

ISSN 2713-3192
DOI 10.15622/ia.2024.23.5
<http://ia.spcras.ru>

ТОМ 23 № 5

**ИНФОРМАТИКА
И АВТОМАТИЗАЦИЯ**

**INFORMATICS
AND AUTOMATION**



СПб ФИЦ РАН

**Санкт-Петербург
2024**



INFORMATICS AND AUTOMATION

Volume 23 № 5, 2024

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

Founded in 2002

Founder and Publisher

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

Editor-in-Chief

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

Editorial Council

A. A. Ashimov	Prof., Dr. Sci., Academician of the National Academy of Sciences of the Republic of Kazakhstan, Almaty, Kazakhstan
I. A. Kalyaev	Prof., Dr. Sci., Academician of RAS, Taganrog, Russia
Yu. A. Merkuryev	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
L. B. Sheremetov	Assoc. Prof., Dr. Sci., Mexico, Mexico

Editor: A.S. Lopotova

Interpreter: Ya.N. Berezina

Art editor: N.A. Dormidontova

Editorial office address

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

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

The journal is indexed in Scopus

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

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

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

Том 23 № 5, 2024

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

Журнал основан в 2002 году

Учредитель и издатель

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

Главный редактор

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

Редакционный совет

А. А. Ашимов	академик Национальной академии наук Республики Казахстан, д-р техн. наук, проф., Алматы, Казахстан
И. А. Каляев	академик РАН, д-р техн. наук, проф., Таганрог, РФ
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 РАО, д-р техн. наук, проф., Санкт-Петербург, РФ
В. А. Соيفер	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. Каия	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, Аizu, 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. Пшихолов	д-р техн. наук, проф., Таганрог, РФ
А. Л. Ронжин	д-р техн. наук, проф., зам. главного редактора, Санкт-Петербург, РФ
Х. Самани	д-р наук, доцент, Плимут, Соединённое Королевство
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», 2024
Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных в составе печатного периодического издания - журнала «ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ» статей по текущим экономическим, политическим, социальным и религиозным вопросам с обязательным указанием имени автора статьи и печатного периодического издания журнала «ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ»

CONTENTS

In Memory of Valentin M. Ponomarev	1283
In Memory of Viktor V. Alexandrov	1287
Artificial Intelligence, Knowledge and Data Engineering	
A. Zuenko, Yu. Oleynik SCHEDULING AS A CONSTRAINT SATISFACTION PROBLEM (USING THE EXAMPLE OF OPEN-PIT MINE PRODUCTION SCHEDULING PROBLEM)	1290
N.V. Hung, N. Tan, N.T.T. Nga, L.T.H. Trang, T.T.T. Hang USING ONTOLOGY TO ANALYZE ENGLISH COMMENTS ON SOCIAL NETWORKS	1311
A. Potyupkin, S. Pilkevich, V. Zaytsev AN APPROACH TO A PRIORI ASSESSMENT OF FUZZY CLASSIFICATION MODELS IN MONITORING TASKS	1339
A.H. Ibrahim, M.A. Boudref, L. Badis CLUSTERING OF NETWORKS USING THE FISH SCHOOL SEARCH ALGORITHM	1367
M. Porhun, M. Vashkevich EFFICIENT IMPLEMENTATION OF GAMMATONE FILTERS BASED ON WARPED COSINE MODULATED FILTER BANK	1398
R. Jenifer, V.J. Prakash RIVEST-SHAMIR-ADLEMAN ALGORITHM OPTIMIZED TO PROTECT IOT DEVICES FROM SPECIFIC ATTACKS	1423
Robotics, Automation and Control Systems	
A. Voevoda, Yu. Filiushov, V. Filiushov DEVELOPMENT OF A LINEAR CONTROL SYSTEM FOR A THROTTLE OF A UAV PROPELLER-MOTOR GROUP	1454
A. Iskandar, A. Hammoud, B. Kovács IMPLICIT UNDERSTANDING: DECODING SWARM BEHAVIORS IN ROBOTS THROUGH DEEP INVERSE REINFORCEMENT LEARNING	1485
A. Shilin, H. Pham Trong, V. Nguyen Vuong SYNTHESIS OF A FUZZY CONTROLLER BY A SECOND-ORDER OBJECT WITH DELAY	1505
T. Fedorova, V. Ryzhov, K. Safronov THE USE OF HYBRID COMMUNICATION ARCHITECTURE IN UNDERWATER WIRELESS SENSOR NETWORKS TO ENHANCE THEIR LIFETIME AND EFFICIENCY	1532

СОДЕРЖАНИЕ

Памяти Пономарёва Валентина Михайловича	1283
Памяти Александрова Виктора Васильевича	1287
Искусственный интеллект, инженерия данных и знаний А.А. Зуенко, Ю.А. Олейник СОСТАВЛЕНИЕ РАСПИСАНИЙ КАК ЗАДАЧА УДОВЛЕТВОРЕНИЯ ОГРАНИЧЕНИЙ (НА ПРИМЕРЕ ПЛАНИРОВАНИЯ ОТКРЫТЫХ ГОРНЫХ РАБОТ)	1290
Н.В. Хунг, Н. Тан, Н.Т.Т. Нга, Л.Х. Транг, Т.Т. Ханг ИСПОЛЬЗОВАНИЕ ОНТОЛОГИИ ДЛЯ АНАЛИЗА АНГЛИЙСКИХ КОММЕНТАРИЕВ В СОЦИАЛЬНЫХ СЕТЯХ	1311
А.А. Потюпкин, С.В. Пилькевич, В.В. Зайцев ПОДХОД К АПРИОРНОМУ ОЦЕНИВАНИЮ НЕЧЕТКИХ КЛАССИФИКАЦИОННЫХ МОДЕЛЕЙ В ЗАДАЧАХ МОНИТОРИНГА	1339
А.Х. Ибрагим, М.А. Будреф, Л. Бадис КЛАСТЕРИЗАЦИЯ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА ПОИСКА КОСЯКОВ РЫБ	1367
М.И. Порхун, М.И. Вашкевич ЭФФЕКТИВНАЯ РЕАЛИЗАЦИЯ ГАММАТОН-ФИЛЬТРОВ НА ОСНОВЕ НЕРАВНОПОЛОСНОГО КОСИНУСНО-МОДУЛИРОВАННОГО БАНКА ФИЛЬТРОВ	1398
Р. Дженифер, В.Д. Пракаш АЛГОРИТМ RIVEST-SHAMIR-ADLEMAN, ОПТИМИЗИРОВАННЫЙ ДЛЯ ЗАЩИТЫ УСТРОЙСТВ ИНТЕРНЕТА ВЕЩЕЙ ОТ КОНКРЕТНЫХ АТАК	1423
Робототехника, автоматизация и системы управления А.А. Воевода, Ю.П. Филошов, В.Ю. Филошов РАЗРАБОТКА ЛИНЕЙНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ТЯГОЙ ВИНТОМОТОРНОЙ ГРУППЫ ДЛЯ БПЛА	1454
А. Искандар, А. Хаммуд, Б. Ковач СКРЫТЫЙ СМЫСЛ: ДЕКОДИРОВКА РОЕВОГО ПОВЕДЕНИЯ РОБОТОВ С ПОМОЩЬЮ ГЛУБОКОГО ОБРАТНОГО ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ	1485
А.А. Шилин, Х. Фам Чонг, В. Нгуен Вонг СИНТЕЗ FUZZY-РЕГУЛЯТОРА ОБЪЕКТОМ ВТОРОГО ПОРЯДКА С ЗАПАЗДЫВАНИЕМ	1505
Т.А. Федорова, В.А. Рыжов, К.С. Сафронов ИСПОЛЬЗОВАНИЕ ГИБРИДНОЙ КОММУНИКАЦИОННОЙ АРХИТЕКТУРЫ ПОДВОДНОЙ БЕСПРОВОДНОЙ СЕНСОРНОЙ СЕТИ ДЛЯ ПОВЫШЕНИЯ ЕЕ ВРЕМЕНИ ЖИЗНИ И ЭФФЕКТИВНОСТИ	1532



ПАМЯТИ ПОНОМАРЁВА ВАЛЕНТИНА МИХАЙЛОВИЧА

1 сентября 2024 года исполнилось 100 лет со дня рождения Валентина Михайловича Пономарева, доктора технических наук, профессора, создателя и первого директора Санкт-Петербургского института информатики и автоматизации Российской академии наук, выдающегося ученого, талантливого педагога, организатора, руководителя и участника крупнейших фундаментальных и прикладных исследований в области теории оптимального управления, создания и повышения эффективности систем управления крылатых и баллистических ракет, космических аппаратов и маневрирующих головных частей, развития вычислительных систем и сетей, автоматизированных систем научных исследований и интегрированных автоматизированных производственных систем, результаты которых явились важным вкладом в отечественную и мировую науку, экономику и образование страны, ее безопасность и развитие международных научных связей.

К семидесятым годам прошлого века в Ленинграде возникла необходимость создания в городе научного учреждения, ориентированного на разработку фундаментальных проблем развития и применения вычислительной техники в интересах повышения эффективности производства и научных исследований. Наряду с этим требовала своего решения и конкретная задача расширения применения ЭВМ и вычислительных методов в ленинградских учреждениях АН СССР. Для решения этой проблемы постановлением Президиума АН СССР был организован Ленинградский научно-исследовательский вычислительный центр АН СССР (ЛНИВЦ).

Пономарев Валентин Михайлович родился на хуторе Черный Сасовского района Рязанской области. После окончания в 1942 г.

в городе Иваново Спецшколы ВВС и краткосрочных курсов летного училища В.М. Пономарев был направлен для прохождения военной службы авиатехником в частях ВВС, а затем в 1943 г. был направлен в Ленинградскую военно-воздушную инженерную академию (ЛВВИА). После окончания академии в 1948-1949 гг. В.М. Пономарев служил в Прикавказском военном округе в должности инженера авиационной эскадрильи.

Научно-педагогическая деятельность В.М. Пономарева началась с должности адъюнкта Военной академии имени А.Ф. Можайского (ВА им. А.Ф. Можайского), где он затем занимал должности преподавателя кафедры динамики полета, старшего научного сотрудника научно-исследовательской лаборатории систем управления летательных аппаратов, старшего преподавателя кафедры систем управления ракет и космических аппаратов, начальника этой кафедры.

После увольнения в запас В.М. Пономарев назначается директором Ленинградского научно-исследовательского вычислительного центра АН СССР ЛНИВЦ, созданного постановлением Президиума АН СССР, при непосредственном участии и благодаря усилиям В.М. Пономарева, который в 1985 г. был преобразован в Ленинградский институт информатики и автоматизации АН СССР (сегодня это Санкт-Петербургский институт информатики и автоматизации РАН).

Под его руководством в 1980 г. была разработана и создана (первая очередь) Региональная вычислительная подсеть «Северо-Запад» Академсеть. Прообраз нынешнего Интернета. Сеть охватывала 30 ленинградских учреждений АН СССР и других ведомств. Благодаря этой сети уже к 1985 г. вычислительными и программными ресурсами ЛНИВЦ пользовались около 2000 специалистов из 82-х организаций, среди которых были академические институты, вузы и научно-производственные организации Ленинграда. При этом ряд ленинградских академических институтов был подключен к сети через локальные автоматизированные системы научных исследований, разработанными в ЛНИВЦ.

В.М. Пономаревым осуществлялось научно-методическое руководство реализацией Государственной территориально-отраслевой программой развития народного хозяйства Ленинграда и Ленинградской области на основе автоматизации и широкого использования вычислительной техники – Интенсификация-90. По этой программе ЛНИВЦ под руководством В.М. Пономарева выполнил ряд фундаментальных и прикладных исследований в области интегральных систем автоматизации и гибких

автоматизированных производственных систем. Результаты этих исследований широко использовались в промышленности.

Благодаря активной и успешной научно-организаторской деятельности В.М. Пономарева ученые ЛНИВЦ активно участвовали в ряде важных и перспективных направлений развития вычислительной техники и технологий ее применения, в работе научных советов АН СССР таких, как совет по комплексной проблеме «Кибернетика», совет по проблемам управления движением и навигации, совет по искусственному интеллекту, а также в таких международных организациях, как Международная федерация по автоматическому управлению (ИФАК) и Международная федерация по обработке информации (ИФИП), Международный институт прикладного системного анализа и др. Под руководством В.М. Пономарева, впервые в нашей стране, были подготовлены и успешно проведены в Ленинграде ряд международных научных мероприятий: I Международное совещание по искусственному интеллекту (1977 г.) и II Международное совещание по искусственному интеллекту (1980 г.), Международная конференция ИФИП/ИФАК по программируемым системам для автоматизации проектирования и технологических процессов в производстве (ПРОЛАМАТ-82), проведен I Международный симпозиум ИФАК по искусственному интеллекту (промышленное применение). В работе этих научных мероприятий участвовало от 150 до 400 специалистов из 15 стран.

Благодаря научным и практическим достижениям В.М. Пономарева в области вычислительной техники и автоматизации в стране оформилось научное направление информатика, что отразилось в названии нового отделения в составе АН СССР (Постановление № 12 от 3 марта 1983 г.) – Отделения информатики, вычислительной техники и автоматизации.

За заслуги и достижения В.М. Пономарев награжден медалью «За победу над Германией в Великой Отечественной войне 1941-1945 гг.» (1946 г.) и медалью «За боевые заслуги» (1953 г.). За научные достижения и педагогические заслуги награжден медалью ордена «За заслуги перед Отечеством» 1 степени (2003 г.) и орденом «Знак почета» (1986 г.), ему присуждена ученая степень доктора технических наук (1964 г.), присвоено ученое звание профессора (1966 г.).

Под руководством В.М. Пономарева подготовлены 18 докторов и 47 кандидатов наук. Он является автором более 300 научных трудов, в том числе 14 монографий, учебников и учебных пособий.

Благодаря широкому научному кругозору В.М. Пономарева, его выдающимся организаторским способностям наш институт стал одним из ведущих научных учреждений в области информатики и автоматизации не только в АН СССР и в Ленинграде – Санкт-Петербурге, но и в целом в стране и за ее пределами.

Научная деятельность В.М. Пономарева и намеченные им, его учениками и последователями цели и пути развития компьютерных технологий во многом обеспечили развитие нового направления – информатизация общества.

Ученики В.М. Пономарева и его коллеги в новых условиях и вызовах XXI века стараются внести свой вклад в получение новых достижений в области информационно-коммуникационных технологий и информатизации общества.

Жизнь Валентина Михайловича Пономарева оборвалась 27 марта 2009 года. Ему было 84 года. Прах В.М. Пономарева находится в Колумбарии Смоленского православного кладбища в Санкт-Петербурге.



ПАМЯТИ АЛЕКСАНДРОВА ВИКТОРА ВАСИЛЬЕВИЧА

10 сентября 2024 г. исполнилось 85 лет со дня рождения Виктора Васильевича Александрова, блестящего ученого, заслуженного деятеля науки Российской Федерации, доктора технических наук, профессора.

Виктор Васильевич не просто стоял у истоков создания Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН), где долгие годы возглавлял лабораторию автоматизации научных исследований; если быть точным, он буквально посвятил многие годы жизни (через век и тысячелетие) созданию, становлению и развитию нашей научной организации.

Окончив в 1954 г. семь классов школы № 196, Дзержинского района Ленинграда, на Моховой улице, поступил в Ленинградский физико-механический техникум, закончил его в 1958 г. и начал работать в Ленинградском научно-исследовательском радиотехническом институте (ЛНИРТИ). Относился ко всему новому с неподдельным энтузиазмом, например, участвовал в летных испытаниях радиоэлектронной аппаратуры – доплеровских и астроинерциальных систем. Продолжал учиться без отрыва от производства и в 1964 г. получил диплом с отличием Ленинградского политехнического института по специальности “Автоматика и телемеханика”. С 1965 г. по 1967 г., став инженером, остался в составе Лаборатории бионики ЛНИРТИ, разработал цифровые методы классификации и распознавания на основе алгоритмов быстрых преобразований Адамара - Уолша.

Большая часть жизни Виктора Васильевича тесно связана с наукой и работой в научных организациях. В 1972-1975 гг. он – успешный доцент кафедры биомедицины ЛЭТИ; с 1975 г. – заведующий

лабораторией вновь созданной организации Академии наук: с 1975 г. – это ЛВЦ (отдел ФТИ им. Иоффе), с 1978 г. – ЛНИИВЦ АН СССР, с 1985 г. – ЛИИАН АН СССР, с 1991 г. – СПИИРАН.

С 1966 г. – по 1975 г. он возглавлял направления, связанные с разработкой цифровых методов классификации и распознавания изображений на основе алгоритмов быстрых преобразований Адамара, впоследствии легших в основу технологий цифровой обработки многомерных сигналов, а также разработкой САПРов для проектирования распознающих оптоэлектронных устройств, оптронов и волоконно-оптической линий связи, вёл активную научно-педагогическую деятельность (кафедра биомедицины Ленинградского электротехнического института).

За выдающиеся успехи в научной сфере в 1990 г. Виктор Васильевич был удостоен премии Центрального Комитета КПСС, а в 1993 г. – почетного звания «Заслуженный деятель науки РФ» и Премии Джона фон Неймана по информатике. Награжден медалью «В память 300-летия Санкт-Петербурга» и медалью ордена «За заслуги перед Отечеством» II степени. За разработку и создание новой техники в 2011 г. ему присуждена Премия Правительства Российской Федерации в области науки и техники. В 2013 г. Правительством РФ отмечен значительный вклад Александрова В.В. в создание прорывных технологий и технологий двойного назначения в рамках Государственных программ разработки современных образцов ВВСТ и присуждена стипендия на основании приказа Президента РФ.

Научные исследования Виктора Васильевича были связаны преимущественно с новым подходом к пространственным преобразованиям структур данных. Этот подход основан на предложенном и развитом им аппарате самоподобного иерархического развертывания n -мерных пространств, использующем принцип заполнения пространство кривых (ЗПК), что послужило основой для разработки адаптивно-динамической структуры данных как компьютерного аналога декартовой системы координат, организации ассоциативной адресации и поиска информации в ЭВМ и построения специализированных ЭВМ пирамидально-рекурсивной архитектуры.

Выдающийся ученый В.В. Александров создал научную школу в области обработки данных, изображений и сигналов. Им предложено и развито новое научное направление имитационного моделирования развивающихся систем, методы рекурсивно-фрактального синтеза. Разработаны теоретические основы программируемой технологии обработки данных и семантического анализа аудио- и видеопотоков данных.

Виктор Васильевич создал и вел постоянный институтский семинар: «Банки данных, информационные и экспертные системы», который много лет был весьма популярен, привлекал ученых из многих научных учреждений России и не только. Виктор Васильевич активно работал в ряде научных и экспертных советов, не оставался равнодушным к оригинальным идеям; в нем естественным образом сочетались высокая требовательность к коллегам, друзьям и доступность, безотказность, готовность поддержать человека в трудной профессиональной или жизненной ситуации. Эти качества также способствовали и успешному развитию профессиональных контактов с учеными из Венгрии, Болгарии, Польши, Германии, США, Канады, Мексики, Иордании, Франции, Финляндии, Великобритании и др., к которым подключалась творческая молодежь. Установившиеся с учеными отношения были надежными, поддерживались многие годы.

Виктора Васильевича всегда влекло неизведанное, а оно есть не только в науке, например, будучи истинным путешественником, он всегда был готов отправиться в новые края, так в 60-е годы XX века сумел проехать на мотоцикле через среднюю Азию, а позднее уже на «Запорожце» и в сопровождении семьи объездил южную и западную части СССР. Стремление открывать было неискоренимым; он сыграл заметную роль в создании секции подводного плавания Ленинградской области, был разработчиком специальных туристических маршрутов и организатором слетов.

Не остались непокоренными и горные вершины Кавказа, Тянь-Шаня, доломитовых Альп и особо любимые северные Хибины; любовь у них была и осталась взаимной, несмотря на то, что однажды он с друзьями попал под сходящую лавину, к счастью, все закончилось хорошо – испытание горами на прочность прошло успешно, а Виктор Васильевич навсегда остался неутомимым горнолыжником.

Результаты исследований Виктора Васильевича Александрова опубликованы в 25 монографиях и более чем 300 научных статьях, часть которых опубликована и в журнале «Информатика и автоматизация» (ранее «Труды СПИИРАН»).

Коллеги помнят Виктора Васильевича как выдающегося ученого, блестящего лектора, футуролога, предвидевшего научные и социальные изменения, и в то же время доброго и отзывчивого человека.

Жизнь Виктора Васильевича Александрова оборвалась 2 ноября 2020 года. Ему был 81 год.

А.А. ЗУЕНКО, Ю.А. ОЛЕЙНИК
**СОСТАВЛЕНИЕ РАСПИСАНИЙ КАК ЗАДАЧА
УДОВЛЕТВОРЕНИЯ ОГРАНИЧЕНИЙ (НА ПРИМЕРЕ
ПЛАНИРОВАНИЯ ОТКРЫТЫХ ГОРНЫХ РАБОТ)**

Зуенко А.А., Олейник Ю.А. Составление расписаний как задача удовлетворения ограничений (на примере планирования открытых горных работ).

Аннотация. Описываемые в статье исследования направлены на развитие методов составления расписаний. Принципиальным недостатком существующих методов смешано-целочисленного линейного программирования в применении к рассматриваемым задачам является то, что они слишком требовательны к объемам оперативной памяти. Сложность же применения процедур локального поиска к подобным задачам высокой размерности состоит в разработке эффективного способа нахождения приемлемого первоначального приближения и определении функции перехода в соседнее состояние, которая бы позволила достаточно быстро достичь оптимума. В теории исследования операций добавление к задаче дополнительных условий может привести к принципиальному изменению используемой схемы решения задачи. Предлагаемые в статье методы реализованы в рамках парадигмы программирования в ограничениях, что позволяет более экономно с точки зрения оперативной памяти представлять зависимости предметной области, а также обеспечивает возможность поэтапного учета разнородных условий задачи без принципиального изменения схемы поиска решений. Существенная часть исследований посвящена использованию методов логического вывода на ограничениях для снижения размерности пространства поиска и ускорения процесса вычислений. Подход к составлению расписаний проиллюстрирован на задаче оптимизации планирования открытых горных работ, которую впервые предложено решать как задачу удовлетворения ограничений. Для нахождения первого допустимого решения предложен метод «жадного» поиска, результат применения которого затем может быть улучшен с помощью разработанного гибридного метода. Оба метода опираются на оригинальные процедуры вывода на ограничениях. Предложенный подход доказал свою эффективность для блочных моделей размерностью в десятки и сотни тысяч блоков.

Ключевые слова: задача удовлетворения ограничений, программирование в ограничениях, распространение ограничений, локальный поиск, смешано-целочисленная оптимизация, планирование открытых горных работ.

1. Введение. Методы составления расписаний в последнее время значительно продвинулись благодаря применению методов и инструментов удовлетворения ограничений [1].

Составление расписаний – это задача назначения множества операций (действий) множеству рабочих ресурсов, на которые (как на операции, так и на ресурсы) накладывается ряд ограничений. Совокупность действий заранее известна и требуется отыскать лишь их порядок и распределение по рабочим ресурсам. Многие задачи составления расписаний являются NP-трудными.

Принципиальным недостатком существующих методов смешано-целочисленного линейного программирования в применении к рассматриваемым задачам является то, что они слишком требовательны к объемам оперативной памяти [2]. Сложность же применения процедур локального поиска к подобным задачам высокой размерности состоит в разработке эффективного способа нахождения приемлемого первоначального приближения и определении функции перехода в соседнее состояние, которая бы позволила достаточно быстро достичь оптимума [3, 4].

Актуальность разработки новых методов решения задач составления расписаний, обуславливается возрастающей потребностью в средствах поддержки принятия решений, обеспечивающих снижение производственных затрат и максимизацию прибыли при анализе пространства альтернатив высокой размерности. В теории исследования операций основное внимание почти всегда уделяется методам решения задач составления расписаний конкретного типа или определению их сложности, а не разработке общего подхода к составлению расписаний. Иногда, добавление к задаче дополнительных условий может привести к принципиальному изменению используемой схемы решения задачи. В результате создано огромное количество алгоритмов для большого количества конкретных задач. Это сильно отличает составление расписаний в рамках традиционного подхода от подхода с использованием технологии программирования в ограничениях, где основное внимание уделяется разработке общих методов решения сложных комбинаторных задач.

Программирование в ограничениях обеспечивает хорошую основу для интеграции методов исследования операций и более общих схем решения комбинаторных задач. Подобная интеграция основана на понятии глобального ограничения. Глобальные ограничения инкапсулируют определенную часть задачи удовлетворения ограничений и, вместо того, чтобы использовать набор ограничений для моделирования конкретной подзадачи, применяется специальное «более крупное» ограничение, которое позволяет лучше анализировать структуру подзадачи. Глобальные ограничения реализуют эффективные алгоритмы для решения четко определенных подзадач. Их можно комбинировать с другими ограничениями, моделирующими некоторые дополнительные условия.

Парадигма программирования в ограничениях (ППО) является мощным средством для решения комбинаторных задач. В неё входит

широкий спектр методов искусственного интеллекта, информатики и исследования операций [5, 6].

Главные принципы ППО, которые позволили авторам осуществить выбор в пользу данной парадигмы как методологии решения задач составления расписаний высокой размерности, заключаются в следующем:

- Для применения ППО решаемая задача должна быть представлена в декларативном виде как задача удовлетворения ограничений (ЗУО, *Constraint Satisfaction Problem*), то есть должны быть заданы множества: переменных, областей определения этих переменных (доменов), а также зависимостей на переменных, называемых ограничениями. *Решением* ЗУО является такое присваивание значений всем переменным, которое не нарушает ни одного ограничения [7].

- Порядок задания ограничений несущественен.

- Обеспечивается возможность унифицированной обработки разнородных ограничений. Ограничения могут задаваться с помощью арифметических выражений; логических формул; таблиц; выражений, формулируемых на языке специализированных теорий; а также с помощью уже упоминаемого механизма глобальных ограничений.

- Любой метод удовлетворения ограничений должен содержать две обязательных компоненты: а) компоненту, реализующую вывод (распространение); б) компоненту, реализующую поиск. Вывод (распространение) реализуется как целенаправленное сужение областей определения переменных. Эвристики, используемые в процедурах поиска, и методы вывода разрабатываются не под конкретную задачу, а под целый класс задач и зачастую являются универсальными.

- Обеспечивается возможность сопровождать модели, открытые для оперативных модификаций, то есть развивающиеся модели. При добавлении/удалении из модели ограничений нет необходимости принципиально изменять схему решения задачи.

В статье преимущества подхода к составлению расписаний на основе технологии программирования в ограничениях проиллюстрированы на таком классе задач исследования операций как оптимизация планирования открытых горных работ (*Open Pit Mine Production Scheduling Problem*). Впервые предложено решать задачи данного класса как задачи удовлетворения ограничений.

2. Постановка задачи ПОГР. Для моделирования и отображения процесса разработки карьера на компьютере часто используется блочная модель карьера (B), представляющая собой набор блоков одинакового размера ($b_i \in B$), на которые разбита порода, планируемая к выемке

за заданное количество периодов (T) [8]. Каждый блок обладает набором свойств, таких как: координаты, процент содержания полезного ископаемого, а также базовая ценность блока. Значение базовой ценности позволяет отнести блок либо к блокам с полезными ископаемыми (ПИ), либо к блокам со вскрышными породами (ВП), на основе этого же параметра рассчитывается итоговая ценность блока, зависящая также от периода добычи блока [9].

Содержательная постановка задачи ПОГР состоит в том, чтобы определить такие положения рабочих бортов по периодам отработки, чтобы объемы ПИ и ВП, заключенные между последовательными положениями, соответствовали заданным с точностью до допустимой погрешности. При этом прибыль (суммарная итоговая ценность блоков) от разработки карьера должна быть максимальной [10, 11].

Развитие карьера (его углубление и расширение) должно происходить с соблюдением требований (ограничений), представленных далее.

К таким ограничениям относятся ограничения на порядок извлечения блоков: для извлечения какого-то определенного блока необходимо, чтобы некоторые другие блоки были извлечены в тот же период или ранее. Эти ограничения носят технологический характер и определяются свойствами горных пород, составляющих карьер. Иногда подобные ограничения задаются с помощью так называемого шаблона выемки, в котором указывается некоторый «ключевой» блок и блоки непосредственно предшествующего ему уровня, подлежащие изъятию до него.

Помимо технологических в задаче следует также учитывать ряд экономических ограничений, которые обусловлены производительностью добывающей техники и перерабатывающих мощностей.

Приведем формализацию базовой задачи ПОГР с помощью линейных уравнений, как это предложено в [12]. Предполагается, что блочная модель задачи состоит из N блоков, планирование производится на T периодов добычи.

Для каждого блока b_i , где $i \in \{1, \dots, N\}$ создается набор переменных $X_{ij} \in \{0, 1\}$, где $j \in \{1, \dots, T\}$. $X_{ij} = 1$ обозначает, что блок i добыт в период j . Переменные X_{ij} должны удовлетворять ограничению:

$$\sum_{j=1}^T X_{ij} = 1, \quad \forall i \in \{1, \dots, N\}. \quad (1)$$

Прибыль от добычи каждого блока i в определенный период задается явно, образуя множество конкретных значений C_{ij} , где $j \in \{1, \dots, T\}$.

Целевая функция с учетом вышенаписанного представляется следующим образом:

$$\sum_{i=1}^N \sum_{j=1}^T (C_{ij} * X_{ij}). \quad (2)$$

Экономические ограничения на объемы добычи принимают вид:

$$\begin{aligned} W_j^{min} &\leq \sum_{i \in I^w} X_{ij} \leq W_j^{max}, & \forall j \in \{1, \dots, T\}, \\ O_j^{min} &\leq \sum_{i \in I^o} X_{ij} \leq O_j^{max}, & I^w, I^o \subseteq \{1, \dots, N\}, \end{aligned} \quad (3)$$

где I^w – множество индексов блоков с ВП, I^o – множество индексов блоков с ПИ, W_j^{min} и W_j^{max} – нижняя и верхняя границы нормы добычи вскрыши в период j , O_j^{min} и O_j^{max} – нижняя и верхняя границы нормы добычи руды в период j .

Для каждого блока b_i блочной модели определим множество $\{b_p\}$, где $p \in P_i$, как множество блоков, извлекаемых непосредственно до блока b_i согласно шаблону выемки. Технологические ограничения в таком представлении разворачиваются во множество линейных уравнений:

$$\left(\sum_{k=1}^t X_{pk} - X_{it} \right) \geq 0, \quad \forall i \in \{1, \dots, N\}, \forall t \in \{1, \dots, T\}, \forall p \in P_i. \quad (4)$$

Уравнения такого вида указывают, что блоки из множества P_i должны быть добыты не позже блока b_i . Если предположить, что шаблон добычи описывается с помощью M блоков, то для описания технологических ограничений потребуются в среднем $N * M * T$ уравнений. Так, например, для простейшего шаблона из 5 блоков на 5 периодов планирования, ограничение на порядок добычи породит в среднем 20 линейных уравнений на каждый блок, что при размерности задачи в 5000 блоков, даст уже 100000 уравнений только для описания технологических ограничений. С увеличением размерности задачи, количества периодов

планирования и сложности шаблона добычи, одно только хранение порожденных линейных уравнений в памяти представляет существенную сложность, не говоря уже о дальнейшей процедуре решения.

Таким образом, использование методов целочисленного программирования не позволяет решать задачи требуемого уровня дискретизации блочной модели (сотни тысяч – миллионы блоков). Добавление дополнительных ограничений только усугубляет данную ситуацию.

В отличие от базовой, в рассматриваемой в настоящей работе постановке задачи ПОГР в качестве еще одного экономического ограничения выступает ограничение на среднее значение содержания полезного компонента в блоках с ПИ, извлекаемых за период. Данное ограничение представляется с помощью линейных уравнений следующим образом:

$$\begin{aligned} \sum_{i=1}^N (g_i - G_j^{max}) X_{ij} &\leq 0, \\ \sum_{i=1}^N (g_i - G_j^{min}) X_{ij} &\geq 0, \end{aligned} \quad \forall j \in \{1, \dots, T\}, \quad (7)$$

где:

g_i – содержание полезного компонента в блоке i ;
 G_j^{min} и G_j^{max} – нижняя и верхняя границы заданного диапазона допустимого среднего содержания полезного компонента в блоках с ПИ извлекаемых в период j .

3. Предлагаемый подход к решению задач ПОГР. В ходе исследований разработан новый подход к представлению и решению задач ПОГР на основе ППО. Особенность подхода состоит в применении процедур рассуждения на ограничениях на большинстве этапов решения рассматриваемого класса задач. Процедуры, реализующие подобные рассуждения, опираются на методы усечения доменов переменных.

Для представления задачи ПОГР как ЗУО каждому блоку b_i карьера сопоставляется соответствующая многозначная переменная ЗУО X_i . В качестве значений доменов переменных выступают номера возможных периодов разработки (от 1 до T). Рассуждения на ограничениях предложено реализовывать с использованием не только типовых, но и авторских методов логического вывода, предназначенных для повышения эффективности унифицированной обработки разнородных требований к проведению горных работ. В частности, для анализа ограничений на

последовательность извлечения блоков (технологических ограничений) разработан метод выявления несовместных значений переменных «*Block sequencing*», который подробно описан ниже [13].

Экономические ограничения задачи формализуются в виде типовых глобальных ограничений *global cardinality constraint*, которые позволяют ограничивать количество переменных, принимающих определенное значение [14]. В рассматриваемой задаче используются 2 таких ограничения: $gcc(X^w, U, Y^w)$ и $gcc(X^o, U, Y^o)$, описывающих нормы добычи для блоков с ВП и ПИ соответственно, где:

X^w и X^o – множества переменных блоков с ВП и ПИ соответственно;

$U = \{1, \dots, T\}$ – возможные номера периодов планирования;

Y^w и Y^o – наборы переменных, имеющих домены:
 $Y_j^w \in \{W_j^{min}, \dots, W_j^{max}\}, Y_j^o \in \{O_j^{min}, \dots, O_j^{max}\}, \forall j \in \{1, \dots, T\}$.

Управляя доменами переменных Y^w и Y^o , можно ограничивать количество переменных из X^w и X^o , принимающих определенные значения.

Особую сложность представляло моделирование требования на среднее значение полезного компонента за период добычи. В ходе исследований найдена такая форма представления данного ограничения, при которой не происходит снижение производительности вычислений по сравнению с вариантом задачи, не включающим это условие. Нормы среднего содержания полезного компонента описываются, парой ограничений *linear* для каждого периода добычи [13]. Эти ограничения представляют по сути линейные уравнения, которые несколько отличаются от уравнений (7) из представленной ранее формализации задачи и имеют вид:

$$\begin{aligned} \sum_{i=1}^N (g_i * X_{ij}) &\leq N * G_j^{max}, \\ \sum_{i=1}^N (g_i * X_{ij}) &\geq N * G_j^{min}, \end{aligned} \quad \forall j \in \{1, \dots, T\}, \quad (8)$$

что записывается как $linear(g, X_j, \ll, N * G_j^{max})$ и $linear(g, X_j, \gg, N * G_j^{min})$ для каждого периода $j \in \{1, \dots, T\}$.

Значение целевой функции в данном представлении рассчитывается как:

$$\sum_{i=1}^N FC_i, \quad (9)$$

где FC_i – итоговая ценность блока i , которая зависит от базовой стоимости блока и периода его добычи, т.е. может быть вычислена с помощью некоторой функции $cost(X_i, C_i)$. Значения FC_i используются достаточно часто в процессе поиска, поэтому каждый раз вычислять их нецелесообразно. Подобную функциональную зависимость можно описать с помощью табличного ограничения (таблица 1), где во втором столбце находятся значения функции $cost$, вычисляемые единожды при построении данного ограничения. Все шаги решения задачи можно разбить на три основных этапа, которые подробно излагаются ниже: а) предварительная обработка данных (предварительное упрощение задачи); б) поиск некоторого допустимого решения; в) улучшение первого полученного решения.

Таблица 1. Зависимость итоговой стоимости от периода разработки

X_i	FC_i
1	$cost(1, C_i)$
2	$cost(2, C_i)$
...	...
T	$cost(T, C_i)$

При формулировании задачи ПОГР в виде ЗВО применяется предварительный анализ блочной модели, позволяющий в некоторых случаях заранее сократить домены некоторых переменных до начала основного процесса решения, уменьшая, таким образом, общее пространство поиска.

Для нахождения за полиномиальное время одного из допустимых решений задачи ПОГР, то есть решения, удовлетворяющего одновременно всем ограничениям, предложен метод «жадного» поиска, использующий разработанные процедуры вывода на ограничениях. Обеспечена возможность одновременного поиска допустимого решения в различных вычислительных потоках. Дальнейшая оптимизация целевой функции реализуется разработанным гибридным методом [15], сочетающим преимущества систематического [16] и локального поиска [17]. Компонента, осуществляющая локальный поиск, на основе первоначально найденного допустимого решения выполняет

формирование подпространств вариантов (альтернатив), где применяется уже систематический поиск решений задачи оптимизации ПОГР. В ходе систематического поиска также используются упомянутые выше методы рассуждений на ограничениях.

4. Предварительная обработка данных (априорное упрощение ЗУО). В ходе предварительного анализа для каждого блока b_i вычисляется минимальное количество рудных и вскрышных блоков, которые необходимо изъять из карьера перед ним. Затем производится уточнение диапазона периодов, когда возможна выемка блока b_i . Если для добычи какого-то блока необходимо предварительно изъять N_o рудных блоков и N_w вскрышных, при этом верхняя норма добычи за период по руде равна O_{max} , а по вскрыше – W_{max} , то нижняя граница домена соответствующего блока сократится до меньшего из значений $1 + (N_o/O_{max})$ и $1 + (N_w/W_{max})$ (деление целочисленное с округлением в меньшую сторону). Для получения для каждого блока b_i значений N_o и N_w необходимо рассмотреть блоки, зависящие друг от друга согласно шаблону, итеративно перебирая слой за слоем вверх.

Пример 1. Для иллюстрации упрощений ЗУО, реализуемых в рамках этапа предварительной обработки, рассмотрим модель, представленную на рисунке 1. Рудные блоки модели обозначены зеленым цветом, вскрышные – белым, внутри каждого блока указан домен соответствующей переменной, ниже модели схематично представлен заданный в рамках примера шаблон добычи.

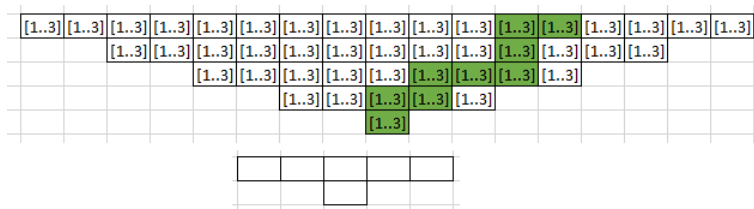


Рис. 1. Двумерная блочная модель карьера и шаблон извлечения

Нормы добычи для простоты будут заданы не в виде диапазона значений, а конкретными значениями: 3 рудных блока за период (O_{max}) и 12 вскрышных (W_{max}). Для каждого блока модели были вычислены значения N_w и N_o , результаты этих вычислений представлены на рисунке 2, где первое значение соответствует N_w , второе – N_o , сам блок также учитываются при подсчете этих значений.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	10	10	10	10	10	10	10	10	10	10	10	01	01	10	10	10	10
B			60	60	60	60	60	60	60	51	42	33	42	42	51		
C					150	150	150	141	132	114	114	114	123				
D							262	244	226	217	226						
E											369						

Рис. 2. Расчет объемов добычи для извлечения каждого блока модели

Например, для блока B_3 множеством блоков, анализируемых согласно шаблону добычи блоков, является множество $S_{B_3} = \{B_3, A_1, A_2, A_3, A_4, A_5\}$, для блока B_4 – множество $S_{B_4} = \{B_4, A_2, A_3, A_4, A_5, A_6\}$ и т.д. Для блока C_5 множество анализируемых блоков можно получить следующим образом: $S_{C_5} = \{C_5\} \cup S_{B_3} \cup S_{B_4} \cup S_{B_5} \cup S_{B_6} \cup S_{B_7}$.

Тогда, на основе анализа множества S_{B_3} , ясно, что при извлечении блока B_3 минимальное количество вынимаемых вскрышных блоков равно 6, а рудных – 0. В результате аналогичных подсчетов для блока B_4 получаются числа 6 и 0, а для блока C_5 – числа 15 и 0. После вычисления значений N_w и N_o для каждого блока карьера, осуществляется сокращение доменов некоторых переменных задачи (рисунок 3).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]
B			[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]
C					[2..3]	[2..3]	[2..3]	[2..3]	[2..3]	[2..3]	[2..3]	[2..3]	[2..3]	[1..3]			
D							3	[2..3]	[2..3]	3	[2..3]						
E											3						

Рис. 3. Сокращение доменов переменных в результате априорного упрощения ЗУО

Вычисления, реализуемые в ходе этапа предварительной обработки данных, позволили сократить начальное пространство поиска демонстрационной задачи с размерности в 3^{45} возможных присваиваний до размерности $3^{30} * 2^{12}$ присваиваний, т.е. более чем в 3500 раз. Для задач большей размерности с доменами переменных большей мощности эффект оказывается еще значительнее.

Далее начинается этап поиска первого решения задачи, за которым следует этап его улучшения. Оба этих этапа в значительной степени опираются на авторский метод вывода на ограничениях, названный «*Block sequencing*».

