive radio ad-hoc networks. *IEEE Wireless Communications and Networking Conference (WCNC)*. 2014. pp. 87–92. DOI: 10.1109/WCNC.2014.6951927.
24. Meghanathan N., Fanuel M. A minimum channel switch routing protocol for cognitive radio ad hoc networks. *12th International Conference on Information Technology-New Generations (ITNG)*. 2015. pp. 280–285.
25. Mirjalili S., Saremi S., Mirjalili S.M., Coelho L.D.S. Multi-objective grey wolf optimizer: a novel algorithm for multi-criterion optimization. *Expert Systems with Applications*. 2016. vol. 47. pp. 106–119.
26. Poornimha J., Kumar A.V.S., Abdullah H.M.A. A New Approach to Improve Energy Consumption Time and Life Time using Energy Based Routing in WSN. *Emerging Trends in Industry 4.0 (ETI 4.0)*. 2021. pp. 1–6. DOI: 10.1109/ETI4.051663.2021.9619412.
27. Yang Qin, Xiaoxiong Zhong, Yuanyuan Yang, Yanlin Li, Li Li. Joint channel assignment and opportunistic routing for maximizing throughput in cognitive radio networks. *IEEE*, 2014. DOI: 10.1109/GLOCOM.2014.7037532.
28. Dutta N., Sarma H.K.D., Polkowski Z. Cluster based routing in cognitive radio adhoc networks: Reconnoitering SINR and ETT impact on clustering. *Computer Communications*. 2018. vol. 115. pp. 10–20. DOI: 10.1016/j.comcom.2017.09.002.
29. Kumar A.V.S., Abdullah H.M.A., Hemashree P. An Efficient Geographical Opportunistic Routing Algorithm Using Diffusion and Sparse Approximation Models for Cognitive Radio Ad Hoc Networks. (Eds.: Smys S., Iliyasa A.M., Bestak R., Shi F.) *New Trends in Computational Vision and Bio-inspired Computing: Selected works presented at the ICCVBIC 2018*. 2020. pp. 323–333. DOI: 10.1007/978-3-030-41862-5_30.
30. Abdullah H.M.A., Kumar A.V.S. Selective Cooperative Jamming Based Relay Selection and Blowfish Encryption for Enhancing Channel and Data Security in CRAHN Routing (Ed.: Elkhodr M.). *Enabling Technologies and Architectures for Next-Generation Networking Capabilities*. 2019. pp. 105–124. DOI: 10.4018/978-1-5225-6023-4.ch005.
31. Mirjalili S., Mirjalili S.M., Lewis A. Grey Wolf Optimizer. *Advances in Engineering Software*. 2014. vol. 69. pp. 46–61. DOI: 10.1016/j.advengsoft.2013.12.007.
32. Li J., Zhang L. Analytical Model for Dynamic Spectrum Decision in Cognitive Radio Ad Hoc Networks: A Stochastic Framework. *Future Intelligent Information Systems*. 2011. pp. 325–332.
33. Keskin R., Aliskan I. MultiObjective Optimisation-based Robust H_{∞} Controller Design Approach for a Multi-Level DC-DC Voltage Regulator. *Elektronika ir Elektrotechnika*. 2023. vol. 29(1). pp. 4–14.
34. Yang X.-S. Cuckoo Search and Firefly Algorithm. Springer Science and Business Media LLC. 2014. vol. 516. 366 p. DOI: 10.1007/978-3-319-02141-6.
35. Emary E., Zawbaa H.M., Grosan C. Experienced Gray Wolf Optimization Through Reinforcement Learning and Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*. 2018. vol. 29(3). pp. 681–694. DOI: 10.1109/TNNLS.2016.2634548.
36. Zhang Q., Li H., Liu C., Hu W. A New Extreme Learning Machine Optimized by Firefly Algorithm. *Sixth International Symposium on Computational Intelligence and Design*. 2013. vol. 1. pp. 133–136. DOI: 10.1109/ISCID.2013.147.

37. Raj R.N., Nayak A., Kumar M.S. QoS-aware routing protocol for Cognitive Radio Ad Hoc Networks. Ad Hoc Networks. 2021. vol. 113. DOI: 10.1016/j.adhoc.2020.102386.

Абдулла Хишам Мохамед Али — Ph.D., доцент, Колледж компьютерных наук – Университет Аль-Джуф. Область научных интересов: специальные сети когнитивного радио, сетевая безопасность, беспроводная сеть. Число научных публикаций — 10. heshammohammedali@gmail.com; Дорога короля Халида, 8013, Сакака, Саудовская Аравия; р.т.: + 966(014)654-4444.

Кумар А.В. Сентхил — Ph.D., профессор, директор, научно-исследовательский отдел компьютерных приложений, Индостанский колледж искусств и науки, Бхаратийский университет. Область научных интересов: интеллектуальный анализ данных, нечеткие экспертные системы, сети, разработка программного обеспечения, информационные системы (бизнес-информатика), искусственный интеллект. Число научных публикаций — 100. avsenthilkumar@yahoo.com; Дорога Аванаша, Пиламеду, Упшилипалаям, 2/23, 641028, Коимбатур, Индия; р.т.: + 91(984)301-3009.

Касем Ахмед Аммар Абдулла — Ph.D., доцент, Колледж компьютерных наук – Университет Аль-Джуф. Область научных интересов: большие данные, облачные вычисления, информационные системы управления, сетевое взаимодействие. Число научных публикаций — 5. ammar.aqahmed@gmail.com; Дорога короля Халида, 8013, Сакака, Алжуф, Саудовская Аравия; р.т.: + 966(552)778-831.

Саид Мослех Мохаммед Абдуллаеф — Ph.D., доцент, Колледж Аль-Фалаах. Область научных интересов: облачные вычисления, базы данных, сети. Число научных публикаций — 7. ma.mosleh2010@gmail.com; улица Джубайль, Думат-эд-Джандай, Саудовская Аравия; р.т.: + 966(535)866-629.

В.А. ЗЕЛЕНЦОВ, И.Ю. ПИМАНОВ, С.А. ПОТРЯСАЕВ
**ИНТЕГРАЦИЯ РАЗНОРОДНЫХ ИНФОРМАЦИОННЫХ
РЕСУРСОВ И ДАННЫХ ДИСТАНЦИОННОГО
ЗОНДИРОВАНИЯ ЗЕМЛИ ПРИ МОНИТОРИНГЕ И
УПРАВЛЕНИИ РАЗВИТИЕМ ТЕРРИТОРИЙ**

Зеленцов В.А., Пиманов И.Ю., Потрясаев С.А. Интеграция разнородных информационных ресурсов и данных дистанционного зондирования Земли при мониторинге и управлении развитием территорий.

Аннотация. Статья посвящена разработке модельно-алгоритмического обеспечения и программных средств для автоматизации процессов интеграции данных дистанционного зондирования Земли (ДЗЗ) и других разнородных информационных ресурсов при решении задач мониторинга и проактивного управления развитием территорий. Отличительной особенностью постановки задачи является включение в состав интегрируемых ресурсов средств моделирования состояния природно-технических объектов, расположенных на анализируемой территории. Основу разработки составляет обоснование технологии интеграции разнородных информационных ресурсов, включающей алгоритм выбора типа архитектуры создаваемого комплекса средств автоматизации, способ описания информационного процесса интеграции данных и их совместной обработки, алгоритм определения наилучшей конфигурации информационных ресурсов при решении тематических задач, а также совокупность программно-технологических решений по интеграции данных ДЗЗ с другими необходимыми данными и их совместному использованию при моделировании. В результате исследований и применения разработанных алгоритмов установлено, что наиболее предпочтительным типом архитектуры систем интеграции разнородных информационных ресурсов является сервис-ориентированная архитектура. Для описания информационного процесса интеграции предложено использование нотации Business Process Model and Notation. Ключевым компонентом разработки в части программно-технологических решений по интеграции разнородных данных является предложенная схема взаимодействия с поставщиками и потребителями данных на основе создания слоя абстракции данных. Применение предложенного решения позволяет привести разнородные данные к единому универсальному формату для последующей совместной обработки на средствах моделирования. Проведенная апробация на конкретных тематических задачах мониторинга и управления развитием территорий показала реализуемость предложенной технологии интеграции и разработанных программно-технологических средств, а также достижение существенного выигрыша в оперативности решения тематических задач.

Ключевые слова: автоматизация, проактивное управление, информационные ресурсы, данные дистанционного зондирования Земли, интеграция данных, комплексное моделирование, сервис-ориентированная архитектура, многокритериальная оптимизация.

1. Введение. В настоящее время при решении задач мониторинга и управления развитием территорий все более широкое применение находят данные дистанционного зондирования Земли (ДЗЗ), получаемые с различных отечественных и зарубежных космических аппаратов (КА). Как в нашей стране, так и за рубежом

развиваются системы доступа к данным ДЗЗ, создаются новые тематические сервисы. Среди российских разработок можно выделить геопортал Роскосмоса [1], информационную систему дистанционного мониторинга Федерального агентства лесного хозяйства России [2], информационные ресурсы ГИС «Каскад» и ГИС «Космоплан» системы космического мониторинга чрезвычайных ситуаций МЧС России [3], геопортал «Вега» Института космических исследований РАН [4], геопортал «Цифровая Земля» [5], тематические порталы на базе ресурсов ряда коммерческих компаний [6 – 8], и другие.

Необходимо отметить, что в статье не рассматриваются такие специальные области интегрированного применения данных космического мониторинга, как моделирование состояния окружающей среды, климата, прогноза погоды, и ряд других, отличающихся глобальным характером наблюдаемых процессов и моделей их описания. Здесь рассматривается класс задач управления развитием территорий локального и регионального масштабов, требующих детального анализа и прогнозирования состояния расположенных на этих территориях различных природно-технических объектов (ПТО). Примерами ПТО являются системы гидроэнергетики, зоны разработки полезных ископаемых, лесные и сельскохозяйственные объекты, и др. К числу наиболее характерных примеров относятся региональные водные объекты и территории распространения наводнений с расположенной на них инфраструктурой.

Несмотря на постоянное совершенствование перечисленных выше систем доступа к данным ДЗЗ и создание большого количества тематических сервисов [9 – 11] для решения рассматриваемых задач, вопросы оперативного доведения пользователям требуемых данных и информации на их основе в значительной степени остаются открытыми. По-прежнему актуальными являются задачи упрощения взаимодействия пользователей-непрограммистов с создаваемыми системами и сокращения длительности обработки разнородных данных для систем мониторинга и управления развитием территорий, что напрямую диктует необходимость дальнейшего повышения уровня автоматизации обработки и подготовки данных и информации для решения задач управления.