5. Глобальное ограничение «*Block sequencing*». В реальных задачах количество ограничений очень велико, и незначительное изменение домена распространителем какого-либо ограничения

запускает целый каскад вызовов алгоритмов распространения для других ограничений. Поэтому целесообразно максимизировать сокращения доменов именно на этапе распространения отдельных ограничений (фильтрации) путем представления набора одноптипных ограничений в виде одного глобального ограничения, для которого разрабатываются эффективные алгоритмы распространения. Так, например, одно глобальное ограничение *alldifferent* для N переменных, которое предписывает, чтобы все значения переменных были различными, заменяет $N(N - 1)/2$ простейших неравенств. Аналогичная тенденция прослеживается и для других глобальных ограничений. Глобальные ограничения помимо ускорения процесса решения еще и упрощают описание ЗУО.

Для повышения эффективности решения комбинаторных задач ПОГР в рамках ППО разработан новый метод анализа несовместных значений переменных (метод вывода на ограничениях) *Block sequencing*. Метод позволяет совместно анализировать условия на последовательность извлечения для совокупности блоков, исключая, таким образом, большее количество неприемлемых значений из доменов переменных. Применение ограничения существенно снижает требования к оперативной памяти, что позволяет решать задачи с требуемым уровнем дискретизации модели месторождения.

Поясним используемые далее обозначения. Переменная X_j с доменом D_j – это переменная обозначающая «ключевой» блок, относительно которого выше по шаблону выемки лежат блоки X_{i_m} , где $i_m \in \{i_1, \dots, i_p\}$, подлежащие выемке непосредственно до «ключевого», а ниже по шаблону находятся блоки X_{k_l} , где $k_l \in \{k_1, \dots, k_r\}$. Очевидно, что $\{i_1, \dots, i_p\} \cap \{k_1, \dots, k_r\} = \emptyset$.

Алгоритм распространения *Block sequencing* состоит в следующем:

Шаг 0. Изначально все переменные модели, помещаются в очередь распространения Q . Затем циклически выполняются шаги 1-7.

Шаг 1. Если очередь пуста, то Шаг 8. Если очередь не пуста, то из нее извлекается переменная X_j .

Шаг 2. Для переменных X_{k_l} , где $k_l \in \{k_1, \dots, k_r\}$, вычисляется $a = \min_{k_l \in \{k_1, \dots, k_r\}} \{right_bound(D_{k_l})\}$, то есть минимальная верхняя граница для всех их доменов D_{k_l} .

Шаг 3. В соответствии с найденным значением a урезается домен переменной X_j (если это возможно) – уточняется его верхняя граница: $D_j^* = \{t : (t \in D_j) \wedge (t \leq a)\}$. Вычисляется $a' = \min\{right_bound(D_j^*), a\}$.

Шаг 4. Уточняются верхние границы доменов переменных $\{X_{i_m}\}$: $D_{i_m}^* = \{t : (t \in D_{i_m}) \wedge (t \leq a')\}$, где $i_m \in \{i_1, \dots, i_p\}$. Те из переменных $\{X_{i_m}\}$, домены которых изменились, помещаются в очередь Q .

Шаг 5. Для переменных X_{i_m} , где $i_m \in \{i_1, \dots, i_p\}$, вычисляется $e = \max_{i_m \in \{i_1, \dots, i_p\}} \{left_bound(D_{i_m})\}$, то есть максимальная нижняя граница для всех их доменов D_{i_m} .

Шаг 6. В соответствии с найденным значением e урезается домен переменной X_j (если это возможно) – уточняется его нижняя граница: $D_j^{**} = \{t : (t \in D_j) \wedge (t \geq e)\}$. Вычисляется $e' = \max\{left_bound(D_j^{**}), e\}$.

Шаг 7. Уточняются нижние границы доменов переменных $\{X_{k_l}\}$: $D_{k_l}^* = \{t : (t \in D_{k_l}) \wedge (t \geq e')\}$, где $k_l \in \{k_1, \dots, k_r\}$. Те из переменных $\{X_{k_l}\}$, домены которых изменились, помещаются в очередь Q . Сама переменная X_j удаляется из очереди Q . Возврат к шагу 1.

Шаг 8. Конец.

При попытке изменения домена любой переменной может возникнуть ситуация, что домен окажется пустым, тогда сообщается о противоречии, и распространитель завершает работу.

Если распространение завершилось без противоречий, производится проверка доменов переменных. Если встречена хотя бы одна переменная, в домене которой осталось больше 1 значения, то распространение достигло неподвижной точки, т.е. для дальнейшего решения задачи требуется запуск распространителей других ограничений или ветвление. Если домены всех переменных сократились до единственного значения, то ограничение считается удовлетворенным.

6. Этап поиска первого допустимого решения: «жадный» поиск решений задачи планирования открытых горных работ. Предлагаемый метод «жадного» поиска решения задачи ПОГР реализует целенаправленное (с использованием специализированных эвристик) движение по пространству поиска. Данное пространство сокращается в процессе движения с помощью механизма распространения ограничений. На каждом шаге осуществляется выбор некоторой переменной и присваивание ей определенного значения из соответствующего домена с последующим запуском распространителей ограничений задачи. Таким образом, из доменов переменных, которым еще предстоит присваивание, исключаются выявленные распространителями заведомо неприемлемые значения.

Далее опишем используемые эвристики выбора переменной и приписываемого ей значения. Значимым признаком упорядочивания

блоков (соответствующих им переменных) является глубина залегания блока в карьере, т.е. координата Z (блоки рассматриваются горизонтальными слоями, начиная с нижнего слоя). Для обеспечения полного порядка на блоках, они упорядочиваются также и по двум остальным координатам в рамках одного горизонтального слоя модели. Были протестированы разные способы подобного упорядочивания. На ряде тестовых задач лучше всего показали следующие эвристики: а) обход блоков уровня зигзагом, б) выбор по расстоянию от проекции дна и с) выбор по расстоянию от среднего рудного блока в горизонтальном слое (рисунок 4).

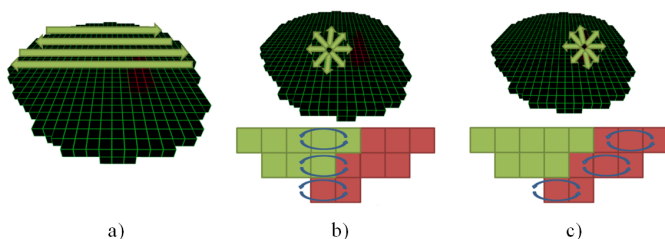


Рис. 4. Тестируемые эвристики «жадного» поиска решений: а) обход зигзагом; б) выбор по расстоянию от проекции дна; с) выбор по расстоянию от среднего рудного блока

При использовании всех трех эвристик, переменной присваивается максимальное из допустимых значений ее домена.

Описанная процедура «жадного» поиска потребляет сравнительно мало машинных ресурсов, поэтому для повышения вероятности нахождения допустимого решения она выполняется в параллельном режиме с использованием всех трех описанных выше эвристик одновременно. В случае нахождения допустимых решений несколькими эвристиками выбирается лучшее решение из найденных. Если ни одна из эвристик «жадного» поиска не позволяет найти допустимое решение, то запускается процедура систематического поиска. При этом для выбора переменной в ходе систематического поиска будет использована та из трех эвристик, с помощью которой «жадному» поиску удалось присвоить значения большему количеству переменных до обнаружения противоречий.

Для иллюстрации принципов работы «жадного» поиска вернемся к представленной ранее демонстрационной задаче из примера 1. Пусть для нее уже проведена предварительная обработка данных, результат которой был показан на рисунке 3. Поскольку блочная модель в

примере является двумерной, то в качестве эвристики для выбора блоков (соответствующих переменных) будет применяться следующее правило: блоки выбираются слева направо, начиная с нижних слоев. В результате применения этой эвристики заполнение клеточек на рисунке 3 начнется снизу, причем сначала в клеточки будет записываться значение «3» (максимальное значение из доменов переменных). Распространение ограничений вызывается после каждого присваивания, однако значимые изменения в доменах станут появляться после присваивания значений всем рудным блокам в нижних двух рядах (*D* и *E*), когда норма добычи руды для третьего периода будет полностью выполнена. Это влечет удаление значения «3» из доменов переменных остальных рудных блоков, в результате чего в доменах рудных блоков ряда *C* остается только значение «2». Таким образом, полностью выбрана норма добычи руды и для второго периода, поэтому значения оставшихся рудных блоков в рядах *A* и *B* становятся равными «1». Далее на основании анализа ограничений на последовательность выемки блоков с помощью глобального ограничения «*Block sequencing*» уточняются домены переменных вскрышных блоков $A_6 - A_{11}$, $A_{14} - A_{16}$, $B_8 - B_{11}$, B_{13} , B_{14} , находящихся выше рудных согласно используемому шаблону (рисунок 5).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..2]	[1..2]	[1..2]	[1..2]	1	1	1	1	1	[1..2]	[1..2]	[1..3]
B		[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..2]	[1..2]	[1..2]	1	[1..2]	[1..2]	[1..3]			
C				[2..3]	[2..3]	[2..3]	[2..3]	[2..3]	2	2	2	[1..3]					
D							3	3	3	3	[2..3]						
E									3								

Рис. 5. Состояние задачи после набора рудных блоков для третьего периода и распространения ограничений

Распространение ограничений на текущем шаге больше не может быть выполнено, поэтому продолжается присваивание значений переменных согласно применяемой эвристике. Следующее значимое изменение произойдет, когда будет выбрана норма добычи вскрышных блоков за третий период. При этом значение «3» удаляется из доменов оставшихся переменных с еще не присвоенными значениями (рисунок 6).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	[1..2]	[1..2]	[1..2]	[1..2]	[1..2]	[1..2]	[1..2]	[1..2]	[1..2]	1	1	1	1	1	[1..2]	[1..2]	[1..2]
B		3	3	[1..2]	[1..2]	[1..2]	[1..2]	[1..2]	[1..2]	[1..2]	1	[1..2]	[1..2]	[1..2]			
C				3	3	3	3	3	3	2	2	2	3				
D							3	3	3	3	3						
E									3								

Рис. 6. Состояние задачи после набора всех блоков для третьего периода и распространения ограничений

Далее, согласно применяемой эвристике, вскрышным блокам будет назначаться значение «2» до момента, пока не будет выбрана норма вскрыши за второй период, после чего оставшимся блокам распространитель назначит значение «1». Полученное в результате этапа «жадного» поиска решение демонстрационной задачи представлено на рисунке 7.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	2	2			1	1	1	1	1	1	1	1	1	1	1	1	1
B			3	3	3	2	2	2	2	2	2	1	2	2	2		
C					3	3	3	3	3	2	2	2	3				
D							3	3	3	3	3						
E									3								

Рис. 7. Полученное допустимое решение задачи

7. Этап улучшения полученного допустимого решения задачи. После получения допустимого решения запускается процедура гибридного поиска. В рамках данного этапа найденное решение задачи модифицируется путем ослабления некоторых его переменных, т.е. возврата их в начальное неприсвоенное состояние. Полученное таким образом новое сравнительно небольшое подпространство поиска затем исследуется систематическим методом с целью выявления новых решений с более высоким значением целевой функции. Систематический поиск в анализируемом подпространстве ограничивается по времени и/или по количеству найденных тупиковых веток дерева поиска. Фактически, параметр допустимого количества тупиковых ветвей задает степень «тщательности» исследования подпространства ослабленного решения и может варьироваться в зависимости от размерности задачи.

В случае нахождения нового решения с более высоким значением целевой функции, оно используется в качестве основы для дальнейшего поиска вместо первого найденного. Если такого решения не обнаружено, то исходное решение снова ослабляется иным образом, т.е. выбираются другие переменные для ослабления.

Признаком остановки процесса решения выступает отводимое на поиск время, а конечным решением задачи будет считаться последнее решение, найденное методом гибридного поиска.

Проиллюстрируем этот этап с помощью продолжения решения двумерной задачи из примера 1. В данном случае для ослабления первого допустимого решения, представленного на рисунке 7, будет выбрана часть переменных верхних двух рядов (ряды A и B), как это показано на рисунке 8.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	2	2	2	1	1	1	1	1	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]
B			3	3	3	2	2	2	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]	[1..3]
C					3	3	3	3	3	2	2	2	3				
D							3	3	3	3	3						
E									3								

Рис. 8. Ослабление части переменных задачи

Теперь необходимо провести процедуру распространения ограничений, что уберет заведомо недопустимые значения из доменов ослабленных переменных. В данном случае, поскольку нормы добычи для третьего периода выбраны в неослабленной части модели, из доменов ослабленных переменных удаляется значение «3», также для второго периода выбраны нормы руды, поэтому ослабленным переменным рудных блоков остается только принять значение «1». Домены остальных переменных сокращаются с использованием глобального ограничения «*Block sequencing*». Состояние задачи после устранения несовместных значений переменных представлено на рисунке 9.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	2	2		1	1	1	1	1	[1..2]	1	1	1	1	1	[1..2]	[1..2]	[1..2]
B			3	3	3	2	2	2	[1..2]	[1..2]	[1..2]	1	[1..2]	[1..2]	[1..2]	[1..2]	[1..2]
C					3	3	3	3	3	2	2	2	3				
D							3	3	3	3	3						
E									3								

Рис. 9. Результат распространения ограничений после ослабления части задачи

Подпространство поиска для состояния задачи, отображенного на рисунке 9, составляет 2^{10} присваиваний, что в $4 * 3^{30}$ раз меньше пространства поиска исходной задачи после предварительной обработки данных, что позволяет применять для исследования результирующего подпространства процедуры систематического поиска. В ходе исследования этого подпространства появляется возможность достаточно быстро найти новые решения задачи. При расчете целевой функции эти решения могут оказаться лучше исходного и стать основой для новой итерации.

8. Оценка эффективности предложенных методов. Для постановки задачи ПОГР, представленной выше в виде совокупности линейных уравнений, авторами проводились расчеты на тестовой машине с объемом оперативной памяти в 32 ГБ с использованием программного обеспечения Google OR-Tools. Машинных ресурсов не хватило даже для расчета задачи размерностью 2000 блоков. В литературе удалось найти пример решения задачи в близкой постановке методами

целочисленного линейного программирования с помощью решателя IBM Cplex, где на схожей по параметрам тестовой машине для размерности уже порядка 20000 блоков не было получено ни одного решения за 48 часов расчетов [18]. Применяя предложенный подход, удалось получить решение для задачи размерностью 82000 блоков, что заняло около получаса, а за 5 часов было построено решение для задачи в 525000 блоков.

По сравнению с методами смешано-целочисленного программирования применение разработанного подхода позволяет существенно снизить требования к оперативной памяти за счет представления зависимостей предметной области с использованием глобальных ограничений, что позволяет решать задачи с требуемым уровнем дискретизации модели месторождения.

9. Заключение. Описываемые в статье исследования направлены на развитие методов составления расписаний. В статье предложено представлять и решать задачу оптимизации ПОГР как задачу удовлетворения ограничений. Разработанные методы реализованы в рамках парадигмы программирования в ограничениях, что позволяет более экономно с точки зрения оперативной памяти представлять зависимости предметной области, а также обеспечивает возможность поэтапного учета разнородных условий задачи без принципиального изменения схемы поиска решений. Существенная часть исследований посвящена использованию методов логического вывода на ограничениях для снижения размерности пространства поиска и ускорения процесса вычислений. Для нахождения допустимого решения предложен метод «жадного» поиска, результат работы которого может быть улучшен с помощью разработанного гибридного метода. Оба метода опираются на оригинальные процедуры вывода на ограничениях. Предложенный подход доказал свою эффективность для блочных моделей размерности в десятки и сотни тысяч блоков.

Литература

1. Baptiste Ph., Le Pape C., Nuijten W. Constraint-based scheduling: applying constraint programming to scheduling problems // Kluwer Academic Publishers. 2001. 198 p.
2. Patidar M., Bhardwaj R., Choudhary S. The study of linear programming approach for optimal scheduling of work in a corporation with different models // Materials Today: Proceedings. 2020. vol. 29. no. 2. pp. 661–667.
3. Ulaga L., Durasevic M., Jakobovic D. Local search based methods for scheduling in the unrelated parallel machines environment // Expert Systems with Applications. 2022. vol. 199.
4. Chen, Y., Lu J., He R., Ou J. An efficient local search heuristic for earth observation satellite integrated scheduling // Applied Sciences. 2020. vol. 10. no. 16.

5. Belaid M., Bessiere C., Lazaar N. Constraint Programming for Association Rules // Proceedings of the International Conference on Data Mining. Society for Industrial and Applied Mathematics. 2019. pp. 127–135.
6. Russell S., Norvig P., Davis E. Artificial intelligence: a modern approach –3 ed. // Upper Saddle River: Prentice Hall. 2010. 1132 p.
7. Narvaez D. Constraint Satisfaction Techniques for Combinatorial Problems // Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. 2018. vol. 32. no. 1. pp. 8028–8029.
8. Alipour A., Khodaiari A., Jafari A., Tavakkoli-Moghaddam R. An integrated approach to open-pit mines production scheduling // Resources Policy. 2022. vol. 75.
9. Novitasari R., Rosyidi C., Aisyati A. A Cut-off Grade Optimization Model in Multi Product Open Pit Mining Considering Reclamation and Valuable Waste Materials // IOP Conference Series: Materials Science and Engineering. 2021. vol. 1096. no. 1. DOI: 10.1088/1757-899X/1096/1/012021.
10. Tolouei K., Moosavi E., Tabrizi A., Afzal P., Bazzazi A. An optimization approach for uncertainty-based long-term production scheduling in open-pit mines using meta-heuristic algorithms // International Journal of Mining, Reclamation and Environment. 2021. vol. 35. no. 2. pp. 115–140.
11. Caccetta L. Application of Optimization Techniques in Open Pit Mining // Handbook of Operations Research in Natural Resources. Boston: Springer US, 2007. vol. 99. pp. 547–559.
12. Espinoza D., Goycoolea M., Moreno E., Newman A. MineLib: a library of open pit mining problems // Annals of Operations Research. 2013. vol. 206. no. 1. pp. 93–114.
13. Oleynik Y., Zuenko A. Open Pit Mine Production Scheduling using New Global Constraint // Proceedings of IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). 2022. pp. 1700–1704.
14. Regin J. Global Constraints: A Survey // Hybrid Optimization: The Ten Years of CPAIOR. 2011. pp. 63–134.
15. Mara S., Norcahyo R., Jodiawan P., Lusiantoro L., Rifai A. A survey of adaptive large neighborhood search algorithms and applications // Computers Operations Research. 2022. vol. 146.
16. Audemard G., Lecoutre C., Prudhomme C. Guiding Backtrack Search by Tracking Variables during Constraint Propagation // Proceedings of the 29th International Conference on Principles and Practice of Constraint Programming (CP 2023). Leibniz International Proceedings in Informatics (LIPIcs). 2023. vol. 280. pp. 9:1–9:17.
17. Kozik M. Solving CSPs using weak local consistency // SIAM Journal on Computing. 2021. vol. 50. no. 4. pp. 1263–1286.
18. Fathollahzadeh K., Mardaneh E., Cigla M., Waqar Ali Asad M. A mathematical model for open pit mine production scheduling with Grade Engineering and stockpiling // International Journal of Mining Science and Technology. 2021. vol. 31(4). pp. 717–728.

Зуенко Александр Анатольевич — канд. техн. наук, ведущий научный сотрудник, лаборатория информационных технологий управления промышленно-природными системами, ИИММ КНЦ РАН. Область научных интересов: программирование в ограничениях, моделирование слабо формализованных предметных областей. Число научных публикаций — 150. zuenko@iimm.ru; улица Ферсмана, 24А, 184209, Апатиты, Россия; р.т.: +7(815-55)79-782.

Олейник Юрий Андреевич — научный сотрудник, лаборатория информационных технологий управления промышленно-природными системами, ИИММ КНЦ РАН. Область научных интересов: программирование в ограничениях, моделирование

природно-промышленных комплексов. Число научных публикаций — 19. yoleynik@iimm.ru; улица Ферсмана, 24А, 184209, Апатиты, Россия; р.т.: +7(815-55)79-248.

Поддержка исследований. Работа выполнена в рамках текущей темы НИР «Разработка теоретических и организационно-технических основ информационной поддержки управления жизнеспособностью региональных критических инфраструктур Арктической зоны Российской Федерации» (регистрационный номер 122022800547-3).

A. ZUENKO , Yu. OLEJNIK
**SCHEDULING AS A CONSTRAINT SATISFACTION PROBLEM
(USING THE EXAMPLE OF OPEN-PIT MINE PRODUCTION
SCHEDULING PROBLEM)**

Zuenko A., Oleynik Yu. Scheduling as a Constraint Satisfaction Problem (Using the Example of Open-Pit Mine Production Scheduling Problem).

Abstract. The research described in the work is aimed at developing methods for Scheduling. The fundamental disadvantage of the existing methods of Mixed-Integer Linear Programming in application to the problems under consideration is the fact that they are too demanding on the amount of RAM. The difficulty of applying local search procedures to such high-dimensional problems is to develop an effective way to find an acceptable initial approximation and determine the neighboring state transition function, which would allow achieving the optimum fast enough. In the Operations Research Theory, adding additional conditions to a problem can lead to a fundamental change in the problem-solving scheme. The methods proposed in the study are implemented within the framework of the Constraint Programming Paradigm which makes it possible to represent the subject domain dependencies saving RAM, as well as to provide the ability step-by-step take into account heterogeneous problem conditions without essentially changing the scheme of finding solutions. A significant part of the research deals with methods of logical inference on constraints to reduce the search space and speed up the computational process. The approach to scheduling is illustrated by the Open-Pit Mine Production Scheduling Problem, which was first proposed to be solved as a Constraint Satisfaction Problem. In order to find the first feasible solution, a «greedy» search method is proposed, the result of which can be improved using the developed hybrid method. Both methods rely on original procedures of inference on constraints. The proposed approach has proven its efficiency for block models with sizes of tens and hundreds of thousands of blocks.

Keywords: constraint satisfaction problem, constraint programming, constraint propagation, local search, mixed-integer optimization, open pit mine production scheduling.

References

1. Baptiste Ph., Le Pape C., Nuijten W. Constraint-based scheduling: applying constraint programming to scheduling problems. Kluwer Academic Publishers. 2001. 198 p.
2. Patidar M., Bhardwaj R., Choudhary S. The study of linear programming approach for optimal scheduling of work in a corporation with different models. *Materials Today: Proceedings*. 2020. vol. 29. no. 2. pp. 661–667.
3. Ulaga L., Durasevic M., Jakobovic D. Local search based methods for scheduling in the unrelated parallel machines environment. *Expert Systems with Applications*. 2022. vol. 199.
4. Chen, Y., Lu J., He R., Ou J. An efficient local search heuristic for earth observation satellite integrated scheduling. *Applied Sciences*. 2020. vol. 10. no. 16.
5. Belaid M., Bessiere C., Lazaar N. Constraint Programming for Association Rules. *Proceedings of the International Conference on Data Mining. Society for Industrial and Applied Mathematics*. 2019. pp. 127–135.
6. Russell S., Norvig P., Davis E. *Artificial intelligence: a modern approach* –3 ed. Upper Saddle River: Prentice Hall. 2010. 1132 p.

7. Narvaez D. Constraint Satisfaction Techniques for Combinatorial Problems. Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. 2018. vol. 32. no. 1. pp. 8028–8029.
8. Alipour A., Khodaiari A., Jafari A., Tavakkoli-Moghaddam R. An integrated approach to open-pit mines production scheduling. Resources Policy. 2022. vol. 75.
9. Novitasari R., Rosyidi C., Aisyati A. A Cut-off Grade Optimization Model in Multi Product Open Pit Mining Considering Reclamation and Valuable Waste Materials. IOP Conference Series: Materials Science and Engineering. 2021. vol. 1096. no. 1. DOI: 10.1088/1757-899X/1096/1/012021.
10. Tolouei K., Moosavi E., Tabrizi A., Afzal P., Bazzazi A. An optimization approach for uncertainty-based long-term production scheduling in open-pit mines using meta-heuristic algorithms. International Journal of Mining, Reclamation and Environment. 2021. vol. 35. no. 2. pp. 115–140.
11. Caccetta L. Application of Optimization Techniques in Open Pit Mining. Handbook of Operations Research in Natural Resources. Boston: Springer US, 2007. vol. 99. pp. 547–559.
12. Espinoza D., Goycoolea M., Moreno E., Newman A. MineLib: a library of open pit mining problems. Annals of Operations Research. 2013. vol. 206. no. 1. pp. 93–114.
13. Oleynik Y., Zuenko A. Open Pit Mine Production Scheduling using New Global Constraint. Proceedings of IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). 2022. pp. 1700–1704.
14. Regin J. Global Constraints: A Survey. Hybrid Optimization: The Ten Years of CPAIOR. 2011. pp. 63–134.
15. Mara S., Norcahyo R., Jodiawan P., Lusiantoro L., Rifai A. A survey of adaptive large neighborhood search algorithms and applications. Computers Operations Research. 2022. vol. 146.
16. Audemard G., Lecoutre C., Prudhomme C. Guiding Backtrack Search by Tracking Variables during Constraint Propagation. Proceedings of the 29th International Conference on Principles and Practice of Constraint Programming (CP 2023). Leibniz International Proceedings in Informatics (LIPIcs). 2023. vol. 280. pp. 9:1–9:17.
17. Kozik M. Solving CSPs using weak local consistency. SIAM Journal on Computing. 2021. vol. 50. no. 4. pp. 1263–1286.
18. Fathollahzadeh K., Mardaneh E., Cigla M., Waqar Ali Asad M. A mathematical model for open pit mine production scheduling with Grade Engineering and stockpiling. International Journal of Mining Science and Technology. 2021. vol. 31(4). pp. 717–728.

Zuenko Alexander — Ph.D., Leading researcher, Laboratory of information technologies for control of industrial-natural systems, IIMM KSC RAS. Research interests: constraint programming, modeling in poorly formalized subject domains. The number of publications — 150. zuenko@iimm.ru; 24A, Fersmana St., 184209, Apatity, Russia; office phone: +7(815-55)79-782.

Oleynik Yurii — Researcher, Laboratory of information technologies for control of industrial-natural systems, IIMM KSC RAS. Research interests: constraint programming, modeling of nature-industrial complexes. The number of publications — 19. yoleynik@iimm.ru; 24A, Fersmana St., 184209, Apatity, Russia; office phone: +7(815-55)79-248.

Acknowledgements. The work was carried out within the framework of the current research topic «Development of theoretical and organizational and technical foundations of information support for managing the viability of regional critical infrastructures of the Arctic zone of the Russian Federation» (registration number 122022800547-3).

N.V. Hung, N. Tan, N.T.T. Nga, L.T.H. Trang, T.T.T. Hang
**USING ONTOLOGY TO ANALYZE ENGLISH COMMENTS ON
SOCIAL NETWORKS**

Nguyen Viet Hung, Nguyen Tan, Nguyen Thi Thuy Nga, Le Thi Huyen Trang, Tran Thi Thuy Hang. **Using Ontology to Analyze English Comments on Social Networks.**

Abstract. Chatbots have become interesting for many users as technology becomes more and more advanced. The need for information exchange among people through computer systems is increasing daily, raising the preference for using chatbots in most countries. Since Vietnam is such a developing country with a variety of ethnic groups, it requires much attention to the proliferation of social networks and the expansion of the cooperative economy. Regarding social networks, the inappropriate use of words in everyday life has become a significant issue. There are mixed reviews of praise and criticism on social networks; and we try to reduce the negative language use and improve the quality of using social networks language. We aim to meet users' needs on social networks, promote economic development, and address social issues more effectively. To achieve these goals, in this paper we propose a deep learning technique using ontology knowledge mining to collect and process comments on social networks. This approach aims to enhance the user experience and facilitate the exchange of information among people by mining opinions in comments. Experimental results demonstrate that our method outperforms the conventional approach.

Keywords: chatbot, ontology, deep learning, machine learning, social network, Vietnam.

1. Introduction. Recently, the development of social networks has promoted the development of the economy as data analysis has partly contributed to improving economic efficiency, especially in improving marketing strategy [1 – 4]. There have been more and more social networking sites used, including Facebook and TikTok, for business purposes. For many businesses, creating a social network website has helped them interact with customers to collect their opinions, helping to get customers' essential needs. Besides, for businesses to be closer to their customers, they desire to receive as many mixed opinions as possible. Chatbots have been no longer strange on social networks recently. It has become more familiar to a great number of users, where consultants or assistants work continuously to give answers for many problems in specific fields. A chatbot system can understand and communicate to people and do a specific task. Chatbots have become automated tools that are able to interact with users online in natural language processing. Besides, chatbots have also been co-operated with other platforms, such as E-learning, to answer all users' questions [5, 6]. This has made online classes more attractive, especially in virtual classrooms during the COVID-19 pandemic time [7]. Despite different requirements and purposes, it is necessary for Chatbots to provide all users with guidance, documents, answers, and even entertaining stories.

Various chatbots [8, 9] are categorized based on actual and domain, goals, input processing, response generation methods, provided services, business support, people support, and methods construction. For example, to help businesses like [4], the author analyzed, classified comments, and evaluated product quality through user comments. As it is a powerful tool, based on the knowledge domain, the authors [10] used an access-based model in which the domain ontology can be built to retrieve information. Besides, several chatbots based on ontology have been built to support autoresponders from different fields, for example, educational and professional orientation [11], shopping in e-commerce [12], and consulting drug information in health [13]. Chatbots have helped a lot in many fields, including helping businesses avoid overloading, and consulting customers in many different areas without being inaccessible to professionals.

On the other hand, feedback generation and input processing methods have helped general models to be closer to users. These models can be referred to as machine learning or deep learning models, such as Long short-term memory [14 – 16], Recurrent neural network [14, 17], Fake Information Recognition [18, 19], Sequence-to-sequence [20, 21], Hierarchical Recurrent Encoder-decoder [22, 23], and other proposed methods. Some of the built popular models are better than teacher consulting services [6]. However, current chatbots are mainly based on a generic model that often gives short answers and needs to train a large data set. This also shows that building a chatbot has become more complicated when the training data is not large enough from the beginning.

In this paper, we propose a method to analyze social network comments for users through a chatbot system. Our method is based on Ontology to collect and analyze online user comments on social networks, thereby providing comments on these to support businesses partly in developing products for sale on social networks and hoping to have positive user reviews. The contribution of this proposed method can be summarized as follows:

- when analyzing social media comments, ontology can help you understand the underlying topics, identify relevant concepts, and establish relationships between them. This can help improve social media data categorization, sentiment analysis, and information retrieval. Furthermore, we employ ontology-specific techniques such as OWL, algorithms, and applications developed specifically for social media comment analysis. Besides, we provide the following. First, a structured and formalized method of representing knowledge is needed. Second, it facilitates knowledge sharing and integration and enables more practical information navigation and organization;

– the proposed method can replace traditional manual methods because it evaluates and provides direct results without requiring manual operations such as selection, classification, etc. The empirical model also performs significantly better than the reference models, up to 80.52%.

The remaining paper is as follows. Section 2 gives an overview of the related work. Next, the proposed model is presented in Section 3. Section 4 provides an evaluation. Finally, the paper is concluded in Section 5.

2. Related work. In this section, we will present the related techniques and some chatbot models being used by ontology.

2.1. Chatbots. There are many built-in definitions of chatbot [9]. However, it is a computer-to-computer communication program simulating human conversation through voice commands, text chat, or both.

There have been a lot of recent chatbot developments, including [17, 24, 25] chatbot models based on the use of neural networks using deep learning technology. The neural network is trained on data sets to generate grammatically responsive and relevant responses, such as sentences and words in the response texts. However, the system's limited analysis, including careful analysis of the data for the question, leads to the fact that the answers are so far mechanical and unfounded.

With the proposed method, our system extracts and analyzes information with a technical approach and ontology features. The system will analyze and evaluate topical content based on essential keywords in the sentence. Besides, our chatbot system also expands the synonyms and related conjunctions we analyze in the file system. This is also important when the system is analyzed more closely regarding comments. Experiments show that the results of the proposed method are superior.

2.2. Ontology. According to research [26–29], Ontology is a Semantic Web tool that defines the real-world semantics of terms and describes data and data relationships. These knowledge representations are very effective because they are applied in many different fields. This is also why many studies use ontologies to represent knowledge bases or retrieve information. Besides, OWL [30] is the primary Web ontology language that meets the requirements for building domain ontology, including its semantic syntax, expression support, and convenience. Therefore, we use the ontology effectively in the chatbot's search engine.

2.3. Existing chatbot models based on ontologies. In this section, we will present some existing models using ontology. Since Ontology is a solution to the desire to understand what one user is expressing, chatbot models based on ontology were born. Ontology-based models are domain knowledge-oriented so that domain-oriented conversations can be generated, in which ontology is

used to store and navigate domain knowledge. Based on the domain knowledge condition, the knowledge base relies on it to build an ontology that can provide information and generate answers to conversations [31]. Of all the benefits of this approach is that the history of mastery is fully preserved during the conversation. Besides, chatbots can be applied for closed domains providing more detailed answers.

Many question-answering systems called chatbots are emerging [32, 33]. Most usual question-answering systems perform three main tasks: analyzing the questions, searching the documents and databases containing the answers, and extracting the answers from them. The system aims to answer users' queries in their natural language, including structured and unstructured free-text databases. The response system will rely on the ontology to turn it into a semantic information inference language in that knowledge domain based on the ontology. Based on this, it does not require training but still naturally asks and answers questions in a particular case.

On the other hand, for better performance than conventional systems like Apache Lucene, it uses a keyword-based text search engine. An educational counseling system [34], which is for admissions counseling, has been proposed. It helps students to identify their future major. Since enrollment is always a big challenge for its time and effort consumption, building this system is very important. There has not been a reasonable response system for this work at universities in Vietnam yet, especially in the context of rapidly developing information technology.

In general, ontology-based models are constrained by a particular domain of knowledge but can be built from many other data sources. Therefore, the research on using ontology to support answering and solving communication problems is significant, especially in university admission counseling.

3. Proposed method

3.1. Research methodology. In this section, we propose a chatbot framework for opinion analysis, following these steps: First, the data is prepared. Next, we proceed to build structure, process, and argument by ontology in Figure 1. The stages are analyzed as follows:

- **step 01:** the user sends comments to the system; here, the system will use ontology to create semantic knowledge;
- **step 02:** the system is based on the stored knowledge bases, from which to analyze the requirements and put the needs into the system;
- **step 03:** the system conducts processing and analysis as required;
- **step 04:** the system understands the user's request;

- **step 05:** the system argues based on the requirements that the system has previously understood;
- **step 06:** the system returns opinions inside comments.

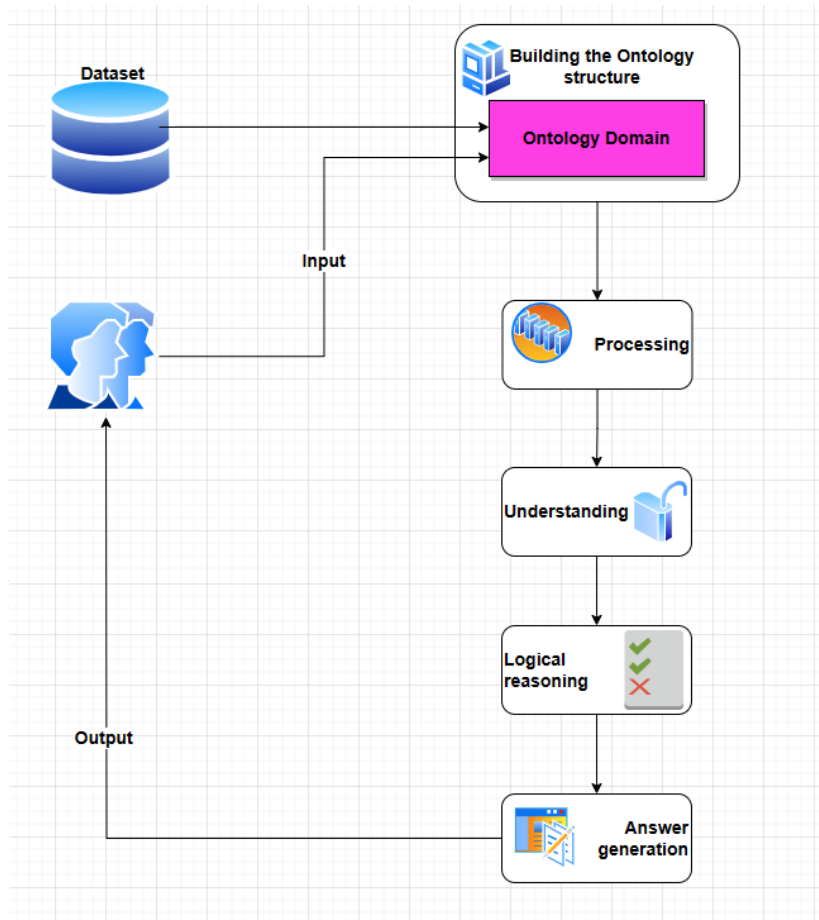


Fig. 1. The proposed communication analysis framework

Overall, this system is built to analyze a simple requirement, such as admissions advice.

3.2. Architecture of our system. This section details the architecture of the chatbot system we built in Figure 2.

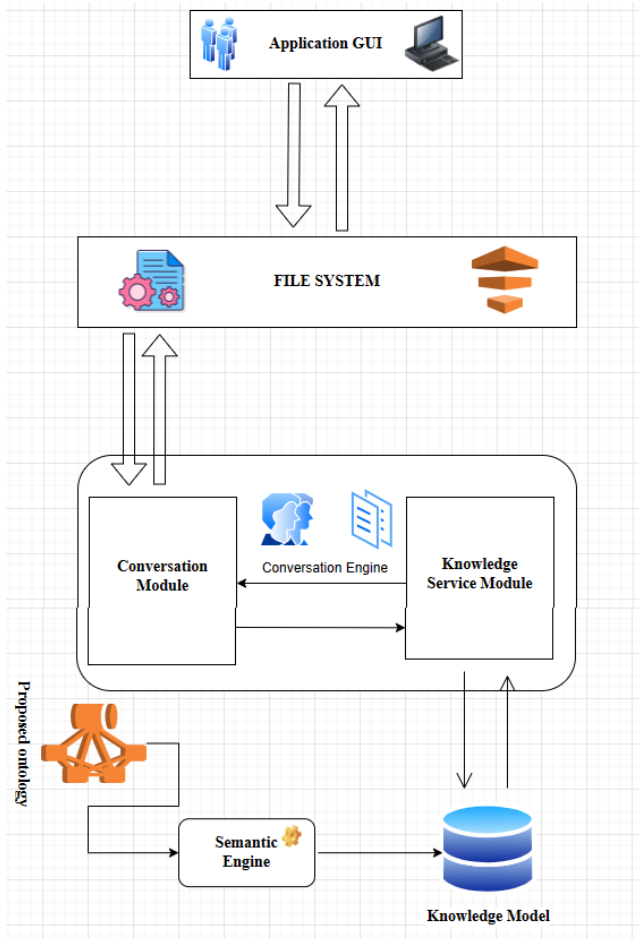


Fig. 2. System architecture diagram

This architecture not only allows interaction with the user but also provides interaction information between the ontology structure and the user's application based on the following criteria.

– **First:** a chatbot will be implemented to exchange information through a User Interface (GUI); the exchanged data is stored and processed through the file system. The file system is responsible for keeping the communication between the user and the system, processing and transmitting information to

the conversation modules inside the system so that the conversation is always processed and ready to respond to the users.

– **Second:** the conversation takes place through two modules that are exchanged based on the dialogue patterns stored in the incoming file system. Based on the user-suggested question, the chatbot can access lead conversation templates that match the user’s intentions. We use a knowledge mining service to get matching answer patterns and send feedback back to achieve this.

– **Third:** this step, which provides and contains the knowledge database, is crucial to get a suitable conversational sample to respond to the user. This knowledge model has both the dialogue patterns that may be relevant to the users and the problems arising from the conversations. The system will analyze and extract these derivatives to make the dialogue natural and reliable.

Furthermore, a further problem is giving users the feeling that the interaction is among humans rather than a computer system. For users, humans are preferred over technological computer systems because of their habits and reliability. A more significant challenge, therefore, is how to get the technology to understand the intent of the sentences and implications in human conversation and represent the authentic answers sent in. Consequently, we designed and exported a chatbot system based on the Neural Network model, called the Sequential model, with 3 Dense layers including:

- the first layer has 128 neurons;
- the second layer has 64 neurons;
- the last layer has the same number of neurons as the number of intents to predict the output.

On the other hand, using the Sequential model allows us to build an End-to-End model from input to creation efficiently. Keras Library supports model building and training thanks to its simplicity, ease of use, and high performance. Dense layers are connected consecutively, in which each layer is connected to all the nodes of the next layer, allowing the model to learn complex features from the data. The Neural Network deep learning model will help chatbots have better learning and prediction capabilities.

3.3. Ontology design our method. In this section, we build and develop the ontology to support the Chatbot system – a computer program is capable of communicating with users in natural language through processing techniques in Figure 3. Specifically, the ontology is designed to model critical concepts and relationships related to natural language’s vocabulary, grammar, and semantics helping our Chatbot system parse the syntax, extract information, classify sentences, and finally understand the meaning of users’ questions or comments on social networks and then give appropriate responses.

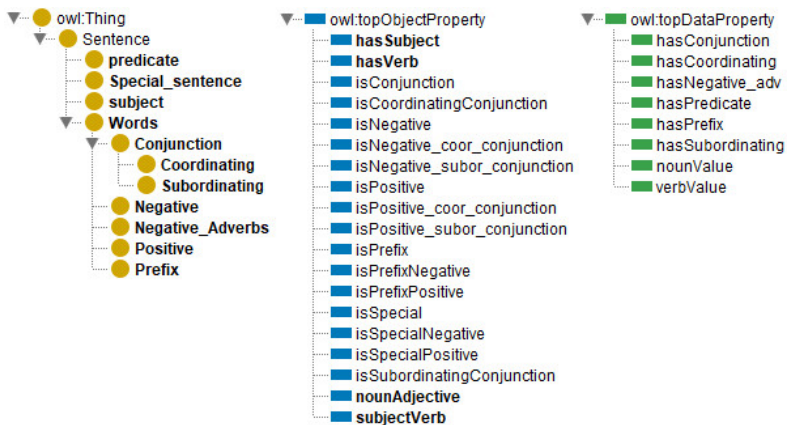


Fig. 3. Flowchart designed using our ontology

Furthermore, in artificial intelligence, ontologies play an essential role in providing general knowledge to computers, helping computers reason and solve problems. In other words, for Chatbot systems, ontology helps provide linguistic knowledge so Chatbots can understand and process natural human language. Therefore, we analyze and build a new model using ontology to evaluate the quality of comments better using the following tricks.

Class. Ontology generally refers to exploring what entities exist, their relation, and the categories and properties that can describe and classify them. Therefore, in our ontology design, a class represents a concept, object type, or a collection of objects with common properties and behavior. A class defines a collection of things with common characteristics, typical properties, and ordinary relationships. For example, when looking at comment sentences, it might be that "positive" represents the positive comments and "Negative" represents the negative class. However, a class can have sub-classes, in which sub-classes are within the same super-class, defining more specific attributes and relationships. For example, the positive category can include simple, compound, and comparative sentences. By defining classes and relationships between them, ontology creates a structure as well as a model representing knowledge about a specific field. This helps clarify concepts and relationships between them, supporting searching, querying, and processing information in that field. In Figure 3, the primary classes of ontology are designed as follows:

Words. Words are an abstract class representing all words in a natural language. The Words class defines the general properties of a word by:

- part-of-speech: shows the word type (noun, verb, adjective, etc.);

- polarity: positivity/negativity of the word;
- sentiment: the emotion the word brings (happy, sad, angry, etc.).

Sentence. Class representing the sentence and its attributes, such as:

- sentiment: general feeling of the whole sentence;
- polarity: the degree of positivity/negativity of the sentence;
- complexity: the complexity of the sentence structure.

Conjunction. This class describes conjunctions, divided into:

- coordinating: conjunction connecting independent clauses (and, or, etc.);
- subordinating: conjunction connecting dependent clauses (if, because, although, etc.).

Positive/Negative. Based on word type to divide into positive or negative classes; however, we need to pay attention to other factors, such as negative or interrogative, etc.:

- class contains words with positive or negative meanings;
- identify words expressing emotions and subjective assessments.

Special sentence. Usually, special sentences are difficult to classify, so for this type of sentence, the system needs more analysis, which can be based on the two criteria below:

- represents sentences whose meaning depends on the context;
- identify sentences that need special treatment based on context rather than literal meaning.

Instances. Instances are particular objects or real-world patterns belonging to a specific ontology class. Each instance is an instance of a class containing particular values for the properties and relationships defined in that class. For example, the attributes can be words in sentences on social comments. Sentences are filled with information through words, which helps to understand the structures and the meaning of sentences. From there, map them to reality and support analysis in different comments.

Properties. Properties in an ontology are used to describe the properties or relationships of a class or instance. Attributes help to identify specific information about objects and their relationships in the ontology. In this design, we divide it into two main categories:

- **Class Property:** class properties describe the common properties of a class. It is applied for all instances of that class. For example, in the class "Positive", the class attribute could be "positive words" to describe the positive comments in the class.

- **Object Property:** object properties describe the relationships between instances of different classes. It defines the relationship or interaction between

instances. For example, in the class "Positive" and class "Negative", the object attribute could be "Positive or negative", describing the relationship between affirmative or negative sentences, such as "good" or "not good".

In general, both of these property types can have constraints and values. Regulations define the rules or conditions that the properties must obey.

Relationship. We use it intending to connect layers, such as the input layer receiving input data and passing it to the following layers for processing. Next, we use an intermediate relationship that receives data from the previous layer and passes it to the next layer to perform calculations and data transformations. Finally, we use the relationship between the last layer in the model and the output. The last layer receives data from the previous layer and performs the final calculation to produce the prediction or final result.

3.4. Proposed model-UOAEN. In this section, we will detail our analysis and evaluation process. The steps are evaluated and analyzed according to our previous ideas and development models. We have changed and developed new methods that use more ontology technology. The steps tested are as follows:

Step 01: Data preprocessing. In this step, we analyze and remove punctuation such as (, *, , @, etc., and eliminate spaces between words in sentences along with "stopwords" from a comment entered into the system or text data file. The process is shown in Figure 4. The algorithm processing process will return an array containing the sentence's words for future comparison and evaluation.

Step 02: Data vectorization algorithm. In this step, we convert the data from non-vector to vector format to apply the natural language processing method in Figure 5. After the data is preprocessed in the above step, vectorizing the data and extracting the features, information, or documents will be encoded into digital vectors that machine algorithms can process and learn. This process is performed as follows.

First, we use word embedding [18], TF-IDF (Term Frequency-Inverse Document Frequency), and word inversion rate to accelerate the weight analysis. We perform the TF calculation by counting the keywords in the input data; if the key is too large, we use the ontology to determine the correlation. Similarly, when calculating IDF , we consider the parameter TF . It reflects whether the keyword is used too frequently or not, such as affecting the cumulative value extreme because it may need to measure the importance of a phrase, not just its frequency in terms of how many times it is used, but also its prepositions, pronouns, conjunctions, etc. A function calculates the reverse because it more accurately reflects the value when the phrase has a higher IDF score. After all, the linear IDF function overestimates the document's score. At the same time,

phrases with high *IDF* scores could include uncommon words and misspelled terms. Below are the two formulas we use to calculate *TF* and *IDF*.

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

$$IDF = \log\left(1 + \frac{\text{Total data}}{\text{Data contains keyword}}\right). \quad (2)$$

Generally, *TF – IDF* is a comprehensive measure instead of the keyword density measure, which only reflects the degree of "stuffing" a specific keyword into text. Furthermore, it reduces the prominence of meaningless words and phrases while increasing the significance of meaningful and uncommon terms.

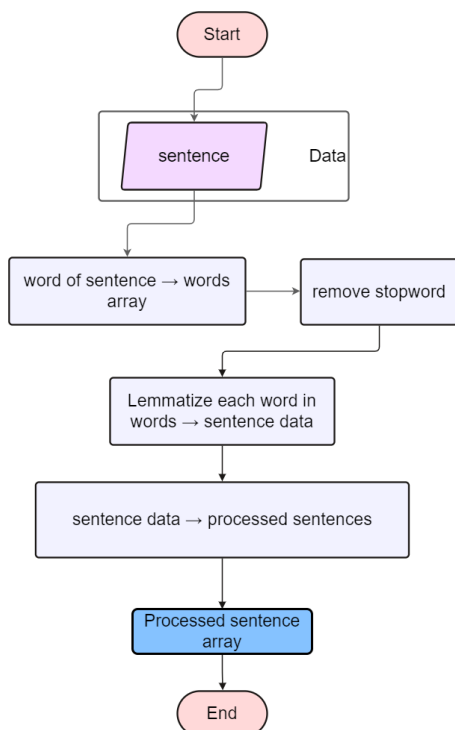


Fig. 4. Algorithmic diagram of the data preprocessing process

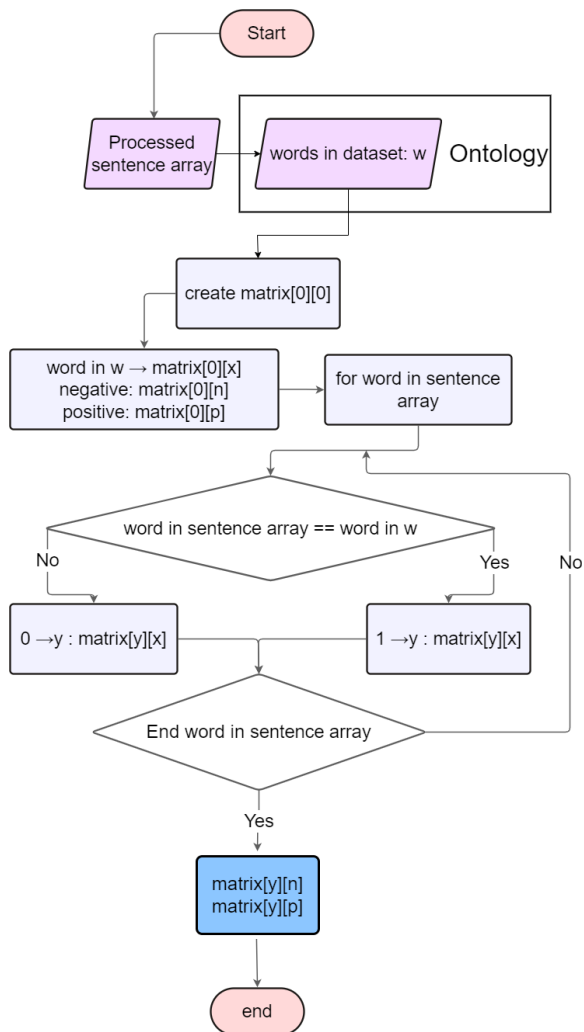


Fig. 5. Data vectorization algorithm

– **Second**, after the data is represented as a feature vector, the data will represent a characteristic or attribute, allowing the algorithm to analyze the data effectively. It helps convert unstructured or textual data into a numerical form to train models and extract information from the data in the future.

On the other hand, to do this, we calculate by matrix by:

- First, create matrix[0][0].
- Second, create two negative and positive matrices of the form [0][1] according to the words in the dataset.
- Third, match each word in the words dataset. If the word in the sentence matches the word, assign 1 to position y in the form [y][x] > [1][1]; otherwise, transfer 0.
- Finally, once the sentence is finished, it will stop and return two matrices.

Step 03: Our Training Model. In this step, we build a model to train the data. This is an important step to evaluate the quality of the proposed model compared to previously researched models. To implement this model, we first perform steps 1 and 2 above. After initial data preprocessing, the next steps are described in Figure 6 as follows:

- First, we use features and labels to train a model and predict based on the corresponding components.
- Second, we initialize the RNN/CNN model through two main libraries, TensorFlow and PyTorch, to learn and create predictions as well as extract the necessary information in order to classify "positive" or "negative" for the training model. Then, they practice with features and labels as objects.
- Third, we use the gradient descent algorithm to find out the standard weight value in order to determine an objective function that needs to be optimized. To do this, we initialize the model parameters with initial values. First, we use the gradient descent algorithm to repeat the following steps until the stopping condition is achieved. We calculate the objective function's gradient at the parameters' current points. The gradient represents the fastest increasing direction of the objective function. We update the model parameters based on the calculated gradient by moving in the opposite direction. Finally, we repeat the process of calculating the gradient and updating the parameters until a stopping condition is reached, such as reaching a sufficient number of iterations (epoch) or convergence of the objective function. Finally, we finish and return to the trained model.

Step 04: Prediction model. To build a model based on prediction and evaluate the results as shown in Figure 7, in this section, we conduct the experimental evaluation with the proposed model as follows:

- Sentence ← preprocess sentence.
- Vec ← vectorize sentence.
- Pred ← predict vec using model.
- Label ← get the label with the highest probability from pred.
- End the function and return the results.

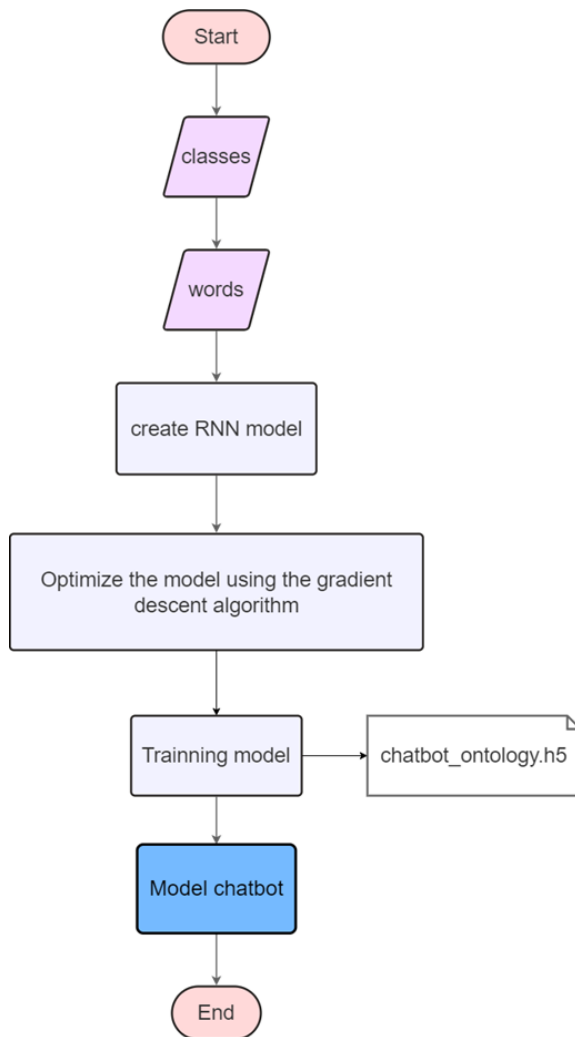


Fig. 6. Domain ontology training flowchart

Overall, in the prediction model building section, we have described it in detail in Figure 7. Based on this model, our results are better than the referenced methods.

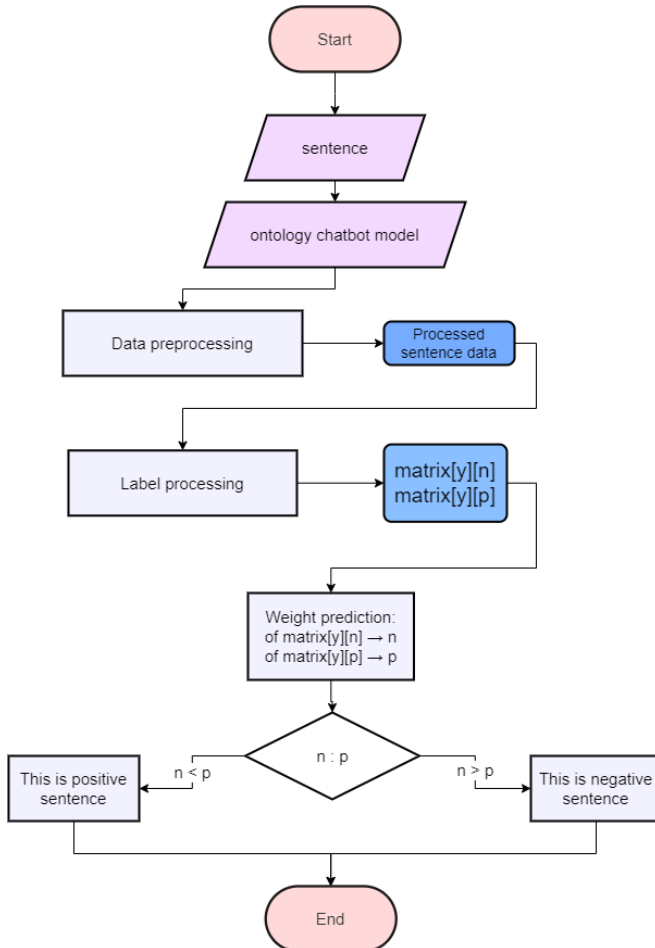


Fig. 7. Prediction model flowchart

4. Evaluation

4.1. Experimental settings. In this section, we will install and test on Python language and run on a 64-bit Windows 11 Pro computer with the following settings: Core i5-6300U (i5-6200U) / 16GB RAM / 512GB SSD / 12.5 inch HD 1366x768 screen. Furthermore, we use negative and positive comments in the dataset from [4].

On the one hand, to test the accuracy of the proposed method with human evaluation, we have created some accurate comments, as shown in Table 1.

Table 1. Some representative review comments between ITEAI (called ITE.), ACCLE (called ACC.), BCSAO (called BCS.), UOAEN (called UOA.) methods, and Human

No.	Some English comments on social networks	Evaluates				
		ITE.	ACC.	BCS.	UOA.	Human
1.	The replacement unit had a problem too: the little switch on the bottom wasn't working properly, making it impossible to select among city forecasts in my area.	t	f	t	t	Negative
2.	Acer has been awful to deal with, they treat their customers like rubbish, and have a miserable repair facility to match their products.	t	f	t	t	Negative
3.	However, since I am in an office, I normally keep them at a pretty low volume, and at low volumes, they sound a little weak.	t	f	f	t	Negative
4.	I am so tired of Netgear 's product quality & customer service that I would NEVER recommend Netgear products to anyone I know.	t	f	f	t	Negative
5.	The Netbook is great and rates 5 stars EXCEPT – Customer service is useless.	t	t	t	t	Negative
6.	I wished that they could make the volume a little bit louder.	t	t	t	t	Negative
7.	Amazing speed and is easy to set up.	f	t	t	t	Positive
8.	So far, I am super happy with this small yet powerful enough laptop.	f	t	t	t	Positive
9.	They are very sturdy, and have no trouble balance-wise, or handling oddly shaped speakers.	f	t	t	t	Positive
10.	This Pinnacle speaker bar is well worth the money.	f	t	t	t	Positive
11.	It is a wonderful little router.	f	t	t	t	Positive
12.	As a final comment, I am happy with the D-Link DIR-655 so far and my home wireless connections are smooth whether I'm using my ASUS netbook or Apple iPad.	f	t	t	t	Positive

We identify that *t* is true for humans, and *f* is false for humans. For example, comment line number 1 in Table 1 is "The replacement unit had a problem too: the little switch on the bottom wasn't working properly, making it impossible to select among city forecasts in my area". Humans evaluate this sentence as "Negative", meaning a negative sentence. ITEAI, BCSAO, and UOAEEN evaluate the following evaluation methods as "Negative" (*t*), while the ACCLE method is evaluated as "Positive" (*f*). Experiments show that our method is successful. The results show that the accuracy of the system we evaluated by manual evaluation has reached 100%, while the remaining methods achieved accuracy rates of ITEAI (67%), ACCLE (50%), and BCSAO (83%), respectively.

On the other hand, to evaluate the proposed method and other methods, in this section, we use three existing methods to assess our approach; the referenced methods include: a chatbot system building to analyze opinions of English opinions (called **UOAEEN**), a Chatbot for Changing Lifestyles in Education (called **ACCLE**), and an interactive Transport Inquiry with AI Chatbot (called **ITEAI**). The following is a summary of those reference methods:

– **BCSAO** [4]: This method is similar to our proposed method. However, the data processing is still done primarily through programming techniques, as opposed to our process, which goes deeper into the use of programming techniques. Creating an ontology is to classify sentences and evaluate in greater depth for each separate topic, then automatically separate based on individual models and optimize comment sentences.

– **ACCLE** [35]: The author proposes a Chatbot system to help teachers and students work together. The system is implemented by having students ask text-based questions to the Chatbot, processing the data using natural language processing and deep learning technology, and then responding to the student. However, this only serves schools without analyzing the respondents' emotions.

– **ITEAI** [36]: Like the ACCLE method, this method creates a Chatbot system that confirms the user's current location and final destination by asking questions. This method design examines the user's query and extracts the relevant entries from the database. It provides the receiver with complete information about the name and the number of the bus. By this, individuals can safely move to the desired location.

On the other hand, to evaluate the performance of the proposed method with the evaluation methods, we use the following formulas to calculate Accuracy and F1-score from [37].

Accuracy. In this research area, we use accuracy measurement to evaluate the performance of a prediction model. It calculates the ratio between

the number of correct predictions and the total amount of predicted data in formula 3. The formula for calculating accuracy is expressed as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3)$$

Besides, we realize that accuracy is sometimes considered as a good measurement for all situations. Especially in data imbalance problems, when one class of data accounts for the majority and another class accounts for only a tiny part, the accuracy can be limited. Therefore, we additionally calculate other measurements, including precision (called Pre) in formula 4, recall (called Rec) in formula 5, and F1-score in formula 6 using the parameters shown in Figure 8, with the following formula:

$$Pre = \frac{TP}{TP + FP}, \quad (4)$$

$$Rec = \frac{TP}{TP + FN}, \quad (5)$$

$$F1-score = 2 * \frac{Pre * Rec}{Pre + Rec}, \quad (6)$$

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.

4.2. Experimental results. Our system outperforms the three methods listed in Table 2. When considering 736 sentences, the results in Table 2 show that the proposed method always accounts for a higher percentage than the other methods. The ACCLE method yields low average results of 50.27%, while the ITEAI and BCSAO methods yield 76.22% and 78.53%, respectively. At 80.52%, the proposed method (UOAEN) achieved the lowest level.

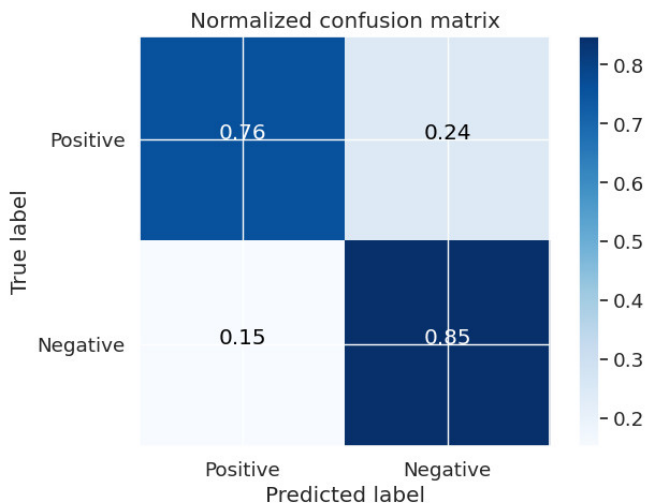


Fig. 8. Confusion matrix

Table 2. Performance of UOAEN and existing methods

Values	ACCLE	ITEAI	BCSAO	UOAEN
Accuracy	50.27%	76.22%	78.53%	80,52%
F1-score	0,82%	73.81%	76.02%	76,72%

On the one hand, we conducted experiments to measure the F1-score value of the algorithms, and the experiments showed that the proposed method achieved better results. Meanwhile, the ACCLE method has an F1-score value of at most 1% since this only focuses on responses based on available data without analyzing in-depth questions.

On the other hand, we create a system like Figure 9. Our program is analyzed through ontology to evaluate comments on social networks to evaluate positive and negative comments. The chatbot system will automatically analyze and automatically respond to each comment.

Overall, the method for exporting has better results than the referenced methods, with deep learning techniques and ontology-based analysis in Figures 10 and 11.

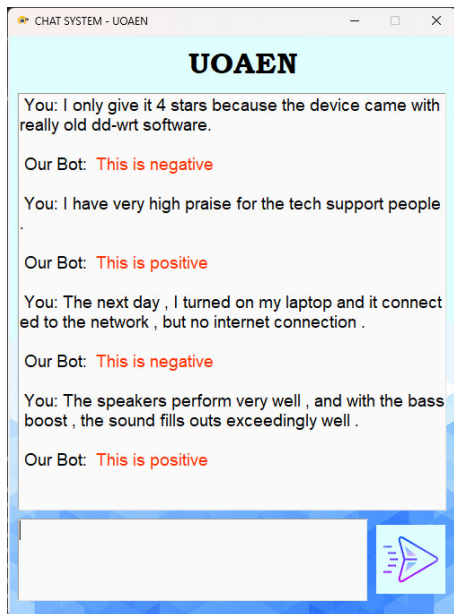


Fig. 9. GUI - UOAEN

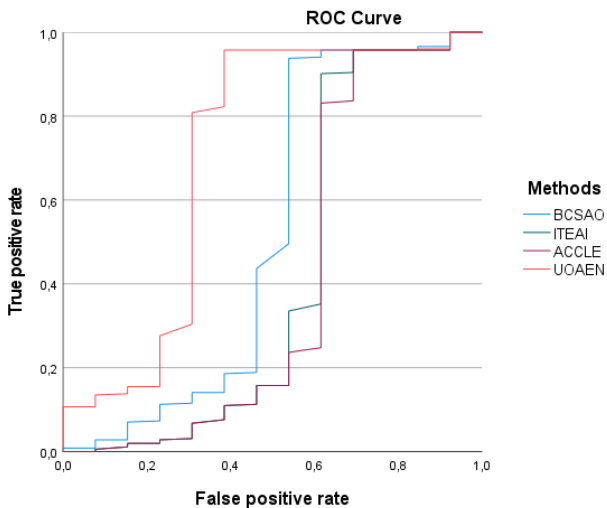


Fig. 10. Accuracy of UOAEN vs. the reference

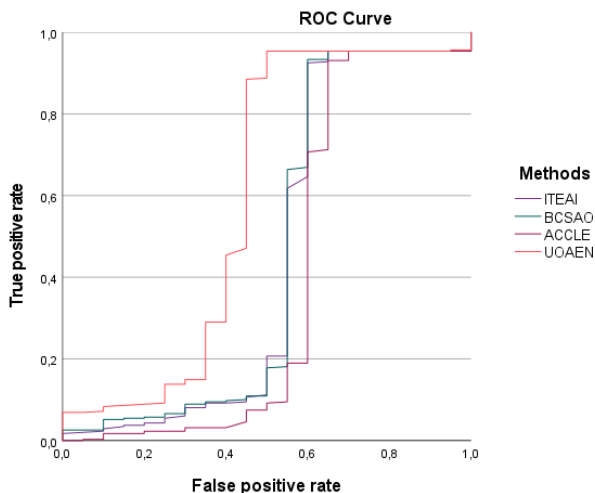


Fig. 11. F1-score of UOAEEN vs. the reference

5. Conclusions. In this paper, we have built a model for evaluating comments on social networks using ontology techniques to determine sentence types such as simple sentences, compound sentences, and complex sentences.

The article has expanded on exploiting user opinions on social networks through the Chatbot system. This technique interests many businesses, especially in Vietnam, when they deal with many customer comments while the number of people serving the company is limited. Furthermore, this new research contributes to significant performance gains compared to the referenced methods. The mentioned improvement shows that the efficiency is up to 80.52%. However, we also realize some existing problems in our studies, including:

- First, testing data is still a significant factor that needs to be expanded not only in English but in other languages also.
- Second, the country's culture also dramatically affects language as well as comment sentence use.

Therefore, we will overcome and supplement the above issues in the future.

References

1. Nasir M., Rehmat U., Ahmad I. Social media analysis of customer emotions in pizza industry. *The Computer Journal*. 2023. vol. 66. no. 7. pp. 1777–1783.
2. Rehmat U., Javed A., Nasir M., Bashir M. Sentimental analysis of beauty brands on social media. *Proc. GS Int. Conf. on Computer Science on Engineering 2020 (GSICCSE 2020)*.

2020. no. 37.
3. Chen Y., Wang L. Misleading political advertising fuels incivility online: A social network analysis of 2020 us presidential election campaign video comments on YouTube. *Computers in Human Behavior*. 2022. vol. 131(37). DOI: 10.1016/j.chb.2022.107202.
 4. Nguyen H., Tan N., Quan N., Huong T., Phat H. Building a chat-bot system to analyze opinions of english comments. *Informatics and Automation*. 2023. vol. 22. no. 2. pp. 289–315.
 5. Fidan M., Gencil N. Supporting the instructional videos with chat-bot and peer feedback mechanisms in online learning: The effects on learning performance and intrinsic motivation. *Journal of Educational Computing Research*. 2022. vol. 60. no. 7. pp. 1716–1741.
 6. Wu E., Lin C.-H., Ou Y.-Y., Liu C.-Z., Wang W.-K., Chao C.-Y. Advantages and constraints of a hybrid model k-12 e-learning assistant chatbot. *Ieee Access*. 2020. vol. 8. pp. 77788–77801.
 7. Surani D., Hamidah H. Students perceptions in online class learning during the covid-19 pandemic. *International Journal on Advanced Science, Education, and Religion*. 2020. vol. 3. no. 3. pp. 83–95.
 8. Suta P., Lan X., Wu B., Mongkolnam P., Chan J. An overview of machine learning in chatbots. *International Journal of Mechanical Engineering and Robotics Research*. 2020. vol. 9. no. 4. pp. 502–510.
 9. Adamopoulou E., Moussiades L. An overview of chatbot technology. *IFIP international conference on artificial intelligence applications and innovations*. Springer, 2020. pp. 373–383.
 10. Nguyen T., Ho D., Nguyen N. An ontology-based question answering system for university admissions advising. *Intelligent Automation and Soft Computing*. 2023. vol. 36. no. 1. pp. 601–616. DOI: 10.32604/iasc.2023.032080.
 11. Zahour O., Eddaoui A., Ouchra H., Hourrane O., Benlahmar E. A system for educational and vocational guidance in morocco: Chatbot e-orientation. *Procedia Computer Science*. 2020. vol. 175. pp. 554–559. DOI: j.procs.2020.07.079.
 12. Hallili A. Toward an ontology-based chatbot endowed with natural language processing and generation. *Processing of the 26th european summer school in logic, language and information*. 2014. 7 p.
 13. Avila C., Calixto A., Rolim T., Franco W., Venceslau A., Vidal V., Pequeno V., Moura F. Medibot: an ontology-based chatbot for Portuguese speakers drug’s users. *International Conference on Enterprise Information Systems*. 2019. vol. 1. pp. 25–36.
 14. Muangkammuen P., Intiruk N., Saikaew K. Automated thai-faq chatbot using rnn-lstm. in *2018 22nd International Computer Science and Engineering Conference (ICSEC)*. IEEE, 2018. pp. 1–4.
 15. Blanc C., Bailly A., Francis E., Guillotin T., Jamal F., Wakim B., Roy P. Flaubert vs. camembert: Understanding patient’s answers by a french medical chatbot. *Artificial Intelligence in Medicine*. 2022. vol. 127.
 16. Hung N., Loi T., Binh N., Nga N., Huong T., Luu D. Building an online learning model through a dance recognition video based on deep learning. *Informatics and Automation*. 2024. vol. 23. no. 1. pp. 101–128.
 17. Dhyani M., Kumar R. An intelligent chatbot using deep learning with bidirectional rnn and attention model. *Materials today: proceedings*. 2021. vol. 34. pp. 817–824.
 18. Hung N., Loi T., Huong N., Hang T., Huong T. Aafndl-an accurate fake information recognition model using deep learning for the vietnamese language. *Informatics and Automation*. 2023. vol. 22. no. 4. pp. 795–825.

19. Suganthi S., Ayoobkhan M., Kumar K., Bacanin N., Venkatachalam K., Hubalovsky S., Trojovsky P. Deep learning model for deep fake face recognition and detection. *PeerJ Computer Science*. 2022. vol. 8. DOI: 10.7717/peerj-cs.881.
20. Rakib A., Rumky E., Ashraf A., Hillas M., Rahman M. Mental healthcare chatbot using sequence-to-sequence learning and bilstm. *International Conference on Brain Informatics*. Springer, 2021. pp. 378–387.
21. Palasundram K., Sharef N., Nasharuddin N., Kasmiran K., Azman A. Sequence to sequence model performance for education chatbot. *International journal of emerging Technologies in Learning (iJET)*. 2019. vol. 14. no. 24. pp. 56–68.
22. Piccini R., Spanakis G. Exploring the context of recurrent neural network based conversational agents. *arXiv preprint arXiv:1901.11462*. 2019.
23. Sun R., Chen B., Zhou Q., Li Y., Cao Y., Zheng H.-T. A non-hierarchical attention network with modality dropout for textual response generation in multimodal dialogue systems. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022. pp. 6582–6586.
24. Karna M., Juliet D., Joy R. Deep learning-based text emotion recognition for chatbot applications. *4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184)*. IEEE, 2020. pp. 988–993.
25. Sperli G. A cultural heritage framework using a deep learning-based chatbot for supporting tourist journey. *Expert Systems with Applications*. 2021. vol. 183. DOI: 10.1016/j.eswa.2021.115277.
26. Hitzler P., Krotzsch M., Parsia B., Patel-Schneider P., Rudolph S., et al. Owl 2 web ontology language primer. *W3C recommendation*. 2009. vol. 27. no. 1.
27. Antoniou G., Groth P., Van Harmelen F., Hoekstra R. *A semantic web primer 3rd ed.* The MIT Press, 2012. 288 p.
28. Casillo M., De Santo M., Mosca R., Santaniello D. An ontology-based chatbot to enhance experiential learning in a cultural heritage scenario. *Frontiers in Artificial Intelligence*. 2022. vol. 5.
29. Chi Y.-L., Sung H.-Y. Building the knowledge base of folk beliefs based on semantic web technology. *International Conference on Human-Computer Interaction*. Springer Nature Switzerland, 2023. pp. 172–182.
30. Owl web ontology language. Available: <https://www.w3.org/TR/owl-features/> (accessed: 2023-08-15).
31. Aceta C., Fernandez I., Soroa A. *Todo: A core ontology for task-oriented dialogue systems in industry 4.0. Further with Knowledge Graphs*. IOS Press, 2021. pp. 1–15.
32. Bouziane A., Bouchiha D., Doumi N., Malki M. Question answering systems: survey and trends. *Procedia Computer Science*. 2015. vol. 73. pp. 366–375.
33. Budler L.C., Gosak L., Stiglic G. Review of artificial intelligence-based question-answering systems in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2023. vol. 13. no. 2.
34. Majid M., Hayat M., Khan F., Ahmad M., Jhanji N., Bhuiyan M., Masud M., AlZain M. Ontology-based system for educational program counseling. *Intelligent Automation and Soft Computing*. 2021. vol. 30. no. 1. pp. 373–386.
35. Kasthuri E., Balaji S. A chatbot for changing lifestyle in education. *Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*. IEEE, 2021. pp. 1317–1322.
36. Dharani M., Jyostna J., Sucharitha E., Likitha R., Manne S. Interactive transport enquiry with ai chatbot. *Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. 2020. pp. 1271–1276.

37. Nguyen H., Dao T., Pham N., Dang T., Nguyen T., Truong T. 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.

Nguyen Viet Hung — Ph.D., Lecturer, Faculty of information technology, East Asia University of 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 — 18. hungnv@eaut.edu.vn; Ky Phu – Ky Anh, Ha Tinh, Viet Nam; office phone: +84(098)911-2079.

Nguyen Tan — Research assistant, East Asia University of Technology. Research interests: applications, data analysis. The number of publications — 2. tan25102000@gmail.com; Trung Dung – Tien Lu, Hung Yen, Viet Nam; office phone: +84(035)919-0216.

Nguyen Thi Thuy Nga — Lecturer, Faculty of information technology, East Asia University of Technology. Research interests: multimedia communications, database management systems, artificial intelligence, applications. The number of publications — 2. ngantt@eaut.edu.vn; Ta Thanh Oai – Thanh Tri, Hanoi, Viet Nam; office phone: +84(097)529-4770.

Le Thi Huyen Trang — 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. tranglth@eaut.edu.vn; Phuong Tri – Thi Tran Phung – Dan Phuong, Hanoi, Viet Nam; office phone: +84(032)8899-334.

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 — 2. hang42c@gmail.com; Mao Khe - Dong Trieu, Quang Ninh, Viet Nam; office phone: +84(090)496-8545.

Н.В. ХУНГ, Н. ТАН, Н.Т.Т. НГА, Л.Х. ТРАНГ, Т.Т. ХАНГ
**ИСПОЛЬЗОВАНИЕ ОНТОЛОГИИ ДЛЯ АНАЛИЗА
АНГЛИЙСКИХ КОММЕНТАРИЕВ В СОЦИАЛЬНЫХ СЕТЯХ**

Вьет Хунг Н., Тан Н., Тхи Туй Нга Н., Хуен Транг Л., Туй Ханг Т. **Использование онтологии для анализа английских комментариев в социальных сетях.**

Аннотация. Чат-боты заинтересовывают многих пользователей по мере того, как технологии становятся все более продвинутыми. Потребность в обмене информацией между людьми через компьютерные системы увеличивается с каждым днем, в результате чего в большинстве стран растет предпочтение использовать чат-боты. Поскольку Вьетнам является развивающейся страной с множеством этнических групп, требуется усиленное внимание к распространению социальных сетей и расширению кооперативной экономики. Серьезной проблемой стало неуместное использование слов в повседневной жизни. В социальных сетях встречаются неоднозначные отзывы с похвалой и критикой о том, что мы пытаемся уменьшить использование негативной лексики и улучшить качество использования языка в социальных сетях. Мы стремимся удовлетворить потребности пользователей в социальных сетях, способствовать экономическому развитию и более эффективно решать социальные проблемы. Для достижения этих целей предлагается метод глубокого обучения, использующий интеллектуальный анализ онтологических знаний для сбора и обработки комментариев в социальных сетях. Этот подход направлен на улучшение пользовательского опыта и облегчение обмена информацией между людьми путем анализа мнений в комментариях. Результаты экспериментов показывают, что наш метод превосходит традиционный подход.

Ключевые слова: чат-бот, онтология, глубокое обучение, машинное обучение, социальная сеть, Вьетнам.

Литература

1. Nasir M., Rehmat U., Ahmad I. Social media analysis of customer emotions in pizza industry. *The Computer Journal*. 2023. vol. 66. no. 7. pp. 1777–1783.
2. Rehmat U., Javed A., Nasir M., Bashir M. Sentimental analysis of beauty brands on social media. *Proc. GS Int. Conf. on Computer Science on Engineering 2020 (GSICSE 2020)*. 2020. no. 37.
3. Chen Y., Wang L. Misleading political advertising fuels incivility online: A social network analysis of 2020 us presidential election campaign video comments on YouTube. *Computers in Human Behavior*. 2022. vol. 131(37). DOI: 10.1016/j.chb.2022.107202.
4. Nguyen H., Tan N., Quan N., Huong T., Phat H. Building a chat-bot system to analyze opinions of english comments. *Informatics and Automation*. 2023. vol. 22. no. 2. pp. 289–315.
5. Fidan M., Gencel N. Supporting the instructional videos with chat-bot and peer feedback mechanisms in online learning: The effects on learning performance and intrinsic motivation. *Journal of Educational Computing Research*. 2022. vol. 60. no. 7. pp. 1716–1741.
6. Wu E., Lin C.-H., Ou Y.-Y., Liu C.-Z., Wang W.-K., Chao C.-Y. Advantages and constraints of a hybrid model k-12 e-learning assistant chatbot. *Ieee Access*. 2020. vol. 8. pp. 77788–77801.

7. Surani D., Hamidah H. Students perceptions in online class learning during the covid-19 pandemic. *International Journal on Advanced Science, Education, and Religion*. 2020. vol. 3. no. 3. pp. 83–95.
8. Suta P., Lan X., Wu B., Mongkolnam P., Chan J. An overview of machine learning in chatbots. *International Journal of Mechanical Engineering and Robotics Research*. 2020. vol. 9. no. 4. pp. 502–510.
9. Adamopoulou E., Moussiades L. An overview of chatbot technology. *IFIP international conference on artificial intelligence applications and innovations*. Springer, 2020. pp. 373–383.
10. Nguyen T., Ho D., Nguyen N. An ontology-based question answering system for university admissions advising. *Intelligent Automation and Soft Computing*. 2023. vol. 36. no. 1. pp. 601–616. DOI: 10.32604/iasc.2023.032080.
11. Zahour O., Eddaoui A., Ouchra H., Hourrane O., Benlahmar E. A system for educational and vocational guidance in morocco: Chatbot e-orientation. *Procedia Computer Science*. 2020. vol. 175. pp. 554–559. DOI: j.procs.2020.07.079.
12. Hallili A. Toward an ontology-based chatbot endowed with natural language processing and generation. *Processing of the 26th european summer school in logic, language and information*. 2014. 7 p.
13. Avila C., Calixto A., Rolim T., Franco W., Venceslau A., Vidal V., Pequeno V., Moura F. Medibot: an ontology-based chatbot for Portuguese speakers drug’s users. *International Conference on Enterprise Information Systems*. 2019. vol. 1. pp. 25–36.
14. Muangkammuen P., Intiruk N., Saikaew K. Automated thai-faq chatbot using rnn-lstm. in *2018 22nd International Computer Science and Engineering Conference (ICSEC)*. IEEE, 2018. pp. 1–4.
15. Blanc C., Bailly A., Francis E., Guillotin T., Jamal F., Wakim B., Roy P. Flaubert vs. camembert: Understanding patient’s answers by a french medical chatbot. *Artificial Intelligence in Medicine*. 2022. vol. 127.
16. Hung N., Loi T., Binh N., Nga N., Huong T., Luu D. Building an online learning model through a dance recognition video based on deep learning. *Informatics and Automation*. 2024. vol. 23. no. 1. pp. 101–128.
17. Dhyani M., Kumar R. An intelligent chatbot using deep learning with bidirectional rnn and attention model. *Materials today: proceedings*. 2021. vol. 34. pp. 817–824.
18. Hung N., Loi T., Huong N., Hang T., Huong T. Aafndl-an accurate fake information recognition model using deep learning for the vietnamese language. *Informatics and Automation*. 2023. vol. 22. no. 4. pp. 795–825.
19. Suganthi S., Ayoobkhan M., Kumar K., Bacanin N., Venkatachalam K., Hubalovsky S., Trojovsky P. Deep learning model for deep fake face recognition and detection. *PeerJ Computer Science*. 2022. vol. 8. DOI: 10.7717/peerj-cs.881.
20. Rakib A., Rumky E., Ashraf A., Hillas M., Rahman M. Mental healthcare chatbot using sequence-to-sequence learning and bilstm. *International Conference on Brain Informatics*. Springer, 2021. pp. 378–387.
21. Palasundram K., Sharef N., Nasharuddin N., Kasmiran K., Azman A. Sequence to sequence model performance for education chatbot. *International journal of emerging Technologies in Learning (iJET)*. 2019. vol. 14. no. 24. pp. 56–68.
22. Piccini R., Spanakis G. Exploring the context of recurrent neural network based conversational agents. *arXiv preprint arXiv:1901.11462*. 2019.
23. Sun R., Chen B., Zhou Q., Li Y., Cao Y., Zheng H.-T. A non-hierarchical attention network with modality dropout for textual response generation in multimodal dialogue systems. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022. pp. 6582–6586.