К ключевым проблемам в рассматриваемой области управления развитием территорий также относятся ограниченные возможности существующих систем по интеграции данных, получаемых с разнотипных КА ДЗЗ, а также ограничения по совместному использованию данных ДЗЗ с другими информационными ресурсами, необходимыми для решения управленческих задач. Вместе с тем,

очевидно, что обработка данных ДЗЗ и поставка их потребителю не являются конечными целями в задачах управления. Главным является качественная информационная поддержка выработки потребителями своевременных и обоснованных решений по управлению развитием территорий, в том числе по действиям в чрезвычайных ситуациях, по планированию природопользованием, и т.д. Решение таких задач требует использования дополнительной информации об анализируемых объектах и территориях – например, результатов наземных измерений, кадастровых и других пространственных и непространственных данных.

Отметим, что понятие «информационный ресурс» рассматривается здесь в широком смысле как совокупность данных, организованных для эффективного получения достоверной информации [12] – в нашем случае, информации, необходимой для анализа, моделирования и прогнозирования развития территорий. К информационным ресурсам будем относить, в том числе, электронные базы данных, содержащие информацию о текущих и прогнозных значениях параметров среды и наблюдаемых объектов, интернет-ресурсы с необходимой для моделирования информацией, и т.д. В таком контексте термины интеграция данных и интеграция информационных ресурсов будут рассматриваться как синонимы.

Важнейшим компонентом информационного обеспечения управления являются также результаты моделирования развивающихся ситуаций и изменения состояния природно-технических объектов на анализируемых территориях. Эта составляющая информационного обеспечения особенно важна для осуществления *проактивного* управления [13, 14] развитием территорий, которое основано на моделировании и прогнозировании изменения состояния ПТО и в целом развивающихся ситуаций на анализируемых территориях.

В настоящее время для различных предметных областей достаточно активно разрабатываются приложения, предполагающие совместное использование данных ДЗЗ и дополнительных данных о территориях [15 – 17]. При этом разработки, нацеленные на совместное использование разнородных данных, включая данные ДЗЗ, и результатов моделирования ПТО, а также автоматизацию этих процессов, находятся лишь в начальной стадии. Как следствие, существующие системы, в основном, предназначены для анализа ситуаций «по фактам» уже случившихся событий и не направлены на их прогнозирование и предупреждение.

При включении средств моделирования в состав интегрируемых ресурсов важно учитывать тот факт, что для адекватного описания ПТО

и высокой достоверности прогнозов требуется применение, как правило, не одной, а семейства моделей или полимодельных комплексов (ПМК) [18, 19]. Каждая из моделей может являться наилучшей в определенном диапазоне внешних условий и/или диапазоне свойств ПТО. Причем эти условия и свойства ПТО могут динамически изменяться во времени. Перспективным направлением обеспечения высокого качества прогнозов на длительном интервале времени или в широком диапазоне свойств ПТО являются параметрическая и структурная адаптация моделей, а также переход от модели к модели (от ПМК к ПМК) в процессе функционирования соответствующих систем информационного обеспечения проактивного управления [19]. Это означает, что автоматизация совместного использования разнородных информационных ресурсов и результатов моделирования требует создания инструментов адаптации и выбора моделей описания ПТО на анализируемых территориях, а также перехода от одной модели к другой в процессе функционирования создаваемых программных комплексов обработки данных. Например, для прогнозирования зон затоплений при речных наводнениях могут использоваться результаты расчетов от нескольких типов гидрологических и гидродинамических моделей, моделей на основе искусственных нейронных сетей, и др. [20 – 23]. В зависимости от внешних факторов и рассматриваемого участка русла реки состав моделей, привлекаемых для расчетов, должен меняться для обеспечения максимальной достоверности прогнозов. Причем сами модели, как правило, принадлежат различным разработчикам и реализованы в виде программных модулей на территориально распределенных вычислительных ресурсах.

Обоснование технологии комплексного применения необходимого модельно-алгоритмического обеспечения и программных средств для сквозной автоматизации процессов интеграции данных ДЗЗ, других разнородных информационных ресурсов и средств моделирования в подобных сложных задачах является целью исследования, результаты которого представлены в данной работе. Основной акцент сделан на формирование автоматически реализующихся процессов полного цикла – от получения необходимых данных ДЗЗ с различных КА, а также других видов данных, до предоставления пользователю конечного результата в виде тематического продукта¹ или сервиса. Ряд частных результатов

¹ Тематический продукт обработки данных дистанционного зондирования Земли из космоса: Продукт, созданный на основе дешифрирования данных дистанционного зондирования Земли из космоса, с применением тематической или аналитической обработки, в том числе с использованием пространственных данных и других видов информации, для решения задач потребителей [24].

по данному направлению исследований представлен в предыдущих публикациях авторов [25 – 27]. В данной статье выполнено обобщение и развитие этих результатов применительно к решению проблемы автоматизации процессов интеграции разнородных информационных ресурсов в целом.

2. Технология интеграции разнородных информационных ресурсов. Перспективным направлением решения рассматриваемой проблемы автоматизации процессов интеграции данных ДЗЗ, других разнородных данных и результатов моделирования является использование и развитие концепции комплексной автоматизации и интеллектуализации задач получения, обработки, анализа и практического использования разнородных данных, поступающих из различных источников [28]. При этом требуемый уровень автоматизации может быть обеспечен путем разработки соответствующей технологии интеграции разнородных информационных ресурсов (ИРИР), в том числе данных ДЗЗ, предполагающей автоматический выбор источников данных в ходе решения каждой конкретной тематической задачи мониторинга и(или) проактивного управления, извлечение требуемых исходных данных из различных источников, их интеграцию и совместную обработку в моделирующих комплексах, опубликование и визуализацию.

Технология ИРИР по своему содержанию относится к информационным технологиям и может быть определена как совокупность алгоритмов и программно-технологических средств, обеспечивающих автоматизацию процессов сбора, интеграции, совместной обработки, хранения разнородных данных о состоянии объектов и территорий, представления и распространения результатов обработки. Соответственно, технология ИРИР включает в свой состав комплекс средств автоматизации (т.е. определенный набор взаимодействующих программных модулей, скомпонованных на базе выбранной архитектуры), а также модели и алгоритмы их совместного функционирования. Это означает, что создание технологии ИРИР должно предусматривать: 1) разработку модельно-алгоритмического обеспечения интеграции разнородных информационных ресурсов; 2) определение состава необходимых программно-технологических средств для решения задач ИРИР при мониторинге и управлении развитием территорий; 3) разработку комплекса средств для автоматизации процессов интеграции и совместной обработки данных, полученных от различных КА ДЗЗ, а также данных из других источников.

Отметим, что в дальнейшем изложении будут различаться понятия комплекса средств автоматизации (КСА) и системы ИРИР в целом. В отличие от КСА, система интеграции включает также информационные ресурсы и средства моделирования, которые являются внешними системами по отношению к КСА и привлекаются в ходе решения конкретной тематической задачи в качестве источников и потребителей разнородных данных. В этом смысле КСА представляет собой универсальное ядро для систем ИРИР, различающихся составом привлекаемых внешних информационных ресурсов и средств моделирования.

С точки зрения разработки модельно-алгоритмического обеспечения могут быть выделены две основные особенности технологии ИРИР. Первая из них состоит в том, что, как уже отмечалось, для осуществления *проактивного* управления развитием территорий в состав интегрируемых ресурсов входят модули (сервисы), осуществляющие моделирование анализируемых ПТО с целью предсказания их состояния в будущие моменты времени. Другой принципиальной особенностью технологии ИРИР является обеспечение возможности использования для описания ПТО не одной, а нескольких моделей (полимодельных комплексов) с реализацией процедуры выбора наилучших моделей (полимодельных комплексов) для каждого этапа функционирования КСА и системы интеграции в целом. На практике процесс выбора моделей заключается в определении состава программных модулей прогнозирования состояния ПТО, включаемых в контур решения тематических задач.

При учете описанных особенностей обобщенная схема технологии интеграции разнородных информационных ресурсов может быть представлена в виде, приведенном на рисунке 1. На данном рисунке в левой части показаны интегрируемые информационные ресурсы, в центре цветом выделены технологические операции ИРИР. В овалах представлены компоненты алгоритмического и программно-технологического обеспечения операций ИРИР.

В соответствии с приведенной на рисунке схемой технология ИРИР включает решение двух типов задач интеграции:

1 – определение состава модулей и информационных ресурсов для решения каждой конкретной тематической задачи, в том числе модулей (сервисов) моделирования ПТО, из доступных альтернативных вариантов;

2 – интеграцию разнородных данных, поступающих из различных источников (информационных ресурсов), в том числе от

разных КА ДЗЗ, для их совместного использования при моделировании и решении тематических задач мониторинга и проактивного управления развитием территорий.

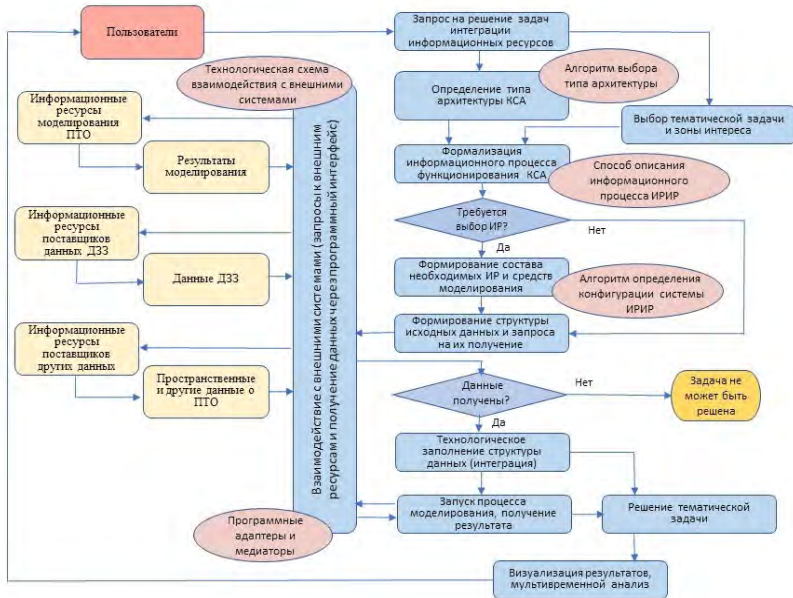


Рис. 1. Обобщённая схема технологии ИРИР (здесь ИР – информационные ресурсы)

В результате решения задачи 1 определяются тип архитектуры КСА, а также для каждой тематической задачи – состав средств моделирования, необходимых им исходных данных, и, соответственно, состав привлекаемых информационных ресурсов. То есть определяется конфигурация системы ИРИР в целом. Далее, на основе привлекаемых и разрабатываемых программно-технологических средств решаются задачи 2 типа по интеграции данных ДЗЗ и других данных для выбранной конфигурации. Совместно результаты решения задач 1 и 2 типов определяют структуру и порядок функционирования КСА и системы ИРИР в целом при решении заданных тематических задач.