24. Karna M., Juliet D., Joy R. Deep learning-based text emotion recognition for chatbot applications. 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184). IEEE, 2020. pp. 988–993.
25. Sperli G. A cultural heritage framework using a deep learning-based chatbot for supporting tourist journey. Expert Systems with Applications. 2021. vol. 183. DOI: 10.1016/j.eswa.2021.115277.
26. Hitzler P., Krotzsch M., Parsia B., Patel-Schneider P., Rudolph S., et al. Owl 2 web ontology language primer. W3C recommendation. 2009. vol. 27. no. 1.
27. Antoniou G., Groth P., Van Harmelen F., Hoekstra R. A semantic web primer 3rd ed. The MIT Press, 2012. 288 p.
28. Casillo M., De Santo M., Mosca R., Santaniello D. An ontology-based chatbot to enhance experiential learning in a cultural heritage scenario. Frontiers in Artificial Intelligence. 2022. vol. 5.
29. Chi Y.-L., Sung H.-Y. Building the knowledge base of folk beliefs based on semantic web technology. International Conference on Human-Computer Interaction. Springer Nature Switzerland, 2023. pp. 172–182.
30. Owl web ontology language. Available: <https://www.w3.org/TR/owl-features/> (accessed: 2023-08-15).
31. Aceta C., Fernandez I., Soroa A. Todo: A core ontology for task-oriented dialogue systems in industry 4.0. Further with Knowledge Graphs. IOS Press, 2021. pp. 1–15.
32. Bouziane A., Bouchiha D., Doumi N., Malki M. Question answering systems: survey and trends. Procedia Computer Science. 2015. vol. 73. pp. 366–375.
33. Budler L.C., Gosak L., Stiglic G. Review of artificial intelligence-based question-answering systems in healthcare. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2023. vol. 13. no. 2.
34. Majid M., Hayat M., Khan F., Ahmad M., Jhanjhi N., Bhuiyan M., Masud M., AlZain M. Ontology-based system for educational program counseling. Intelligent Automation and Soft Computing. 2021. vol. 30. no. 1. pp. 373–386.
35. Kasthuri E., Balaji S. A chatbot for changing lifestyle in education. Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, 2021. pp. 1317–1322.
36. Dharani M., Jyostna J., Sucharitha E., Likitha R., Manne S. Interactive transport enquiry with ai chatbot. Proceedings of the 4th International Conference on Intelligent Computing and Control Systems (ICICCS). 2020. pp. 1271–1276.
37. Nguyen H., Dao T., Pham N., Dang T., Nguyen T., Truong T. 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.

Вьет Хунг Нгуен — Ph.D., преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, сетевая безопасность, искусственный интеллект, управление трафиком в сетях нового поколения, гарантия QoE/QoS для сетевых сервисов, экологически чистые сети и приложения. Число научных публикаций — 18. hungnv@caut.edu.vn; Ки Фу – Ки Ань, Хатинь, Вьетнам; p.t.: +84(098)911-2079.

Тан Нгуен — научный сотрудник, Восточноазиатский технологический университет. Область научных интересов: прикладные программы, анализ данных. Число научных публикаций — 2. tan25102000@gmail.com; Чунг Зунг – Тьен Лу, Хынгйен, Вьетнам; p.t.: +84(035)919-0216.

Тхи Туй Нга Нгуен — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов:

мультимедийные коммуникации, системы управления базами данных, искусственный интеллект, приложения. Число научных публикаций — 2. ngantt@eaut.edu.vn; Та Тхань Оай – Тханьчи, Ханой, Вьетнам; р.т.: +84(097)529-4770.

Хуен Транг Ле Тхи — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, системы управления базами данных, искусственный интеллект, приложения. Число научных публикаций — 1. trangleth@eaut.edu.vn; Фуонг Три – Тхи Чан Пхунг – Дан Фуонг, Ханой, Вьетнам; р.т.: +84(032)8899-334.

Туй Ханг Тран Тхи — преподаватель, факультет информационных технологий, Восточноазиатский технологический университет. Область научных интересов: мультимедийные коммуникации, системы управления базами данных, искусственный интеллект, приложения. Число научных публикаций — 2. hang42c@gmail.com; Мао Кхе - Донг Чиу, Куанг Нин, Вьетнам; р.т.: +84(090)496-8545.

А.А. ПОТЮПКИН, С.В. ПИЛЬКЕВИЧ, В.В. ЗАЙЦЕВ
**ПОДХОД К АПРИОРНОМУ ОЦЕНИВАНИЮ НЕЧЕТКИХ
КЛАССИФИКАЦИОННЫХ МОДЕЛЕЙ В ЗАДАЧАХ
МОНИТОРИНГА**

Потюпкин А.А., Пилькевич С.В., Зайцев В.В. Подход к априорному оцениванию нечетких классификационных моделей в задачах мониторинга.

Аннотация. Статья посвящена проблемам применения средств автоматизации при решении задач мониторинга и управления в части, касающейся оценки качества нечетких классификационных моделей, для которых порядок классификации реализуется на основании знаний (правил) в условиях отсутствия обучающей выборки. Предложен подход к получению априорных оценок качества классификации на основе исследования чувствительности используемой модели к изменению значений внутренних параметров в ходе соответствующего моделирования. Получена интерпретация результатов моделирования в виде оценки рисков, вызванных несовершенством самих классификационных моделей. В статье приведен пример нечеткой классификационной модели, в основе которой лежит сравнение текущего состояния объекта мониторинга, описываемого с использованием нечетких признаков, с набором заранее заданных типовых состояний, формирующих соответствующие нечетко равные между собой (близкие) состояния (ситуации мониторинга). Сравнение осуществляется с использованием операции нечеткой импликации при условии соблюдения требуемой достоверности. На примере данной модели показано, каким образом вид операции импликации, а также внутренние особенности модели влияют на результаты классификации, предложены соответствующие показатели, которые являются как интерпретацией общепринятых показателей оценки качества классификации, так и уникальными, присущими именно рассмотренной модели. Проведены вычислительные эксперименты, которые позволили получить графики изменения показателей оценки качества классификации для рассматриваемой модели и ее модификации, наглядно отобразить влияние внутренних параметров модели на результаты ее применения. Предложены ряд показателей, позволяющих провести априорную оценку рисков, возникающих вследствие применения модели, до ее фактического применения.

Ключевые слова: нечеткие множества, классификация, оценка качества, риск.

1. Введение. Применение средств автоматизации является важным аспектом успешного решения задач мониторинга и управления в условиях роста количества обрабатываемой информации, увеличения числа контролируемых параметров, динамичности изменения состояния объекта наблюдения и условий его функционирования. Одной из задач, возникающих в ходе сбора и обработки информации, является задача классификации состояний объекта наблюдения (мониторинга). Под классификацией понимается отнесение рассматриваемого состояния объекта к одному из заранее сформированных классов состояний по заданным правилам [1]. Использование результатов классификации при решении задач

мониторинга приводит к тому, что достигаемое качество работы классификатора оказывает непосредственное влияние на достоверность принимаемых решений, их адекватность складывающейся обстановке. В связи с этим применение классификаторов в схемах принятия решений предполагает заинтересованность соответствующих специалистов и экспертов в обладании априорной информацией о качестве их функционирования. Такая информация позволяет провести оценивание последствий применения классификатора и рисков различного рода, источником которых является несовершенство классификаторов. В целом оценка влияния используемых средств автоматизации на качество принимаемых решений является важной исследовательской задачей. В работе [1] приводится два подхода к осуществлению автоматизированной классификации:

1. Классификация на основе обучения. Предполагает наличие обучающей выборки и соответствующего алгоритма обучения. Правила, по которым осуществляется классификация, формируются в ходе обучения в соответствии с используемым алгоритмом.

2. Классификация на основе знаний (правил) экспертов. Данный подход предполагает, что правила классификации формируются экспертами вручную на основе имеющихся знаний и информации об особенностях объекта наблюдения, при этом обучающая выборка отсутствует.

В первом случае такая информация может быть получена путем анализа соответствующих количественных показателей качества классификатора, полученных по результатам его обучения и тестирования: полнота, точность, F-мера и т.д. [2, 3]. Во втором случае эксперты не располагают результатами обучения, и именно о таких моделях в дальнейшем будет говориться в статье при рассмотрении классификационных моделей. Априорно информация о качестве работы такого классификатора может быть получена либо путем интерпретации представлений самих экспертов об адекватности используемой модели классификации объекту мониторинга, его достоинствах и недостатках, либо путем формирования экспертами размеченной тестовой выборки для оценки качества классификации по аналогии с обучаемыми классификаторами. Получение такой выборки является наиболее трудоемкой задачей и не всегда представляется возможным.

Особую актуальность данное обстоятельство приобретает в сфере поддержки принятия решений в условиях, когда моделирование поведения объекта мониторинга осуществляется с использованием

конструкций естественного языка, а для формализованного описания его состояния применяются методы из теории нечетких множеств, в частности, описание при помощи лингвистических переменных [4 – 8]. В этом случае принадлежность определенного состояния объекта к одному из классов состояний оценивается при помощи соответствующих нечетких классификационных моделей, результат применения которых интерпретируется с использованием категорий достоверности и нечеткости (размытости). Вследствие этого получение точных и объективных оценок качества классификации для такого рода нечетких моделей представляется весьма затруднительным.

В этой связи предлагается рассматривать результат исследования внутренних закономерностей нечетких классификационных моделей, формируемых на основе экспертных правил, а именно оценку чувствительности нечеткой модели к изменению внутренних параметров, как априорную информацию о качестве классификации, которая может быть дополнительно интерпретирована в интересах экспертов. В статье предлагается на примере нечеткой классификационной модели и ее модификации рассмотреть подход к получению таких оценок и их интерпретации в виде оценок рисков различного рода, а также провести соответствующие вычислительные эксперименты.

2. Описание модели. В основе рассматриваемого в статье классификатора лежит модель, подробно описанная в работе [8], а также в работах [9 – 12]. Предполагается, что состояние объекта мониторинга описывается с некоторого множества признаков $Z = \{z_i, i = 1 \div N\}$, N – общее число признаков, значения которых представляют собой вербальные величины, т.е. конструкции естественного языка. Каждый признак формализованно представлен соответствующей лингвистической переменной $L_i = \langle z_i, T_i, D_i, \Omega_i, \Psi_i \rangle$, где $T_i = \{T_{ij}, j = 1 \div M_i\}$ – терм-множество переменной, M_i – мощность терм-множества T_i , D_i – базовое (опорное) множество значений признака z_i , Ω_i – синтаксические правила, порождающие названия термов для z_i , Ψ_i – семантические правила, задающие функции принадлежности нечетких термов, порожденных правилами Ω_i [5]. Для описания термов T_{ij} используются нечеткие переменные

$\tilde{L}_{ij} = \langle T_{ij}, D_i, \tilde{C}_{ij} \rangle$, где $\tilde{C}_{ij} = \left\{ \mu_{C_{ij}}(x) / x, x \in D_i \right\}$ – нечеткое подмножество D_i , x – элемент базового множества.

Рассмотрим пример использования такой модели. Пусть решается задача выработки ситуационным центром Правительства Российской Федерации (РФ) рекомендаций высшему руководству страны по действиям в условиях возникновения социально-политических кризисов в соседних с РФ регионах. Подробно история создания, особенности функционирования и перспективы развития системы ситуационных центров различного уровня в РФ рассмотрены в работах [13 – 15].

В качестве объекта мониторинга выступает социально-политическая обстановка, а для описания ее состояния используется два признака: z_1 – удаленность кризисного региона от границ РФ (км) и z_2 – продолжительность кризиса по времени (сут.), которым соответствуют лингвистические переменные $L_1 = \langle \text{Удаленность}, T_1, X \rangle$, где $T_1 = \{ \text{Неприемлемая}, \text{Приемлемая} \}$, $X = [0; 1000]$ и $L_2 = \langle \text{Продолжительность}, T_2, Y \rangle$, где $T_2 = \{ \text{Малая}, \text{Большая} \}$, $Y = [0; 30]$. Лингвистические переменные представлены в сокращенном виде без описания соответствующих параметров Ω_i, Ψ_i , т.к. это не влияет на особенности рассматриваемого в статье подхода. Количество признаков в данном примере обусловлено удобством визуального отображения результатов вычислительных экспериментов.

Каждому классу состояний обстановки сопоставляется сценарий соответствующих мер и действий, принимаемых руководством страны: перевод силовых структур на усиленный режим работы, проведение внеплановых сборов и тренировок, уточнение и корректировка планов по реагированию на кризисные ситуации, усиление миграционного режима и т.д. По итогу классификации Ситуационный центр выдает рекомендации в виде сценария предполагаемых действий, соответствующих классу состояния обстановки. В целом, актуальность рассматриваемой в примере задачи и необходимость выработки соответствующих мер реагирования наглядно показаны в анализе современных социально-политических кризисов, проведенном в работе [16].

В рамках рассматриваемого примера для формализованного описания лингвистических переменных будем использовать сигмоидальные функции принадлежности вида:

$$\mu_{T_{11}}(x) = \frac{1}{1 + e^{0,01(x-600)}}; \mu_{T_{12}}(x) = \frac{1}{1 + e^{-0,01(x-600)}}, \quad (1)$$

где T_{11} – *Неприемлемая*, T_{12} – *Приемлемая* – термы лингвистической переменной L_1 и

$$\mu_{T_{21}}(y) = \frac{1}{1 + e^{0,5(y-14)}}; \mu_{T_{22}}(y) = \frac{1}{1 + e^{-0,5(y-14)}}, \quad (2)$$

где T_{21} – *Малая*, T_{22} – *Большая* – термы лингвистической переменной L_2 . Графики функций принадлежности (1) и (2) представлены на рисунке 1.

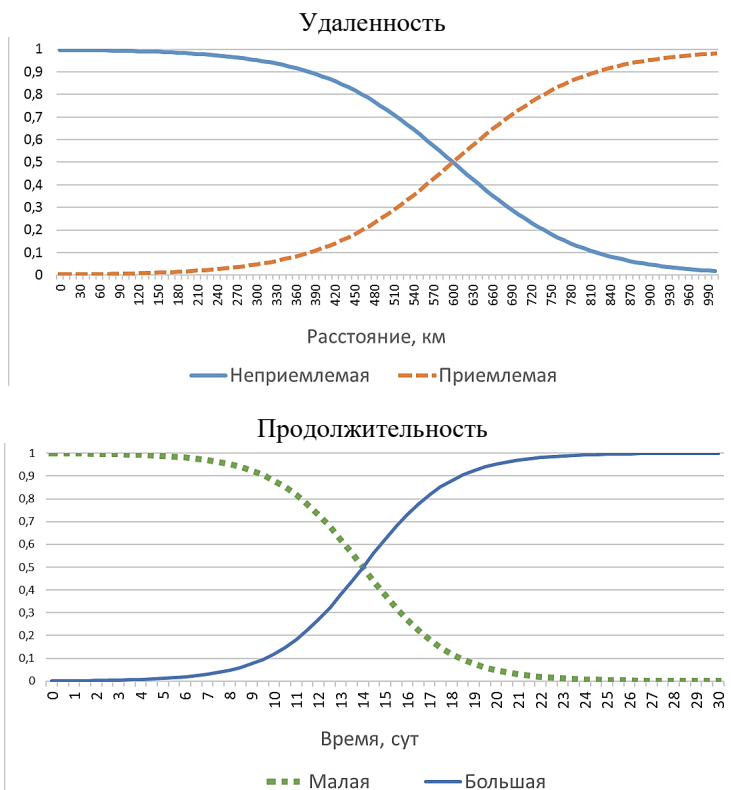


Рис. 1. Вид графиков функций принадлежности для L_1 и L_2

Значения $\mu_{T_{11}}$, $\mu_{T_{12}}$ и $\mu_{T_{21}}$, $\mu_{T_{22}}$ определяют достоверность нечетких высказываний о том, что лингвистические переменные L_1 и L_2 приняли те или иные значения из соответствующих термножеств T_1 и T_2 . Функции принадлежности $\mu_{T_{11}}(x)$ и $\mu_{T_{21}}(y)$ относятся к Z -образным, а $\mu_{T_{12}}(x)$ и $\mu_{T_{22}}(y)$ – к S -образным функциям, которые используются для формализованного описания нечетких признаков с низкой (Z -образные) и высокой (S -образные) степенью проявления того или иного свойства [5], что соответствует характеру используемых в примере признаков. Кроме того, функции (1-2) отвечают требованиям, предъявляемым к функциям принадлежности, рассмотренным в работе [8], в связи с чем их использование в данном примере обоснованно.

В рамках модели состояние объекта мониторинга представляется в виде нечеткого множества:

$$\tilde{S} = \{ \mu_s(z_i) / z_i, z_i \in Z \}, \quad (3)$$

где значение $\mu_s(z_i)$ определяется как:

$$\mu_s(z_i) = \left\{ \mu_{\mu_s(z_i)}(T_{ij}) / T_{ij}, i = 1 \div N, j = 1 \div M_i \right\}. \quad (4)$$

Нечеткое множество \tilde{S} , представленное в выражении (3), будем называть «нечеткой ситуацией мониторинга» (ситуацией мониторинга) [7, 8]. Примером такой ситуации в рамках рассматриваемой задачи мониторинга состояния социально-политической обстановки может быть нечеткое множество

$$\tilde{S} = (\langle 0,7 / \text{Неприемлемая}; 0,3 / \text{Приемлемая} \rangle / \text{"Удаленность"}, \langle 0,9 / \text{Малая}; 0,1 / \text{Большая} \rangle / \text{"Продолжительность"}).$$

Каждый класс состояний объекта мониторинга $G_k \in G$, где $G = \{G_k, k = 1 \div N_G\}$ – множество заданных классов состояний общим количеством N_G , представляется в виде совокупности схожих между собой ситуаций и формализованно описывается в виде нечеткой

ситуации-эталона $\tilde{\gamma}_k$, типовой для данного класса, которой, в свою очередь, нечетко равны все ситуации, входящие в этот класс. Для обозначения множества всех ситуаций-эталонов будем использовать $\Gamma = \{\tilde{\gamma}_k, k = 1 \div N_G\}$.

Произвольные ситуации \tilde{S} и $\tilde{\gamma}$ считаются нечетко равными между собой, если степень нечеткого равенства $\mu(\tilde{S}, \tilde{\gamma})$ превосходит заданное пороговое значение достоверности $\mu_{\Pi\Gamma}$. Выбор конкретных значений $\mu_{\Pi\Gamma}$ зависит от особенностей объекта мониторинга, требований к достоверности принимаемых решений и т.д., обычно это значения из диапазона $[0,6; 1]$. Степень нечеткого равенства $\mu(\tilde{S}, \tilde{\gamma})$ определяется поэтапно по каждому из признаков $z_i \in Z$ (3) в соответствии с выражением:

$$\mu(\tilde{S}, \tilde{\gamma}) = \min_{z_i \in Z} \mu(\mu_s(z_i), \mu_\gamma(z_i)), \quad (5)$$

где $\mu(\mu_s(z_i), \mu_\gamma(z_i))$ – степень нечеткого равенства по z_i признаку, определяемая через степень нечеткого равенства соответствующих нечетких переменных (4):

$$\mu(\mu_s(z_i), \mu_\gamma(z_i)) = \min_{T_{ij} \in T_i} C(\mu_{\mu_s(z_i)}(T_{ij}), \mu_{\mu_\gamma(z_i)}(T_{ij})), \quad (6)$$

где $C(\mu_{\mu_s(z_i)}(T_{ij}), \mu_{\mu_\gamma(z_i)}(T_{ij}))$ рассчитывается из следующего условия:

$$C(\mu_{\mu_s(z_i)}(T_{ij}), \mu_{\mu_\gamma(z_i)}(T_{ij})) = \begin{cases} \mu_{\mu_s(z_i)}(T_{ij}) \leftrightarrow \mu_{\mu_\gamma(z_i)}(T_{ij}), f = 1 \\ 1, f = 0 \end{cases}. \quad (7)$$

Обозначим как $I_{\Pi\Gamma} = (1 - \mu_{\Pi\Gamma}; \mu_{\Pi\Gamma})$ – пороговый интервал сравнения. Тогда параметр f рассчитывается, исходя из условия:

$$f = \begin{cases} 1, \text{ если } \mu_{\mu_s(z_i)}(T_{ij}), \mu_{\mu_\gamma(z_i)}(T_{ij}) \notin I_{\Pi\Gamma} \\ 0, \text{ если } \mu_{\mu_s(z_i)}(T_{ij}) \in I_{\Pi\Gamma} \vee \mu_{\mu_\gamma(z_i)}(T_{ij}) \in I_{\Pi\Gamma} \end{cases}. \quad (8)$$

Необходимость условия (8) в работе [8] определяется следующим образом. При сравнении между собой текущей ситуации и ситуации-эталона принято условие, что если достоверность значений нечетких признаков $\mu_{\mu_s(z_i)}(T_{ij})$ или $\mu_{\mu_\gamma(z_i)}(T_{ij})$ попадают в интервал I_{III} , такие признаки считаются «плохо выраженными», «размытыми», поэтому они из общего сравнения исключаются и не берутся в расчет. В связи с тем, что в выражениях (6) и (7) используется операция минимума, то результатам сравнения таких пар, где $\mu_{\mu_s(z_i)}(T_{ij})$ или $\mu_{\mu_\gamma(z_i)}(T_{ij})$ находятся в этом интервале, присваивается значение 1, которое не влияет на итоговый результат. Иными словами, при определении степени нечеткого равенства ситуаций все «плохо выраженные» признаки исключаются из рассмотрения.

Обозначим для краткости записи $\mu_{\mu_s(z_i)}(T_{ij})$ и $\mu_{\mu_\gamma(z_i)}(T_{ij})$ как μ_s и μ_γ соответственно. Тогда в выражении (7) операция $\mu_s \leftrightarrow \mu_\gamma = \min(\mu_s \rightarrow \mu_\gamma, \mu_\gamma \rightarrow \mu_s)$, где « \rightarrow » – операция нечеткой импликации. В рассматриваемой модели [8 – 11] в качестве операции импликации используется нечеткая импликация по Гёделлю $\mu_s \rightarrow \mu_\gamma = \max(1 - \mu_s, \mu_\gamma)$. Тогда имеет место выражение:

$$\mu_s \leftrightarrow \mu_\gamma = \min(\max(1 - \mu_s, \mu_\gamma), \max(1 - \mu_\gamma, \mu_s)). \quad (9)$$

В результате сравнения по всем $\tilde{\gamma}_k \in \Gamma$ с использованием выражений (5-9) определяется ситуация-эталон $\tilde{\gamma}^*$, которой текущая \tilde{S} наиболее близка:

$$\tilde{\gamma}^* : \mu(\tilde{S}, \tilde{\gamma}^*) = \max_{\tilde{\gamma}_k \in \Gamma} \mu(\tilde{S}, \tilde{\gamma}_k), \quad (10)$$

при обязательном условии $\mu(\tilde{S}, \tilde{\gamma}^*) \geq \mu_{III}$, и соответствующий класс состояния объекта мониторинга G_k . Невыполнение условия $\mu(\tilde{S}, \tilde{\gamma}^*) \geq \mu_{III}$ говорит о том, что ситуация не может быть классифицирована с требуемой достоверностью. Такие ситуации считаются нештатными, неизвестными для данного классификатора, а результаты такой классификации не могут быть использованы при

принятии решений и формировании рекомендаций по сценарию действий.

Таким образом, в рассматриваемой модели (3-10) результатом классификации текущего состояния объекта мониторинга является наиболее близкая к текущей нечеткой ситуации \tilde{S} ситуация-эталон \tilde{y} , соответствующая одному из заранее заданных классов состояний, при условии соблюдения требований по минимально допустимой достоверности классификации. Далее в рамках статьи данную модель будем рассматривать как базовую, определяющую основной порядок и принципы проведения классификации, относительно которой возможны различного рода модификации, не меняющие ее основную суть. Для различения базовой модели и модификаций обозначим ее как «модель по Гёделю».

3. Описание подхода. Для рассматриваемой модели априорные оценки качества классификации позволяют судить о непротиворечивости правил, формирующих алгоритм классификации, с учетом заданных способов формализованного описания классификационных признаков, видов функций принадлежности и др. параметров, а также выбранного множества классов, иначе говоря, «качестве» самой классификационной модели. В связи с тем, что для рассматриваемой модели предполагается отсутствие обучающей выборки, то для ее получения необходимо сформировать тестовую размеченную экспертами выборку возможных входных наборов данных (ситуаций) $\Theta^i \in \Theta$, где Θ – множество всех возможных нечетких ситуаций. Особенность множества Θ^i заключается в том, что для $\tilde{S} \in \Theta^i$ не известны истинные исходы классификации, можно лишь утверждать, что для модели с используемой операцией нечеткой импликации по Гёделю (9) результат классификации для «пограничных» (радикальных) примеров в виде самих ситуаций-эталонов \tilde{y} должен принадлежать диапазону $[\mu_{\text{IT}}; 1]$. Это делает весьма затруднительным формирование размеченной выборки, и, как следствие, применение традиционно используемых оценок качества классификации (полнота и точность классификации, F-мера и др.).

В связи с этим в статье предлагается на примере нечеткой классификационной модели рассмотреть подход к получению априорных оценок качества классификации на основе анализа чувствительности модели к изменению значений ее внутренних параметров, а также интерпретации полученных результатов. Последовательность действий, реализуемых в рамках подхода,

является обобщением результатов вычислительных экспериментов с рассматриваемой классификационной моделью, которая может быть также применена и для других моделей, функционирующих на основе правил, задаваемых экспертами. Отличительной особенностью предлагаемого подхода является решение по интерпретации полученных результатов в виде оценок рисков различного рода, источником которых является несовершенство классификаторов. Необходимо отметить, что несмотря на применение «модели по Гёделю» для решения разноплановых технических задач [8 – 12], на сегодняшний момент в материалах открытой печати не было обнаружено публикаций, посвященных оценке ее качества, в связи с чем приводимые в статье положения и результаты вычислительных экспериментов представляют интерес.

Применение подхода предполагает выполнение следующих этапов:

- 1) Анализ особенностей модели, выявление возможных противоречий и ограничений в правилах и алгоритмах.
- 2) Определение внутренних параметров модели, которые оказывают существенное влияние на работу классификатора.
- 3) Выбор показателей оценки влияния.
- 4) Моделирование работы классификатора и оценка показателей.
- 5) Интерпретация результатов моделирования и полученных показателей.

Для рассматриваемой модели реализация первого и второго этапа позволяет сделать вывод о наличии ряда особенностей, которые могут оказывать существенное влияние на работу классификатора:

- 1) В результате классификации возможны исходы, когда рассматриваемая ситуация не может быть с требуемой достоверностью отнесена ни к одному из заданных классов, т.е. результатом работы классификатора будет пустое множество, а сама ситуация признается нештатной. Условие нештатности имеет вид:

$$\tilde{S} \in \Lambda : \forall G_k \in G \mu(\tilde{S}, \tilde{\gamma}_k) < \mu_{III}, \quad (11)$$

где Λ – множество нештатных ситуаций.

- 2) Наличие условий (7-8) создает возможность некорректной классификации, вызванной случаем, когда для ситуаций S и γ для всех сравниваемых пар значений $\mu_{\mu_S(z_i)}(T_{ij})$ и $\mu_{\mu_\gamma(z_i)}(T_{ij})$ (6-7) одно из

них будет принадлежать интервалу I_{III} , и тогда параметр $f = 0$. В таком случае, как следует из условий (5-8), все результаты сравнения $C(\cdot)$ будут равны 1 и $\mu(\tilde{S}, \tilde{\gamma}) = 1$, в том числе и для всех остальных $G_k \in G$. Ситуация \tilde{S} , которая одновременно принадлежит к нескольким классам с достоверностью 1, не может быть классифицирована. С учетом введенных ранее обозначений, условие некорректности для нечеткой ситуации \tilde{S} и класса G_k (ситуации-эталона $\tilde{\gamma}_k$) имеет вид:

$$\tilde{S} \in H : \forall C(\mu_{\mu_S(z_i)}(T_{ij}), \mu_{\mu_{\gamma}(z_i)}(T_{ij})) = 1, \quad (12)$$

где H – множество ситуаций, отнесенных к некорректным.

В результате моделирования работы классификатора на наборе данных Θ^t может быть подсчитано количество нештатных и некорректно классифицированных ситуаций и оценены следующие показатели:

$$\eta = \frac{|\Lambda|}{|\Theta^t|}, \quad (13)$$

где η – оценка доли нештатных ситуаций от общего числа входных ситуаций Θ^t , и

$$\xi = \frac{|H|}{|\Theta^t|}, \quad (14)$$

где ξ – оценка доли некорректно классифицированных от общего количества входных ситуаций Θ^t . Показатель ξ может быть интерпретирован как показатель ошибки классификации [1]. Для отдельно взятого класса G_k ошибка классификации рассчитывается исходя из выражения вида

$$E_k = 1 - \frac{TP + TN}{U} = 1 - \frac{U - (FP + FN)}{U}, \quad (15)$$

где TP – число ситуаций, правильно отнесённых к классу G_k , TN – число ситуаций, правильно не отнесённых к классу G_k , U – общее число ситуаций, при этом $U = TP + TN + FP + FN$, а FP и FN – количество ситуаций, ошибочно отнесённых и ошибочно не отнесённых к классу G_k соответственно. При условии, что все корректные исходы классификации в результате реализации правил (5-10) составляют совокупность исходов TP и TN для каждого класса, а ошибочное отнесение или неотнесение ситуации к классу возможно вследствие некорректной работы классификатора (12), то выражение (15) принимает вид:

$$E_k = 1 - \frac{U - (FP + FN)}{U} = 1 - \frac{|\Theta'| - |H_k|}{|\Theta'|} = \frac{|H_k|}{|\Theta'|}, \quad (16)$$

т.е. представляет собой значение показателя ξ по отношению к классу G_k . Тогда на основе выражений (13-15) может быть подсчитан показатель средней ошибки классификации ξ_{cp} :

$$\xi_{cp} = \frac{1}{N_G} \sum_{k=1}^{N_G} \xi_k = \frac{1}{N_G} \sum_{k=1}^{N_G} \frac{|H_k|}{|\Theta'|}, \quad (17)$$

а также, по аналогии с (16), показатели максимальной ξ_{max} и минимальной ошибок ξ_{min} классификации из всех ξ_k . Отметим, что конкретный вид и значения параметров, используемых для описания нечётких признаков функций принадлежности, окажет влияние на значение предложенных показателей (13) и (14), однако не повлияет на порядок их расчета, что позволяет использовать их для различных модификаций рассматриваемой в статье модели.

3) Существенное влияние на работу модели и результат классификации оказывает выбор значения пороговой достоверности $\mu_{ПГ}$ и вид используемой операции нечеткой импликации.

В работах [5, 17] проведен сравнительный анализ применения различных видов операций нечеткой импликации для нечеткого моделирования процессов управления. В соответствии с приведенными в этих работах данными только основных, наиболее часто используемых видов операций нечеткой импликации

насчитывается более 10. В данной статье с целью изучения и демонстрации влияния вида операции импликации на результат классификации в качестве примеров были выбраны следующие:

- импликация по Заде $\mu_s \rightarrow \mu_\gamma = \max(\min(\mu_s, \mu_\gamma), 1 - \mu_\gamma)$;
- импликация по Лукасевичу $\mu_s \rightarrow \mu_\gamma = \min(1, 1 - \mu_s + \mu_\gamma)$;
- импликация по Гёделю $\mu_s \rightarrow \mu_\gamma = \max(1 - \mu_s, \mu_\gamma)$;
- импликация по Вади $\mu_s \rightarrow \mu_\gamma = \max(\mu_s \cdot \mu_\gamma, 1 - \mu_s)$.

Наглядно данное влияние отображено на рисунке 2, где приведены сравнительные графики применения различных видов импликации для вычисления значения выражения $\mu_s \leftrightarrow \mu_\gamma$ (9) при $\mu_s = \mu_\gamma$ и $\mu_\gamma, \mu_s \in [0; 1]$. Для импликации по Заде на рисунке 2 введена дополнительная шкала справа, выделенная цветом.

Результатом операции $\mu_s \leftrightarrow \mu_\gamma$ является достоверность того, что обе сравниваемые величины нечетко равны, которая, на первый взгляд, должна быть всегда равна 1. Однако полученные графики свидетельствуют, что для различных видов используемых импликаций, достоверность этого нечеткого равенства варьируется в диапазоне от 0,4 до 1 (для импликации по Вади), а для рассматриваемой модели – от 0,5 до 1 (импликация по Гёделю), в зависимости от величины сравниваемых одинаковых значений. Этим, в частности, объясняется необходимость дополнительного использования порогового интервала I_{III} . В отношении же импликации по Лукасевичу следует вывод, что результат сравнения не зависит от самих сравниваемых значений, а его достоверность равна 1.

В работе [12] рассмотрена модификация модели из п.2 (3-10), в которой как раз используется импликация по Лукасевичу, далее будем обозначать ее как «модель по Лукасевичу». В целом, она предполагает проведение классификации нечеткой ситуации в соответствии с теми же выражениями (3-10), но с тем отличием, что вид оператора $C(\cdot)$ в выражениях (6-7) не зависит от условия (8) и с учетом введенных сокращений представляется как:

$$C(\mu_s, \mu_\gamma) = \mu_s \leftrightarrow \mu_\gamma, \quad (18)$$

а вместо выражения (9) используется выражение:

$$\mu_s \leftrightarrow \mu_\gamma = \min\left(\min(1, 1 - \mu_s + \mu_\gamma), \min(1, 1 - \mu_\gamma + \mu_s)\right). \quad (19)$$

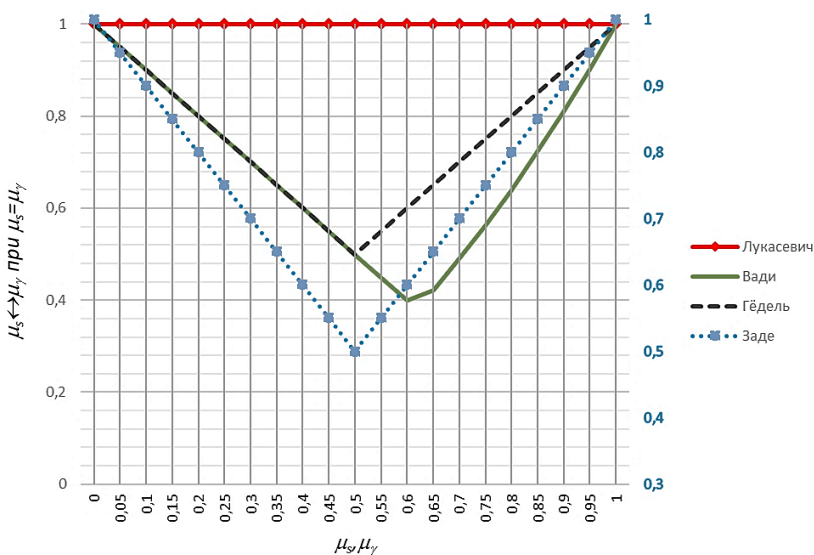


Рис. 2. Графики результатов применения различных видов операции нечеткой импликации

В рамках применения описываемого подхода в качестве основного влияющего параметра была выбрана пороговая достоверность классификации $\mu_{ПГ}$, а в качестве оценки чувствительности классификационной модели к данному параметру использовались зависимости $\eta(\mu_{ПГ})$ и $\xi(\mu_{ПГ})$ при $\mu_{ПГ} \in [0; 1]$. Конкретный вид зависимостей был получен в ходе соответствующего моделирования работы классификатора, реализованного в рамках следующего этапа. При этом были проведены исследования как для исходной модели («по Гёделю»), так и для ее модификации («по Лукасевичу»).

4. Вычислительный эксперимент. Для моделирования работы классификатора и получения зависимостей $\eta(\mu_{ПГ})$ и $\xi(\mu_{ПГ})$ было сформировано множество классов состояний социально-политической обстановки $G = \{G_1, G_2, G_3, G_4\}$, описываемой с применением лингвистических переменных L_1, L_2 и функций принадлежности (1) и (2), графики которых представлены на рисунке 1. При этом каждому

классу поставлен в соответствие некоторый комплекс рекомендаций по возможным решениям в связи со сложившейся обстановкой. Формальное описание предлагаемого множества классов в виде ситуаций-эталонов имеет вид:

$$\tilde{\gamma}_1 = (\langle 0,7 / \text{Неприемлемая}; 0,3 / \text{Приемлемая} \rangle / z_1, \\ \langle 0,9 / \text{Малая}; 0,1 / \text{Большая} \rangle / z_2),$$

$$\tilde{\gamma}_2 = (\langle 0,9 / \text{Неприемлемая}; 0,1 / \text{Приемлемая} \rangle / z_1, \\ \langle 0,2 / \text{Малая}; 0,8 / \text{Большая} \rangle / z_2),$$

$$\tilde{\gamma}_3 = (\langle 0,15 / \text{Неприемлемая}; 0,85 / \text{Приемлемая} \rangle / z_1, \\ \langle 0,75 / \text{Малая}; 0,25 / \text{Большая} \rangle / z_2),$$

$$\tilde{\gamma}_4 = (\langle 0,3 / \text{Неприемлемая}; 0,7 / \text{Приемлемая} \rangle / z_1, \\ \langle 0,15 / \text{Малая}; 0,85 / \text{Большая} \rangle / z_2),$$

при этом $\tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\gamma}_3, \tilde{\gamma}_4 \in \Theta^t$. Множество классов было сформировано искусственно таким образом, чтобы каждый из используемых признаков имел как минимум одно «хорошо выраженное» значение, т.е. превосходящее пороговое $\mu_{\text{ПГ}}$, для каждого из описывающих его термов. Таким образом, для 2-х признаков, описываемых с использованием 2-х термов каждый, было сформировано 4 класса состояний обстановки. Пороговое значение $\mu_{\text{ПГ}}$ было декларативно определено равным 0,7.

Получение зависимостей $\eta(\mu_{\text{ПГ}})$ и $\xi(\mu_{\text{ПГ}})$ осуществлялось в рамках вычислительного эксперимента, который проводился в два этапа:

- 1) Оценка чувствительности нечеткой классификационной модели с нечеткой импликацией по Гёделлю.
- 2) Оценка чувствительности нечеткой классификационной модели с нечеткой импликацией по Лукасевичу.

В рамках первого этапа эксперимента был сформирован входной набор нечетких ситуаций Θ^t и проведено моделирование работы классификатора. Тестовое множество Θ^t было искусственно

сгенерировано исходя из того, что в соответствии с выражениями (3-5) каждая нечеткая ситуация, \tilde{S} представляет собой совокупность значений соответствующих функций принадлежности $\mu_{\mu_S(z_i)}(T_{ij})$ от значений аргументов $x, y \in D_{XY}$, где $D_{XY} = X \times Y$ – «опорное» декартово множество парных значений признаков z_1, z_2 . Тогда, поскольку функции $\mu_{\mu_S(z_i)}(T_{ij})$ являются непрерывными и не имеют разрывов на соответствующих им опорных множествах X и Y , то для исследования классификационной модели для всех возможных значений множества D_{XY} может быть сформирована выборка пар $D_{XY}^t \in D_{XY}$, равномерно распределенных на шкалах, соответствующих X и Y . Следовательно, входной набор данных может быть интерпретирован как совокупность элементов множества D_{XY}^t .

В ходе моделирования для каждой нечеткой ситуации $\tilde{S} \in \Theta^t$ был произведен расчет степени равенства каждому из классов из множества G и выбрано максимальное значение $\mu = \max_{\gamma_k \in G} \mu(S, \gamma_k)$.

Полученная совокупность точек с координатами (x, y, μ) образует поверхность $\mu = f(x, y)$ – поверхность «максимума достоверности», где значение μ есть максимальная достоверность отнесения рассматриваемой ситуации хотя бы к одному из классов множества G . Значение μ вычисляется с использованием выражений (5-10) и представляет собой грубую оценку принципиальной возможности классификации входной ситуации с требуемой достоверностью. Построенная поверхность $f(x, y)$ позволяет наглядно определить те области нечетких ситуаций в проекции на множество D_{XY} , которые являются нештатными, а также выявить случаи некорректной классификации.

Общий вид поверхности $f(x, y)$ для заданного множества классов G и нечетких признаков, описываемых при помощи лингвистических переменных L_1, L_2 , представлен на рисунке 3. Обращает на себя внимание тот факт, что вид поверхности заметно меняется при изменении параметра μ_{IT} . Кроме того, в результате моделирования не было выявлено нештатных исходов классификации,

т.е. все ситуации, соответствующие входному набору данных D_{XY} , были классифицированы с требуемой достоверностью и $\eta(\mu_{ПГ}) = 0 \forall \mu_{ПГ} \in [0; 1]$, что также можно видеть на графиках (рисунок 3).

Вид графиков зависимостей $\xi_{\min}(\mu_{ПГ})$, $\xi_{\max}(\mu_{ПГ})$, $\xi_{\text{ср}}(\mu_{ПГ})$, полученных в результате моделирования при $\mu_{ПГ} \geq 0,5$ представлен на рисунке 4 (для $\mu_{ПГ} < 0,5$ значения всех показателей равны нулю). Мощность тестового множества Θ^t составила 250 ситуаций.

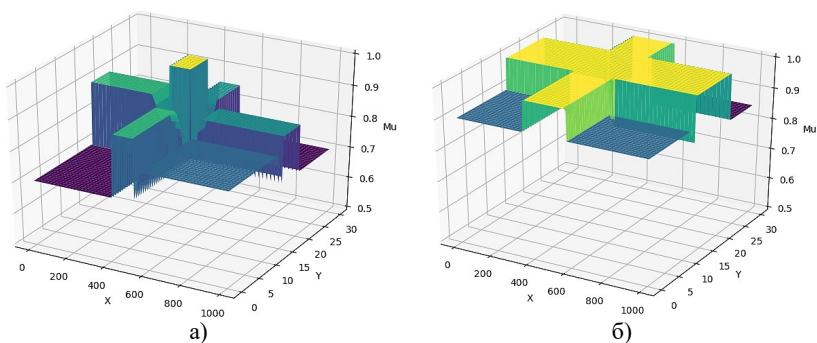


Рис. 3. Вид поверхности $\mu = f(x, y)$ при различных значениях пороговой достоверности: а) $\mu_{ПГ} = 0,7$; б) $\mu_{ПГ} = 0,85$

Целью второго этапа вычислительного эксперимента явилась демонстрация влияния вида используемой операции импликации на функционирование нечеткой классификационной модели рассматриваемого типа и получение зависимостей $\eta(\mu_{ПГ})$ и $\xi(\mu_{ПГ})$. Для этого исследовалась модель с нечеткой импликацией по Лукасевичу [12]. Получение результатов эксперимента осуществлялось по тому же принципу, как и в предыдущем этапе вычислительного эксперимента.

Интерес к тестированию модели по Лукасевичу был связан с тем, что в ней отсутствуют предпосылки к случаям описанной выше некорректной классификации. В результате моделирования установлено, что вид получаемой поверхности $\mu = f(x, y)$ не зависит от пороговой достоверности классификации (рисунок 5).

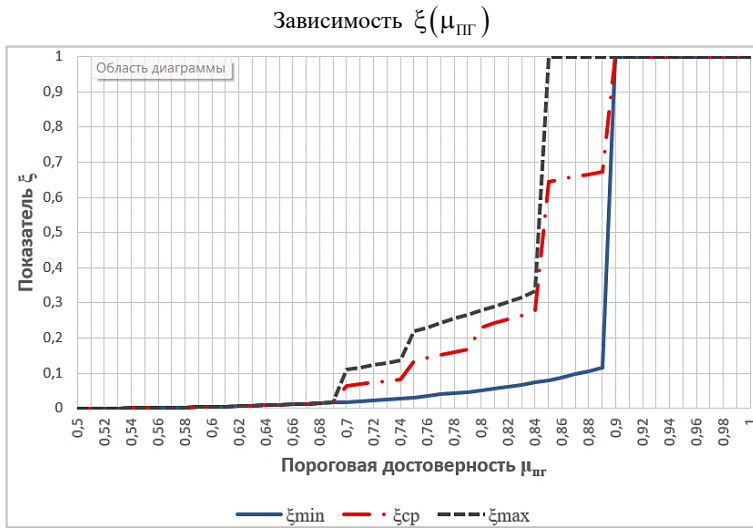


Рис. 4. Вид графиков зависимостей $\xi(\mu_{ПГ})$

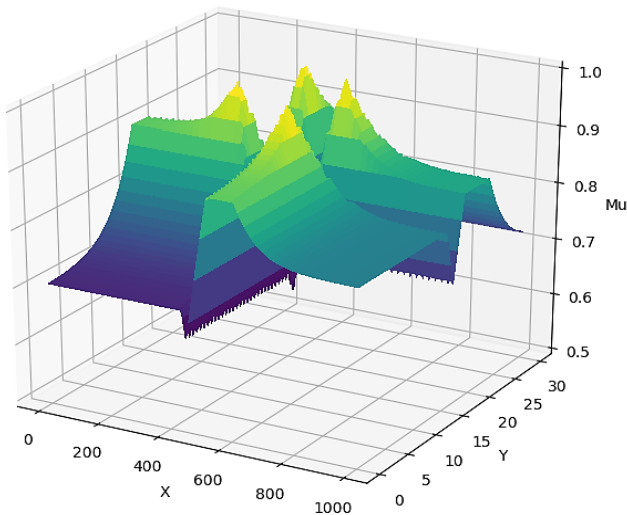


Рис. 5. Вид поверхности $\mu = f(x, y)$ для модели с нечеткой импликацией по Лукасевичу

В тоже время, моделирование позволило выявить, что в отличие от модели по Гёделю, параметр $\mu_{ПГ}$ имеет влияние на принятие решения о том, может ли входная ситуация быть классифицирована или является нештатной (11). В результате проведенных тестов для модели по Лукасевичу был также получен график зависимости $\eta(\mu_{ПГ})$, отражающий изменение доли нештатных ситуаций при варьировании значения $\mu_{ПГ}$ (рисунок 6). Отмечено, что для выбранного набора классов при $\mu_{ПГ} < 0,5$ все $\eta(\mu_{ПГ}) = 0$.



Рис. 6. Вид графика зависимости $\eta(\mu_{ПГ})$

5. Подходы к интерпретации полученных результатов.

Проведенный вычислительный эксперимент позволил для рассматриваемой нечеткой классификационной модели с учетом ее модификации наглядно отобразить влияние внутренних параметров модели на результат ее функционирования и на примере множества классов $G = \{G_1, G_2, G_3, G_4\}$ получить зависимости $\eta(\mu_{ПГ})$ и $\xi(\mu_{ПГ})$ для различных видов используемых операций нечеткой импликации. Использование показателей η и ξ для фиксированных параметров модели, а также графиков зависимостей $\eta(\mu_{ПГ})$ и $\xi(\mu_{ПГ})$ в качестве априорной информации о достигаемом качестве классификации может

дать соответствующее представление как о самой модели, так и о возможных последствиях ее применения.

В частности, это позволяет оценить поведение модели в динамично изменяющихся условиях решения задачи мониторинга, требующих повышения пороговой достоверности классификации $\mu_{ПГ}$. Показатель ξ может рассматриваться как общепринятый показатель ошибки классификации, в то время как показатель η является уникальным, учитывающим особенности рассматриваемой в статье нечеткой классификационной модели.

Отдельно $\eta(\mu_{ПГ})$ и $\xi(\mu_{ПГ})$ могут быть интерпретированы как характеристики функционирования классификатора наподобие амплитудно-частотных характеристик приборов и устройств в электро-технике. В статье в рамках завершающего этапа подхода к получению и интерпретации оценок качества классификации предлагается рассматривать полученные зависимости и показатели в рамках оценивания рисков, связанных с использованием подобных моделей.

В соответствии с ГОСТ Р ИСО 31000-2019 «Менеджмент риска. Принципы и руководство» [18] под риском понимается следствие влияния неопределенности на достижение поставленных целей. При этом, риск представляется как сочетание вероятности наступления события (неблагоприятного) и его последствий [19]. Подходы к оценке вероятности события варьируются в зависимости от рассматриваемой предметной области. В частности, в статьях [20, 21] для этого используются разновидности байесовских сетей, а в статье [22] расчет показателя риска основан на применении нечеткого логического вывода. При этом, если в [20] в качестве последствий рассматривается ущерб от столкновения морских судов, а в [21] причинение вреда здоровью человека, то в работе [23] само наступление неблагоприятного события рассматривается как неприемлемое безотносительно к любому возможному ущербу.

В случае применения рассматриваемой в статье нечеткой классификационной модели под неблагоприятным событием следует понимать невозможность классифицировать текущую ситуацию вследствие некорректности классификации или нештатности самой ситуации. В качестве последствий неблагоприятного события следует рассматривать ущерб, который может быть нанесен вследствие непринятия решений или ошибочного принятия решений, не соответствующих складывающейся обстановке, реальному состоянию объекта мониторинга. Примерами понесенного ущерба могут быть: экономические потери, людские потери, снижение рейтинга

поддержки политического руководства среди населения, потеря количества мест в выборных органах страны на выборах и т.д.

Для рассмотренной в статье классификационной модели с учетом ее различных модификаций значение риска может быть количественно оценено с использованием показателя, концептуальный вид которого представляет собой функционал вида:

$$R \sim f(\xi, \eta, \rho, \mu_{\text{ПГ}}), \quad (20)$$

где R – показатель риска, ρ – величина ущерба от наступления неблагоприятных событий. Строго формально, данная функция не позволяет говорить о вероятности наступления неблагоприятного события в аксиоматическом смысле, однако дает некоторую оценку вероятности наступления такого события на основе обработки статистических данных, которая тем точнее, чем больше соответствующая тестовая выборка (при условии ее репрезентативности), в качестве которой выступает Θ^t . Дальнейшее изложение предлагаемых подходов к расчету показателя риска R в статье приведены с учетом указанного допущения.

В соответствии с рассмотренными в статье особенностями классификационной модели задачи по оценке риска могут быть разделены на две группы по виду источника риска:

1. Некорректная классификация текущей ситуации.
2. Нештатность текущей ситуации.

К первой группе относятся события, когда текущая ситуация некорректно отнесена к какому-то одному классу $G_k \in G$, т.е. условие (12) выполняется только для одного класса, и, вследствие ошибочного определения состояния объекта мониторинга и последовавших за этим ошибочных действий, возникает ущерб ρ_k . В рамках приведенного в статье примера такой исход возможен при ранжировании классов ситуаций по срочности принимаемых мер (сверхсрочный сценарий, срочный сценарий, обычный сценарий, отмена всех действий). Тогда, если известны или есть возможность оценить последствия ошибочно принятых решений по причине некорректной классификации, то при заданном для модели значении пороговой достоверности для оценивания риска может использоваться следующее выражение:

$$R_{cp} = \frac{1}{N_G} \sum_{k=1}^{N_G} \xi_k \cdot \rho_k, \quad (21)$$

где R_{cp} – показатель риска в виде средневзвешенного ущерба, возникающего вследствие некорректной классификации, ρ_k – ущерб от ошибочного выполнения действий, соответствующих классу G_k . Здесь и далее будем подразумевать, что значение показателей ξ_k и показателя η_k рассчитано для конкретного значения пороговой достоверности $\mu_{ПГ}$.

Второй группой задач является оценка рисков в случае, когда неблагоприятное событие наступает из-за невозможности классифицировать текущую ситуацию. К таким событиям относятся исходы, когда текущая нечеткая ситуация вследствие ошибки некорректно, т.е. с достоверностью 1, одновременно отнесена к двум и более классам $G_k \in G$ (12) или не может с требуемой достоверностью быть отнесена ни к одному из классов (11). В обоих случаях модель не позволяет определить состояние объекта мониторинга. Примерами последствий такого события является невыполнение требований по оперативности представления сведений, утрата актуальности представляемой информации, несоблюдение требований по ограничениям на критически важные показатели деятельности. В этом случае, если существует возможность оценить последствия таких событий, то показатель риска может быть представлен в виде:

$$R_{НК} = \left(\xi_{|G^*| \geq 2} + \eta \right) \cdot \rho_{НК}, \quad (22)$$

где $R_{НК}$ – показатель риска в виде взвешенного ущерба, возникающего по причине невозможности классификации текущей ситуации, $\xi_{|G^*| \geq 2}$ – показатель некорректной классификации для 2-х классов и более, G^* – множество результатов классификации, $\rho_{НК}$ – ущерб от невозможности классификации ситуации.

Если независимо от ущерба сам факт неблагоприятного события из-за некорректной классификации или невозможности классификации ситуации считается неприемлемым, то для оценки вероятности наступления такого события может использоваться показатель вида:

$$R_{НЕП} = \xi_{\max} + \eta, \quad (23)$$

где $R_{НЕПР}$ – показатель риска в виде оценки вероятности наступления неприемлемого события, $R_{НЕПР} \in [0; 1]$, ξ_{\max} – максимальная оценка доли некорректно классифицированных ситуаций, $\xi_{\max} = \max_{G_k \in G} \xi_k$.

6. Заключение. Таким образом, в статье рассмотрен подход к получению и интерпретации априорных оценок качества классификации и его применение на примере нечеткой классификационной модели и ее модификации, рассмотренных в работах [8 – 11] и [12]. Главной особенностью предлагаемого подхода является получение оценок качества классификации на основе исследования чувствительности модели к изменению значений внутренних параметров в ходе соответствующего моделирования, а также интерпретация полученных результатов в виде оценки рисков, вызванных несовершенством самого классификатора. В статье предложены показатели оценки чувствительности модели, а также различные виды показателей рисков и способы их расчета. Применение подхода и реализация различных его этапов проиллюстрированы результатами вычислительных экспериментов. По состоянию на текущий момент в материалах открытой печати не обнаружено публикаций с подобными исследованиями в отношении рассмотренной в статье нечеткой классификационной модели.

Литература

1. Кривенко М.П., Васильев В.Г. Методы классификации данных большой размерности // М.: ИПИ РАН. 2013. 208 с.
2. Hartmann J., Huppertz J., Schamp C., Heitmann M. Comparing automated text classification methods. *International Journal of Research in Marketing*. 2019. vol. 36(1). pp. 20–38.
3. Belyadi H., Haghghat A. *Machine Learning Guide for Oil and Gas Using Python*. Gulf Professional Publishing, 2021. 476 p.
4. Заде Л. Понятие лингвистической переменной и ее применение к принятию приближенных решений // М.: Мир. 1976. 167 с.
5. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH // СПб.: БХВ-Петербург, 2005. 736 с.
6. Tsoukalas L. *Fuzzy Logic: Applications in Artificial Intelligence, Big Data, and Machine Learning* // McGraw Hill. 2023. 176 p.
7. van Krieken E., Acar E., van Harmelen F. Analyzing Differentiable Fuzzy Logic Operators. *Artificial Intelligence*. 2022. vol. 302. DOI: 10.1016/j.artint.2021.103602.
8. Мелихов А.Н., Бернштейн Л.С., Коровин С.Я. Ситуационные советующие системы с нечеткой логикой // М.: Наука, 1990. 272 с.
9. Борисов В.В., Авраменко Д.Ю. Нечеткое ситуационное управление сложными системами на основе их композиционного гибридного моделирования // Системы управления, связи и безопасности. 2021. № 3. С. 207–237.
10. Бакасов С.Р., Санаева Г.Н., Воронин Ю.А., Пророков А.Е., Богатиков В.Н. Управление технологической безопасностью промышленных процессов на

- основе мультиагентного моделирования // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2019. № 4. С. 37–45.
11. Кривов М.В., Асламова Е.А., Асламова В.С. Система выработки стратегий управления промышленной безопасностью // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 59. С. 55–65.
 12. Мелихова О.А. Приложение матлогики к проблемам моделирования // Известия ЮФУ. Технические науки. 2014. № 7(156). С. 204–214.
 13. Старовойтов В.Г. Ситуационный центр как эффективный механизм в системе управления // Национальная безопасность / nota bene. 2021. № 5. С. 22–29.
 14. Шедий М.В. Становление и тенденции развития системы ситуационных центров как ключевого фактора повышения эффективности государственного управления // Вестник НГУЭУ. 2021. № 3. С. 8–18.
 15. Avdeeva Z., Kovriga S., Lepskiy V., Raikov A., Slavin B., Zatsarinny A. The Distributed Situational Centers System as an Instrument of State and Corporate Strategic Goal-Setting in the Digital Economy. IFAC-PapersOnLine. 2020. vol. 53(2). pp. 17499–17504.
 16. Kriesi H., Lorenzini J., Wuest B., Hausermann S. Contention in Times of Crisis: Recession and Political Protest in Thirty European Countries. Cambridge University Press. 2020. 300 p.
 17. Борисов В.В., Круглов В.В., Федулов А.С. Нечеткие модели и сети. 2-е изд., стереотип. // М.: Горячая линия – Телеком. 2012. 284 с.
 18. ГОСТ Р ИСО 31000-2019. Менеджмент риска. Принципы и руководство // М.: Госстандарт России. 2020.
 19. ISO 14971:2019. Medical devices – Application of risk management to medical devices. 2019.
 20. Fu S., Zhang Y., Zhang M., Han B., Wu Z. An object-oriented Bayesian network model for the quantitative risk assessment of navigational accidents in ice-covered Arctic waters. Reliability Engineering & System Safety. 2023. vol. 238. DOI: 10.1016/j.ress.2023.109459.
 21. Hunte J., Neil M., Fenton N. A hybrid Bayesian network for medical device risk assessment and management. Reliability Engineering & System Safety. 2024. vol. 241. 16 p.
 22. Jiang M., Liu Y., Lu J., Qu Z., Yang Z. Risk assessment of maritime supply chains within the context of the Maritime Silk Road. Ocean & Coastal Management. 2023. vol. 231. 14 p.
 23. Молоканов Г.Г., Пинчук А.В., Потюпкин А.А. Нечётко-множественный подход к оцениванию целевых рисков при управлении развитием орбитальной группировки космических аппаратов // Труды ВНИИЭМ. Вопросы электромеханики. 2015. Т. 148. № 5. С. 19–23.

Потюпкин Александр Александрович — канд. техн. наук, слушатель, Военно-космическая академия имени А.Ф. Можайского. Область научных интересов: ситуационное управление, применение методов теории нечетких множеств для управления сложными организационно-техническими системами, методы анализа данных и обработки информации. Число научных публикаций — 35. vka@mil.ru; улица Ждановская, 13, 197198, Санкт-Петербург, Россия; р.т.: +7(812)237-1249.

Пилькевич Сергей Владимирович — д-р техн. наук, доцент, начальник кафедры, Военно-космическая академия имени А.Ф. Можайского. Область научных интересов: информационная безопасность, дискретная математика, моделирование сложных

социотехнических систем. Число научных публикаций — 150. vka@mil.ru; улица Ждановская, 13, 197198, Санкт-Петербург, Россия; р.т.: +7(812)347-9654.

Зайцев Всеволод Владимирович — заместитель начальника кафедры, Военно-космическая академия имени А.Ф. Можайского. Область научных интересов: информационная безопасность, системы искусственного интеллекта, компьютерная лингвистика. Число научных публикаций — 7. vka@mil.ru; улица Ждановская, 13, 197198, Санкт-Петербург, Россия; р.т.: +7(812)347-9654; факс: +7(812)347-9654.

A. POTYUPKIN, S. PILKEVICH, V. ZAYTSEV
**AN APPROACH TO A PRIORI ASSESSMENT OF FUZZY
CLASSIFICATION MODELS IN MONITORING TASKS**

Potyupkin A., Pilkevich S., Zaytsev V. An Approach to a Priori Assessment of Fuzzy Classification Models in Monitoring Tasks.

Abstract. The article addresses the problems of using automation tools to perform monitoring and management tasks applicable to assessing the quality of fuzzy classification models, where the classification procedure is implemented on the basis of knowledge (rules) in the absence of the training set. An approach is proposed to obtain a priori assessments of the classification quality based on the study of the used model sensitivity to changes in the values of internal parameters during the corresponding modeling. The interpretation of the modeling results in the form of risk assessment caused by the self-imperfection of the classification models is obtained. The article provides an example of a fuzzy classification model based on a comparison of the current state of a monitoring object described using fuzzy features with a set of predefined typical states, which form corresponding fuzzy equal (close) states (monitoring situations). The comparison is carried out using the fuzzy implication operation provided that the required reliability is met. The example of this model demonstrates how the type of implication operation, as well as the internal features of the model, affect the results of classification, and appropriate indicators are proposed, which are both an interpretation of generally accepted indicators for assessing the classification quality, and unique, inherent in the considered model. Computational experiments were carried out, which made it possible to obtain graphs of changes in classification quality assessment indicators for the considered model and its modification and visualize the influence of internal parameters of the model on the results of its application. A number of indicators are proposed that allow for an a priori assessment of the risks arising from the application of the model before its actual application.

Keywords: fuzzy sets, classification, quality assessment, risk.

References

1. Krivenko M., Vasil'ev V. *Metody klassifikacii dannyh bol'shoj razmernosti* [Classification methods of big scale data]. Moscow: IPI RAN, 2013. 208 p. (In Russ.).
2. Hartmann J., Huppertz J., Schamp C., Heitmann M. Comparing automated text classification methods. *International Journal of Research in Marketing*. 2019. vol. 36(1). pp. 20–38.
3. Belyadi H., Haghghat A. *Machine Learning Guide for Oil and Gas Using Python*. Gulf Professional Publishing, 2021. 476 p.
4. Заде Л. Понятие лингвистической переменной и ее применение к принятию приближенных решений. М.: Мир. 1976. 167 с.
5. Leonenkov A. *Nechetkoe modelirovanie v MATLAB i fuzzyTECH* [Fuzzy modeling in MATLAB and fuzzyTECH]. Saint-Petersburg: BHV-Peterburg, 2005. 736 p. (In Russ.).
6. Tsoukalas L. *Fuzzy Logic: Applications in Artificial Intelligence, Big Data, and Machine Learning*. McGraw Hill. 2023. 176 p.
7. van Krieken E., Acar E., van Harmelen F. Analyzing Differentiable Fuzzy Logic Operators. *Artificial Intelligence*. 2022. vol. 302. DOI: 10.1016/j.artint.2021.103602.
8. Melihov A., Bernshtejn L., Korovin S. *Situacionnye sovetujushhie sistemy s nechetkoj logikoj* [Situational advising systems with fuzzy logic]. Moscow: Nauka. Gl. red. fiz.-mat.lit, 1990. 272 p. (In Russ.).

9. Borisov V., Avramenko D. [Fuzzy situational control of complex systems based on composite hybrid modeling] *Sistemy upravleniya, svyazi i bezopasnosti – Systems of Control, Communication and Security*. 2021. № 3. pp. 207-237. (In Russ.).
10. Bakasov S., Sanaeva G., Voronin Yu., Prorokov A., Bogatkov V. [Management of technological safety of industrial processes based on multi-agent modeling] *Vestnik Astrahanskogo gosudarstvennogo tehničeskogo universiteta. Serija: Upravlenie, vychislitel'naja tehnika i informatika – Bulletin of the Astrakhan State Technical University. Series: Management, Computer Engineering and Computer Science*. 2019. № 4. pp. 37–45. (In Russ.).
11. Krivov M., Aslamova E., Aslamova V. [A system for developing industrial safety management strategies] *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naja tehnika i informatika – Bulletin of Tomsk State University. Management, computer engineering and computer science*. 2022. № 59. pp. 55–65. (In Russ.).
12. Melihova O. [Application of mathematical logic to modeling problems] *Izvestija JuFU. Tehničeskije nauki – Izvestiya SFedU. Engineering sciences*. 2014. no. 7(156). pp. 204–214. (In Russ.).
13. Starovojtov V. [The situation center as an effective mechanism in the management system] *Nacional'naja bezopasnost' / nota bene – National Security / nota bene*. 2021. № 5. pp. 22–29. (In Russ.).
14. Shedij M. [Formation and development trends of the system of situational centers as a key factor in improving the efficiency of public administration] *Vestnik NGUJeU – Vestnik NSUEM*. 2021. no. 3. pp. 8–18. (In Russ.).
15. Avdeeva Z., Kovriga S., Lepskiy V., Raikov A., Slavin B., Zatsarinny A. The Distributed Situational Centers System as an Instrument of State and Corporate Strategic Goal-Setting in the Digital Economy. *IFAC-PapersOnLine*. 2020. vol. 53(2). pp. 17499–17504.
16. Kriesi H., Lorenzini J., Wuest B., Hausermann S. *Contention in Times of Crisis: Recession and Political Protest in Thirty European Countries*. Cambridge University Press. 2020. 300 p.
17. Borisov V., Kruglov V., Fedulov A. *Nechetkie modeli i seti [Fuzzy models and networks]*. Moscow: Goryachaya liniya – Telekom. 2012. 284 p. (In Russ.).
18. GOST R ISO 31000-2019. [Risk management. Principles and Guide]. Moscow: Gosstandart Rossii [Russian Government Standard]. 2020. (In Russ.).
19. ISO 14971:2019. *Medical devices – Application of risk management to medical devices*. 2019.
20. Fu S., Zhang Y., Zhang M., Han B., Wu Z. An object-oriented Bayesian network model for the quantitative risk assessment of navigational accidents in ice-covered Arctic waters. *Reliability Engineering & System Safety*. 2023. vol. 238. DOI: 10.1016/j.ress.2023.109459.
21. Hunte J., Neil M., Fenton N. A hybrid Bayesian network for medical device risk assessment and management. *Reliability Engineering & System Safety*. 2024. vol. 241. 16 p.
22. Jiang M., Liu Y., Lu J., Qu Z., Yang Z. Risk assessment of maritime supply chains within the context of the Maritime Silk Road. *Ocean and Coastal Management*. 2023. vol. 231. 14 p.
23. Molokanov G., Pinchuk A., Potyupkin A. [Fuzzy multiple approach to target risk assessment during control of satellite constellation evolution] *Trudy VNIEM. Voprosy elektromehaniki – Electromechanical matters. VNIEM studies*. 2015. vol. 148. no. 5. pp. 19–23. (In Russ.).

Potyupkin Alexander — Ph.D., Student, Mozhaysky Military Space Academy. Research interests: situational management, application of fuzzy sets methods in complex organizational and technical systems control tasks, data analysis and information processing methods. The number of publications — 35. vka@mil.ru; 13, Zhdanovskaya St., 197198, St. Petersburg, Russia; office phone: +7(812)237-1249.

Pilkevich Sergey — Ph.D., Dr.Sci., Associate Professor, Chief of the department, Mozhaysky Military Space Academy. Research interests: information security, discrete mathematics, social systems modeling. The number of publications — 150. vka@mil.ru; 13, Zhdanovskaya St., 197198, St. Petersburg, Russia; office phone: +7(812)347-9654.

Zaytsev Vsevolod — Deputy chief of the department, Mozhaysky Military Space Academy. Research interests: information security, artificial intelligence systems, computer linguistics. The number of publications — 7. vka@mil.ru; 13, Zhdanovskaya St., 197198, St. Petersburg, Russia; office phone: +7(812)347-9654; fax: +7(812)347-9654.

A.H. IBRAHIM, M.A. BOUDREF, L. BADIS
**CLUSTERING OF NETWORKS USING THE FISH SCHOOL
SEARCH ALGORITHM**

Ibrahim A.H., Boudref M.A., Badis L. Clustering of Networks Using the Fish School Search Algorithm.

Abstract. A network is an aggregation of nodes joined by edges, representing entities and their relationships. In social network clustering, nodes are organized into clusters according to their connectivity patterns, with the goal of community detection. The detection of community structures in networks is essential. However, existing techniques for community detection have not yet utilized the potential of the Fish School Search (FSS) algorithm and modularity principles. We have proposed a novel method, clustering with the Fish School Search algorithm and modularity function (FSC), that enhances modularity in network clustering by iteratively partitioning the network and optimizing the modularity function using the Fish School Search Algorithm. This approach facilitates the discovery of highly modular community structures, improving the resolution and effectiveness of network clustering. We tested FSC on well-known and unknown network structures. Also, we tested it on a network generated using the LFR model to test its performance on networks with different community structures. Our methodology demonstrates strong performance in identifying community structures, indicating its effectiveness in capturing cohesive communities and accurately identifying actual community structures.

Keywords: clustering, fish school search algorithm, modularity function, network structures.

1. Introduction. In network theory, a graph or network refers to a group of nodes linked through edges or lines, representing different entities and their relationships or interactions [1]. For example, in metabolic networks, nodes could represent biochemical compounds or enzymes, and edges could represent metabolic reactions or pathways [2]. In the case of the Internet, nodes could represent web pages or computers, and edges could represent hyperlinks or network connections [3]. The representation of a network is a graph $G(V, E)$, where V refers to the set of nodes or vertices and E refers to the set of edges or links [4]. The number of vertices and edges determines the size of a graph. In the case of a social network, the size of the graph depends on the number of people in the network and their relationships or interactions [5]. In social network clustering, nodes are grouped into clusters based on their connectivity patterns, which are determined by the links or edges that connect them. This clustering process relies on community recognition, where a community refers to a group of nodes that are densely connected with only a few connections to nodes outside the community [6]. The objective of community detection is to identify clusters or communities of nodes within a network, where the nodes

within a community exhibit denser connections among themselves compared to nodes in other communities [7].

Clustering a social network can help us identify groups of individuals that are more closely connected, revealing social substructures such as cliques, subgroups, or communities. Understanding these substructures can provide insights into the dynamics of social interactions and help us identify influential individuals or groups within the network [8]. For example, recognizing communities and clusters can help identify groups of proteins that work together to perform specific functions in the cell [9]. In the case of a transportation network, clustering can help identify clusters of nodes that are more interconnected in traffic planning and routing [10]. Several methods can structure networks [11], including modularity-based methods [12], spectral clustering [13], hierarchical clustering [14], and modularity optimization [15, 16].

Maximizing modularity is a challenging task due to its computational intractability, as finding an optimal solution has been proven to be NP-complete [17]. Heuristic algorithms are often used, but they may not always produce the best partitions. Modularity maximization is a promising approach for community detection, aiming to maximize within-community interconnectedness while minimizing between-community interconnectedness.

Our main contribution is the enhancement of modularity in network clustering through the innovative application of the Fish School Search Algorithm [18, 19]. This algorithm, inspired by the collective behavior and intelligence of fish schools, serves as a powerful metaheuristic for navigating the complex landscape of community detection. By iteratively partitioning the network and judiciously pruning edges, our approach focuses on maximizing the modularity function [20], which quantifies the strength of the division of a network into modules or communities.

The Fish School Search Algorithm's ability to simulate the dynamic and collaborative [21] search strategies of fish enables the discovery of highly modular community structures, thereby improving the resolution and effectiveness of network clustering. This method not only aligns with the principles of modularity optimization but also introduces a novel perspective to the field, leveraging natural processes to address computational challenges inherent in the task of community detection.

In the clustering process, the modularity function serves as the objective function, assessing the quality of a clustering solution [22]. The proposed method works by initializing a population of fish, each representing a potential clustering solution. Until they reach an optimal

solution, the fish iteratively evaluate and update their solution based on the modularity function.

The method was tested on well-known and unknown network structures and compared to other commonly used clustering methods [23]. The findings demonstrate the method's superior performance in both modularity and computational efficiency. The paper is organized in a logical and easy-to-follow manner. The second section, titled Background and Related Works, is further divided into four subsections. The first subsection provides background on community detection and its importance. The second subsection discusses modularity and modularity maximization. The third subsection introduces the Fish School Search (FSS) algorithm and its use in the proposed network clustering method, while the fourth subsection reviews related works in the field. The third section presents the proposed approach, while the fourth section of the paper presents a performance assessment of our method. The fifth section discusses the results and their implications, and the sixth section concludes the study and suggests future work.

2. Background and related works

2.1. Community detection. Community detection in complex networks is the process of identifying clusters of nodes that share common properties, depicted as nodes and edges in a graph. These algorithms group nodes based on denser connections within groups compared to connections between groups. Communities can represent groups of individuals with shared interests and interactions within human society, and they can be identified in networks based on connection patterns [24].

Hierarchical clustering is a popular approach to community detection that builds a hierarchy of partitions either by merging smaller communities into bigger ones or by dividing larger communities into smaller ones based on similarity measures. Initially, each node is considered a separate community, and similarity measures between pairs of communities are calculated using factors such as shared neighbors or connection strength. The algorithm then merges the most similar communities iteratively, creating a hierarchical structure [25].

Hierarchical clustering allows flexible exploration of community structure at different granularity levels, ranging from broad to specific communities. This approach provides a comprehensive view of the network's community structure, enabling the analysis of nested communities and different levels of detail. The FSC proposed was developed to solve community detection problems by maximizing modularity, and is a widely recognized and commonly used approach for discovering communities in structure networks. It involves finding the split of the network into clusters

that maximize the modularity value, which is a measure of the quality of the division.

2.2. Modularity. Modularity, a concept pioneered by [20], is a metric for assessing the efficacy of partitioning a graph into distinct communities. While its original formulation was designed for undirected graphs, subsequent research has broadened its applicability to encompass directed and weighted graphs [26 – 28]. The notion of modularity within a partition is expressed through a singular numerical measure (with a range reaching 1) used to evaluate the level of interconnectedness among components residing within communities, as opposed to the connections linking different communities together. A heightened positive modularity value indicates an enhanced organization of communities.

Modularity maximization is a widely used approach in network analysis for resolving community detection problems. Modularity quantifies the quality of the community structure in a network by comparing the number of edges each community has to the number expected by chance. The goal of modularity maximization is to find a partition of nodes into communities that maximize the modularity score. This involves iteratively assigning nodes to communities and evaluating the change in modularity until no further enhancement can be performed.

Modularity maximization has several advantages for resolving community detection problems. It is a flexible and scalable method that can be used in networks of various sizes and types. It can find the optimal division of the network into communities without prior knowledge of their number or size, allowing for an unbiased exploration of the community structure.

Maximizing modularity to find the best partition of a network into communities is not feasible for large and complex networks due to the NP-complete [29] nature of the problem. Several heuristic algorithms have been proposed to approximate the optimal solution. Despite its drawbacks, such as the resolution limit and high computational requirements, modularity maximization remains a popular and effective approach for community detection. The FSC performance will be evaluated by comparing the communities obtained using different algorithms with the ground truth and assessing the quality and reliability of community detection results through modularity scores.

2.3. Artificial Fish School Search Algorithm. The study of collective behavior in decentralized, self-organized systems, encompassing both natural and artificial systems, falls within the realm of swarm intelligence [30]. The inspiration behind swarm intelligence stems from observing the collective behavior exhibited in natural societies, including the coordinated movements of birds, fish, ants, bees, termites, and other

species. This behavior emerges from the interactions of numerous individuals acting in unison. It is a form of distributed problem-solving in which a group of individuals cooperates to achieve a common goal [31]. Algorithm 1, known as Artificial Fish School Search (FSS) [32], is a prime example of swarm intelligence, drawing inspiration from the collective behavior observed in fish schools. The foundational algorithmic structure of the FSS optimization algorithm, as outlined in [32], is as follows:

Algorithm 1. FSS optimization algorithm for community detection

1. Input:
2. Graph ($G = (V, E)$): The input graph with vertices (V) and edges (E).
3. Maximum Cycle Number: The maximum number of iterations for the algorithm.
4. Maximum CPU Time: The maximum allowed CPU time for the optimization process.
5. Output:
6. π : The resulting cluster assignment that maximizes modularity Q
7. Do
8. Attraction fish phase:
9. Move each fish towards the center of mass in the school.
10. Aggregation fish phase:
11. Increase the step size of each fish
12. Movement fish phase:
13. Move each fish based on its current position and step size.
14. Evaluate the modularity Q of each fish's cluster assignment
15. Save the best solution (cluster assignment π) found so far that maximizes Q .
16. While Cycle < Maximum Cycle Number and CPU Time < Maximum CPU Time
17. Return the best cluster assignment π that maximizes modularity Q .

The FSS algorithm is based on the idea that fish in a school move together, following some simple rules of behavior such as attraction, repulsion, and alignment. Within the domain of swarm intelligence, the Artificial Fish School Search (FSS) algorithm stands out as a notable exemplar. Taking cues from the coordinated behavior displayed by fish in their schools, the FSS algorithm derives its principles from this collective phenomenon. Like their aquatic counterparts, the algorithm envisions a scenario where virtual fish move in unison, guided by simple yet impactful behavioral rules encompassing attraction, repulsion, and alignment [33]. By harnessing these fundamental principles, the FSS algorithm aims to address complex problems through the power of collective intelligence. The process of fish locating their food typically unfolds in three stages: attraction, aggregation, and coordinated movement [34]. Initially, in the attraction

phase, fish employ a myriad of sensory cues, ranging from chemical to visual and auditory signals, to sense the availability of food. This phase often sees fish swimming in the direction of a positive chemical gradient, which is a reliable indicator of food's existence in the water. A plethora of fish species possess specialized sensory apparatus, like olfactory organs, that are exceptionally responsive to waterborne chemical cues. This sensitivity enables fish to discern even minor chemical traces released by potential food sources and subsequently follow this scent trail to discover food [35]. In the aggregation phase, fish may form schools or shoals to more efficiently locate and access the food source. In these groups, individuals may benefit from the presence of others in terms of increased feeding opportunities and protection from predators. In the last stage, referred to as the coordinated movement phase, fish within the collective meticulously adjust their movements to ensure the maintenance of group unity and synchronously approach the food source [36]. This often entails each fish attentively responding to their neighbors' actions, facilitating a seamless and efficient group movement. The creation of collective wisdom in fish schools and other kinds of animal swarms heavily hinges on effective information exchange. Fish utilize diverse sensory signals to communicate and disseminate information among each other. This information exchange often involves the use of auditory, visual, and tactile cues [37]. Based on the concept of information exchange, the Artificial Fish School Search (FSS) algorithm can be divided into four stages to find maximum modularity. These stages include an initialization phase, where the FSS algorithm starts by randomly searching for a food source, which is considered a potential solution. Following the initialization phase, the process transitions into the individual swimming phase, which is analogous to conducting a local search in the current location of each fish or solution. Each instance of individual swimming generates a novel candidate solution, guided randomly and exhibiting distinct values. Upon the conclusion of the individual phase, an evaluation or update of the fitness function is performed. If there's no enhancement in the fish's position, it is assumed that this specific fish or solution remains static. Only those fish or solutions showing improvements in their fitness functions will transition to a new position. During the instinctive-collective phase, the overall positioning of the school of fish or set of solutions is adjusted, taking into consideration the alterations in the fitness function of each fish or solution from the prior iteration [38]. The process culminates with the collective-volitional phase, where the fish or solutions are moved if there has been an improvement. In algorithmic terms, this signifies a refinement of the solution for the optimization problem. If there hasn't been an enhancement in the position of the entire school or set

of solutions compared to its previous position, it indicates the need for another round of food search or, in algorithmic terms, a return to the initialization phase.

2.4. Related Works. Community detection is a rapidly evolving field, with researchers continuously proposing new and innovative approaches to identify and analyze community structures in complex networks. These approaches range from leveraging the concept of modularity to using artificial intelligence algorithms for clustering. In this review, we will discuss some notable works in the literature that focus on harnessing the power of modularity and artificial intelligence algorithms for community detection.

The authors in [39] proposed a method for module partitioning in complex products using stable overlapping community detection and component allocation.

It effectively handles intricate component correlations, demonstrating superiority over existing methods on a CNC grinding machine. The study model (CDFSE) [40] presents a dynamic model simulating fish school behavior to showcase the formation of larger, stable groups based on shared attributes. Additionally, the study [41] introduces a new network clustering method using the Bee Colony Algorithm and Modularity Function. This method involves iterative edge removal for network partitioning and uses the Bee Colony Algorithm to optimize the modularity function, revealing optimized clusters. The study [42] highlights encord's role in enhancing the efficiency and accuracy of medical data annotation in cancer subtype classification for identifying patient clusters, advancing targeted therapies, and biomarker discovery. Automated and collaborative features in encord improve annotation speed and consistency, benefiting AI-driven diagnostics and treatment planning. Lastly, the study [43] introduces an algorithm for disjoint communities in complex networks. It's designed for undirected, unweighted graphs and uses cosine similarity for weight determination. It starts with individual nodes as communities and merges them based on modularity values.

Our proposed approach, FSC-Fish School Clustering, addresses gaps in previous work by offering a comprehensive solution for community detection in unipartite, unweighted, and undirected networks. While existing methods often have limited scope, our approach widens their applicability by specifically targeting these types of networks. Moreover, we introduce a dynamic modeling aspect inspired by fish school behavior, enabling the formation of larger, stable groups based on shared attributes. To overcome optimization challenges, we integrate the Fish School Search Algorithm with the Modularity function, enhancing accuracy and efficiency

in community detection. Additionally, scalability concerns are addressed by leveraging the efficiency of the Fish School Search Algorithm, making our approach suitable for handling large-scale networks. By bridging these gaps, our proposal aims to advance the field of network analysis and provide more accurate and efficient tools for understanding complex network structures.

The FSC provides a unique approach to optimization problems, drawing inspiration from natural phenomena. Its strengths lie in its adaptability and collective problem-solving power. It mimics fish school behavior to navigate complex search spaces effectively. However, it may lag behind modern network assembly methods in computational efficiency and scalability. While FSC is ideal for intricate problems requiring exploration, modern methods suit tasks demanding precision and large-scale network management. The choice between the two depends on the specific needs of the optimization challenge, balancing the strengths of FSC against the requirements for efficiency and control in network assembly.

3. The proposed approach (FSC-Fish School Clustering).

Hierarchical methods are clustering algorithms that can be either divisive or agglomerative. Agglomerative methods recursively merge clusters into larger clusters, while divisive methods recursively split clusters into smaller clusters. Divisive methods in community detection remove edges based on various criteria. For example, Edge Density [44] involves removing edges based on the density of connections to identify densely connected subgraphs as communities. Topological [45] measures utilize measures like centrality or clustering coefficient to determine edge importance; high centrality edges connecting communities are preserved, while low centrality edges are removed. Additionally, structural properties [46] involve removing edges that bridge distinct clusters to promote clearer separation between communities and identify internally connected communities. Some graph-splitting methods create separate partitions by deleting edges that connect vertices with very high or very low weights [47]. In this study, we propose a novel network clustering approach, referred to as FSC, which uses the Fish School Search Algorithm and the modularity function introduced by [20].

In this study, we propose a novel network clustering method that focuses on identifying community structures in unipartite, unweighted, and undirected networks. We utilize the Fish School Search Algorithm to iteratively split the network and remove edges while maximizing the modularity function. During this splitting phase, the algorithm discovers a set of clusters that represent the community structure. This process is repeated iteratively until each cluster consists of a single vertex. FSC leverages the modularity function as an objective metric to measure the

power of the community structure. It guides the determination of the optimal number of network communities (clusters). By maximizing modularity, FSC aims to identify the community structure that best fits the given network. This is done through the use of the Fish School Search Algorithm and the Modularity function. For a graph $G(V, E)$, our method identifies the most optimal community structure $\pi = \{c_1, \dots, c_{nc}\}$, where:

- $\bigcup_{i=1}^{nc} c_i = V$,
- $c_i \neq \emptyset$,
- $c_i \cap c_j = \emptyset$ if $i \neq j$,
- nc : number of clusters.