3. Модельно-алгоритмическое обеспечение технологии ИРИР. Решение задач 1 типа производится на основе применения и (или) разработки совокупности моделей и алгоритмов, к которым относятся показанные на рисунке 1:

- алгоритм выбора архитектуры КСА;
- способ описания информационного процесса ИРИР для его визуального отображения и последующей алгоритмизации задач выбора состава интегрируемых информационных ресурсов и средств моделирования;
- алгоритм выбора состава средств моделирования и необходимых информационных ресурсов.

Задачи выбора, для решения которых предназначены алгоритмы из данного перечня, являются, как правило, дискретными и многокритериальными. Конкретная реализация алгоритмов, в общем случае, зависит от типа постановки задач управления развитием территорий. Подробно возможные методы решения задач многокритериального выбора в различных постановках рассмотрены, например, в [19]. Ниже в данной работе кратко описывается применение разработанных алгоритмов для достаточно широкого круга задач интеграции ИРИР в статической постановке.

3.1. Алгоритм выбора типа архитектуры КСА. Выбор типа архитектуры КСА при формировании технологии ИРИР, как показано в [27], может быть основан на оценивании различных применимых архитектур по совокупности показателей качества и последующем решении задачи многокритериального выбора. К наиболее распространенным типам архитектуры, в принципе пригодным для создания КСА процессов интеграции разнородных информационных ресурсов, относятся:

- монолитная архитектура (*mon*);
- модульная архитектура (*mod*);
- компонентная архитектура (*com*);
- клиент-серверная архитектура (*csa*);
- сервис-ориентированная архитектура (*soa*).

Основными показателями качества архитектур при проведении сравнения являются: показатель оперативности, показатель модульности, показатель допустимой гетерогенности, показатель многопользовательского режима, показатель масштабируемости. Множество сравниваемых архитектур оценивается набором значений вышеперечисленных показателей, каждый из которых представляет собой лингвистическую переменную.

Алгоритм выбора базируется на научно-методическом аппарате, предполагающем совместное использование методов экспертного опроса, вербального анализа решений и методов преобразования качественных показателей в количественные, теории нечетких

множеств, отношений и мер, а также теории планирования эксперимента. Он включает следующие основные этапы:

1. Конструирование лингвистических шкал для частных показателей, перевод их значений в интервал $[-1, +1]$.

2. Проведение экспертного опроса на основе построенного ортогонального плана для выявления влияния частных показателей на результирующий показатель качества архитектур.

3. Формирование и расчет значений результирующего показателя качества F .

В результате реализации процедуры многокритериального оценивания в соответствии с разработанным алгоритмом получены следующие значения результирующего показателя качества: $F_{mon}=0,711$; $F_{mod}=0,754$, $F_{com}=0,813$, $F_{csa}=0,891$, $F_{soa}=0,931$. Таким образом, наилучшее значение результирующего показателя качества достигается при использовании СОА, что позволяет сделать обоснованный выбор в пользу ее использования для реализации технологии ИРИР.

Сервис-ориентированная архитектура реализует модульный подход к разработке программного обеспечения, основанный на использовании распределённых, слабо связанных заменяемых компонентов, оснащённых стандартизированными интерфейсами для взаимодействия по стандартизированным протоколам [29, 30]. При использовании СОА для реализации технологии ИРИР каждый из информационных ресурсов, либо модулей прогнозирования состояния ПТО, участвующих в процессе обработки разнородных данных, может оставаться на своем месте локализации, без физического перемещения на рабочее место пользователя при решении конкретной тематической задачи. Тем самым решается проблема необходимости отчуждения универсального ПО, разработанного для моделирования ПТО, от разработчика, что значительно упрощает построение систем интеграции ИРИР.

3.2. Способ описания информационного процесса интеграции разнородных информационных ресурсов. Применение СОА определяет правила взаимодействия разнородных информационных ресурсов и средств моделирования. При этом вопросы организации самого информационного процесса обработки данных и определения состава необходимых средств обработки требуют отдельного рассмотрения.

Функционирование систем, разрабатываемых для интеграции разнородных информационных ресурсов, как следует из рисунка 1, заключается в реализации информационного процесса, состоящего из

упорядоченных операций получения, обработки, хранения и передачи данных и информации. Для автоматизации процессов интеграции разнородных информационных ресурсов необходима разработка способа формального описания реализуемого информационного процесса с включением в него всех функционально необходимых операций (действий) по обработке данных, и с описанием связей между ними. Наличие такого описания позволит в явном виде отобразить функциональную конфигурацию систем интеграции, перейти к количественной оценке различных конфигураций, выбору наилучших конфигураций, и к разработке программных средств автоматизации процессов интеграции.

При формальном описании должен быть отражен тот факт, что каждое из действий по обработке данных может в общем случае быть реализовано различными способами, с привлечением различных ресурсов – информационных, или осуществляющих моделирование ПТО. Построенная с учетом этого конфигурация системы интеграции является необходимым компонентом задачи выбора оптимального состава средств моделирования и информационных ресурсов, используемых для решения той или другой тематической задачи.

Среди потенциально применимых подходов для описания информационного процесса функционирования рассматриваемых систем интеграции могут быть выделены: сети Петри [31]; графоаналитические подходы [32]; стандарты IDEF (Integrated DEFinition) [33]; диаграммы UML [34]; стандарт eEPC (Extended Event Driven Process Chain) [35]; нотация BPMN (Business Process Model and Notation) [36].

Сравнительный анализ показал, что нотация BPMN, изначально предназначенная для описания диаграмм бизнес-процессов, предоставляет, по сравнению с другими подходами, расширенные возможности для формального описания взаимодействия компонентов сложных систем и модели распределенных процессов [37, 38]. Именно данные отличия BPMN в наибольшей степени соответствуют особенностям рассматриваемой технологии ИРИР. Кроме того, благодаря использованию сервис-ориентированной архитектуры и нотации BPMN становится возможным визуальное конструирование процессов функционирования КСА и системы интеграции в целом за счет применения языка исполнения бизнес-процессов (Business Process Execution Language, BPEL) [39, 40], который позволяет описать и организовать логику взаимодействия веб-сервисов, используя, в том числе, визуальный редактор.

Описание информационного процесса с использованием BPMN на примере тематической задачи прогнозирования наводнений приведено на рисунке 2. Часть из показанных на рисунке действий может быть реализована альтернативными способами, с привлечением различных средств моделирования. Например, на рисунке 2 такие операции выделены цветом. Для расчета расхода воды могут быть привлечены модули, выполняющие табличный расчет (А-1), либо статистический расчет (А-2), либо реализующие нейросетевые подходы (А-3). Примерами альтернативных подходов для расчета глубин в точках являются модели среза (Б-1), модели на основе уравнений Сен-Венана (Б-2), и нейросетевые модели (Б-3).

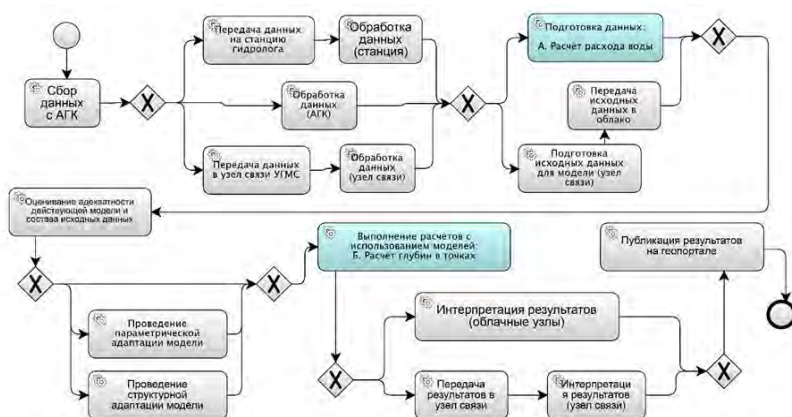


Рис. 2. Представление процесса решения задачи прогнозирования наводнений с использованием BPMN (АГК – автоматизированный гидрологический комплекс, УГМС – Управление гидрометеослужбы)

3.3. Алгоритм определения состава средств моделирования и информационных ресурсов. Определение конкретного состава средств моделирования и информационных ресурсов – источников необходимых данных, то есть выбор конфигурации системы ИРИР при решении тематических задач, относится к классу задач структурного синтеза. Она может быть решена на основе методов многокритериальной оптимизации.

Данная задача может быть формализовано описана следующим образом. Введем обозначения:

$A = \{A, i \in N\}$, $N = \{1, 2, \dots, n\}$ – множество наименований средств моделирования (СрМ) и (или) информационных ресурсов, каждое из которых может быть реализовано одним из альтернативных способов;

$B = \{B_j^i, j \in \{1, 2, \dots, m_i\}, i \in N\}$ – множество потенциально возможных вариантов реализации СрМ (B_j^i – j -й вариант реализации i -го СрМ);

$c_{ij}, i \in N, j \in \{1, 2, \dots, m_i\}$ – стоимость j -го варианта i -го СрМ;

$t_{ij}, i \in N, j \in \{1, 2, \dots, m_i\}$ – оперативность (длительность работы) j -го варианта реализации i -го СрМ;

$p_{ij}, i \in N, j \in \{1, 2, \dots, m_i\}$ – целевой показатель качества работы j -го варианта реализации i -го СрМ – обычно это показатели достоверности, точности моделирования (прогнозирования), или другие показатели, характеризующие степень удовлетворения пользователя качеством моделирования (прогнозирования состояния ПТО);

$X = \|x_{ij}\|$ – вариант конфигурации системы ИРИР, где $x_{ij} = 1$, если в качестве A_i используется СрМ B_j^i , и $x_{ij} = 0$, в противном случае. При этом $\sum_{j=1}^{m_i} x_{ij} = 1, \forall i \in N$, т.к. в проектируемой конфигурации в качестве СрМ A_i может использоваться какой-то один элемент B_j^i из имеющейся номенклатуры. Данные ограничения определяют область Δ допустимых вариантов конфигурации системы ИРИР.

В качестве целевых функций задачи, в соответствии с положениями теории эффективности сложных систем [41], целесообразно использовать три группы показателей: затраты на реализацию варианта конфигурации, оперативность, под которой понимается длительность решения требуемой тематической задачи, и целевой показатель, отражающий качество полученного решения – достоверность или точность результата.

С учетом введенных обозначений задача формирования конфигурации системы ИРИР сводится к многокритериальной задаче выбора на дискретном множестве допустимых альтернатив следующего вида:

$$\min_{X \in \Delta} C(X), \min_{X \in \Delta} T(X), \max_{X \in \Delta} P(X), \quad (1)$$

где $C(X)$ – стоимость использования средств моделирования и информационных ресурсов в конфигурации X ; $T(X)$ – длительность

решения тематической задачи при этой конфигурации; $P(X)$ – целевой показатель качества работы системы ИРИР.