Modularity assesses the cohesion within different segments of a network by assigning a numerical value to each segment, known as a community structure. Modularity Q ranges between 0 and 1, where higher scores typically indicate better-defined partitions, while lower scores may imply less cohesive groupings. However, it's important to note that the quality of community structures can vary depending on the specific context. The modularity function is formulated around the comparison between the observed fraction, denoted as $e(c_i)$ of edges within communities, and the expected fraction, denoted as $a(c_i)$, of edges within the same communities. The modularity, represented as Q , is calculated using the expression:

$$Q = \sum_{c_i} e(c_i) - a(c_i)^2. \quad (1)$$

Consider an undirected, unweighted graph G , with n vertices and m edges, and a partition denoted $\pi = (c_1, \dots, c_{nc})$.

In this context, the modularity [48], can be expressed as follows:

$$Q = \frac{1}{(2m)} \sum_{vw} \left[A[v, w] - \frac{K_v K_w}{(2m)} \right] \delta(v, w), \quad (2)$$

where





$$\delta(v, w) = \begin{cases} 1 & \text{if } v \text{ and } w \text{ belong to the same community} \\ 0 & \text{otherwise} \end{cases},$$

where:

- v, w – are vertices of G ,
- K_v, K_w – the degree of v and w which is the number of edges linked to each vertex,
- m – total number of edges in the graph G ,
- A – the adjacency matrix of the graph G .

The overarching framework of our algorithm designed to uncover community structures within networks is outlined as follows:

Algorithm 2. FSC Algorithm

1. Input: $G = (V, E)$ – The input graph with vertices V and edges E .
2. Result cluster: π , maximizes Q , the resulting cluster assignment cluster.
3. Cluster $\leftarrow G$
4. $\pi \leftarrow$ Cluster
5.  repeat
6. Set_Cluster $\leftarrow \pi$
7.  for $i \in |$ Set_Cluster $|$ do
8. Cluster = Set_Cluster[i]
9. $\pi' \leftarrow$ Fish School Search (Cluster)
10. Divide Cluster based on π'
11. Update π'
12.  end for
13.  until $| \pi | \geq |V|$
14. return cluster that maximizes Q

The FSC algorithm, utilizing the Artificial Fish School Search (FSS) algorithm, iteratively partitions a network to maximize modularity. Starting with the entire graph as a single cluster, the algorithm refines cluster assignments through FSS, aiming to enhance modularity within each cluster. This iterative process continues until each cluster comprises a single node, culminating in the selection of the assignment with the highest modularity. The resulting clusters represent distinct communities, offering flexibility to tailor the number of communities. Additionally, we construct a tree diagram to visually depict the hierarchical organization of communities, elucidating the structural composition of the network.

4. Performance Assessment. We tested FSC on known and unknown network topologies. These networks are well-studied and have been used to benchmark other community detection algorithms. This means that we can compare our results to those of other algorithms and see how FSC performs. Networks tested include Karate Club, College Football, US Politics Books, and Dolphins. We additionally created a network using the Lancichinetti et al. random graph model, specifically designed to generate networks with desired community structures. This enables us to assess the performance of FSC on diverse network structures with varying communities. We conducted a comparative analysis of our approach against several established methods, such as:

- **The Infomap method** [49] identifies community structure by minimizing the description length of a random walker's navigation, using iterative node agglomeration to achieve improved partitions.

– **The label propagation method** [50] efficiently detects communities in complex networks by assigning unique labels to nodes and adopting the most prevalent label among neighbors, but its accuracy may suffer due to its random nature and potential tie situations.

– **The Louvain method** [51] optimizes modularity by iteratively reassigning nodes between communities to enhance network division strength.

– **The Fast Greedy method** [52] utilizes community analysis by greedily enhancing modularity scores through the continuous merging of community pairs until maximal improvements are achieved. The Fish School Search (FSS) algorithm chosen for network clustering stands out due to its unique nature-inspired approach that simulates the behavior of a school of fish, enabling it to explore the search space more effectively and efficiently. Unlike other algorithms such as Infomap, Label Propagation, Louvain, and Fast Greedy, FSS does not require a predefined number of clusters, making it highly adaptable to various datasets. It further optimizes the modularity function, a unique feature that quantifies the strength of the division of a network into modules or communities, enhancing the resolution and effectiveness of network clustering. Demonstrating robust performance, FSS excels in identifying community structures in both well-known and unknown network structures, making it a compelling, versatile, and reliable choice for network clustering. This dynamic modeling aspect, inspired by fish school behavior, forms larger, stable groups based on shared attributes, addressing gaps in previous work. It also integrates the Fish School Search Algorithm with the Modularity function, enhancing accuracy and efficiency in community detection and making it suitable for handling large-scale networks. In this part, we assess the performance and effectiveness of our method. This required two measures to compare and evaluate it against existing methods: normalized mutual information and computer-generated networks.

4.1. Normalized Mutual Information (NMI). We can leverage the NMI scale to compare the outcomes of our approach with those of various community detection methods. It is a widely used measure of community quality and is a good predictor of human-annotated ground-truth communities. In [53] proposed the utilization of the Normalized Mutual Information (NMI) metric for the comparative assessment of community-based screening techniques. NMI requires a confusion matrix M , where the rows in M indicate the real communities and the columns in M indicate the detected communities. Each element M, M_{ij} in the matrix represents the count of nodes belonging to the actual community i that are present in the discovered community j . This confusion matrix allows us to quantify the

overlap and agreement between the two sets of communities. The NMI measure employs information theory principles to evaluate the similarity between the partitions. It calculates the mutual information between the real and found communities, taking into account the distribution of nodes across the communities. The NMI measure captures the shared information and dependence between the partitions by considering the probabilities of nodes belonging to specific communities. The formula derived from data analysis principles [53] for measuring the likeness between two community structures A and B is as follows:

$$I(A, B) = \frac{-2 \sum_{i=1}^{c_A} \sum_{j=1}^{c_B} M_{ij} \log(M_{ij}M/M_i M_j)}{\sum_{i=1}^{c_A} M_i \log(M_i/M) + \sum_{j=1}^{c_B} M_j \log(M_j/M)}. \quad (3)$$

Here, c_A reflects the number of actual communities, while c_B represents the number of discovered communities. The total of the elements in the row i vector of the matrix M_{ij} is denoted by M_i , and the summation over column j is denoted by M_j . When our method's community structure perfectly aligns with the actual community structure, $I(A, B)$ achieves its maximum value of 1. This means that the two sets of clusters are a perfect fit, and the NMI value will be 1, indicating a complete match between the clusters. Conversely, if the detected communities have no resemblance to the actual communities in the network, the NMI value will be 0, signifying a complete lack of similarity. This means that the two sets of clusters have no connection or resemblance. When the discovered community structure is similar but not identical to the actual community structure, it means that the clusters share some commonalities while also having some differences; the NMI value will fall between 0 and 1, capturing the partial agreement between the two sets of clusters.

4.2. Computer-Generated Networks. To validate the effectiveness of our method, we can employ computer-generated grids as a means of testing. These computer-generated networks possess a predetermined community structure, which renders them highly suitable for evaluating the accuracy and robustness of community detection algorithms. By leveraging these networks, we can thoroughly assess the performance and reliability of FSC in accurately identifying and uncovering community structures. To create networks with a community structure, we utilized the LFR model proposed by [54]. The LFR model is a widely used framework for generating networks with diverse properties, including power-law degree distributions, community structure, and even overlapping communities. In our study, we employed the LFR model to generate a network consisting of 128 nodes. The

degree exponent distribution was set to 2, determining the number of links per node. The community size distribution exponent was 3, governing the sizes of the communities. The average degree was 16, showing the ratio of the actual connections to the potential connections in a network. The network consisted of 3 communities, and the mixing parameter varied from 0.1 to 0.9, influencing the interconnectivity between communities. In our method, the fitness function is defined as the modularity function. We set 500 (the number of fish) as food sources and specified the maximum number of iterations as 1000 for the optimization FSC.

5. Results and Discussion. In this section, we present an elaborate examination of the authentic social networks utilized to assess the effectiveness of our method. We have tested our method on the following networks: The Zachary Club Network [55], American College Football [56], The Dolphin Social Network [57], The Book about US Politics Network [58] Facebook [59], Amazon [60], Les Miserable Network [61], The Jazz Collaboration Network [62], The HIV Network [63], and The Contiguous USA [64] Network. Table 1 summarizes the fundamental characteristics of real benchmark networks, including the number of nodes ($|V|$), the number of edges ($|E|$), the average degree ($\langle k \rangle$), and whether the community structure (CS) is known or unknown.

Table 1. Summarizes characteristics of Benchmark Social Networks

Networks	CS	$ V $	$ E $	$\langle k \rangle$	Description
Zachary	Known	34	78	4.58	Derived from a university karate club.
Football	Known	115	613	10.66	Fall 2000 Division IA college games.
Dolphin	Known	62	159	5.12	62 New Zealand dolphins' interactions.
Book	Known	105	441	8.4	2004 election US politics book list.
Facebook	Unknown	4039	88234	18.02	Facebook social network dataset.
Amazon	Unknown	334863	925872	5.52	Amazon product co-purchasing network.
Miserable	Unknown	77	245	6.36	Victor Hugo's character network.
Jazz	Unknown	198	2742	27.69	Jazz collaborations from 1912-1940.
HIV	Unknown	40	41	2.05	Early HIV spread in USA contacts.
USA	Unknown	48	107	4.45	US shared border network with DC.

Table 2 presents a comparative analysis of the modularity metric for our proposed algorithm against four established algorithms. The modularity metric is calculated using the formula (2) (Section 3).

In Table 3, we present two metrics:

1. The number of clusters ($|C|$) is the number of clusters obtained after the iterative splitting process.

2. The normalization mutual information (NMI) is a measure of similarity between the clusters obtained using formula (3) (See section 4.1).

Table 2. Modularity for real networks with community structures

Methods	Q					
	Karate	Football	Books	Dolphins	Facebook	Amazon
Infomap	0.37	0.60	0.52	0.52	0.05	0.82
Label Propagation	0.37	0.57	0.47	0.51	0.65	0.78
Louvain	0.41	0.60	0.52	0.52	0.68	0.92
Fast Greedy	0.38	0.54	0.50	0.49	0.64	0.87
FSC	0.40	0.64	0.54	0.52	0.72	0.94

Table 3. NMI for real networks with community structures

Methods	Karate		Football		Books		Dolphins		Facebook		Amazon	
	$ C $	NMI	$ C $	NMI	$ C $	NMI	$ C $	NMI	$ C $	NMI	$ C $	NMI
Infomap	2	0.59	2	0.92	6	0.49	5	0.53	239	0.09	17296	0.1
Label Propagation	2	0.1	10	0.83	3	0.48	4	0.47	15	0.18	22496	0.01
Louvain	4	0.50	9	0.85	5	0.50	4	0.49	11	0.18	240	0.02
Fast Greedy	3	0.69	5	0.65	3	0.53	3	0.41	25	0.1	1532	0.1
FSC	2	0.87	7	0.1	3	0.60	4	0.81	150	0.74	260	0.1

To evaluate the performance of our proposed algorithm, we conducted experiments in a controlled environment using the Python programming language. Our algorithm's code was developed from scratch, and after multiple executions, we recorded the best modularity value achieved. In contrast, the baseline algorithms were sourced from the igraph library, which is well-known for its comprehensive collection of network analysis tools.

According to Table 2 and Table 3, FSC performed well, achieving high NMI and Q values. For example, on the Karate network, it detected 2 clustering with an NMI of 0.87 and a Q value of 0.40. On the football network, it detected 7 clusters with an NMI of 0.1 and a Q value of 0.64. FSC achieved the highest or second-highest NMI and Q scores on all four networks, suggesting it is more effective than other methods considered.

FSC identified meaningful communities in the karate network (Figure 1) and detected three clusters in the network of books about US politics (Figure 2), with high modularity and NMI values indicating good performance. This provides insights into social dynamics.

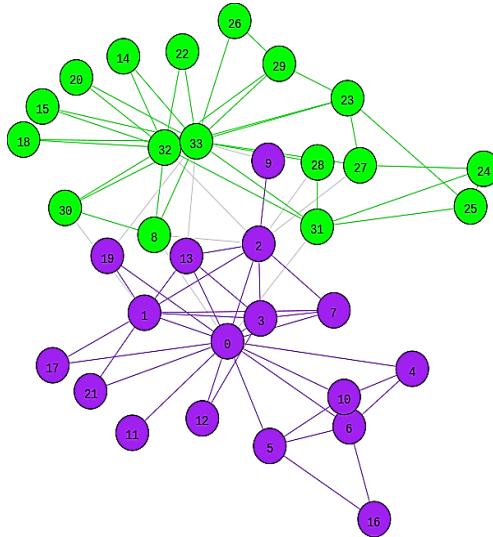


Fig. 1. FSC identifies 2 clusters in the Zachary club network structure

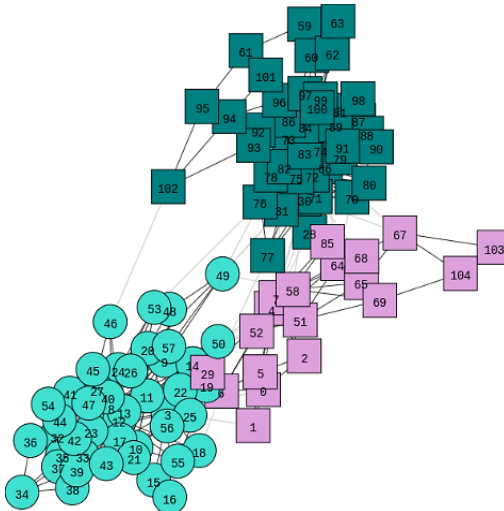
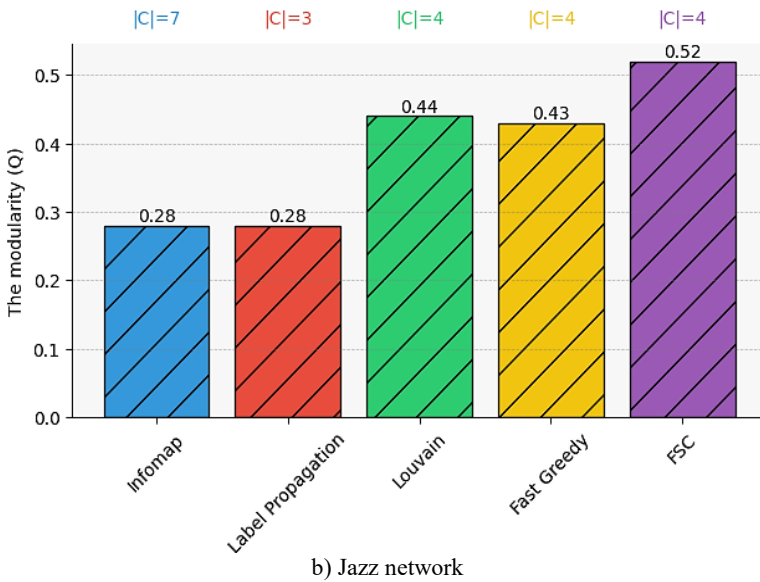
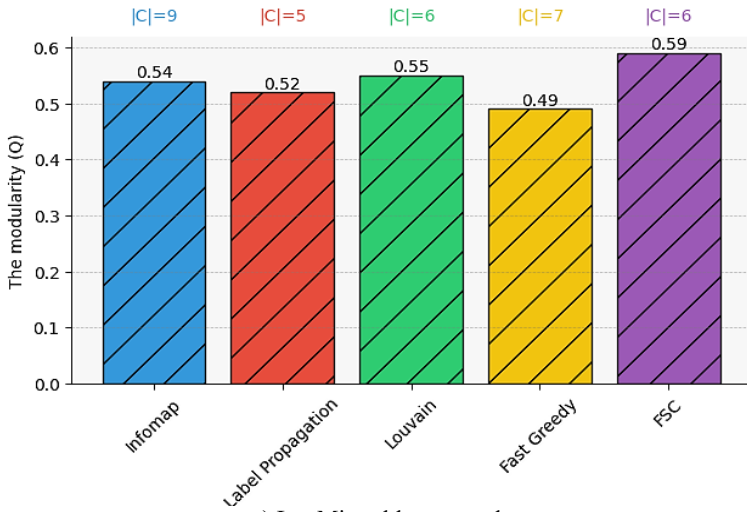
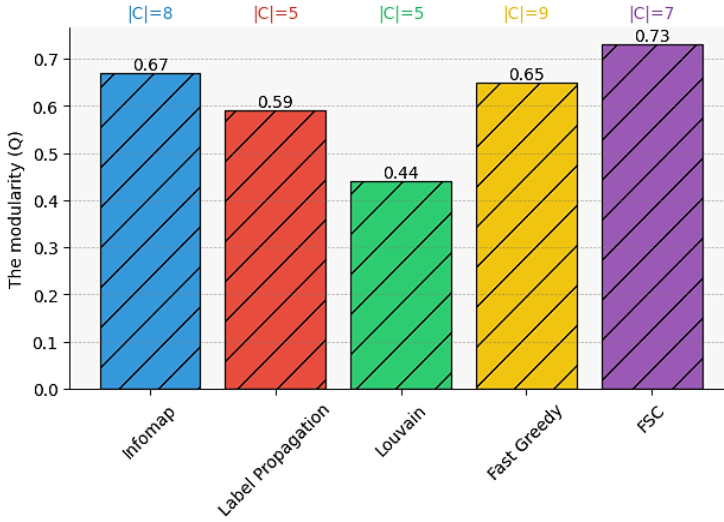


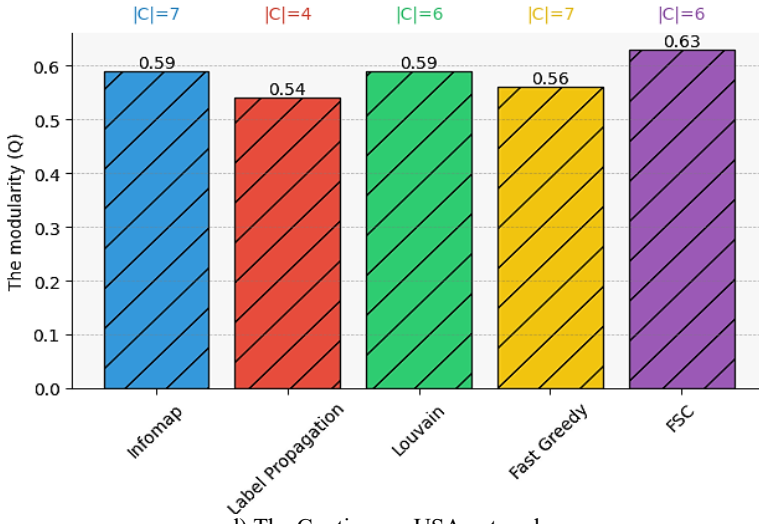
Fig. 2. FSC identifies 3 clusters in the structure of the Books about US politics

We tested FSC on various unknown networks (Les Miserable, Jazz, HIV, USA, Facebook, and Amazon), as shown in Figure 3.

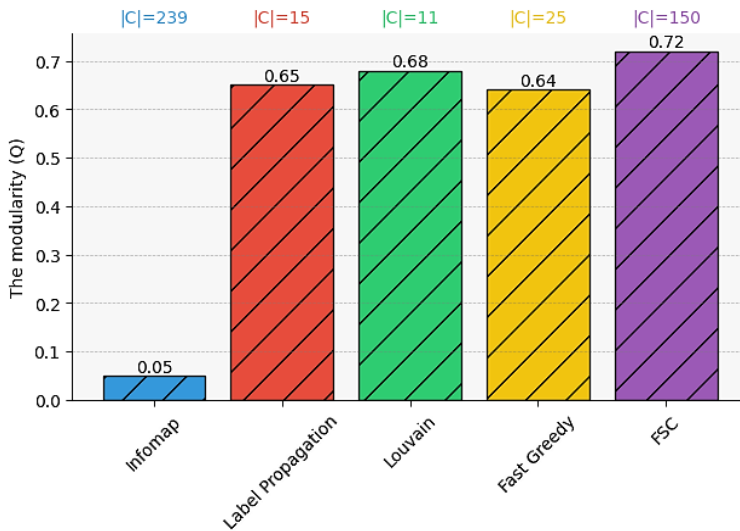




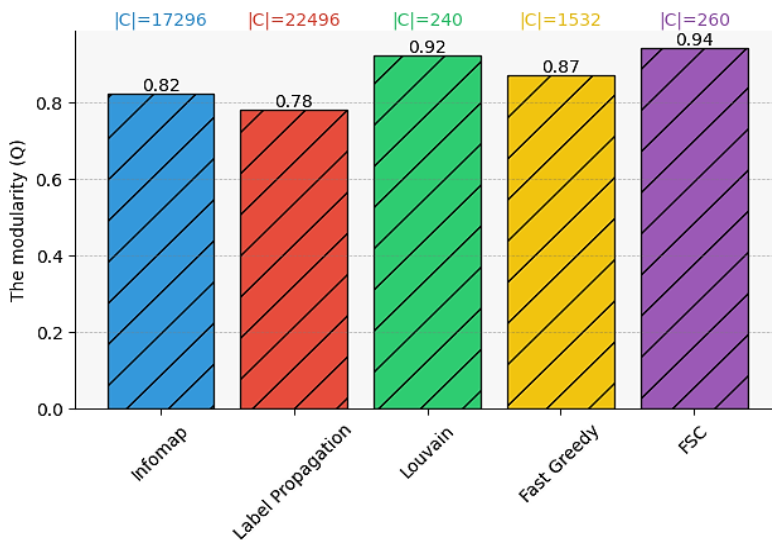
c) The HIV network



d) The Contiguous USA network



e) Facebook network



f) Amazon network

Fig. 3. Compares FSC with different unknown networks: a) Les Miserable Network; b) Jazz Network; c) HIV Network; d) USA Network; e) Facebook Network; f) Amazon Network

FSC outperformed alternative approaches in several cases, although the specific networks where it excelled varied across different contexts. The FSC approach has demonstrated exceptional performance when applied to the intricate networks of Facebook and Amazon, exhibiting commendable quality in community detection. Its success in the Amazon network particularly stands out, showcasing unparalleled excellence with Q index values surpassing 0.90. Moreover, within the Facebook network, the FSC method has achieved a remarkable Q index value of 0.72, significantly eclipsing the performance of alternative algorithms, as illustrated in the accompanying Figure 4. Figure 4 shows our novel method for partitioning the contiguous USA into six clusters, providing a new way to analyze regional patterns. Using a mixing parameter denoted as μ , FSC accurately partitions the graph and uncovers clear communities (Figure 5). This value of $\mu = 0.1$ is commonly used in community detection within networks because it indicates a clear community structure where only 10% of the edges are between different communities, while 90% of the edges are within the same community. The choice of $\mu = 0.1$ provides stability and robustness for accurate detection and has been empirically successful in various network studies [65]. By using this mixing parameter value, the FSC algorithm can effectively identify distinct communities within the graph, as shown in Figure 5.

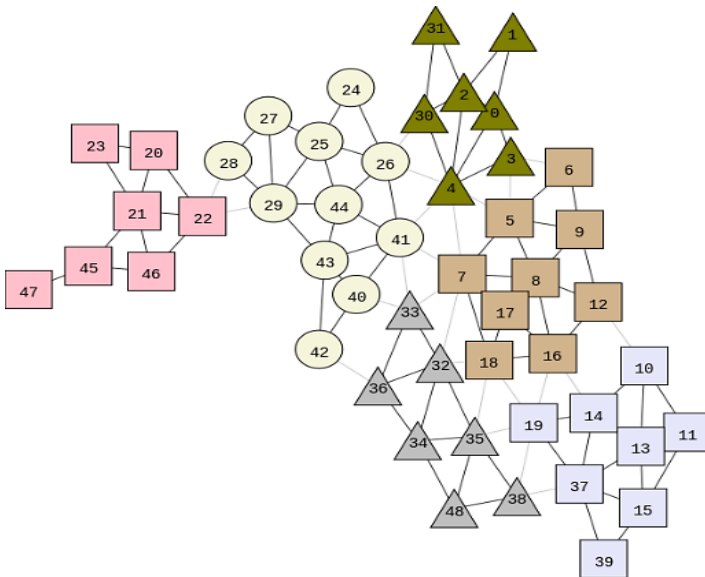


Fig. 4. FSC identifies 6 clusters in the Contiguous USA network

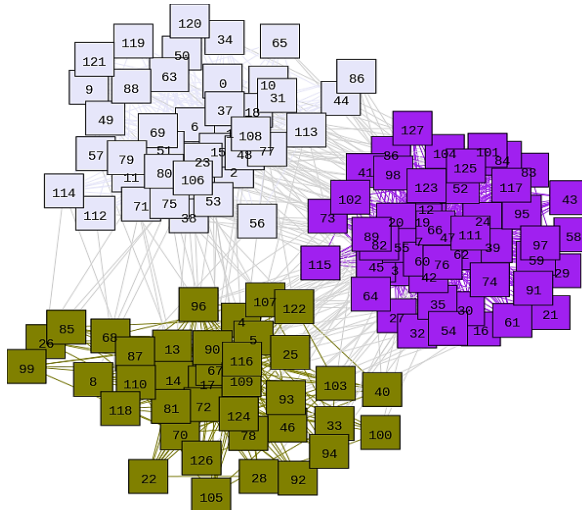


Fig. 5. FSC detects computer-generated networks at $\mu = 0.1$

Figure 6 shows the FSC achievement of a maximum NMI value of 1 when the mixing parameter ranges from 0 to 0.4, indicating the successful identification of robust and well-defined community structures. The connections within communities are denser compared to other methods, highlighting our approach's effectiveness. However, other methods like Fast Greedy and Louvain have NMI values less than 1, indicating accuracy in identifying true communities, particularly for mixing parameters between 0 and 0.3. As the parameter exceeds 0.3, both Infomap and Label Propagation encounter challenges in defining distinct communities. Overall, as the mixing parameter increases, all methods face difficulties in accurately uncovering the true community structure. This decline in performance suggests that as more edges are added between different communities, making the network more interconnected, the task of identifying clear divisions and separating the network into distinct communities becomes increasingly challenging for all methods. According to [52], a network's community structure is significant when its modularity exceeds 0.3. FSC observed a high modularity value, indicating a strong community structure and the ability to uncover densely connected and cohesive communities (Figure 7). The higher modularity values imply strong internal connections and fewer interconnections with other communities, indicating that FSC is effective in capturing meaningful community divisions and that the community structure is robust.

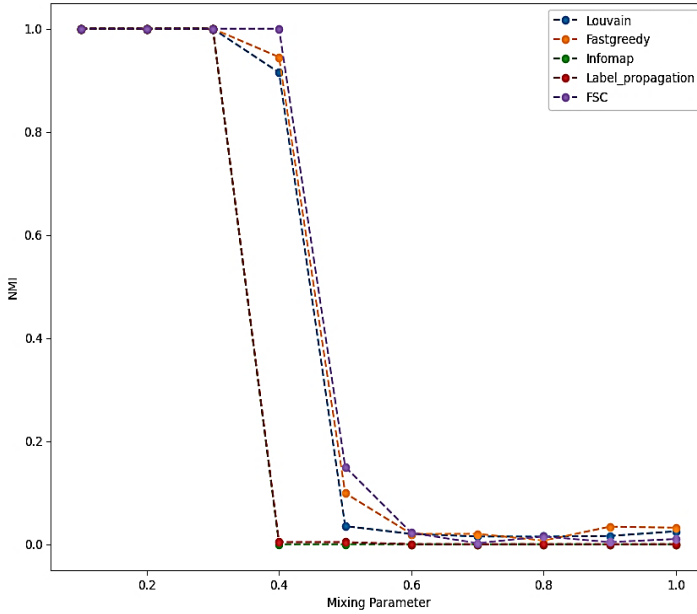


Fig. 6. How NMI varies with the level of community mixing

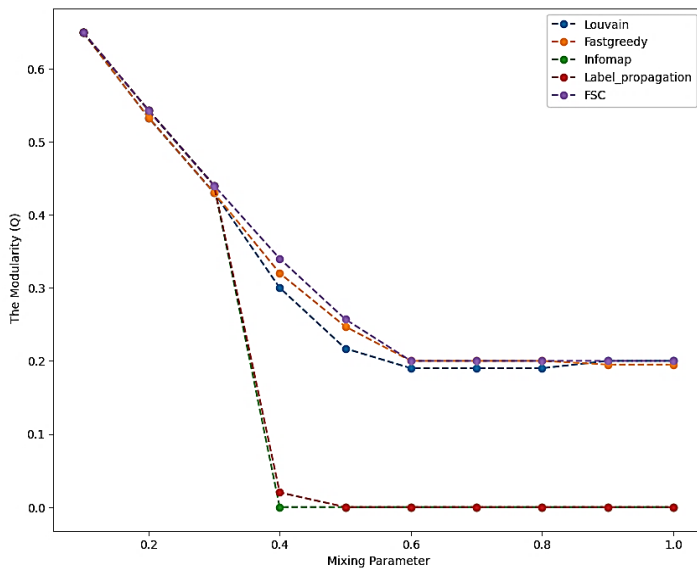


Fig. 7. How modularity changes with the degree of inter-community mixing

6. Conclusion and future prospects. We introduce a new method for clustering networks that uses the Fish School Search Algorithm to enhance the modularity function. The approach iteratively removes edges from the network to maximize the modularity function. Through rigorous testing on both established and novel network structures, including those generated by the LFR model, our FSC methodology has demonstrated exceptional prowess in discerning cohesive communities and pinpointing authentic community structures. This breakthrough signifies a leap forward in network analysis, offering a potent tool for researchers and practitioners alike to navigate the ever-evolving landscape of networks. As we stand on the brink of this new horizon, it is clear that the FSC method is more than just an algorithm; it is a beacon that guides us toward a deeper comprehension of the networks that encompass and connect us all. The implications are vast, from enhancing social network analysis to optimizing transportation systems and even unraveling the collaborative networks within our very cells. The future of network clustering is bright, and it is our conviction that the FSC method will be at the forefront of this transformative journey. In future work, FSC can be extended to handle diverse network types, including directed, bipartite, or weighted networks. Additionally, research can explore alternative optimization algorithms or modularity functions to enhance FSC's performance in identifying community structures. Incorporating side information, such as node attributes or interactions with other nodes, is also a promising avenue for improvement.

References

1. Pavlopoulos G. Using graph theory to analyze biological networks. Pavlopoulos et al. *BioData Mining*. 2011. vol. 4. pp. 1–27.
2. Sorokina M., Medigue C., Vallenet D. A new network representation of the metabolism to detect chemical transformation modules. *BMC Bioinformatics*. 2015. vol. 16. DOI: 10.1186/s12859-015-0809-4.
3. Clauset A., et al. The Structure and Function of Complex Networks. *Trends in Ecology and Evolution*. 2011. vol. 45. no. 2. p. 78.
4. Mathew J., Rejikumar K. Communication Structures, its graph representation and decomposition possibilities. *Journal of Physics: Conference Series*. 2020. DOI: 10.1088/1742-6596/1706/1/012050.
5. Sarker I. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN Computer Science*. 2021. vol. 2. no. 5. DOI: 10.1007/s42979-021-00765-8.
6. Lee M., Buckley C., Zhang X., Louhivuori L., Uhlen P., Wilson C., McCarron J. Small-world connectivity dictates collective endothelial cell signaling. *Proceedings of the National Academy of Sciences*. 2022. vol. 119. no. 18. DOI: 10.1073/PNAS.2118927119.
7. Hoffman M., Steinley D., Gates K., Prinstein M., Brusco M. Detecting Clusters/Communities in Social Networks. *Multivariate behavioral research*. 2018. vol. 53. no. 1. p. 57–73. DOI: 10.1080/00273171.2017.1391682.

8. Felmlee D., Kreager D. The invisible contours of online dating communities: A social network perspective. *Journal of Social Structure*. 2017. vol. 18. no. 1. pp. 1–28. DOI: 10.21307/JOSS-2018-004.
9. Inuwa-Dutse I., Liptrott M., Korkontzelos I. A multilevel clustering technique for community detection. *Neurocomputing*. 2021. vol. 441. pp. 64–78. DOI: 10.1016/J.NEUCOM.2021.01.059.
10. Bramson A., Hori M., Zha B., Inamoto H. Scoring and classifying regions via multimodal transportation networks. *Applied Network Science*. 2019. vol. 4. no. 1. DOI: 10.1007/S41109-019-0191-7.
11. Copic J., Jackson M., Kirman A. Identifying Community Structures from Network Data via Maximum Likelihood Methods. *The BE Journal of Theoretical Economics*. 2009. vol. 9. no. 1. DOI: 10.2202/1935-1704.1523.
12. Yang L., Cao X., He D., Wang C., Wang X., Zhang W. Modularity based community detection with deep learning. *International Joint Conference on Artificial Intelligence*. 2016. vol. 2016. pp. 2252–2258.
13. Roy S. Spectral clustering of graphs. 2018. pp. 1–65.
14. Cohen-Addad V., Kanade V., Mallmann-trenn F., Mathieu C., Hierarchical Clustering: Objective Functions and Algorithms. *Journal of the ACM*. 2019. vol. 66. no. 4. DOI: 10.1145/3321386.
15. Ghosh S., Halappanavar M., Tumeo A., Kalyanaraman A. Scaling and Quality of Modularity Optimization Methods for Graph Clustering. 2019. pp. 1–6. DOI: 10.1109/HPEC.2019.8916299.
16. Zhang X. et al. Modularity optimization in community detection of complex networks. *Europhysics Letters*. 2009. vol. 87. no. 3. DOI: 10.1209/0295-5075/87/38002.
17. Brandes U., Delling D., Gaertler M., Gorke R., Hoefer M., Nikoloski Z., Wagner D. Maximizing Modularity is hard. *arXiv preprint physics/0608255*. 2006.
18. Bastos Filho C., de Lima Neto F., Lins A., Nascimento A., Lima M. Fish school search. Nature-inspired algorithms for optimisation. 2009. vol. 193. pp. 261–277. DOI: 10.1007/978-3-642-00267-0_9.
19. Bastos-Filho C., Guimarães A. Multi-Objective Fish School Search. *International Journal of Swarm Intelligence Research*. 2015. vol. 6. no. 1. pp. 23–40. DOI: 10.4018/IJSIR.2015010102.
20. Newman M., Girvan M. Finding and evaluating community structure in networks. *Physical Review E*. 2004. vol. 69. no. 2. DOI: 10.1103/PhysRevE.69.026113.
21. Lukac Z. Metaheuristic optimization. *Proceedings of the 11th International Symposium on Operational Research in Slovenia, SOR 2011*. 2011. pp. 17–22. DOI: 10.4249/SCHOLARPEDIA.11472.
22. Kim J., Luo S., Cong G., Yu W. DMCS: Density Modularity based Community Search. *Proceedings of the ACM SIGMOD International Conference on Management of Data*. 2022. pp. 889–903. DOI: 10.1145/3514221.3526137.
23. Tantardini M., Ieva F., Tajoli L., Piccardi C. Comparing methods for comparing networks. *Scientific Reports* 2019vol. 9. no. 1. DOI: 10.1038/s41598-019-53708-y.
24. Rustamaji H., Kusuma W., Nurdianti S., Batubara I. Community detection with greedy modularity disassembly strategy. *Scientific Reports*. 2024. vol. 14. no. 1. DOI: 10.1038/s41598-024-55190-7.
25. Drazdilová P., Prokop P., Platoš J., Snášel V. A hierarchical overlapping community detection method based on closed trail distance and maximal cliques. *Information Sciences*. 2024. vol. 662. DOI: 10.1016/J.INS.2024.120271.
26. Newman M. Fast algorithm for detecting community structure in networks. *Physical Review E – Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*. 2004. vol. 69. no. 6. DOI: 10.1103/PhysRevE.69.066133.

27. Dorfler F., Bullo F. Kron reduction of graphs with applications to electrical networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2012. vol. 60. no. 1. pp. 150–163. DOI: 10.1109/TCSI.2012.2215780.
28. Bickel P., Chen A. A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences*. 2009. vol. 106. no. 50. pp. 21068–21073. DOI: 10.1073/pnas.0907096106.
29. Brandes U., Dellinger D., Gaertler M., Hofer R., M Nikoloski Z., Wagner D. On Modularity Clustering. *IEEE Transactions on knowledge and data engineering*. 2008. vol. 20. no. 2. pp. 172–188. DOI: 10.1109/TKDE.2007.190689.
30. Hu Y. Swarm intelligence Ant colony optimization algorithms. 2012. pp. 1–11.
31. Dehuri S., Ghosh S., Coello C. An introduction to swarm intelligence for multi-objective problems. *Swarm Intelligence for Multi-objective Problems in Data Mining. Studies in Computational Intelligence*. 2009. vol. 242. DOI: 10.1007/978-3-642-03625-5_1.
32. Herbert-Read J., Perna A., Mann R., Schaerf T., Sumpter D., Ward A. Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences*. 2011. vol. 108. no. 46. pp. 18726–18731. DOI: 10.1073/PNAS.1109355108.
33. A novel search algorithm based on fish school behavior. 2008 IEEE International Conference on Systems, Man and Cybernetics. Available at: <https://scihub.se/10.1109/ICSMC.2008.4811695> (accessed: 14/05/2024).
34. Molina D., Poyatos J., Del Ser J., García S., Hussain A., Herrera F. *Comprehensive Taxonomies of Nature- and Bio-inspired Optimization: Inspiration Versus Algorithmic Behavior, Critical Analysis Recommendations. Cognitive Computation*. 2020. vol. 12. pp. 897–939. DOI: 10.1007/S12559-020-09730-8.
35. Abangan A., Kopp D., Faillettaz R. Artificial intelligence for fish behavior recognition may unlock fishing gear selectivity. *Front Mar Sci*. 2023. vol. 10. DOI: 10.3389/FMARS.2023.1010761.
36. Krongauz D., Lazebnik T. Collective evolution learning model for visionbased collective motion with collision avoidance. *PLoS One*. 2023. vol. 18. no. 5. DOI: 10.1371/JOURNAL.PONE.0270318.
37. Malone T. Collective intelligence. 2007 International Symposium on Collaborative Technologies and Systems. 2008. DOI: 10.1109/CTS.2007.4621716.
38. Huang Z., Application of the Artificial Fish School Algorithm and Particle Filter Algorithm in the Industrial Process Control Particle Filtering Algorithm for Industrial Process Control. *Math Probl Eng*. 2020. vol. 2020. DOI: 10.1155/2020/3070539.
39. Liu Z., Zhong P., Liu H., Jia W., Sa G., Tan J. Module partition for complex products based on stable overlapping community detection and overlapping component allocation. *Research in Engineering Design*. 2024. vol. 35. pp. 269–288. DOI: 10.1007/S00163-024-00432-Y.
40. Sun Y., Sun Z., Chang X., Pan Z., Luo L. Community Detection Based on Fish School Effect. *IEEE Access*. 2022. vol. 10. pp. 48523–48538. DOI: 10.1109/ACCESS.2022.3172298.
41. Saoud B. Networks clustering with bee colony. *Artificial Intelligence Review*. 2019. vol. 52. no. 2. pp. 1297–1309. DOI: 10.1007/s10462-018-9657-8.
42. Encord E. Available at: https://encord.com/try-it-free/?&utm_campaign=cta-blog-medical-dark (accessed: 10/03/2024).
43. Cetin P., Amrahov Ş. A new network-based community detection algorithm for disjoint communities. *Turkish Journal of Electrical Engineering and Computer Sciences*. 2022. vol. 30. no. 6. pp. 2190–2205. DOI: 10.55730/1300-0632.3933.
44. Bonchi F., García-Soriano D., Miyauchi A., Tsourakakis C. Finding densest k-connected subgraphs. *Discrete Applied Mathematics*. 2021. vol. 305. pp. 34–47. DOI: 10.1016/J.DAM.2021.08.032.

45. Oldham S., Fulcher B., Parkes L., Arnatkeviciute A., Suo C., Fornito A. Consistency and differences between centrality measures across distinct classes of networks. *PLoS One*. 2019. vol. 14. no. 7. DOI: 10.1371/JOURNAL.PONE.0220061.
46. Leskovec J., Lang K., Dasgupta A., Mahoney M. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. *Internet Mathematics*. 2008. vol. 6. no. 1. pp. 29–123.
47. Adoni W., Nahhal T., Krichen M., El byed A., Assayad I. DHPV: a distributed algorithm for large-scale graph partitioning. *Journal of Big Data*. 2020. vol. 7. DOI: 10.1186/S40537-020-00357-Y.
48. Newman M. Modularity and community structure in networks. Available at: <https://sci-hub.se/10.1073/pnas.0601602103> (accessed: 14/05/2024).
49. Rosvall M., Bergstrom C. Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A*. 2008. vol. 105. no. 4. pp. 1118–1123. DOI: 10.1073/PNAS.0706851105.
50. Cordasco G., Gargano L. Label propagation algorithm: a semi-synchronous approach. *International Journal of Social Network Mining*. 2012. vol. 1. no. 1. DOI: 10.1504/IJSNM.2012.045103.
51. Blondel V., Guillaume J.-L., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*. 2008. vol. 2008. no. 10. DOI: 10.1088/1742-5468/2008/10/P10008.
52. Clauset A., Newman M., Moore C. Finding community structure in very large networks. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics*. 2004. vol. 70. no. 6. DOI: 10.1103/PhysRevE.70.066111.
53. Danon L., Diaz-Guilera A., Duch J., Arenas A. Comparing community structure identification. *Journal of statistical mechanics: Theory and experiment*. 2005. vol. 2005. no. 09. DOI: 10.1088/1742-5468/2005/09/P09008.
54. Lancichinetti A., Fortunato S., Radicchi F. Benchmark graphs for testing community detection algorithms. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics*. 2008. vol. 78. no. 4. DOI: 10.1103/PHYSREVE.78.046110.
55. Zachary W. An Information Flow Model for Conflict and Fission in Small Groups. *Journal of anthropological research*. 1977. vol. 33. no. 4. pp. 452–473. DOI: 10.1086/JAR.33.4.3629752.
56. Girvan M., Newman M. Community structure in social and biological networks. *Proceedings of the national academy of sciences*. 2002. vol. 99. no. 12. pp. 7821–7826. DOI: 10.1073/pnas.122653799.
57. Lusseau D., Schneider K., Boisseau O., Haase P., Slooten E., Dawson S. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: Can geographic isolation explain this unique trait? *Behavioral Ecology and Sociobiology*. 2003. vol. 54. pp. 396–405. DOI: 10.1007/S00265-003-0651-Y.
58. Newman M. Network data. Network data collection. 2013. Available at: <http://www-personal.umich.edu/~mej>. (accessed 26.10.2023).
59. Leskovec J., Mcauley J. Learning to Discover Social Circles in Ego Networks. *Advances in neural information processing systems*. 2012. vol. 25.
60. Leskovec J., Lada A., Huberman B. The Dynamics of Viral Marketing. *ACM Transactions on the Web (TWEB)*. 2007. vol. 1. no. 1. DOI: 10.1145/1232722.1232727.
61. Knuth D. *The Stanford GraphBase: a platform for combinatorial computing*. New York: ACM Press, Addison-Wesley Publishing Company. 2009. 577 p.
62. Gleiser P., Danon L. Community structure in jazz. *Advances in complex systems*. 2003. vol. 6. no. 04. pp. 565–573.

63. Kunegis J. Konect: The Koblenz Network Collection. Proceedings of the 22nd international conference on world wide web. 2013. pp. 1343–1350. DOI: 10.1145/2487788.248817.
64. Knuth D. The Art of Computer Programming: Volume 4 Fascicle 0: Introduction to Combinatorial Algorithms and Boolean Functions. Addison-Wesley Professional, 2008. 228 p.
65. Yang Z., Algesheimer R., Tessone C. A comparative analysis of community detection algorithms on artificial networks. Scientific reports. 2016. vol. 6. no. 1 DOI: 10.1038/srep30750.

Ibrahim Abuzer — Researcher, Department of computer science, Lim Laboratory, Akli Mohand Oulhadj University of Bouira. Research interests: computer networking, network communication, networking, network architecture, routing, information security. The number of publications — 1. h.abouzer@univ-bouira.dz; Rue Drissi Yahia Bouira, 10000, Bouira, Algeria; office phone: +213(558)606-511.

Boudref Mohamed — Researcher, Department of mathematics, Lim Laboratory, Akli Mohand Oulhadj University of Bouira. Research interests: computer networking, network communication, networking, network architecture, routing, information security. The number of publications — 3. m.boudref@univ-bouira.dz; Rue Drissi Yahia Bouira, 10000, Bouira, Algeria; office phone: +213(550)251-333.

Badis Lyes — Researcher, Department of computer science, Lim Laboratory, Akli Mohand Oulhadj University of Bouira. Research interests: computer networking, network communication, networking, network architecture, routing, information security. The number of publications — 7. lbadis@univ-bouira.dz; Rue Drissi Yahia Bouira, 10000, Bouira, Algeria; office phone: +213(659)564-064.

А.Х. ИБРАГИМ, М.А. БУДРЕФ, Л. БАДИС
**КЛАСТЕРИЗАЦИЯ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ
АЛГОРИТМА ПОИСКА КОСЯКОВ РЫБ**

Ибрагим А.Х., Будреф М.А., Бадис Л. Кластеризация сетей с использованием алгоритма поиска косяков рыб.

Аннотация. Сеть представляет собой совокупность узлов, соединенных ребрами, которые представляют сущности и их взаимосвязи. В кластеризации социальных сетей узлы организованы в кластеры в соответствии с их шаблонами соединений с целью обнаружения сообществ. Выявление структур сообществ в сетях является важным. Однако существующие методы обнаружения сообществ еще не использовали потенциал алгоритма поиска косяков рыб (FSS) и принципов модулярности. Мы предложили новый метод, основанный на кластеризации с использованием алгоритма поиска рыбной школы и функции модулярности (FSC), который улучшает модулярность в кластеризации сети путем итерационного разбиения сети и оптимизации функции модулярности. Этот подход облегчает обнаружение высокомодулярных структур сообществ, улучшая разрешение и эффективность кластеризации сети. Мы протестировали FSC на известных и неизвестных структурах сетей. Также мы протестировали его на сети, сгенерированной с использованием модели LFR, чтобы проверить его производительность на сетях с различными структурами сообществ. Наша методология демонстрирует высокую эффективность в выявлении структур сообществ, что указывает на ее способность эффективно захватывать сплоченные сообщества и точно определять фактические структуры сообществ.

Ключевые слова: кластеризация, алгоритм поиска косяков рыб, функция модулярности, сетевые структуры.

Литература

1. Pavlopoulos G. Using graph theory to analyze biological networks. Pavlopoulos et al. *BioData Mining*. 2011. vol. 4, pp. 1–27.
2. Sorokina M., Medigue C., Vallenet D. A new network representation of the metabolism to detect chemical transformation modules. *BMC Bioinformatics*. 2015. vol. 16. DOI: 10.1186/s12859-015-0809-4.
3. Clauset A., et al. The Structure and Function of Complex Networks. *Trends in Ecology and Evolution*. 2011. vol. 45. no. 2. p. 78.
4. Mathew J., Rejikumar K. Communication Structures, its graph representation and decomposition possibilities. *Journal of Physics: Conference Series*. 2020. DOI: 10.1088/1742-6596/1706/1/012050.
5. Sarker I. Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN Computer Science*. 2021. vol. 2. no. 5. DOI: 10.1007/s42979-021-00765-8.
6. Lee M., Buckley C., Zhang X., Louhivuori L., Uhlen P., Wilson C., McCarron J. Small-world connectivity dictates collective endothelial cell signaling. *Proceedings of the National Academy of Sciences*. 2022. vol. 119. no. 18. DOI: 10.1073/PNAS.2118927119.
7. Hoffman M., Steinley D., Gates K., Prinstein M., Brusco M. Detecting Clusters/Communities in Social Networks. *Multivariate behavioral research*. 2018. vol. 53. no. 1. p. 57–73. DOI: 10.1080/00273171.2017.1391682.

8. Felmlee D., Kreager D. The invisible contours of online dating communities: A social network perspective. *Journal of Social Structure*. 2017. vol. 18. no. 1. pp. 1–28. DOI: 10.21307/JOSS-2018-004.
9. Inuwa-Dutse I., Liptrott M., Korkontzelos I. A multilevel clustering technique for community detection. *Neurocomputing*. 2021. vol. 441. pp. 64–78. DOI: 10.1016/J.NEUCOM.2021.01.059.
10. Bramson A., Hori M., Zha B., Inamoto H. Scoring and classifying regions via multimodal transportation networks. *Applied Network Science*. 2019. vol. 4. no. 1. DOI: 10.1007/S41109-019-0191-7.
11. Copic J., Jackson M., Kirman A. Identifying Community Structures from Network Data via Maximum Likelihood Methods. *The BE Journal of Theoretical Economics*. 2009. vol. 9. no. 1. DOI: 10.2202/1935-1704.1523.
12. Yang L., Cao X., He D., Wang C., Wang X., Zhang W. Modularity based community detection with deep learning. *International Joint Conference on Artificial Intelligence*. 2016. vol. 2016. pp. 2252–2258.
13. Roy S. Spectral clustering of graphs. 2018. pp. 1–65.
14. Cohen-Addad V., Kanade V., Mallmann-trenn F., Mathieu C., Hierarchical Clustering: Objective Functions and Algorithms. *Journal of the ACM*. 2019. vol. 66. no. 4. DOI: 10.1145/3321386.
15. Ghosh S., Halappanavar M., Tumeo A., Kalyanaraman A. Scaling and Quality of Modularity Optimization Methods for Graph Clustering. 2019. pp. 1–6. DOI: 10.1109/HPEC.2019.8916299.
16. Zhang X. et al. Modularity optimization in community detection of complex networks. *Europhysics Letters*. 2009. vol. 87. no. 3. DOI: 10.1209/0295-5075/87/38002.
17. Brandes U., Delling D., Gaertler M., Gorke R., Hoefer M., Nikoloski Z., Wagner D. Maximizing Modularity is hard. *arXiv preprint physics/0608255*. 2006.
18. Bastos Filho C., de Lima Neto F., Lins A., Nascimento A., Lima M. Fish school search. Nature-inspired algorithms for optimisation. 2009. vol. 193. pp. 261–277. DOI: 10.1007/978-3-642-00267-0_9.
19. Bastos-Filho C., Guimarães A. Multi-Objective Fish School Search. *International Journal of Swarm Intelligence Research*. 2015. vol. 6. no. 1. pp. 23–40. DOI: 10.4018/IJSIR.2015010102.
20. Newman M., Girvan M. Finding and evaluating community structure in networks. *Physical Review E*. 2004. vol. 69. no. 2. DOI: 10.1103/PhysRevE.69.026113.
21. Lukac Z. Metaheuristic optimization. *Proceedings of the 11th International Symposium on Operational Research in Slovenia, SOR 2011*. 2011. pp. 17–22. DOI: 10.4249/SCHOLARPEDIA.11472.
22. Kim J., Luo S., Cong G., Yu W. DMCS: Density Modularity based Community Search. *Proceedings of the ACM SIGMOD International Conference on Management of Data*. 2022. pp. 889–903. DOI: 10.1145/3514221.3526137.
23. Tantardini M., Ieva F., Tajoli L., Piccardi C. Comparing methods for comparing networks. *Scientific Reports* 2019vol. 9. no. 1. DOI: 10.1038/s41598-019-53708-y.
24. Rustamaji H., Kusuma W., Nurdianti S., Batubara I. Community detection with greedy modularity disassembly strategy. *Scientific Reports*. 2024. vol. 14. no. 1. DOI: 10.1038/s41598-024-55190-7.
25. Drazdilová P., Prokop P., Platoš J., Snášel V. A hierarchical overlapping community detection method based on closed trail distance and maximal cliques. *Information Sciences*. 2024. vol. 662. DOI: 10.1016/J.INS.2024.120271.
26. Newman M. Fast algorithm for detecting community structure in networks. *Physical Review E – Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*. 2004. vol. 69. no. 6. DOI: 10.1103/PhysRevE.69.066133.

27. Dorfler F., Bullo F. Kron reduction of graphs with applications to electrical networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2012. vol. 60. no. 1. pp. 150–163. DOI: 10.1109/TCSI.2012.2215780.
28. Bickel P., Chen A. A nonparametric view of network models and Newman-Girvan and other modularities. *Proceedings of the National Academy of Sciences*. 2009. vol. 106. no. 50. pp. 21068–21073. DOI: 10.1073/pnas.0907096106.
29. Brandes U., Dellinger D., Gaertler M., Hofer R., M Nikoloski Z., Wagner D. On Modularity Clustering. *IEEE Transactions on knowledge and data engineering*. 2008. vol. 20. no. 2. pp. 172–188. DOI: 10.1109/TKDE.2007.190689.
30. Hu Y. Swarm intelligence Ant colony optimization algorithms. 2012. pp. 1–11.
31. Dehuri S., Ghosh S., Coello C. An introduction to swarm intelligence for multi-objective problems. *Swarm Intelligence for Multi-objective Problems in Data Mining. Studies in Computational Intelligence*. 2009. vol. 242. DOI: 10.1007/978-3-642-03625-5_1.
32. Herbert-Read J., Perna A., Mann R., Schaerf T., Sumpter D., Ward A. Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences*. 2011. vol. 108. no. 46. pp. 18726–18731. DOI: 10.1073/PNAS.1109355108.
33. A novel search algorithm based on fish school behavior. 2008 IEEE International Conference on Systems, Man and Cybernetics. Available at: <https://scihub.se/10.1109/ICSMC.2008.4811695> (accessed: 14/05/2024).
34. Molina D., Poyatos J., Del Ser J., García S., Hussain A., Herrera F. *Comprehensive Taxonomies of Nature- and Bio-inspired Optimization: Inspiration Versus Algorithmic Behavior, Critical Analysis Recommendations. Cognitive Computation*. 2020. vol. 12. pp. 897–939. DOI: 10.1007/S12559-020-09730-8.
35. Abangan A., Kopp D., Faillettaz R. Artificial intelligence for fish behavior recognition may unlock fishing gear selectivity. *Front Mar Sci*. 2023. vol. 10. DOI: 10.3389/FMARS.2023.1010761.
36. Krongauz D., Lazebnik T. Collective evolution learning model for visionbased collective motion with collision avoidance. *PLoS One*. 2023. vol. 18. no. 5. DOI: 10.1371/JOURNAL.PONE.0270318.
37. Malone T. Collective intelligence. 2007 International Symposium on Collaborative Technologies and Systems. 2008. DOI: 10.1109/CTS.2007.4621716.
38. Huang Z., Application of the Artificial Fish School Algorithm and Particle Filter Algorithm in the Industrial Process Control Particle Filtering Algorithm for Industrial Process Control. *Math Probl Eng*. 2020. vol. 2020. DOI: 10.1155/2020/3070539.
39. Liu Z., Zhong P., Liu H., Jia W., Sa G., Tan J. Module partition for complex products based on stable overlapping community detection and overlapping component allocation. *Research in Engineering Design*. 2024. vol. 35. pp. 269–288. DOI: 10.1007/S00163-024-00432-Y.
40. Sun Y., Sun Z., Chang X., Pan Z., Luo L. Community Detection Based on Fish School Effect. *IEEE Access*. 2022. vol. 10. pp. 48523–48538. DOI: 10.1109/ACCESS.2022.3172298.
41. Saoud B. Networks clustering with bee colony. *Artificial Intelligence Review*. 2019. vol. 52. no. 2. pp. 1297–1309. DOI: 10.1007/s10462-018-9657-8.
42. Encord E. Available at: https://encord.com/try-it-free/?&utm_campaign=cta-blog-medical-dark (accessed: 10/03/2024).
43. Cetin P., Amrahov Ş. A new network-based community detection algorithm for disjoint communities. *Turkish Journal of Electrical Engineering and Computer Sciences*. 2022. vol. 30. no. 6. pp. 2190–2205. DOI: 10.55730/1300-0632.3933.
44. Bonchi F., García-Soriano D., Miyauchi A., Tsourakakis C. Finding densest k-connected subgraphs. *Discrete Applied Mathematics*. 2021. vol. 305. pp. 34–47. DOI: 10.1016/J.DAM.2021.08.032.

45. Oldham S., Fulcher B., Parkes L., Arnatkeviciute A., Suo C., Fornito A. Consistency and differences between centrality measures across distinct classes of networks. *PLoS One*. 2019. vol. 14. no. 7. DOI: 10.1371/JOURNAL.PONE.0220061.
46. Leskovec J., Lang K., Dasgupta A., Mahoney M. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. *Internet Mathematics*. 2008. vol. 6. no. 1. pp. 29–123.
47. Adoni W., Nahhal T., Krichen M., El byed A., Assayad I. DHPV: a distributed algorithm for large-scale graph partitioning. *Journal of Big Data*. 2020. vol. 7. DOI: 10.1186/S40537-020-00357-Y.
48. Newman M. Modularity and community structure in networks. Available at: <https://sci-hub.se/10.1073/pnas.0601602103> (accessed: 14/05/2024).
49. Rosvall M., Bergstrom C. Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A*. 2008. vol. 105. no. 4. pp. 1118–1123. DOI: 10.1073/PNAS.0706851105.
50. Cordasco G., Gargano L. Label propagation algorithm: a semi-synchronous approach. *International Journal of Social Network Mining*. 2012. vol. 1. no. 1. DOI: 10.1504/IJSNM.2012.045103.
51. Blondel V., Guillaume J.-L., Lambiotte R., Lefebvre E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*. 2008. vol. 2008. no. 10. DOI: 10.1088/1742-5468/2008/10/P10008.
52. Clauset A., Newman M., Moore C. Finding community structure in very large networks. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics*. 2004. vol. 70. no. 6. DOI: 10.1103/PhysRevE.70.066111.
53. Danon L., Diaz-Guilera A., Duch J., Arenas A. Comparing community structure identification. *Journal of statistical mechanics: Theory and experiment*. 2005. vol. 2005. no. 09. DOI: 10.1088/1742-5468/2005/09/P09008.
54. Lancichinetti A., Fortunato S., Radicchi F. Benchmark graphs for testing community detection algorithms. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics*. 2008. vol. 78. no. 4. DOI: 10.1103/PHYSREVE.78.046110.
55. Zachary W. An Information Flow Model for Conflict and Fission in Small Groups. *Journal of anthropological research*. 1977. vol. 33. no. 4. pp. 452–473. DOI: 10.1086/JAR.33.4.3629752.
56. Girvan M., Newman M. Community structure in social and biological networks. *Proceedings of the national academy of sciences*. 2002. vol. 99. no. 12. pp. 7821–7826. DOI: 10.1073/pnas.122653799.
57. Lusseau D., Schneider K., Boisseau O., Haase P., Slooten E., Dawson S. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: Can geographic isolation explain this unique trait? *Behavioral Ecology and Sociobiology*. 2003. vol. 54. pp. 396–405. DOI: 10.1007/S00265-003-0651-Y.
58. Newman M. Network data. Network data collection. 2013. Available at: <http://www-personal.umich.edu/~mej>. (accessed 26.10.2023).
59. Leskovec J., McAuley J. Learning to Discover Social Circles in Ego Networks. *Advances in neural information processing systems*. 2012. vol. 25.
60. Leskovec J., Lada A., Huberman B. The Dynamics of Viral Marketing. *ACM Transactions on the Web (TWEB)*. 2007. vol. 1. no. 1. DOI: 10.1145/1232722.1232727.
61. Knuth D. *The Stanford GraphBase: a platform for combinatorial computing*. New York: ACM Press, Addison-Wesley Publishing Company. 2009. 577 p.
62. Gleiser P., Danon L. Community structure in jazz. *Advances in complex systems*. 2003. vol. 6. no. 04. pp. 565–573.

63. Kunegis J. Konect: The Koblenz Network Collection. Proceedings of the 22nd international conference on world wide web. 2013. pp. 1343–1350. DOI: 10.1145/2487788.248817.
64. Knuth D. The Art of Computer Programming: Volume 4 Fascicle 0: Introduction to Combinatorial Algorithms and Boolean Functions. Addison-Wesley Professional, 2008. 228 p.
65. Yang Z., Algesheimer R., Tessone C. A comparative analysis of community detection algorithms on artificial networks. Scientific reports. 2016. vol. 6. no. 1 DOI: 10.1038/srep30750.

Ибрагим Абзур Хусейн — научный сотрудник, факультет компьютерных наук, Университет Буира – Акли Моханд Улхадж. Область научных интересов: компьютерные сети, сетевая связь, сети, сетевая архитектура, маршрутизация, информационная безопасность. Число научных публикаций — 1. h.abouzer@univ-bouira.dz; Рю Дрисси Яхья Буира, 10000, Буира, Алжир; р.т.: +213(558)606-511.

Будреф Мохамед Ахмед — научный сотрудник, факультет математики, Университет Буира – Акли Моханд Улхадж. Область научных интересов: компьютерные сети, сетевая связь, сети, сетевая архитектура, маршрутизация, информационная безопасность. Число научных публикаций — 3. m.boudref@univ-bouira.dz; Рю Дрисси Яхья Буира, 10000, Bouira, Алжир; р.т.: +213(550)251-333.

Бадис Лайес — научный сотрудник, факультет компьютерных наук, Университет Буира – Акли Моханд Улхадж. Область научных интересов: компьютерные сети, сетевая связь, сети, сетевая архитектура, маршрутизация, информационная безопасность. Число научных публикаций — 7. l.badis@univ-bouira.dz; Рю Дрисси Яхья Буира, 10000, Буира, Алжир; р.т.: +213(659)564-064.

М.И. ПОРХУН, М.И. ВАШКЕВИЧ
**ЭФФЕКТИВНАЯ РЕАЛИЗАЦИЯ ГАММАТОН-ФИЛЬТРОВ НА
ОСНОВЕ НЕРАВНОПОЛОСНОГО
КОСИНУСНО-МОДУЛИРОВАННОГО БАНКА ФИЛЬТРОВ**

Порхун М.И., Вашкевич М.И. Эффективная реализация гамматон-фильтров на основе неравнополосного косинусно-модулированного банка фильтров.

Аннотация. В работе представлена эффективная реализация банка гамматон-фильтров (БГФ) на основе неравнополосного косинусно-модулированного банка фильтров (НКМБФ), использующего фазовое преобразование. Рассмотрены примеры практических задач, в которых применяется банк гамматон-фильтров, проанализированы его основные особенности и недостатки. Приведено описание равнополосного косинусно-модулированного банка фильтров, а также показан процесс синтеза НКМБФ из его равнополосного аналога при помощи фазового преобразования. Разработан оптимизационный метод проектирования фильтра-прототипа НКМБФ для аппроксимации частотных характеристик БГФ. В основе метода лежит мультипликативная модель импульсной характеристики фильтра-прототипа, использующая логистические сигмоидальные функции. Суть предлагаемого метода заключается в оптимизации фильтра-прототипа с целью минимизации среднеквадратичной ошибки между АЧХ БГФ и НКМБФ для каждого канала. Выполнена программная реализация на языке Python с использованием библиотеки PyTorch. Проведены экспериментальные исследования предложенного метода. Результаты экспериментов показали, что НКМБФ можно использовать для аппроксимации частотных характеристик БГФ, а результирующая АЧХ имеет монотонные спады за счёт использования логистических сигмоидальных функций. Проведён анализ зависимости результирующей ошибки аппроксимации частотных характеристик банка гамматон-фильтров от количества сигмоид, используемых для синтеза фильтра-прототипа НКМБФ на базе мультипликативной модели импульсной характеристики. Выполнен анализ вычислительной сложности НКМБФ, показано как зависит число операций сложения и умножения от длины импульсной характеристики фильтра-прототипа и числа каналов банка фильтров. Сделан вывод, что использование НКМБФ для реализации банка гамматон-фильтров позволяет существенно уменьшить вычислительные затраты на реализацию гамматон-фильтров по сравнению с прямой реализацией.

Ключевые слова: банк гамматон-фильтров, неравнополосный косинусно-модулированный банк фильтров, слуховые фильтры, полифазная реализация, фазовое преобразование, PyTorch.

1. Введение. Слуховая система человека представляет собой сложный, нелинейный и адаптивный анализатор звука. Одним из её наиболее важных принципов работы является частотная избирательность слуха, представляющая собой механизм, позволяющий человеку различать и отделять друг от друга частотные компоненты в звуковом сигнале [1]. Модели слухового восприятия описывают как частотную избирательность слуха, так и некоторые другие феномены. Подобные модели используются во многих практических задачах: кодировании

аудио [2], распознавании речи [3 – 5], распознавании диктора [6], верификации диктора [7], детектировании и локализации звуковых событий [8], при улучшении разборчивости речи с использованием информации о частоте основного тона [9], при моделировании слуховых патологий [10 – 12] и проектировании кохлеарных имплантов [13].

Общая структура модели слухового восприятия, используемая в задачах цифровой обработки сигналов показана на рисунке 1. Входной сигнал $x(n)$ раскладывается на M субполос при помощи банка слуховых фильтров. Далее, в зависимости от решаемой задачи, выполняется соответствующая субполосная обработка сигналов. В результате на выходе образуются информационные признаки, используемые для решения прикладных задач.

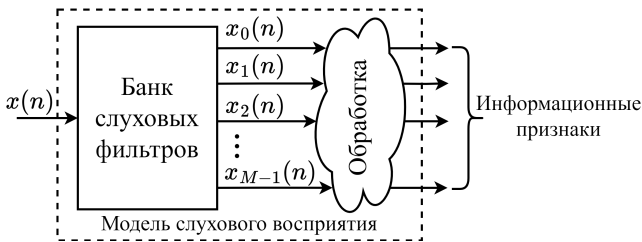


Рис. 1. Общая схема применения моделей слухового восприятия

Далее рассматривается несколько практических примеров использования моделей слухового восприятия. На рисунке 2 показана типовая схема аудиокодера, которая включает банк фильтров анализа, выполняющий декомпозицию сигнала на равнополосные частотные компоненты, которые затем квантуются и кодируются. Также важной частью аудиокодера является модель слухового восприятия (англ. *psychoacoustic model*), которая используется для определения слышимости частотных компонент сигнала. Выходные данные модели используются при квантовании субполосных компонент аудиосигнала для уменьшения объема передаваемой информации [2].

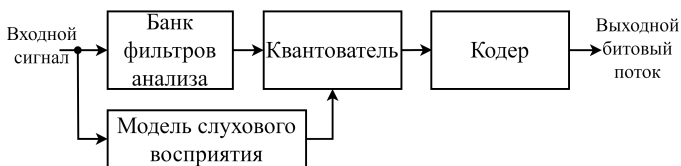


Рис. 2. Общая схема аудиокодера

В работе [9] модель слухового восприятия используется в системе улучшения разборчивости речи (рисунок 3) совместно с оценщиком частоты основного тона (F0). В процессе обработки центральные частоты слуховых фильтров выбираются кратными частоте основного тона, а субполосные сигналы после обработки суммируются для формирования широкополосного сигнала с повышенной разборчивостью. Следует заметить, что в схеме на рисунке 1 центральные полосы банка слуховых фильтров считаются фиксированными и заданными на этапе проектирования, а в схеме на рисунке 3 центральные полосы изменяются в процессе обработки поступающего сигнала и зависят от частоты основного тона.

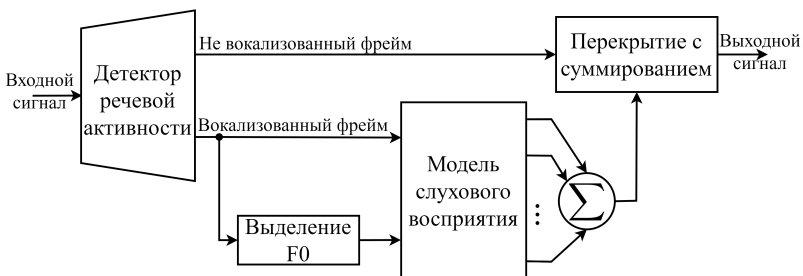


Рис. 3. Общая схема метода улучшения разборчивости речи из работы [9]

Особую роль модели слухового восприятия имеют при построении систем моделирования потери слуха. Такие системы используются на этапе разработки методов коррекции слуха [11]. При тестировании пожилых слабослышащих людей достаточно сложно выделить основную причину ухудшения разборчивости речи, которая может быть связана с работой периферического слуха, слухового нерва либо с нарушением когнитивных функций. Системы моделирования потери слуха частично позволяют решить данную проблему [10], поскольку они имитируют только дисфункции периферического слуха, такие как повышение абсолютного порога слышимости, эффект ухудшения частотной избирательности, феномен ускоренного нарастания громкости и др.

При моделировании потери слуха [10 – 12] модель слухового восприятия используется для разделения звукового сигнала на субполосы, согласованные с работой слуховой системы человека (рисунок 4). Настройка модели выполняется на основе аудиограммы конкретного (слабослышащего) человека. Для каждой субполосы выполняется частотно-зависимое ослабление в соответствии с данными аудиограммы.

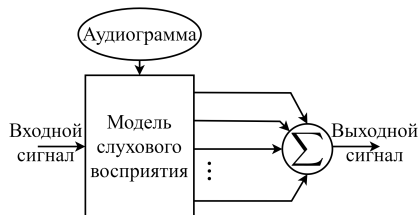


Рис. 4. Общая схема модели потери слуха

Таким образом, основу моделей слухового восприятия составляют частотно-временные методы анализа и синтеза звуковых сигналов, базирующиеся на принципах психоакустики. Общей идеей функционирования данных моделей является декомпозиция звукового сигнала на субполосы, ширина которых согласована с работой слуховой системы. В большинстве случаев для достижения данной цели используются слуховые (англ. *auditory filters*) фильтры. Наиболее часто используются банки гамматон- [14] и гаммачирп-фильтров [10, 15]. Данные фильтры моделируют работу базилярной мембраны улитки человеческого уха. Одной из особенностей использования таких фильтров в системах моделирования эффекта потери слуха является то, что они работают без децимации выходных сигналов, а число каналов в банке колеблется от в промежутке от 20 до 100 [8, 10, 11]. В связи с этим на практике возникает проблема высокой вычислительной сложности при реализации подобных моделей [3, 10]. В данной работе предлагается эффективная реализация гамматон-фильтров при помощи неравнополосного косинусно-модулированного банка фильтров (НКМБФ) [16]. Аппроксимация АЧХ гамматон-фильтров достигается за счёт оптимизации фильтра-прототипа НКМБФ. Помимо того, практическая реализация НКМБФ требует значительно меньше вычислительных ресурсов, чем прямая реализация банка гамматон-фильтров (БГФ).

2. Банк гамматон-фильтров (БГФ). БГФ представляет собой гребёнку гамматон-фильтров, каждый из которых связан с определенной характеристической частотой [17, 18]. Импульсная характеристика гамматон-фильтра с центральной частотой f_c описывается выражением:

$$g(t) = t^{(l-1)} \exp^{-2\pi b \text{ERB}(f_c)t} \cos(2\pi f_c t), t > 0, \quad (1)$$

где t – время, l – порядок фильтра, b – параметр, регулирующий ширину полосы фильтра, f_c – центральная частота фильтра,

$ERB(f_c)$ – эквивалентная прямоугольная полоса пропускания слухового фильтра. На практике часто используются значения параметров $l = 4$ и $b = 1.019$ [17].

Обычно центральные частоты f_c банка фильтров распределены равномерно относительно ERB-шкалы (англ. *equivalent rectangular bandwidth*). ERB-шкала подобна шкале критических полос слуха человека. Зная частоту f (в Гц) можно перейти к ERB-шкале используя выражение [19]:

$$ERB(f_c) = 24,673(0,004368f + 1). \quad (2)$$

При решении практических задач импульсную характеристику (1) гамматон-фильтров дискретизируют и ограничивают по времени умножением на прямоугольное окно [20]. Рассмотрим основные причины, препятствующие эффективной реализации БГФ:

1. Длинные импульсные характеристики. Из-за длительности импульсных характеристик требуются значительные вычислительные затраты для реализации фильтра [20]. Например, на рисунке 5 показаны импульсные характеристики (левая часть рисунка) и АЧХ гамматон-фильтров (в правой части рисунка) с центральными частотами $f_1 = 100$ Гц, $f_2 = 500$ Гц и $f_3 = 1000$ Гц.

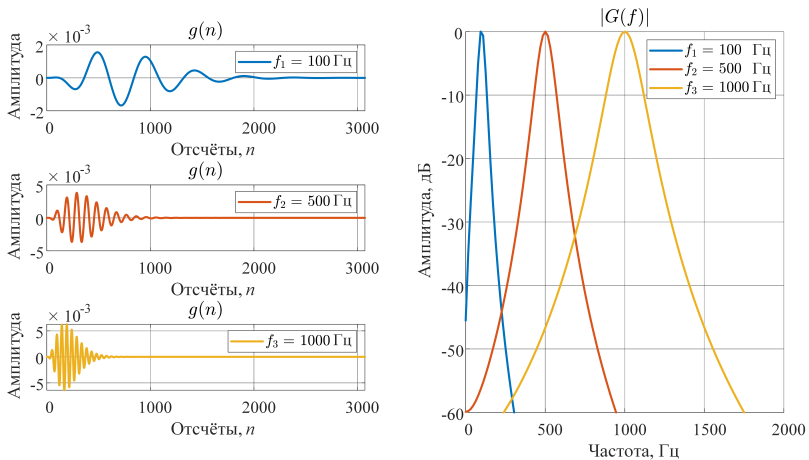


Рис. 5. Импульсные характеристики и АЧХ гамматон-фильтров с центральными частотами $f_1 = 100$ Гц, $f_2 = 500$ Гц и $f_3 = 1000$ Гц

Видно, что с уменьшением центральной частоты длина импульсной характеристики увеличивается. Так, если для $f_3 = 1000$ Гц эффективная длина импульсной характеристики составляет порядка 800 отсчётов, то для $f_1 = 100$ Гц эффективная длина приближается к 3000 отсчётов. Таким образом, прямая реализация КИХ-фильтра с низкой центральной частотой потребует значительных вычислительных ресурсов (около 3к MAC-операций на один входной отсчёт).

2. Требуемое количество фильтров. На практике, при использовании банка гамматон-фильтров число каналов может доходить до ста [10]. Этот факт с учётом длинных импульсных характеристик на низких частотах на порядок повышает вычислительные затраты на реализацию моделей, использующих гамматон-фильтры.

Таким образом, актуальна задача разработки эффективной реализации банка гамматон-фильтров. Целью настоящей работы является разработка эффективной реализации банка гамматон-фильтров. Одним из возможных вариантов решения данной задачи может быть использование НКМБФ, построенного на базе фазового преобразования [16, 23].

На рисунке 6 показаны АЧХ 20-канального БГФ ($|G_{GFB}(f)|$) и НКМБФ ($|G_{WCMFB}(f)|$). Центральные частоты банков фильтров одинаковы и выбраны в соответствии с психоакустической шкалой Барков. АЧХ каждого фильтра с соответствующей центральной частотой обозначена своим цветом. Видно, что частотные характеристики банков фильтров различны, соответственно прямая замена БГФ на НКМБФ не имеет смысла. Таким образом, в для достижения поставленной цели необходимо решить задачу аппроксимации АЧХ БГФ при помощи НКМБФ.

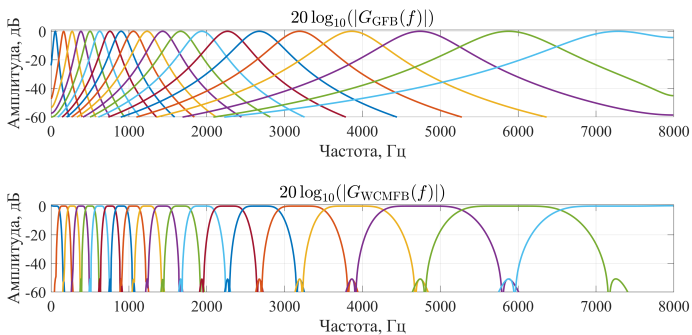


Рис. 6. АЧХ 20-канального БГФ ($|G_{GFB}(f)|$) и НКМБФ ($|G_{WCMFB}(f)|$)

3. Неравнополосный косинусно-модулированный банк фильтров. Косинусно-модулированный банк фильтров (КМБФ) образуется путём применения косинусной модуляции к фильтру-прототипу, представляющему собой КИХ-фильтр нижних частот $h(n)$ с линейной фазо-частотной характеристикой и частотой среза π/M , где M – количество каналов. Таким образом, синтез банка фильтров сводится к расчёту только одного фильтра-прототипа [21]. Импульсные характеристики M -канального КМБФ анализа $h_k(n)$ описываются выражением:

$$h_k(n) = 2h(n) \cos \left(\frac{\pi(k+0,5)}{M} \left(n - \frac{(N-1)}{2} \right) + \frac{(-1)^k \pi}{4} \right), \quad (3)$$

где $N = 2mM$ – порядок фильтра-прототипа, m – положительное число, $k = 0, \dots, M-1$ – номер канала, $n = 0, \dots, N-1$ – временной индекс, $h(n)$ – коэффициенты фильтра-прототипа.

Передаточную функцию k -го фильтра анализа можно записать как

$$H_k(z) = a_k b_k H \left(z^{-1} W_{2M}^{(k+0,5)} \right) + \bar{a}_k \bar{b}_k H \left(z^{-1} W_{2M}^{-(k+0,5)} \right), \quad (4)$$

где $H(z)$ – передаточная функция фильтра-прототипа, $a_k = e^{j(-1)^k \pi/4}$, $W_M = e^{-j2\pi/M}$, $b_k = W_{2M}^{\frac{N-1}{2}(k+0,5)}$, а верхняя черта означает комплексное сопряжение.

НКМБФ образуется из равнополосного КМБФ путём применения фазового преобразования, суть которого заключается в замене элементов задержки в структуре фильтра на фазовые звенья $z^{-1} \rightarrow A(z)$ [16].

Воспользовавшись заменой $z^{-1} \rightarrow A(z)$, перепишем $H_k(z)$:

$$H_k(z) = a_k b_k H \left(A(z) W_{2M}^{(k+0,5)} \right) + \bar{a}_k \bar{b}_k H \left(A(z) W_{2M}^{-(k+0,5)} \right). \quad (5)$$

В данной работе рассматривается фазовое звено первого порядка:

$$A(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad |\alpha| \leq 1. \quad (6)$$

Амплитудно-частотная характеристика фазового звена определяется как:

$$A(e^{j\omega}) = e^{j\Theta_\alpha(\omega)}, \quad \Theta_\alpha(\omega) = \omega - 2 \arctan \left(\frac{\alpha \sin(\omega)}{\alpha \cos(\omega) - 1} \right), \quad (7)$$

где α – коэффициент фазового звена, а ω – нормированная круговая частота в рад/отсчет.

В результате замены элементов задержки в структуре равнополосного КМБФ на фазовые звенья происходит деформация (отображение) частотной оси $\omega \rightarrow \Theta_\alpha(\omega)$ (рисунок 7).

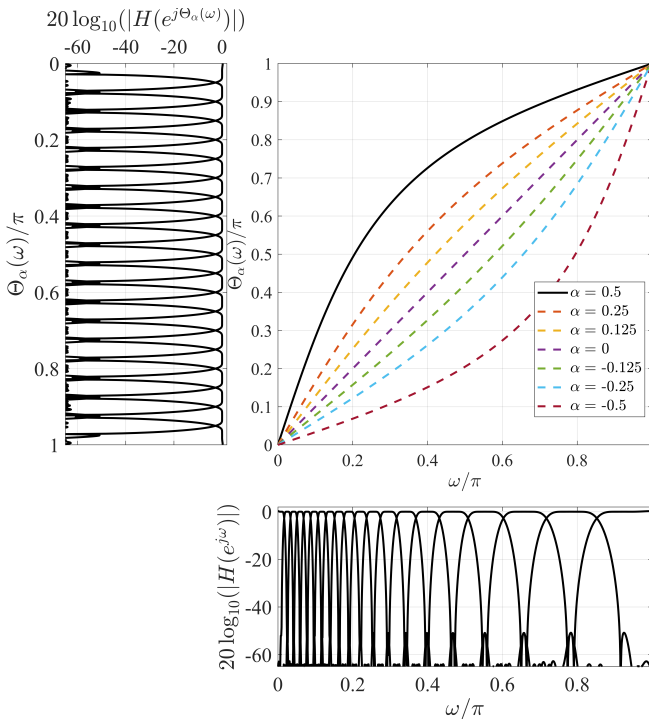


Рис. 7. НКМБФ на основе деформации частотной сетки

Степень деформации частот зависит только от параметра α . Следует отметить, что в результате преобразования банк фильтров

становится равномерным относительно деформированной оси частот $\Omega = \Theta_\alpha(\omega)$, однако относительно исходной оси частот ω он становится неравнополосным. В дальнейшем нас будут интересовать положение центральных полос неравнополосного банка фильтров ω_k^c . Чтобы их найти обозначим через Ω_k^c центральные частоты равнополосного банка фильтров, тогда центральные полосы неравнополосного банка фильтров можно найти как $\omega_k^c = \Theta_\alpha^{-1}(\Omega_k^c)$. Можно показать, что обратная функция Θ_α^{-1} равна $\Theta_{-\alpha}$ и, таким образом, выражение для центральных частот неравнополосного банка фильтров принимает вид:

$$\omega_k^c = \Theta_{-\alpha}(\Omega_k^c). \quad (8)$$

4. Полифазная структура НКМБФ. В работе [21] для эффективной реализации КМБФ предложено использовать полифазное представление фильтра-прототипа:

$$H(z) = \sum_{r=0}^{N-1} h(n)z^{-n} = \sum_{q=0}^{2M-1} \sum_{r=0}^{m-1} h(q + 2rM)z^{-(q+2rM)}. \quad (9)$$

Используя выражение (9) и методику, описанную в работе [21], можно применить данное выражение для эффективной реализации НКМБФ [22]:

$$H_k(z) = \sum_{q=0}^{2M-1} c_{k,q} \sum_{r=0}^{m-1} (-1)^r h(q + 2rM)A^{(q+2rM)}(z), \quad (10)$$

где $c_{k,l} = 2 \cos\left(\frac{(2k+1)\pi}{2M} \left(\frac{l-(N-1)}{2}\right) + (-1)^k \frac{\pi}{4}\right)$.

Введём обозначение $g(q + 2rM) = (-1)^r h(q + 2rM)$.

На основании выражения (10) можно получить структуру НКМБФ анализа (рисунок 8). На рисунке 8 блок косинусной модуляции обозначен как $\mathbf{C} = [c_{k,l}]_{0 \leq k < M, 0 \leq l < 2M}$. В данном блоке выполняется умножение выходов полифазных компонент на матрицу косинусной модуляции \mathbf{C} . Выход $x_k(n)$ соответствует фильтру $H_k(z)$.