Система ИРИР, в общем случае, может иметь сетевую структуру, подобную, показанной на рисунке 2. В этой сети можно сформировать множество кратчайших путей успешного функционирования системы ИРИР, для каждого пути $k = 1, 2, \dots, K$ обозначим через Π_k множество номеров СрМ, входящих в него. Тогда анализируемые показатели могут быть определены следующим образом:

– затраты на реализацию цикла расчетов для решения тематической задачи:

$$C(X) = \sum_{i=1}^n \sum_{j=1}^{m_i} c_{ij} \cdot x_{ij} \quad , \quad (2)$$

– длительность расчетов:

$$T(X) = \max_{k=1, \dots, K} \sum_{i \in \Pi_k} \sum_{j=1}^{m_i} t_{ij} \cdot x_{ij} \quad , \quad (3)$$

– целевой показатель:

$$P(X) = \min_{k=1, \dots, K} \min_{i \in \Pi_k} \sum_{j=1}^{m_i} p_{ij} \cdot x_{ij} \quad . \quad (4)$$

Содержательно постановка задачи выбора интерпретируется следующим образом: необходимо определить такой состав средств моделирования и информационных ресурсов для решения конкретной тематической задачи мониторинга и управления развитием территорий, при котором обеспечивается минимизация затрат на функционирование выбранной конфигурации, минимизация длительности решения тематической задачи и максимизация целевого показателя качества.

Алгоритм решения подобных задач, как правило, включает определение множества недоминируемых (Парето-оптимальных) вариантов конфигурации и нахождение на этом множестве единственного решения, выбранному критерию. Для выбора единственного решения, в частности, апробированным подходом является нахождение компромиссного варианта, при котором

достигается минимизация максимального из взвешенных относительных отклонений от оптимумов по частным показателям [42]:

$$X^k = \arg \min_{X \in \Delta} \max_{l \in L} \rho_l w_l(X), \quad (5)$$

где: L – множество частных показателей качества; ρ_l – весовые коэффициенты их относительной важности, $\rho_l > 0, \sum_{l \in L} \rho_l = 1$; $w_l(X)$ – относительные отклонения значений показателей от оптимумов, $w_l(X) = (f_l(\bullet) - f_l^0) / (f_l^* - f_l^0)$, где f_l^0 и f_l^* – соответственно наилучшее и наихудшее значения l -го показателя. Применительно к рассматриваемой задаче (1) множество L состоит из трех показателей, и частные показатели $f_l, l \in L$, имеют следующую интерпретацию: показатель f_1 представляет собой $C(X)$, f_2 – это $T(X)$, а f_3 – $P(X)$.

Например, для схемы, представленной на рисунке 2 и исходных данных по альтернативным вариантам ИР, представленных в таблице 1, возможно 9 вариантов конфигурации системы интеграции – таблица 2. В ячейках таблицы 2 случаю использованию модуля соответствует 1, не использованию – 0. Значения относительных взвешенных отклонений от оптимумов для возможных конфигураций приведены в таблице 3, причем считается, что весовые коэффициенты частных показателей принимают значения: 0,3 – для показателя точности, 0,3 – для показателя стоимости, 0,4 – для показателя оперативности реализации алгоритма.

Таблица 1. Значения частных показателей качества альтернативных вариантов использования средств моделирования

Средства моделирования	Значения показателей качества		
	Точность	Стоимость (усл. ед.)	Длительность (сек.)
Модуль А-1	0,8	0,05	60
Модуль А-2	0,6	0,15	10
Модуль А-3	0,9	0,34	5
Модуль Б-1	0,7	1,2	90
Модуль Б-2	0,75	2	70
Модуль Б-3	0,79	3,5	60

Таблица 2. Варианты конфигурации системы интеграции

Средства моделирования	Варианты конфигурации								
	1	2	3	4	5	6	7	8	9
Модуль А-1	0	0	1	0	0	1	0	0	1
Модуль А-2	0	1	0	0	1	0	0	1	0
Модуль А-3	1	0	0	1	0	0	1	0	0
Модуль Б-1	0	0	0	0	0	0	1	1	1
Модуль Б-2	0	0	0	1	1	1	0	0	0
Модуль Б-3	1	1	1	0	0	0	0	0	0

Таблица 3. Относительные взвешенные отклонения от оптимумов

Показатели	Значения относительных взвешенных показателей для вариантов конфигурации								
	1	2	3	4	5	6	7	8	9
Точность	0,000	0,300	0,000	0,063	0,300	0,063	0,142	0,300	0,142
Стоимость	0,300	0,278	0,266	0,126	0,104	0,093	0,034	0,012	0,000
Оперативность	0,000	0,024	0,259	0,047	0,071	0,306	0,141	0,165	0,400
Макс. отклонение	0,300	0,300	0,266	0,126	0,300	0,306	0,142	0,300	0,400

В соответствии с используемым критерием выбора, наилучшим вариантом является конфигурация 4 из таблицы 2, при котором для операции «Расчёт расхода воды» выбран сервис «Нейросетевой расчёт», а для операции «Расчёт глубин в точках» – сервис «Уравнения Сен-Венана».

Как видно из приведенного примера, в практических задачах число n элементов конфигурации системы ИРИР, допускающих альтернативные варианты реализации, весьма ограничено, а число потенциально возможных вариантов реализации каждого такого средства моделирования не является большим (m_i варьируется от 2-3 до 5-7). Поэтому размерность задачи выбора даже при использовании полного перебора вариантов (как в приведенном примере) не является высокой, и для современных средств вычислительной техники не является фактором, который может сколько-нибудь заметно влиять на оперативность решения тематических задач мониторинга и управления развитием территорий.

4. Программно-технологические решения по интеграции данных ДЗЗ и других информационных ресурсов. Выбор конфигурации информационных ресурсов и привлекаемых средств моделирования однозначно определяет состав данных, в том числе, данных ДЗЗ, которые необходимы для решения заданной тематической задачи. Совместное использование получаемых

разнородных данных обеспечивается применением существующих и созданием новых программно-технологических средств, т.е. решением задач интеграции 2-го типа, в соответствии с введенной выше классификацией.

Ключевым вопросом обеспечения интеграции разнородных данных является организация показанного на рисунке 1 взаимодействия с информационными ресурсами и сервисами, являющимися источниками и потребителями интегрируемых данных. Эти ресурсы и сервисы являются, в общем случае, внешними системами по отношению к разрабатываемому КСА.

Функционально взаимодействие с внешними системами включает ряд операций: 1) формирование структуры необходимых исходных данных, индивидуальных для каждой тематической задачи мониторинга и проактивного управления; 2) выполнение запросов к поставщикам данных в соответствии со сформированными структурами; 3) получение запрошенных данных; 4) заполнение структуры данных, их нормализация и группирование.

Схема взаимодействия с внешними системами, учитывающая разнородность интегрируемых ИР и различия в типах поставляемых данных, показана на рисунке 3. В верхних строках рисунка приведены интегрируемые разнородные данные и интерфейсы, используемые для организации взаимодействия с внешними системами. В качестве примера приведены информационные ресурсы, используемые в задачах прогнозирования речных наводнений [20 – 23]. С точки зрения общности предлагаемого решения принципиальным является рассмотрение 4-х типов источников и потребителей разнородных данных: 1) различные КА ДЗЗ (или различные типы съемочной аппаратуры); 2) ресурсы с пространственными данными; 3) ресурсы с непространственными данными (например, результаты измерений с датчиков и др.); 4) модули (сервисы) моделирования ПТО. При решении других тематических задач конкретный состав интегрируемых информационных ресурсов может различаться, однако предложенный набор типов таких источников, влияющий на способ интеграции данных, является достаточно общим.

Отличительная особенность предложенной схемы и основное решение, обеспечивающее выполнение интеграции разнородных данных, состоит в создании дополнительного слоя абстракции данных (выделен пунктиром). Создание этого слоя позволяет заменить сложные отношения М:М между поставщиками данных и их потребителями (сервисами моделирования) на простые соотношения М:1. То есть большое количество разнообразных связей каждой

модели с каждым поставщиком данных заменяется простой связью моделей со слоем абстракции данных, безотносительно к тому, каким образом эти данные были получены и в каком формате. При необходимости применяемые модели или ПМК снабжаются программными обёртками, которые могут работать со слоем абстракции данных и извлекать требуемую информацию.

При этом внешние информационные ресурсы – поставщики разнородных данных, должны удовлетворять ряду требований, основными из которых являются:

- Наличие у внешней системы программного интерфейса.
- Наличие разделения данных и их представления.
- Использование для передачи текстовых данных общепризнанных форматов XML или JSON.
- Наличие механизмов автоматической аутентификации.



Рис. 3. Порядок взаимодействия с внешними системами при запросе и получении данных

В случае если поставщики данных не удовлетворяют перечисленным требованиям, каждый частный формат получаемых исходных данных приводится к предложенному универсальному формату с помощью системы адаптеров и медиаторов, показанных на рисунке 1 в качестве необходимых компонентов реализации технологии ИРИР. Пример подобного адаптера, выполняющего функции каталогизатора и предварительной обработки данных ДЗЗ, приведен в [43].

Применение предложенного подхода в случае универсализации описания типов данных позволяет интегрировать данные из различных источников не только на физическом и логическом уровнях, но и на семантическом уровне. Это возможно, если

для всех источников подбираются типы данных из справочника Schema.org [44]. При их отсутствии в справочнике для создания нового типа возможно использование механизма расширения Schema.org. Пример семантического описания гидропоста в формате JSON-LD для задачи прогнозирования наводнений приведен в листинге 1.

5. Структура комплекса средств автоматизации. Как уже отмечалось, внедрение технологии ИРИР подразумевает разработку комплекса средств автоматизации, основанного на сервис-ориентированной архитектуре и реализующего рассмотренные выше алгоритмы и решения.

Общий состав КСА должен позволять осуществлять интеграцию разнородных данных, среди которых особая роль принадлежит данным ДЗЗ. Данные ДЗЗ, получаемые с различных КА, являются важнейшим компонентом информационного обеспечения проактивного управления развитием территорий. Особое значение они имеют для удаленных и обширных территорий, когда отсутствует возможность использования других средств наблюдения. Создание программно-технологических средств для их автоматического извлечения и подготовки к использованию совместно с другими типами данных в сервисах моделирования является на сегодняшний день малоисследованной задачей. Поэтому в рассматриваемых КСА должны быть предусмотрены соответствующие средства для подготовки данных ДЗЗ.

```
{
  "@context": "https://schema.org",
  "@type": "CivicStructure",
  "name": "Каликино",
  "geo": {
    "@type": "GeoCoordinates",
    "latitude": "46.16",
    "longitude": "60.44"
  },
  "waterLevel": {
    "@type": "Observation",
    "measuredValue": "885",
    "observationDate": "2020-07-25T11:32:31"
  }
}
```

Листинг 1. Пример семантического описания гидропоста «Каликино»

При реализации технологии ИРИР работа с данными ДЗЗ и их обработка совместно с другими данными включает несколько этапов:

1) формирование требований к составу и характеристикам данных ДЗЗ, необходимых для решения каждой конкретной задачи по мониторингу и управлению развитием территорий; формирование на этой основе структуры данных для выполнения запроса в соответствии со схемой на рисунке 3;

2) поиск, подбор, загрузка материалов ДЗЗ на ресурсах поставщиков данных ДЗЗ, подходящих под заданные условия (дата, тип, облачность, угол солнца, разрешение на местности и т.д.), размещение и каталогизацию данных ДЗЗ;

3) предобработка материалов космической съемки, извлечение метаданных, формирование файлов быстрого просмотра и тематических слоев с рассчитанными значениями индексов заданного типа;

4) интеграцию с другими разнородными данными для моделирования состояния объектов и территорий в интересах решения тематической задачи;

5) решение тематических задач на основе обработки данных ДЗЗ, других данных, и результатов моделирования изменения состояния ПТО;

6) публикацию и визуализацию результатов обработки данных ДЗЗ совместно с другими данными, включая возможность выполнения мультивременного анализа для исследования динамики происходящих изменений.

Соответствующий комплекс средств для реализации технологии ИРИР с использованием данных ДЗЗ включает следующие компоненты, показанные на рисунке 4:



Рис. 4. Структура КСА

- программный модуль экспертного формирования параметров исходных данных ДЗЗ – реализует этап 1;
- программный модуль каталогизации космических снимков – реализует операции этапа 2;
- программный модуль предобработки космических снимков – реализует этап 3;
- программный модуль мультитременного анализа данных – реализует ключевые функции этапа 6.

Задачи этапа 4 выполняются через модуль взаимодействия с внешними системами, этап 5 реализуется с помощью внешних по отношению к КСА ресурсов.

В состав минимально необходимых программно-технологических компонентов, требуемых для реализации технологии ИРИР, также входят сервисная шина и программные адаптеры. Кроме того, необходимыми компонентами при работе с пространственными данными являются геосервер, база данных и модуль визуализации геоданных. Для взаимодействия с потребителями создается пользовательский интерфейс.

Перечисленные модули функционируют согласованно под управлением сервисной шины, в соответствии сценариями интеграции информационных ресурсов. Сценарии определяются в результате использования описанных выше алгоритмов выбора состава привлекаемых информационных ресурсов и средств моделирования ПТО. Взаимодействие с внешними системами для получения исходных данных, используемых при моделировании и решении тематических задач, осуществляется через программный интерфейс в соответствии с описанной выше технологической схемой взаимодействия и алгоритмами функционирования программных адаптеров.

6. Экспериментальная проверка реализуемости технологии ИРИР. Тестирование реализуемости технологии ИРИР выполнено при решении ряда тематических задач мониторинга и проактивного управления развитием территорий, отвечающих следующему требованию: решаемые задачи должны показывать возможность интеграции разнородных данных для 3-х случаев:

- 1 – интеграция данных ДЗЗ от разных поставщиков и разных КА;
- 2 – интеграция данных ДЗЗ с другими типами данных (пространственными и непространственными);
- 3 – интеграция данных ДЗЗ с другими типами данных (пространственными и непространственными), а также с результатами

расчетов по моделям изменения состояния ПТО на анализируемой территории с одновременным выбором конкретного сервиса моделирования из числа нескольких имеющихся.

Исходя из этого, определен следующий перечень тематических задач для отработки технологии ИРИР:

1) Мультивременной визуальный анализ динамики изменения индекса биомассы (NDVI – Normalized Difference Vegetation Index) на заданной территории. Решение задачи демонстрирует возможности КСА по интеграции данных ДЗЗ от различных поставщиков.

2) Выявление и оценивание площади лесных гарей после пожаров на основе индекса ВАИ (Burn Area Index). Решение задачи демонстрирует возможности КСА по интеграции разнородных данных от различных поставщиков данных ДЗЗ, а также данных другого типа – картографических материалов о границах контуров лесных кварталов, и др.

3) Прогнозирование изменения границ природных объектов (например, болот, природно-климатических зон в тундре, и т.п.) Решение задачи демонстрирует возможности КСА для реализации полимодельного подхода к управлению развитием территорий на основе интеграции разнородных данных от различных поставщиков и результатов расчетов по прогнозным моделям.

Вид пользовательского интерфейса КСА при решении тематических задач приведен на рисунке 5.

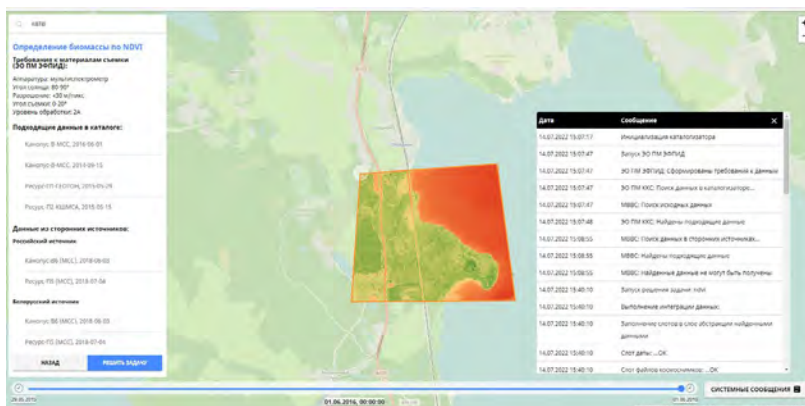


Рис. 5. Визуализация расчёта индекса NDVI

В левой части экрана отображаются: тип решаемой задачи; сформированные автоматически требования к данным ДЗЗ для этой

задачи; поставщики выбранных данных ДЗЗ; необходимые для решения задачи данные других типов; источник данных ДЗЗ для визуализируемых результатов.

В нижней части экрана отображается временная шкала, перемещая которую пользователь может просматривать результаты, относящиеся к разным моментам времени, и анализировать динамику изменений на территории.

В основном поле демонстрируются результаты решения, а также выводятся (по желанию пользователя) служебные сообщения о функционировании КСА и его отдельных модулей.

Для удобства пользования реализован целый ряд дополнительных функций: изменение прозрачности слоев на карте, изменение порядка слоев, показ квиклуквов (файлов быстрого просмотра снимков), и др.

На рисунках 6 и 7 приведены результаты решения задачи 1 для случая, когда необходимо проанализировать изменение индекса биомассы за 2015–2016 гг. В качестве источников данных, подходящих под сформированные требования, определены КА Ресурс П1-ГЕОТОН, снимок от 29.05.2015, и КА Канопус-В-МСС, снимок от 01.01.2016 (с меньшим, по сравнению со снимком КА Ресурс П1, пространственным разрешением).

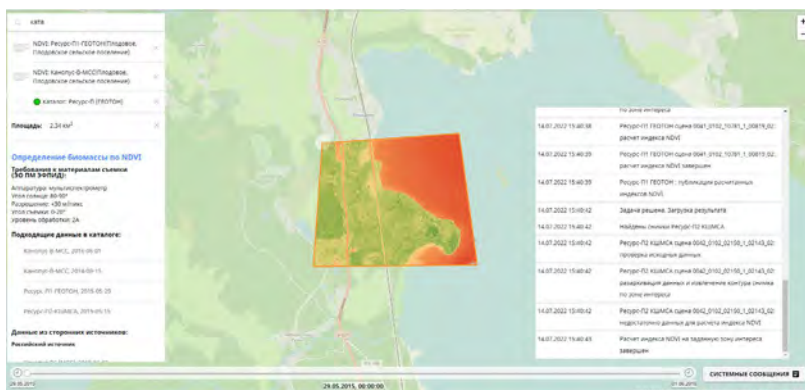


Рис. 6. Результат визуализации индекса NDVI по снимку КА Ресурс П1-ГЕОТОН

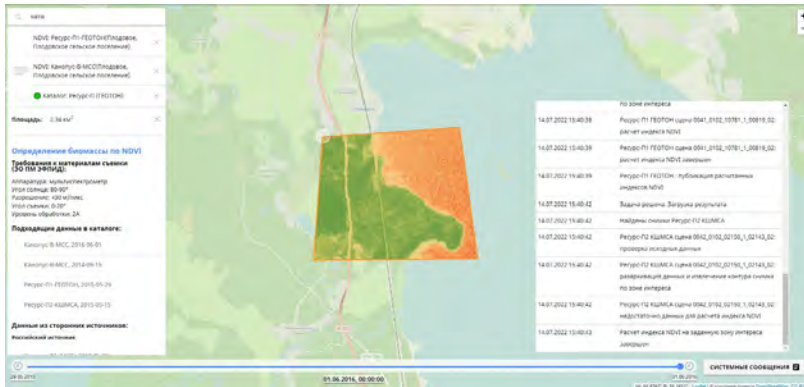


Рис. 7. Результат визуализации индекса NDVI по снимку КА Канопус-В-МСС

В ходе функционирования КСА в автоматическом режиме определены требования к данным ДЗЗ, найдены соответствующие им данные от разных поставщиков и различных КА, произведены расчеты индексов, обеспечена возможность мультитременного анализа результатов путем перемещения ползунка шкалы и изменения прозрачности слоев.

Результаты решения задачи 2 приведены на рисунке 8 для случая, когда необходимо выявить лесные гари на заданном участке территории за 2021 г.



Рис. 8. Визуализация участков с лесными гари

Здесь, как и в первом случае, в автоматическом режиме определены требования к данным ДЗЗ, найдены соответствующие им данные, а также выявлены участки с гарями, найдены и загружены данные по границам лесных кварталов, соответствующие этим участкам, рассчитана площадь гарей. В рассматриваемом примере отображаются четыре слоя (рисунок 8): «Лесные гари (Участок 1)», «Лесные гари (Участок 2)», «Лесной квартал № 94», «Лесной квартал № 107»).

В качестве задачи 3 рассмотрен достаточно простой пример прогнозирования изменения границ болот на заданном участке территории.

После выбора типа задачи и участка территории также в автоматическом режиме подбираются и загружаются необходимые данные ДЗЗ, выполняется их обработка для идентификации границ природного объекта. Отличительной особенностью задачи является привлечение сторонних моделей для прогнозирования изменения границ болота. Используются два типа моделей с различными значениями целевых показателей: первая из них позволяет проводить расчеты за минимальное время, но с меньшей точностью, при этом у второй модели лучшие показатели по точности, но хуже по оперативности. В зависимости от задаваемых итоговых предпочтений пользователя осуществляется расчет либо по первой, либо по второй модели. Тем самым осуществляется интегрированное использование как разнородных данных, так и результатов расчетов по моделям.

Результат решения задачи 3 представляется в виде, показанном на рисунке 9. Более темным цветом выделен контур природного объекта на дату съемки, определенный в результате обработки данных ДЗЗ, более светлым – результат прогноза. Применяя ползунок временной шкалы, можно осуществлять анализ развития ситуации в будущие моменты времени.

Необходимо отметить, что все полученные результаты не только визуализируются, но и доступны для передачи потенциальным пользователям через стандартные программные интерфейсы, в том числе описанные выше в данной статье.

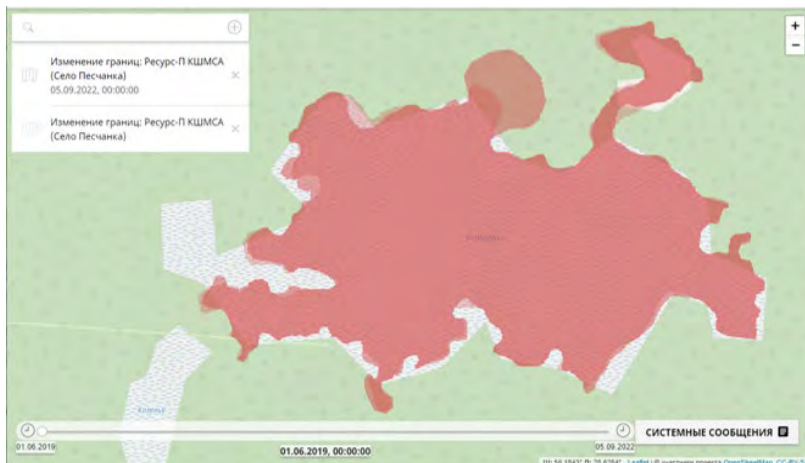


Рис. 9. Визуализация результатов расчета и прогнозирования изменения границ природных объектов

В целом, результаты отработки и апробации КСА показали работоспособность предложенных решений по реализации технологии ИРИР и возможность достижения высокой степени автоматизации при решении тематических задач мониторинга и проактивного управления развитием территорий. Достигнутая степень автоматизации работы с разнородными данными подтверждается результатами экспериментальной проверки выигрыша в оперативности решения тематических задач. В ходе эксперимента проведено сравнение длительностей выполнения операций работы с данными ДЗЗ, выполняемых специалистами по традиционной технологии, с длительностью тех же операций при использовании КСА. Средняя суммарная длительность выполнения операций работы с данными в случае использования предложенных решений уменьшается не менее, чем в 10 раз без снижения качества получаемых результатов.

7. Заключение. Рассмотренные модельно-алгоритмическое обеспечение и программно-технологические решения позволяют автоматизировать процессы интеграции разнородных информационных ресурсов, включающих, в первую очередь, данные ДЗЗ, а также средства моделирования изменения состояния ПТО, при мониторинге и проактивном управлении развитием территорий. Это достигается за счет решения задач выбора состава информационных ресурсов и сервисов моделирования, разработки новых способов и средств интеграции разнородных данных, и создания соответствующих комплексов программного обеспечения.

Представленные результаты отличаются, прежде всего, комплексным использованием ранее совместно не применявшихся алгоритмов формирования структуры систем интеграции и выбора моделей ПТО, а также программно-технологических средств для автоматического формирования массивов разнородных исходных данных, необходимых для выбранных моделей. Такой подход и его реализация в виде целостной технологии ИРИР создают методическую базу и программно-технологическую основу применения методологии и методов квалиметрии моделей и полимодельных комплексов в рассматриваемых задачах проактивного управления развитием территорий.

Проведенные апробация и эксперименты подтвердили реализуемость предложенного подхода к интеграции разнородных информационных ресурсов, возможность сквозной автоматизации всех этапов интеграции, и получение за счет этого значительного выигрыша в оперативности решения задач мониторинга и управления развитием территорий. При этом существенно снижаются требования к пользователям создаваемых средств автоматизации, а также создается принципиальная возможность реализации проактивности при управлении развитием территорий за счет включения моделей прогнозирования состояния ПТО в контур решения тематических задач.

Результаты отраженных в статье исследований показывают принципиальную возможность и перспективность разработок по созданию технологий автоматизации для применения методов квалиметрии моделей и полимодельных комплексов, требующих использования разнородных данных. Направления дальнейших исследований связаны с расширением состава показателей для оценивания и выбора конкретных моделей прогнозирования состояния ПТО, в соответствии с положениями квалиметрии моделей [19], а также с дальнейшей разработкой средств интеграции разнородных данных на семантическом уровне.

Литература

1. Геопортал Роскосмоса. URL: <https://gptl.ru/> (дата обращения: 06.02.2023).
2. Информационная система дистанционного мониторинга Федерального агентства лесного хозяйства России. URL: <http://www.nffc.aviales.ru/> (дата обращения: 06.02.2023).
3. Алексеев Я.В., Фахми Ш.С. Применения информационной системы космического мониторинга МЧС России для обеспечения эффективных управленческих решений по защите населения и территории от чрезвычайных ситуаций природного и техногенного характера // Пятнадцатая Всероссийская открытая конференция "Современные проблемы дистанционного зондирования

- Земли из космоса". URL: <http://conf.rse.geosmis.ru/thesisshow.aspx?page=144&thesis=6014> (дата обращения 06.02.2023).
4. ВЕГА-PRO. URL: <http://pro-vega.ru/> (дата обращения: 06.02.2023).
 5. Компания «Совзонд». Геоинформационные онлайн-сервисы. URL: <https://sovzond.ru/products/online-services/> (дата обращения: 06.02.2023).
 6. Тематические сервисы Scanex. URL: <https://www.scanex.ru/cloud/> (дата обращения: 06.02.2023).
 7. «ТерраТех»: геоинформационные решения на основе анализа данных дистанционного зондирования. URL: <https://terratech.ru/services/> (дата обращения: 06.02.2023).
 8. Цифровая Земля. URL: <https://dgearth.ru/> (дата обращения: 06.02.2023).
 9. Кучейко А.А. Мировой опыт обеспечения открытого доступа к данным ДЗЗ. Экономические и технологические аспекты // Дистанционное зондирование Земли из космоса в России. 2020. № 2. С. 50–64.
 10. Заичко В.А., Шведов Д.О., Кутумов А.А. О состоянии и развитии российской государственной космической системы дистанционного зондирования Земли // Дистанционное зондирование Земли из космоса в России. 2022. № 2. С. 6–17.
 11. Kurbanov E., Vorobiev O., Sha J., Li X., Gitas I., Minakou C., Gabdelkhakov A., Martynova M. A survey on the use of GIS and remote sensing for sustainable forestry and ecology in Russia and China // Current problems in remote sensing of the Earth from space. 2020. vol. 17. no 5. pp. 9–22.
 12. Воройский Ф.С. Информатика. Новый систематизированный толковый словарь-справочник: Введение в современные информационные и телекоммуникационные технологии в терминах и фактах. 3-е изд., перераб. и доп. // М.: Физматлит. 2003. 760 с.
 13. Охтилев М.Ю., Охтилев П.А., Соколов Б.В., Юсупов Р.М. Методологические и методические основы проактивного управления жизненным циклом сложных технических объектов // Изв. высш. учебн. заведений: Приборостроение. 2022. № 11. С. 781–788.
 14. Sokolov B.V., Potryasaev S.A., Yusupov R.M. Proactive Management of Information Processes in the Industrial Internet // Journal of Physics: Conference Series. 2021. vol. 1864(1). DOI: 10.1088/1742-6596/1864/1/012007.
 15. Lasaponara R., Aromando A., Cardettin G., Proto M. Fire Risk Estimation at Different Scales of Observations: An Overview of Satellite Based Methods // Computational Science and Its Applications (ICCSA 2018). Lecture Notes in Computer Science. Springer, Cham. 2019. vol. 10964. pp. 375–378.
 16. Zhuo L. Satellite Remote Sensing of Soil Moisture for Hydrological Applications: A Review of Issues to Be Solved // (Eds.: Scozzari A., Mounce S., Han D., Soldovieri F., Solomatine D.). ICT for Smart Water Systems: Measurements and Data Science. The Handbook of Environmental Chemistry. Springer, Cham. 2019. vol. 102. pp. 259–281.
 17. Previtali M. A GIS and Remote Sensing Approach for Desertification Sensitivity Assessment of Cultural Landscape in Apulia Region (Italy) // (Eds.: Ioannides M., Fink E., Cantoni L., Champion E.). Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. Lecture Notes in Computer Science. Springer, Cham. 2021. vol. 12642. pp. 138–149.
 18. Sokolov B.V., Potryasaev S.A., Zakharov V.V., Pavlov A.N. Methodology and Technologies of the Complex Objects Proactive Intellectual Situational Management and Control in Emergencies // Advances in Intelligent Systems and Computing. Proceedings of the Fourth International Scientific Conference “Intelligent Information Technologies for Industry” (ITI¹⁹). 2020. pp. 234–243.

19. Микони С.В., Соколов Б.В., Юсупов Р.М. Квалиметрия моделей и полимодельных комплексов // М.: РАН. 2018. 314 с.
20. Brown E., Bachmann D., Cranston M., et al. Methods and tools to support real time risk-based flood forecasting – a UK pilot application // FLOODrisk 2016 – 3rd European Conference on Flood Risk Management. E3S Web of Conferences. 2016. vol. 7(18019). 8 p.
21. Зеленцов В.А., Алябян А.М., Крыленко И.Н., Пиманов И.Ю., Пономаренко М.Р., Потрясаев С.А., Семёнов А.Е., Соболевский В.А., Соколов Б.В., Юсупов Р.М. Модельно-ориентированная система оперативного прогнозирования речных наводнений // Вестник Российской академии наук. 2019. Т. 89. № 8. С. 831–843.
22. Kornilova E.D., Krylenko I.N., Rets E.P., Motovilov Y.G., Bogachenko E.M., Krylenko I.V., Petrakov D.A. Modeling of Extreme Hydrological Events in the Baksan River Basin, the Central Caucasus, Russia // Hydrology. 2021. vol. 8(1). 24 p.
23. Chomba I.C., Banda K.E., Winsemius H.C., Chomba M.J., Mataa M., Ngwenya V., Sichingabula H.M., Nyambe I.A., Ellender B. A Review of Coupled Hydrologic-Hydraulic Models for Floodplain Assessments in Africa: Opportunities and Challenges for Floodplain Wetland Management // Hydrology. 2021. vol. 8(1). no. 44. 12 p.
24. ГОСТ Р 59082-2020. Данные дистанционного зондирования Земли из космоса. Продукты обработки данных дистанционного зондирования земли из космоса тематические. Типы задач, решаемых на основе тематических продуктов // М.: Госстандарт России. 2021. 16 с.
25. Зеленцов В.А., Потрясаев С.А. Архитектура и примеры реализации информационной платформы для создания и предоставления тематических сервисов с использованием данных дистанционного зондирования Земли // Труды СПИИРАН. 2017. Т. 6. № 55. С. 86–113.
26. Пиманов И.Ю. Автоматизация выбора функциональной структуры системы комплексного моделирования чрезвычайных ситуаций // Информатизация и связь. 2021. № 2. С. 15–21.
27. Зеленцов В.А., Потрясаев С.А., Пиманов И.Ю. Выбор архитектуры систем интеграции разнородных информационных ресурсов при комплексном моделировании природно-технических объектов // Информатизация и связь. 2021. № 7. С. 72–77.
28. Ахметов Р.Н., Васильев И.Е., Капитонов В.А., Охтилев М.Ю., Соколов Б.В. Концепция создания и применения перспективной АСУ подготовкой и пуском ракеты космического назначения «Союз-2»: новые подходы к интеграции, интеллектуализации, управлению // Авиакосмическое приборостроение. 2015. № 4. С. 3–54.
29. Wojkić M., Pržulj Đ., Stefanović M. Possible Application of Service-Oriented Architecture in Domain of Land Administration System // Proceedings on 18th International Conference on Industrial Systems (IS'20). 2022. pp. 496–502.
30. Hustad E., Olsen D.H. Creating a sustainable digital infrastructure: The role of service-oriented architecture // Procedia Computer Science. 2021. vol. 181. pp 597–604.
31. Мараховский В.Б., Розенблюм Л.Я., Яковлев А.В. Моделирование параллельных процессов. Сети Петри // СПб.: Профессиональная литература. 2014. 400 с.
32. Мельников Г.П. Системология и языковые аспекты кибернетики // М.: Советское радио. 1978. 368 с.
33. IDEF. URL: <https://www.idef.com/> (дата обращения: 06.02.2023).
34. UML. URL: <https://www.uml.org/> (дата обращения: 06.02.2023).
35. Scheer A.W. The Development Lines of Process Automation // (Eds.: Bergener K., Räckers M., Stein A.). The Art of Structuring. Springer, Cham. 2019. pp. 213–220.

36. BPMN. URL: <https://www.bpmn.org/> (дата обращения: 06.02.2023).
37. Троицкий Д.В., Городецкий В.И. Сценарная модель знаний и язык описания процессов для оценки и прогнозирования ситуаций // Труды СПИИРАН. 2009. № 8. С. 94–127.
38. Flávio E.A. Horita, Porto de Albuquerque J., Marchezini V., Mendiondo E.M. Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil // Decision Support Systems. 2017. vol. 97. pp. 12–22.
39. Vasiliev Y. SOA and WS-BPEL: Composing Service-Oriented Solution with PHP and ActiveBPEL // Packt Publishing. 2007. 316 p.
40. Юдицкий С.А. Сценарный подход к моделированию поведения бизнес-систем // М.: СИНТЕГ. 2001. 112 с.
41. Петухов Г.Б. Основы теории эффективности целенаправленных процессов. Ч1. Методология, методы, модели // М.: Изд-во МО СССР. 1989. 635 с.
42. Михалевиц В.С., Волкович В.Л. Вычислительные методы исследования и проектирования сложных систем // М.: Наука. 1982. 286 с.
43. Зеленцов В.А., Потрясаев С.А., Пиманов И.Ю., Семенов А.Е. Программное обеспечение «Каталог-В» для автоматической каталогизации космических снимков. Свидетельство № 2017612870. Зарегистрировано в реестре программ для ЭВМ 03.03.2017.
44. CivicStructure. URL: <https://schema.org/CivicStructure> (дата обращения: 06.02.2023).

Зеленцов Вячеслав Алексеевич — д-р техн. наук, профессор, главный научный сотрудник, лаборатория информационных технологий в системном анализе и моделировании, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: теоретические основы обеспечения надежности и управления эксплуатацией сложных систем, методы и информационные технологии поддержки принятия решений на основе интегрированной обработки наземно-аэрокосмических данных. Число научных публикаций — 350. v.a.zelentsov@gmail.com; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-0103.

Пиманов Илья Юрьевич — канд. техн. наук, младший научный сотрудник, лаборатория информационных технологий в системном анализе и моделировании, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: системы поддержки принятия решений, геоинформационные системы и сервисы. Число научных публикаций — 65. pimeno@list.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-0103.

Потрясаев Семен Алексеевич — д-р техн. наук, ведущий научный сотрудник, лаборатория информационных технологий в системном анализе и моделировании, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН). Область научных интересов: системный анализ и исследование операций, теория управления структурной динамикой сложных организационно-технических систем. Число научных публикаций — 120. semp@mail.ru; 14-я линия В.О., 39, 199178, Санкт-Петербург, Россия; р.т.: +7(812)328-0103.

Поддержка исследований. Исследования по разделам 1, 2, 3.2, 3.3, 4-6 выполнены в рамках бюджетной темы FFZF-2022-0004. Исследования по разделу 3.1 выполнены за счет гранта Российского научного фонда (проект № 22-19-00767, <https://rscf.ru/project/22-19-00767>).

[19-00767](#)). Примеры расчетов выполнены с использованием ресурсов Центра коллективного пользования «Северо-Западный центр мониторинга и прогнозирования развития территорий», <https://ckp-rf.ru/catalog/ckp/2079622/>. Авторы выражают благодарность В.Ф. Мочалову за помощь в обработке данных дистанционного зондирования Земли при тестировании программного обеспечения.

V. ZELENTSOV, I. PIMANOV, S. POTRYASAYEV
**INTEGRATION OF HETEROGENEOUS INFORMATION
RESOURCES AND EARTH REMOTE SENSING DATA IN
MONITORING AND MANAGEMENT OF TERRITORIAL
DEVELOPMENT**

Zelentsov V., Pimanov I., Potryasayev S. **Integration of Heterogeneous Information Resources and Earth Remote Sensing Data in Monitoring and Management of Territorial Development.**

Abstract. The article is devoted to the development of model-algorithmic support and software tools for automating the integration of Earth remote sensing data and other heterogeneous information resources in solving problems of monitoring and proactive management of territories development. A distinctive feature of the problem statement is the inclusion of tools for modeling the state of natural and technical objects located in the analyzed territory into the resources should be integrated. The development is based on the justification of the technology for integrating heterogeneous information resources, which includes an algorithm for choosing the type of architecture for the created automation tool complex, a method for describing the information process of integrating data and their joint processing, an algorithm for determining the best configuration of information resources when solving thematic problems, as well as a set of software and technological solutions for integration of remote sensing data with other necessary data and their joint use in modeling. As a result of research and developed algorithms application, it has been established that the most preferred type of systems' architecture for integrating heterogeneous information resources is a service-oriented architecture. To describe the information integration process, it is proposed to use a Business Process Model and Notation. The key component of the development in terms of software and technological solutions for the integration of heterogeneous data is the proposed interaction scheme with data providers and consumers based on data abstraction layer creation. The application of the proposed solution allows you to bring heterogeneous data to a single format suitable for further processing on modeling tools. The testing carried out on specific thematic tasks of monitoring and managing the territories' development showed the feasibility of the proposed integration technology and the developed software tools, as well as the achievement of a significant gain in the rapidness of solving thematic tasks.

Keywords: automation, proactive management, information resources, Earth remote sensing data, data integration, complex modeling, service-oriented architecture, multi-objective optimization.

References

1. Geoportal Roskosmosa [Geoportal of Roscosmos]. Available at: <https://gptl.ru/>. (accessed 06.02.2023). (In Russ.).
2. Informatsionnaya sistema dstantsionnogo monitoringa Federal'nogo agentstva lesnogo khozyaystva Rossii [Information system for remote monitoring of the Federal Forestry Agency of Russia]. Available at: <http://www.nffc.aviales.ru/>. (accessed 06.02.2023). (In Russ.).
3. Alekseyenko Ya.V., Fakhmi Sh.S. [Application of the space monitoring information system of the Ministry of Emergency Situations of Russia to ensure effective management decisions to protect the population and the territory from natural and man-made emergencies]. Pyatnadsataya Vserossiyskaya otkrytaya konferentsiya

- "Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa" [Fifteenth All-Russian open conference "Modern problems of the Earth remote sensing"]. Available at: <http://conf.rse.geosmis.ru/thesisshow.aspx?page=144&thesis=6014>. (accessed 06.02.2023). (In Russ.).
4. VEGA-PRO [VEGA-PRO]. Available at: <http://pro-vega.ru/>. (accessed 06.02.2023). (In Russ.).
 5. Kompaniya «Sovzond». Geoinformatsionnyye onlayn-servisy. [Sovzond Company. Geoinformation online services]. Available at: <https://sovzond.ru/products/online-services/>. (accessed 06.02.2023). (In Russ.).
 6. Tematicheskiye servisy Scanex [Scanex thematic services]. Available at: <https://www.scanex.ru/cloud/>. (accessed 06.02.2023). (In Russ.).
 7. «TerraTekh»: geoinformatsionnyye resheniya na osnove analiza dannykh distantsionnogo zondirovaniya ["TerraTech": geoinformation solutions based on the analysis of remote sensing data]. Available at: <https://terratech.ru/services/>. (accessed 06.02.2023). (In Russ.).
 8. Tsifrovaya Zemlya [Digital Earth]. Available at: <https://dgeth.ru/>. (accessed 06.02.2023). (In Russ.).
 9. Kucheiko A.A. [World experience in providing open access to remote sensing data. Economic and technological aspects]. Distsionnoye zondirovaniye Zemli iz kosmosa v Rossii. [Remote sensing of the Earth from space in Russia]. 2020. no. 2. pp. 50–64. (In Russ.).
 10. Zaichko V.A., Shvedov D.O., Kutumov A.A. [On the state and development of the Russian state space system for the Earth remote sensing]. Distsionnoye zondirovaniye Zemli iz kosmosa v Rossii. [Remote sensing of the Earth from space in Russia]. 2022. no. 2. pp. 6–17. (In Russ.).
 11. Kurbanov E., Vorobiev O., Sha J., Li X., Gitas I., Minakou C., Gabdelkhakov A., Martynova M. A survey on the use of GIS and remote sensing for sustainable forestry and ecology in Russia and China. Current problems in remote sensing of the Earth from space. 2020. vol. 17. no. 5. pp. 9–22.
 12. Voroyksiy F.S. Informatika. Novyy sistematizirovannyi tolkovyy slovar'-spravochnik: Vvedeniye v sovremennyye informatsionnyye i telekommunikatsionnyye tekhnologii v terminakh i faktakh. 3-ye izd., pererab. i dop. [Computer science. New systematized explanatory dictionary-reference book: Introduction to modern information and telecommunication technologies in terms and facts. 3rd ed., revised, and additional] Moscow: Fizmatlit, 2003. 760 p. (In Russ.).
 13. Okhtilev M.Yu., Okhtilev P.A., Sokolov B.V., Yusupov R.M. [Methodological and methodological foundations of proactive life cycle management of complex technical objects]. Izv. vyssh. uchebn. zavedeniy: Priborostroyeniye – Proceedings of the higher educational institutions: Instrumentation. 2022. no. 11. pp. 781–788. (In Russ.).
 14. Sokolov B.V., Potryasaev S.A., Yusupov R.M. Proactive Management of Information Processes in the Industrial Internet. Journal of Physics: Conference Series. 2021. vol. 1864(1). DOI: 10.1088/1742-6596/1864/1/012007.
 15. Lasaponara R., Aromando A., Cardetin, G., Proto M. Fire Risk Estimation at Different Scales of Observations: An Overview of Satellite Based Methods. Computational Science and Its Applications (ICCSA 2018). Lecture Notes in Computer Science. Springer, Cham. 2019. vol. 10964. pp. 375–378.
 16. Zhuo L. Satellite Remote Sensing of Soil Moisture for Hydrological Applications: A Review of Issues to Be Solved. (Eds.: Scozzari A., Mounce S., Han D., Soldovieri F., Solomatine D.). ICT for Smart Water Systems: Measurements and Data Science. The Handbook of Environmental Chemistry. Springer, Cham. 2019. vol. 102. pp. 259–281.

17. Previtali M. A GIS and Remote Sensing Approach for Desertification Sensitivity Assessment of Cultural Landscape in Apulia Region (Italy). (Eds.: Ioannides M., Fink E., Cantoni L., Champion E.). Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. Lecture Notes in Computer Science. Springer, Cham. 2021. vol. 12642. pp. 138–149.
18. Sokolov B.V., Potryasaev S.A., Zakharov V.V., Pavlov A.N. Methodology and Technologies of the Complex Objects Proactive Intellectual Situational Management and Control in Emergencies. Advances in Intelligent Systems and Computing. Proceedings of the Fourth International Scientific Conference “Intelligent Information Technologies for Industry” (ITI’19). 2020. pp. 234–243.
19. Mikoni S.V., Sokolov B.V., Yusupov R.M. Kvalimetriya modeley i polimodel'nykh kompleksov. [Qualimetry of models and polymodel complexes]. Moscow: RAS, 2018. 314 p. (In Russ.).
20. Brown E., Bachmann D., Cranston M., et al. Methods and tools to support real time risk-based flood forecasting – a UK pilot application. FLOODrisk 2016 – 3rd European Conference on Flood Risk Management. E3S Web of Conferences. 2016. vol. 7(18019). 8 p.
21. Zelentsov V.A., Alabyan A.M., Krylenko I.N. et al. A model-oriented system for operational forecasting of river floods. Herald of the Russian Academy of Sciences. 2019. vol. 89(4). pp. 405–417.
22. Kornilova E.D., Krylenko I.N., Rets E.P., Motovilov Y.G., Bogachenko E.M., Krylenko I.V., Petrakov D.A. Modeling of Extreme Hydrological Events in the Baksan River Basin, the Central Caucasus, Russia. Hydrology. 2021. vol. 8(1). 24 p.
23. Chomba I.C., Banda K.E., Winsemius H.C., Chomba M.J., Mataa M., Ngwenya V., Sichingabula H.M., Nyambe I.A., Ellender B. A Review of Coupled Hydrologic-Hydraulic Models for Floodplain Assessments in Africa: Opportunities and Challenges for Floodplain Wetland Management. Hydrology. 2021. vol. 8(1). no. 44. 12 p.
24. GOST R 59082-2020. [Earth remote sensing data from space. Earth remote sensing data processing products themed. Types of tasks solved on the basis of thematic products]. M.: Gosstandart Rossii, 2021. 16 p. (In Russ.).
25. Zelentsov V.A., Potryasaev S.A. [Architecture and examples of implementation of the informational platform for creation and provision of thematic services with use of Earth remote sensing data]. Trudy SPIIRAN – SPIIRAS Proceedings. 2017. vol. 6. no. 55. pp. 86–113. (In Russ.).
26. Pimanov I.Yu. [Automation of selecting the functional structure of a system for complex emergency simulation]. Informatizatsiya i svyaz' – Informatization and communication. 2021. no. 2. pp. 15–21. (In Russ.).
27. Zelentsov V.A., Potryasaev S.A., Pimanov I.Yu. [Choice of architecture for heterogeneous information resources integration systems in natural-technical objects complex modeling]. Informatizatsiya i svyaz' – Informatization and communication. 2021. no. 7. pp. 72–77. (In Russ.).
28. Akhmetov R.N., Vasil'yev I.Ye., Kapitonov V.A., Okhtilev M.Yu., Sokolov B.V. [The concept of creating and using a promising automated control system for the preparation and launch of the Soyuz-2 space rocket: new approaches to integration, intellectualization, and control]. Aviakosmicheskoye priborostroyeniye – Aerospace Instrumentation. 2015. no. 4. pp. 3–54. (In Russ.).
29. Bojčić M., Pržulj Đ., Stefanović M. Possible Application of Service-Oriented Architecture in Domain of Land Administration System. Proceedings on 18th International Conference on Industrial Systems (IS'20). 2022. pp. 496–502.

30. Hustad E., Olsen D.H. Creating a sustainable digital infrastructure: The role of service-oriented architecture. *Procedia Computer Science*. 2021. vol. 181. pp 597–604.
31. Marakhovskiy V.B., Rozenblyum L.Ya., Yakovlev A.V. Modelirovaniye parallelnykh protsessov. Seti Petri. [Simulation of parallel processes. Petri nets]. SPb.: Professional'naya literature, 2014. 400 p. (In Russ.).
32. Melnikov G.P. Sistemologiya i yazykovyye aspekty kibernetiki. [Systemology and linguistic aspects of cybernetics]. M.: Sovetskoye radio, 1978. 368 c. (In Russ.).
33. IDEF. Available at: <https://www.idef.com/> (accessed 06.02.2023).
34. UML. Available at: <https://www.uml.org/> (accessed 06.02.2023).
35. Scheer A.W. The Development Lines of Process Automation. (Eds.: Bergener K., Räckers M., Stein A.). *The Art of Structuring*. Springer, Cham. 2019. pp. 213–220.
36. BPMN. Available at: <https://www.bpmn.org/> (accessed 06.02.2023).
37. Trotskiy D.V., Gorodetskiy V.I. [Scenario knowledge model and process description language for assessing and predicting situations]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2009. no. 8. pp. 94–127. (In Russ.).
38. Flávio E.A. Horita, Porto de Albuquerque J., Marchezini V., Mendiondo E.M. Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil. *Decision Support Systems*. 2017. vol. 97. pp. 12–22.
39. Vasiliev Y. SOA and WS-BPEL: Composing Service-Oriented Solution with PHP and ActiveBPEL. Packt Publishing. 2007. 316 p.
40. Yuditskiy S.A. Stenarnyy podkhod k modelirovaniyu povedeniya biznes-sistem. [Scenario approach to modeling the behavior of business systems]. M.: SINTEG, 2001. 112 p. (In Russ.).
41. Petukhov G.B. Osnovy teorii effektivnosti tselenappavlennykh protsessov. Chast'1. Metodologiya, metody, modeli. [Fundamentals of the purposeful processes' efficiency theory. Part 1. Methodology, methods, models]. M.: Izd-vo MO SSSR, 1989. 635 p. (In Russ.).
42. Mikhalevich V.S., Volkovich V.L. Vychislitel'nyye metody issledovaniya i proyektirovaniya slozhnykh sistem. [Computational methods for research and design of complex systems]. M.: Nauka, 1982. 286 p. (In Russ.).
43. Zelentsov V.A., Potryasayev S.A., Pimanov I.Yu., Semenov A.Ye. Programmnoye obespecheniye «Katalog-V» dlya avtomaticheskoy katalogizatsii kosmicheskikh snimkov. [Software "Catalog-B" for automatic cataloging of space images]. Svidetel'stvo o gosudarstvennoy registratsii programmy dlya EVM № 2017612870. (In Russ.).
44. CivicStructure: Available at: <https://schema.org/CivicStructure> (accessed 06.02.2023).

Zelentsov Viacheslav — Ph.D., Dr.Sci., Professor, Chief researcher, Laboratory of information technologies in system analysis and modeling, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: system analysis, reliability theory, methods and information technologies for decision support in complex organizational and technical systems using aerospace data. The number of publications — 350. v.a.zelentsov@gmail.com; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-0103.

Pimanov Ilya — Ph.D., Junior researcher, Laboratory of information technologies in system analysis and modeling, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: geographic information systems, web cartography, Earth remote sensing. The number of publications — 65. pimen@list.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-0103.

Potryasayev Semyon — Ph.D., Dr.Sci., Leading researcher, Laboratory of information technologies in system analysis and modeling, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS). Research interests: system analysis and operations research, theory of managing the structural dynamics of complex organizational and technical systems. The number of publications — 120. semp@mail.ru; 39, 14-th Line V.O., 199178, St. Petersburg, Russia; office phone: +7(812)328-0103.

Acknowledgements. The research on sections 1, 2, 3.2, 3.3, 4-6 was performed within the budgetary theme No FFZF-2022-0004. The research on Section 3.1 was supported by the Russian Science Foundation grant (project No. 22-19-00767, <https://rscf.ru/project/22-19-00767>). Examples of calculations were performed using the resources of the Core Shared Research Facility «North-West Center for Monitoring and Forecasting the Development of Territories», <https://ckp-rf.ru/catalog/ckp/2079622/>. The authors are grateful to Victor Mochalov for his help in remote sensing data processing when software testing.

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 personal account on the website of the journal "Informatics and Automation" <http://ia.spcras.ru/>. At the registration the authors are requested to fill out all data fields in the proposed form. The submissions should be prepared using MS Word 2007, LaTeX. The text of the paper in the main part should not exceed 30 pages. Pages are not numbered; hyphenations are 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 06.07.2023. Passed for print 01.08.2023.

Printed in Publishing center GUAP.

Address: 67 litera A, B. Morskaya, St. Petersburg, 190000, Russia

Founder and Publisher: SPC RAS.

Address: 39 litera A, 14th Line V.O., St. Peterburg, 199178, Russia.

The journal is registered in the Federal Service for Supervision of Communications, Information Technology, and Mass Media, Registration Certificate (registration number) ПИ № ФС77-79228 dated September 25, 2020 Subscription Index П5513, Russian Post Catalog

Подписано к печати 06.07.2023. Дата выхода в свет 01.08.2023.

Формат 60×90 1/16. Усл. печ. л. 13,1. Заказ № 317. Тираж 300 экз., цена свободная.

Отпечатано в Редакционно-издательском центре ГУАП.

Адрес типографии: Б. Морская, д. 67, лит. А, г. Санкт-Петербург, 190000, Россия

Учредитель и издатель: СПб ФИЦ РАН.

Адрес учредителя и издателя: 14-я линия В.О., д. 39, лит. А, г. Санкт-Петербург, 199178, Россия

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

Подписной индекс П5513 по каталогу «Почта России»