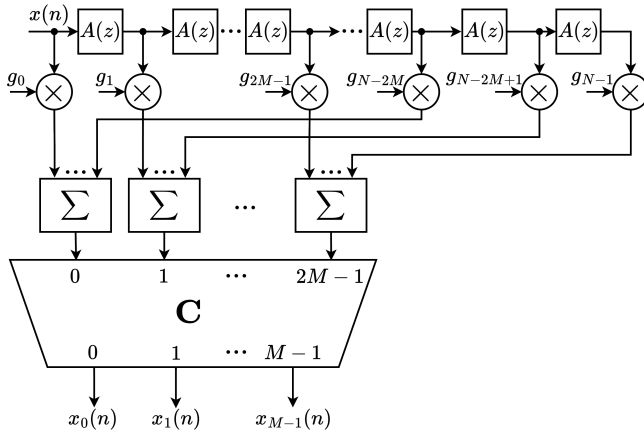


Рис. 8. Полифазная структура НКМБФ

5. Расчётные выражения. В данном разделе будет показано, как выразить зависимость частотных характеристик НКМБФ от коэффициентов фильтра-прототипа при помощи векторно-матричных операций. Данные выражения будут в дальнейшем использоваться при разработке метода оптимизации фильтра-прототипа.

Частотную характеристику КИХ-фильтра чётного порядка N с линейной фазо-частотной характеристикой можно записать в виде:

$$H(e^{j\omega}) = e^{-j(N-1)\omega/2} \mathbf{C}^T(\omega) \mathbf{h}, \quad (11)$$

где

$$\mathbf{C}(\omega) = \left[2 \cos\left(\frac{\omega}{2}\right) 2 \cos\left(\frac{3\omega}{2}\right) \dots 2 \cos\left(\frac{(N-1)\omega}{2}\right) \right]^T, \quad (12)$$

$$\mathbf{h} = \left[h\left(\left\lfloor \frac{N}{2} \right\rfloor\right) h\left(\left\lfloor \frac{N}{2} \right\rfloor + 1\right) \dots h(N-1) \right]^T, \quad (13)$$

где надстрочный знак T означает транспонирование, $[a]$ – операция округления числа a до ближайшего целого числа, которое меньше или равно a .

Используя (11) запишем выражение для расчёта частотной характеристики КМБФ:

$$H_k(e^{j\omega}) = a_k b_k e^{-j(N-1)\omega/2} \mathbf{C}^T(\omega) \mathbf{h} W_{2M}^{(k+0,5)} + \bar{a}_k \bar{b}_k e^{-j(N-1)\omega/2} \mathbf{C}^T(\omega) \mathbf{h} W_{2M}^{-(k+0,5)}. \quad (14)$$

Ранее было показано, что НКМБФ получается из КМБФ путём отображения частотной оси, следовательно его частотную характеристику можно записать в виде:

$$H_k(e^{j\omega}) = a_k b_k e^{-j(N-1)\Theta_\alpha(\omega)/2} \mathbf{C}^T(\Theta_\alpha(\omega)) \mathbf{h} W_{2M}^{(k+0,5)} + \bar{a}_k \bar{b}_k e^{-j(N-1)\Theta_\alpha(\omega)/2} \mathbf{C}^T(\Theta_\alpha(\omega)) \mathbf{h} W_{2M}^{-(k+0,5)}. \quad (15)$$

Выражение (15) можно использовать для создания процедуры оптимизации фильтра-прототипа НКМБФ для аппроксимации АЧХ БГФ. В данном случае настроечными параметрами будут коэффициенты фильтра-прототипа \mathbf{h} . Такой подход был ранее применен в работе [23].

При анализе вида АЧХ БГФ и НКМБФ (рисунок 6) можно заметить, что частотные характеристики гамматон-фильтров монотонно спадают относительно центральной частоты. В то же время АЧХ НКМБФ имеют колебания в полосе заграждения. Таким образом, для аппроксимации АЧХ БГФ с использованием НКМБФ необходимо обеспечить монотонность его частотных характеристик. Для достижения данной цели, в соответствии со свойством преобразования Фурье, импульсная характеристика фильтра-прототипа НКМБФ должна быть монотонна и не пересекать уровень нуля.

Для создания фильтра-прототипа НКМБФ предлагается использовать функцию логистического сигмоида, часто используемую в машинном обучении:

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \quad (16)$$

Выбор данной функции обусловлен её монотонностью, а также отсутствием переходов через нуль. Также стоит заметить, что фактически требуется получить только половину реальной импульсной характеристики (выражение (13)), вторая половина образуется симметричным отображением. Таким образом, предлагается формировать половину фильтра-прототипа НКМБФ с использованием следующей мультипликативной модели:

$$p(t) = A \prod_{r=1}^R \sigma(k_r t + b_r), \quad t \geq 0, \quad A > 0. \quad (17)$$

Очевидно, что $p(t)$ является монотонно-убывающей функцией, поскольку она представляет собой произведение монотонно-убывающих логистических сигмоидов. В связи с тем, что $p(t)$ является непрерывной, то необходимо выполнить её дискретизацию согласно выражению:

$$p(n) = p(n\Delta T), \quad \Delta T = \frac{1}{f_p}, \quad n = 0, 1, \dots, \frac{N}{2} - 1. \quad (18)$$

где f_p – параметр дискретизации, определяющий расстояние между соседними отсчетами дискретизированной функции (в дальнейших примерах использовалось значение $f_p = 24$), $p(n)$ – половина импульсной характеристики фильтра прототипа:

$$\mathbf{h} = [p(0)p(1) \dots p(N/2 - 1)]^T. \quad (19)$$

На рисунке 9 показан пример обычного фильтра-прототипа $h_{orig}(n)$ с немонотонной импульсной характеристикой и фильтр, полученный с использованием мультипликативной сигмоидной функции – $h_{sigm}(n)$.

Можно заметить, что для $h_{orig}(n)$ частотная характеристика содержит множество нулей и немонотонна в полосе задерживания. В случае с использованием сигмоидной функции и импульсная, и частотная характеристики обладают свойством монотонности. Параметризация (17) будет использована в процессе оптимизации фильтра-прототипа при помощи метода градиентного спуска. Естественно, что градиент можно было бы вычислять непосредственно для каждого отсчета

импульсной характеристики (выражение (19)), однако в этом случае нельзя гарантировать монотонность результирующей импульсной характеристики. Кроме того, введенная параметризация (17) значительно уменьшает число оптимизируемых параметров.

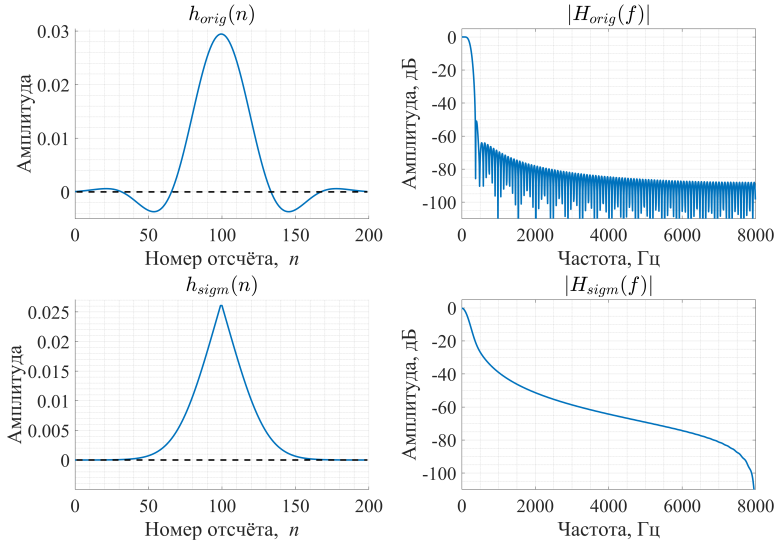


Рис. 9. Импульсные характеристики и АЧХ $h_{orig}(n)$ и $h_{sig}(n)$

6. Оптимизация фильтра-прототипа. Оптимизация фильтра-прототипа НКМБФ основана на минимизации квадратичной ошибки (англ. *RSS – residual sum of squares*) между АЧХ БГФ $|H_k^g(e^{j\omega})|$ и НКМБФ $(|H_k^c(e^{j\Theta(\omega)})|)$ в каждом канале:

$$RSS = \sum_{\omega} \sum_{k=1}^{M-2} \left(|H_k^g(e^{j\omega})| - |H_k^c(e^{j\Theta(\omega)})| \right)^2. \quad (20)$$

При расчёте ошибки предлагается не учитывать первый ($k = 0$) и последний каналы ($k = M - 1$), поскольку у НКМБФ они соответствуют фильтру нижних и верхних частот, а у БГФ – полосовым фильтрам, что приводит к некорректной задаче аппроксимации полосового фильтра при помощи ФНЧ и ФВЧ, соответственно. Минимизация (20) выполняется

относительно параметров A , k_r , b_r мультипликативной сигмоидной функции фильтра-прототипа (17).

Общая схема предлагаемого метода расчёта фильтра-прототипа показана на рисунке 10.

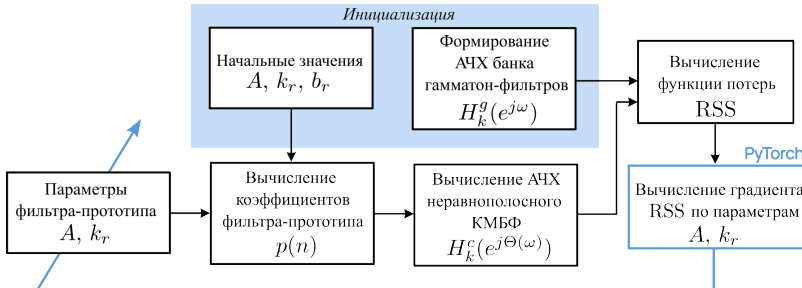


Рис. 10. Схема метода расчета фильтра-прототипа

На начальном этапе происходит формирование АЧХ банка гаммафон-фильтров $|H_k^g(e^{j\omega})|$. Далее выполняется инициализация параметров модели A , k_r , b_r функции (17) и вычисление коэффициентов фильтра-прототипа $p(n)$ (выражение (18)). Следующим шагом рассчитывается АЧХ НКМБФ $|H_k^c(e^{j\Theta(\omega)})|$ на основании текущих значений $p(n)$. Далее вычисляется значение функции ошибки RSS и последующая её минимизация средствами библиотеки PyTorch. Возможность автоматического дифференцирования PyTorch используется для вычисления градиента функции ошибки RSS по параметрам A , k_r сигмоидной функции. Стоит обратить внимание, что по параметру b_r значение градиента не вычислялось, данный параметр фиксировался в процессе инициализации. Минимизация функции ошибки RSS выполняется итеративно при помощи метода градиентного спуска.

Процесс инициализации выполнялся следующим образом. Начальное значение параметра A устанавливалось равным 0,08. Значения параметра k_r выбирались из диапазона $[-\frac{1}{4} \dots \frac{1}{4}]$ с шагом $1/2R$. Параметры b_r устанавливались равными 1,5.

Предложенный метод оптимизации фильтра-прототипа был реализован программно на языке Python.

7. Экспериментальные исследования. В качестве примера синтезировался 20-канальный банк фильтров, у которого центральные частоты выбирались в соответствии с психоакустической шкалой Барков. 20 критических полос (т.е. 20 Барков) занимают частотный диапазон от 0 до 6400 Гц, а без учёта первой и последней полосы, которые не

участвовали в процессе оптимизации фильтра-прототипа, итоговый частотный диапазон составил [100, 4400] Гц. Порядок фильтра-прототипа $N = 2mM = 2 \cdot 5 \cdot 20 = 200$, частота дискретизации $f_s = 16$ кГц, количество сигмоид R в модели фильтра-прототипа изменялось от 2 до 14 с шагом 2.

В процессе обучения для минимизации функции ошибки (20) использовался адаптивный метод градиентного спуска Adam [24], а также адаптивный метод Adadelta [25]. Вначале 100 эпох оптимизация выполнялась при помощи Adam, а следующие 50 эпох – при помощи Adadelta. Для оптимизатора Adam параметр скорости обучения задавался равным $\eta = 0,005$, а для Adadelta $\eta = 0,15$. Выбор двух разных оптимизаторов обусловлен тем, что Adam позволяет эффективно найти область с локальным минимумом функции ошибки, а Adadelta – позволяет выполнить «тонкую» настройку оптимизируемых параметров.

В результате проведения экспериментов были получены значения функции ошибки RSS для разного количества сигмоид R (таблица 1). Если обозначить через RSS_R значение ошибки, получаемое при использовании R сигмоид, то используя выражение

$$D_R = \frac{RSS_2 - RSS_R}{RSS_2} \times 100\%,$$

можно оценить в процентах снижение ошибки по отношению к «базовой» модели фильтра-прототипа на основе двух сигмоид. Полученные значения D_R приведены в таблице 1.

Таблица 1. Результаты экспериментального исследования

Число сигмоид R	2	4	6	8	10	12	14
RSS	6,297	6,120	6,101	6,093	6,086	6,082	6,079
D_R	–	2,81%	3,11%	3,24%	3,35%	3,41%	3,46%

Из таблицы 1 видно, что при увеличении количества сигмоид значение целевой функции ошибки RSS уменьшается. В то же время, после значения $R = 8$ ошибка уменьшается незначительно (менее 0,12% при увеличении числа сигмоид на два), поэтому в данном случае значение $R = 8$ можно считать рациональным.

На рисунке 11 показан график изменения функции ошибки RSS в зависимости количества используемых сигмоид R . Из рисунка видно, что после 90 эпох ошибка снижается, но незначительно.

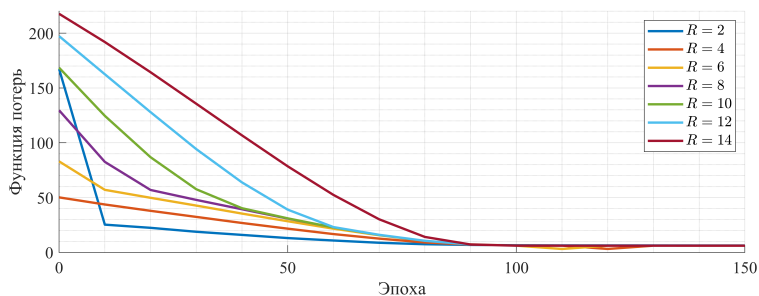
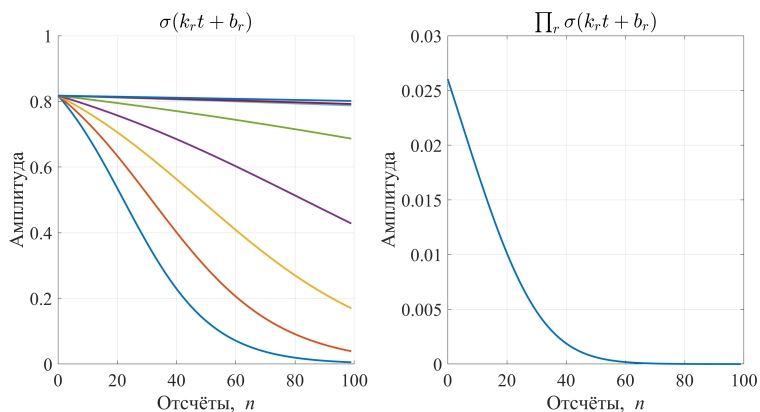


Рис. 11. Изменение функции ошибки

На рисунке 12 показан процесс «конструирования» фильтра-прототипа из сигмоид при $R = 8$. Из рисунка видно, что «нижние» 5 сигмоид вносят больший вклад в итоговую модель фильтра-прототипа, чем 3 сигмоиды, близкие к прямой на уровне около 0,8.

Рис. 12. Логистические сигмоиды и импульсная характеристика фильтра-прототипа при $R = 8$

Незначительное снижение значения ошибки RSS при $R > 8$ обусловлено тем, что при увеличении значения R новые сигмоиды после оптимизации целевых параметров становятся близкими к постоянным значениям и находятся довольно близко, а следовательно, вносят малый вклад в итоговый вид фильтра-прототипа.

Таким образом, рациональным количеством сигмоид для построения фильтра-прототипа является $R = 8$, при этом значение функции ошибки составляет $RSS = 6,093$.

В работе [23] проводилось схожее исследование по возможности аппроксимации АЧХ БГФ при помощи НКМБФ. В контексте данной задачи также выполнялась оптимизация фильтра-прототипа НКМБФ. При этом начальное приближение фильтра-прототипа синтезировалось классическим оконным методом. Для решения поставленной задачи минимизировалась функция ошибки (20). В работе [23] для минимизации (20) использовалась MATLAB-функция *fminunc*, в которой не учитывалась информация о градиенте функции ошибки. В качестве оптимизируемых параметров выступали непосредственно коэффициенты фильтра-прототипа. В результате были получены частотные характеристики НКМБ близкие по виду к БГФ (рисунок 13).

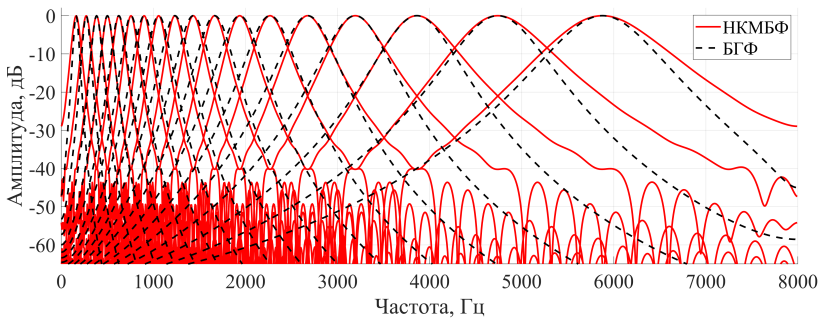


Рис. 13. АЧХ БГФ и оптимизированная по методу [23] АЧХ НКМБФ

Аппроксимация АЧХ БГФ полученная с использованием предложенного метода для $R = 8$ представлена на рисунке 14. Видно, что в результате оптимизации коэффициентов фильтра-прототипа АЧХ НКМБФ довольно близка к требуемой АЧХ БГФ. Для сравнения данного результата с предыдущим [23] (рисунок 13) вычислена функция ошибки (20). Для аппроксимации БГФ по методу [23] $RSS = 5.752$, а для предложенного метода $RSS = 6.093$. Это свидетельствует о том, что величина RSS в обоих случаях сопоставима. Однако по виду АЧХ можно сделать вывод, что предложенный метод позволяет получить аппроксимацию БГФ с монотонными спадами частотных характеристик. Фильтры, полученные по методу [23], имеют колебания в полосе ослабления на уровне -43 дБ, что является негативным фактором с точки зрения последующего использования банка фильтров для частотного анализа в моделях слухового восприятия.

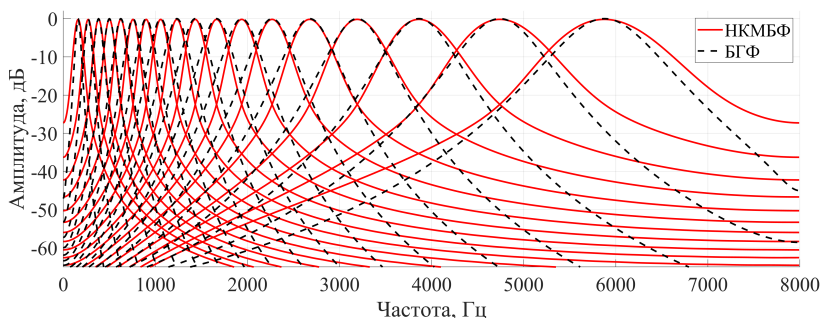


Рис. 14. АЧХ БГФ и оптимизированная АЧХ НКМБФ предложенным методом

На рисунке 15 показаны импульсные характеристики БГФ ($g(n)$) и полученного НКМБФ ($h(n)$) с центральными частотами $f_1 = 161$ Гц, $f_2 = 500$ Гц и $f_3 = 1057$ Гц. Из рисунка видно, что импульсные характеристики схожи по своему виду, однако у НКМБФ они сдвинуты во времени, что говорит о большей групповой задержке. Данная особенность может иметь значение в приложениях, которые ориентированы на обработку сигнала в реальном времени с низкой задержкой (например, в слуховых аппаратах).

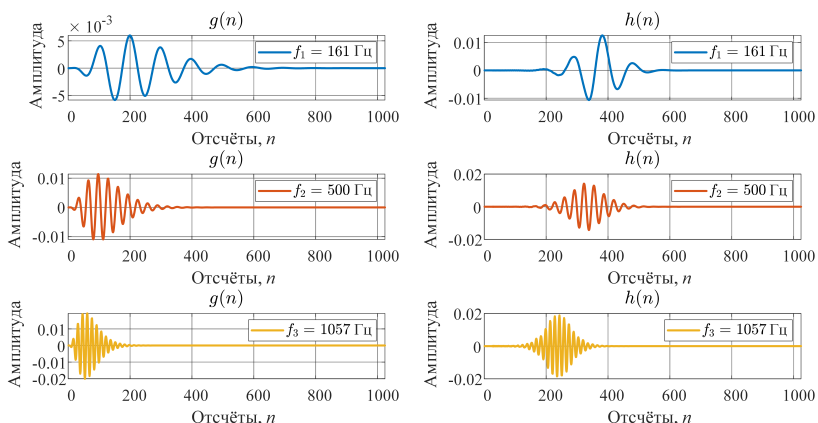


Рис. 15. Импульсные характеристики БГФ и полученного НКМБФ

8. Анализ вычислительной сложности. В работе [22] приведены расчёты вычислительной сложности реализации НКМБФ. Затраты по числу умножений и сложений при реализации НКМБФ показаны в таблице 2.

Таблица 2. Затраты на реализацию НКМБФ

Блок	Число сложений	Число умножений
Цепочка фазовых звеньев	$2(N - 1)$	$N - 1$
Полифазный фильтр-прототип	$N - 2M$	N
Блок модуляции	$\frac{1}{2} \log_2(M) + M$	$\frac{3}{2} \log_2(M) + 2M$
Всего	$3N - M + \frac{1}{2} \log_2(M) - 2$	$2N + \frac{3}{2} \log_2(M) + 2M - 1$

Для сравнения прямой реализации БГФ и НКМБФ проведён эксперимент. Для этого был синтезирован 100-канальный БГФ с центральными частотами в соответствии с психоакустической шкалой Барков, частота дискретизации $f_s = 16$ кГц. Теоретически импульсные характеристики БГФ бесконечны, однако при практической реализации они должны быть ограничены. Прямая реализация БГФ выполняется в виде набора КИХ-фильтров с различной длиной импульсной характеристики. С увеличением временного индекса n амплитудные значения импульсных характеристик гамматон-фильтра затухают и стремятся к нулю (рисунок 5). Ограничение импульсных характеристик БГФ выполнялось путём отбрасывания отсчётов с абсолютной амплитудой ниже уровня 10^{-7} . Суммарная длина импульсных характеристик БГФ после усечения составила 41100. Таким образом, для прямой реализации такого банка фильтров необходимо 41,1 тыс. МАС-операций.

Для сравнения вычислительной сложности, рассмотрим теперь затраты на реализацию 100-канального НКМБФ ($N = 2mM = 2 \cdot 5 \cdot 100 = 1000$). Подставляя значения N и M в расчётные выражения из таблицы 2 получим итоговые затраты на реализацию данного НКМБФ: 2209 операций умножения и 2901 сложений.

Таким образом, считая операцию умножения более «дорогостоящей» с вычислительной точки зрения, реализация НКМБФ в $41100/2209 = 18,6$ раз эффективнее, чем прямая реализация БГФ.

Также известен метод аппроксимации БГФ с использованием гребенки БИХ-фильтров, состоящих из каскада биквадратных звеньев [26]. Полноценное сравнение предложенной аппроксимации БГФ с методом, представленным в [26] выходит за рамки данного исследования. Тем не менее, можно отметить, что использование БИХ-фильтров особенно при реализации в арифметике с фиксированной запятой приводит к деградации частотных характеристик фильтров из-за ошибок квантования. Более того,

предлагаемый подход аппроксимации БГФ с использованием НКМБФ имеет преимущество в том отношении, что положение центральных частот гамматон-фильтров зависит от значения одного параметра α и может изменяться в процессе работы. Изменение центральных частот гамматон-фильтров при использовании метода [26] приведет к необходимости пересчета коэффициентов всех БИХ-фильтров в гребенке.

9. Заключение. В работе рассмотрена эффективная реализация БГФ при помощи НКМБФ. Предложен метод построения фильтра-прототипа для НКМБФ, новизна метода заключается в использовании мультипликативной сигмоидной функции, позволяющей аппроксимировать характеристики БГФ. Предложен способ оптимизации фильтра-прототипа за счёт минимизации функции ошибки между АЧХ БГФ и НКМБФ в каждом канале. Для проверки эффективности предложенного метода выполнена его программная реализация на Python с использованием библиотеки PyTorch. Результаты эксперимента показали, что полученные АЧХ достаточно точно аппроксимируют АЧХ БГФ. Помимо того, предложенный метод расчёта БГФ, в отличие от существующих подходов, позволяет получать монотонные спады частотных характеристик банка фильтров. Выполнен анализ вычислительной сложности реализации БГФ предложенным методом и в прямой форме. Расчёты показали, что предложенная реализация в 18,6 эффективнее с вычислительной точки зрения в случае, когда число каналов банка фильтров равно 100.

Литература

1. Lass N.J., Donai J.J. Hearing science fundamentals. San Diego: Plural Publishing. 2021. 355 p.
2. Herre J., Dick S. Psychoacoustic models for perceptual audio coding – A tutorial review // Applied Sciences. 2019. vol. 9. no. 14. DOI: 10.3390/app9142854.
3. Schilling A., Gerum R., Metzner C., Maier A., Krauss P. Intrinsic noise improves speech recognition in a computational model of the auditory pathway // Frontiers in Neuroscience. 2022. vol. 16. DOI: 10.3389/fnins.2022.908330.
4. Schadler M.R., Hulsmeier D., Warzybok A., Kollmeier B. Individual aided speech-recognition performance and predictions of benefit for listeners with impaired hearing employing FADE // Trends in Hearing. 2020. vol. 24. DOI: 10.1177/2331216520938929.
5. Luo Q. The improving effect of intelligent speech recognition System on english learning // Advances in Multimedia. 2022. DOI: 10.1155/2022/2910859.
6. Krobb A., Debyeche M., Selouani S. A. A novel hybrid feature method based on Caelen auditory model and gammatone filterbank for robust speaker recognition under noisy environment and speech coding distortion // Multimedia Tools and Applications. 2023. vol. 82. no. 11. pp. 16195–16212.
7. Krobb A., Debyeche M., Selouani S. A. Mixture linear prediction Gammatone Cepstral features for robust speaker verification under transmission channel noise // Multimedia Tools and Applications. 2020. vol. 79. pp. 18679–18693.

8. Jacome K.G.R., Grijalva F.L., Masiero B.S. Sound events localization and detection using bio-inspired gammatone filters and temporal convolutional neural networks // IEEE/ACM transactions on audio, speech, and language processing. 2023. vol. 31. pp. 2314–2324.
9. Queiroz A., Coelho R. F0-based gammatone filtering for intelligibility gain of acoustic noisy signals // IEEE Signal Processing Letters. 2021. vol. 28. pp. 1225–1229.
10. Irino T. Hearing Impairment Simulator Based on Auditory Excitation Pattern Playback: WHIS // IEEE Access. 2023. vol. 11. pp. 78419–78430.
11. Kates J.M., Arehart K.H. The hearing-aid speech perception index (HASPI) version 2 // Speech Communication. 2021. vol. 131. pp. 35–46.
12. Порхун М.И., Вашкевич М.И. Метод моделирования эффекта ухудшения частотного разрешения слуха у больных нейросенсорной тугоухостью // Информатика. 2021. Т. 18. № 3. С. 68–82. DOI: 10.37661/1816-0301-2021-18-3-68-82.
13. Islam R., Tarique M. Investigating the performance of gammatone filters and their applicability to design cochlear implant processing system // Designs. 2024. vol. 8. no. 1. DOI: 10.3390/designs8010016.
14. Mourgela A., Reiss J., Agus T.R. Investigation of a real-time hearing loss simulation for use in audio production // Audio Engineering Society Convention 149. 2020.
15. Nagae M., Irino T., Nisimura R., Kawahara H., Patterson R.D. Hearing impairment simulator based on compressive gammachirp filter // Signal and Information Processing Association Annual Summit and Conference (APSIPA'2014). 2014. DOI: 10.1109/APSIPA.2014.7041579.
16. Вашкевич М.И., Азаров И.С. Сравнение частотно-временных преобразований: Фурье-анализ, вейвлеты и банки фильтров на основе фазового преобразования // Цифровая обработка сигналов. 2020. № 2. С. 13–26.
17. Patterson R.D., Nimmo-Smith I., Holdsworth J., Rice P. An efficient auditory filterbank based on the gammatone function // Meeting of the IOC Speech Group on Auditory Modelling at RSRE. 1987. vol. 2. no. 7.
18. Irino T., Patterson R.D. The gammachirp auditory filter and its application to speech perception // Acoustical Science and Technology. 2020. vol. 41. no. 1. pp. 99–107.
19. Wang D., Brown G. Computational Auditory Scene Analysis: Principles, Algorithms, and Applications. New Jersey: Wiley-IEEE Press. 2008. 395 p.
20. Feldbauer C., Kubin G., Kleijn W.B. Anthropomorphic Coding of Speech and Audio: A Model Inversion Approach // EURASIP Journal on Advances in Signal Processing. 2005. no. 9. pp. 1334–1349.
21. Koilpillai R.D., Vaidyanathan P.P. Cosine-modulated FIR filter banks satisfying perfect reconstruction // IEEE Trans. on Signal Processing. 1992. vol. 4. no. 4. pp. 770–783.
22. Вашкевич М.И., Петровский А.А. Неравнополосные банки фильтров для слуховых аппаратов: анализ алгоритмов, автоматизация проектирования // Автоматизация проектирования дискретных систем: материалы 7-й междунар. конф. (г. Минск, 16-17 ноября). 2010. С. 53–60.
23. Порхун М.И., Вашкевич М.И. Моделирование частотной характеристики банка гамматон-фильтров при помощи неравнополосного косинусно-модулированного банка фильтров // Цифровая обработка сигналов и её применение: труды 24-й междунар. конф. (г. Москва, 30 марта – 1 апреля). 2022. Т. 1. С. 53–57.
24. Kingma D.P., Ba J. Adam: A method for stochastic optimization // arXiv preprint. 2014. arXiv: 1412.6980. 15 p. URL: <https://arxiv.org/pdf/1412.6980.pdf> (дата обращения: 15.07.2024).

25. Zeiler M.D. ADADELTA: An adaptive learning rate method // arXiv preprint. 2012. arXiv:1212.5701. 6 p. URL: <https://arxiv.org/pdf/1212.5701> (дата обращения: 15.07.2024).
26. Malcolm S. An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank // Apple Computer Technical Report. 1993. vol. 35. no. 8. 42 p.

Порхун Максим Игоревич — магистр технических наук, старший преподаватель, кафедра электронных вычислительных средств факультета компьютерных систем и сетей, Белорусский государственный университет информатики и радиоэлектроники (БГУИР). Область научных интересов: цифровая обработка сигналов, алгоритмы обработки речевых сигналов для слуховых аппаратов, методы моделирования слуховых патологий, микропроцессорные средства управления и обработки сигналов. Число научных публикаций — 15. porhun@bsuir.by; улица П. Бровки, 6, 220013, Минск, Беларусь; р.т.: +375(17)293-8420.

Вашкевич Максим Иосифович — д-р техн. наук, профессор кафедры, кафедра электронных вычислительных средств факультета компьютерных систем и сетей, Белорусский государственный университет информатики и радиоэлектроники (БГУИР). Область научных интересов: методы и алгоритмы акустического анализа голоса для выявления патологий, методы и алгоритмы обработки речевых сигналов для слуховых аппаратов, синтез быстрых алгоритмов цифровой обработки сигналов, аппаратная и программная реализация алгоритмов цифровой обработки сигналов. Число научных публикаций — 85. vashkevich@bsuir.by; улица П. Бровки, 6, 220013, Минск, Беларусь; р.т.: +375(17)293-8478.

M.I. PORHUN , M.I. VASHKEVICH
**EFFICIENT IMPLEMENTATION OF GAMMATONE FILTERS
BASED ON WARPED COSINE MODULATED FILTER BANK**

Porhun M.I., Vashkevich M.I. Efficient implementation of gammatone filters based on warped cosine modulated filter bank.

Abstract. The paper presents an effective implementation of a gammatone filter bank (GFB) based on a warped cosine modulated filter bank (WCMFB) using an allpass transform. Examples of practical tasks in which a GFB is used are considered, and its main features and disadvantages are analyzed. A description of a uniform cosine-modulated filter bank is given, and the process of synthesis of a WCMFB using all-pass transform is shown. An optimization method for designing a WCMFB prototype filter to approximate the frequency characteristics of GFB has been developed. The method is based on a multiplicative model of the impulse response of the prototype filter using logistic sigmoid functions. The essence of the proposed method is to optimize the prototype filter in order to minimize the RMS error between the frequency response of the GFB and WCMFB for each channel. A software implementation in Python using the PyTorch library has been performed. Experimental studies of the proposed method have been carried out. The experimental results showed that the WCMFB can be used to approximate the frequency characteristics of the GFB, and the resulting frequency response has monotonic declines due to the use of logistic sigmoid functions. The resulting GFB frequency characteristics approximation error dependence on the number of sigmoids used in the prototype filter is analyzed. The analysis of the computational complexity of the WCMFB is performed, and it is shown how the number of addition and multiplication operations depends on the length of the impulse response of the prototype filter and the number of channels of the filter bank. It is concluded that the use of the WCMFB for the implementation of the GFB can significantly reduce the computational costs of implementing gammatone filters compared with direct implementation.

Keywords: gammatone filter bank, warped cosine modulated filter bank, auditory filters, polyphase implementation, all-pass transform, PyTorch.

References

1. Lass N.J., Donai J.J. Hearing science fundamentals. San Diego: Plural Publishing. 2021. 355 p.
2. Herre J., Dick S. Psychoacoustic models for perceptual audio coding – A tutorial review. Applied Sciences. 2019. vol. 9. no. 14. DOI: 10.3390/app9142854.
3. Schilling A., Gerum R., Metzner C., Maier A., Krauss P. Intrinsic noise improves speech recognition in a computational model of the auditory pathway. Frontiers in Neuroscience. 2022. vol. 16. DOI: 10.3389/fnins.2022.908330.
4. Schadler M.R., Hulsmeier D., Warzybok A., Kollmeier B. Individual aided speech-recognition performance and predictions of benefit for listeners with impaired hearing employing FADE. Trends in Hearing. 2020. vol. 24. DOI: 10.1177/2331216520938929.
5. Luo Q. The improving effect of intelligent speech recognition System on english learning. Advances in Multimedia. 2022. DOI: 10.1155/2022/2910859.
6. Krobb A., Debyeche M., Selouani S. A. A novel hybrid feature method based on Caelen auditory model and gammatone filterbank for robust speaker recognition under noisy environment and speech coding distortion. Multimedia Tools and Applications. 2023. vol. 82. no. 11. pp. 16195–16212.

7. Krobba A., Debyeche M., Selouani S. A. Mixture linear prediction Gammatone Cepstral features for robust speaker verification under transmission channel noise. *Multimedia Tools and Applications*. 2020. vol. 79. pp. 18679–18693.
8. Jacome K.G.R., Grijalva F.L., Masiero B.S. Sound events localization and detection using bio-inspired gammatone filters and temporal convolutional neural networks. *IEEE/ACM transactions on audio, speech, and language processing*. 2023. vol. 31. pp. 2314–2324.
9. Queiroz A., Coelho R. F0-based gammatone filtering for intelligibility gain of acoustic noisy signals. *IEEE Signal Processing Letters*. 2021. vol. 28. pp. 1225–1229.
10. Irino T. Hearing Impairment Simulator Based on Auditory Excitation Pattern Playback: WHIS. *IEEE Access*. 2023. vol. 11. pp. 78419–78430.
11. Kates J.M., Arehart K.H. The hearing-aid speech perception index (HASPI) version 2. *Speech Communication*. 2021. vol. 131. pp. 35–46.
12. Porhun M.I., Vashkevich M.I. [A method for simulation the effect of the reduced frequency resolution of the ear in patients with sensorineural hearing loss]. *Informatika – Informatics*. 2021. vol. 18. no. 3. pp. 68–82. (In Russ.).
13. Islam R., Tarique M. Investigating the performance of gammatone filters and their applicability to design cochlear implant processing system. *Designs*. 2024. vol. 8. no. 1. DOI: 10.3390/designs8010016.
14. Mourgela A., Reiss J., Agus T.R. Investigation of a real-time hearing loss simulation for use in audio production. *Audio Engineering Society Convention 149*. 2020.
15. Nagae M., Irino T., Nisimura R., Kawahara H., Patterson R.D. Hearing impairment simulator based on compressive gammachirp filter. *Signal and Information Processing Association Annual Summit and Conference (APSIPA'2014)*. 2014. DOI: 10.1109/APSIPA.2014.7041579.
16. Vashkevich M.I., Azarov I.S. [Comparison of time-frequency transforms: fourier analysis, wavelets and allpass transformed filter banks]. *Cifrovaja obrabotka signalov – Digital signal processing*. 2020. no. 2. pp. 13–26. (In Russ.).
17. Patterson R.D., Nimmo-Smith I., Holdsworth J., Rice P. An efficient auditory filterbank based on the gammatone function. *Meeting of the IOC Speech Group on Auditory Modelling at RSRE*. 1987. vol. 2. no. 7.
18. Irino T., Patterson R.D. The gammachirp auditory filter and its application to speech perception. *Acoustical Science and Technology*. 2020. vol. 41. no. 1. pp. 99–107.
19. Wang D., Brown G. *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. New Jersey: Wiley-IEEE Press. 2008. 395 p.
20. Feldbauer C., Kubin G., Kleijn W.B. Anthropomorphic Coding of Speech and Audio: A Model Inversion Approach. *EURASIP Journal on Advances in Signal Processing*. 2005. no. 9. pp. 1334–1349.
21. Koilpillai R.D., Vaidyanathan P.P. Cosine-modulated FIR filter banks satisfying perfect reconstruction. *IEEE Trans. on Signal Processing*. 1992. vol. 4. no. 4. pp. 770–783.
22. Vashkevich M.I., Petrovsky A.A. [Warped filter banks for hearing aids: algorithm analysis, design automation]. *Avtomatizacija proektirovaniya diskretnyh system: materialy 7-j mezhdunar. konf. [Automation of discrete system design: materials of the 7th international. conf.]*. Minsk, 2010. pp. 53–60. (In Russ.).
23. Porhun M.I., Vashkevich M.I. [Gammatone filter bank magnitude response approximation using non-uniform cosine-modulated filter bank]. *Cifrovaja obrabotka signalov i ejo primenenie: trudy 24-j mezhdunar. konf. [Digital signal processing and its applications: proceedings of the 24th int. conf.]*. Moscow, 2022. vol. 1. pp. 53–57. (In Russ.).
24. Kingma D.P., Ba J. Adam: A method for stochastic optimization. *arXiv preprint*. 2014. arXiv: 1412.6980. 15 p. Available at: <https://arxiv.org/pdf/1412.6980.pdf> (accessed 15.07.2024).

25. Zeiler M.D. ADADELTA: An adaptive learning rate method. arXiv preprint. 2012. arXiv:1212.5701. 6 p. Available at: <https://arxiv.org/pdf/1212.5701> (accessed обращения: 15.07.2024).
26. Malcolm S. An Efficient Implementation of the Patterson-Holdsworth Auditory Filter Bank. Apple Computer Technical Report. 1993. vol. 35. no. 8. 42 p.

Porhun Maxim — M.S., Senior lecturer, Electronic computing facilities department, faculty of computer systems and networks, Belarusian State University of Informatics and Radioelectronics (BSUIR). Research interests: digital signal processing, speech signal processing algorithms for hearing aids, hearing loss simulation methods, microprocessor control and signal processing tools. The number of publications — 15. porhun@bsuir.by; 6, P. Brovky St., 220013, Minsk, Belarus; office phone: +375(17)293-8420.

Vashkevich Maxim — Dr.Sci., Professor of the department, Electronic computing facilities department, faculty of computer systems and networks, Belarusian State University of Informatics and Radioelectronics (BSUIR). Research interests: methods and algorithms of digital signal processing, application of mathematics and computer science methods in medical diagnostics. The number of publications — 85. vashkevich@bsuir.by; 6, P. Brovky St., 220013, Minsk, Belarus; office phone: +375(17)293-8478.

R. JENIFER, V.J. PRAKASH
**RIVEST-SHAMIR-ADLEMAN ALGORITHM OPTIMIZED
TO PROTECT IOT DEVICES FROM SPECIFIC ATTACKS**

Jenifer R., Prakash V.J. Rivest-Shamir-Adleman Algorithm Optimized to Protect IoT Devices from Specific Attacks.

Abstract. IoT devices are crucial in this modern world in many ways, as they provide support for environmental sensing, automation, and responsible resource conservation. The immense presence of IoT devices in everyday life is inevitable in the smart world. The predominant usage of IoT devices lurks the prying eyes of intentional hackers. Though there are several precautionary security systems and protocols available for generic wireless networks, it is observed that there is a need to formulate a state-of-the-art security mechanism exclusively for IoT network environments. This work is submitted here for the betterment of IoT network security. Three dedicated contributions are integrated in this work to achieve higher security scores in IoT network environments. Fast Fuzzy Anomaly Detector, Legacy Naïve Bayes Attack Classifiers, and Variable Security Schemer of Rivest-Shamir-Adleman algorithm are the novel modules introduced in this work abbreviated as ASORI (Attack Specific Security Optimized RSA for IoT). Captivating the advantages of the onboard IoT certification mechanism and selecting a dynamic security strategy are the novelties introduced in this work. ASORI model is tested with industrial standard network simulator OPNET to ensure the improved security along with vital network performance parameter betterments.

Keywords: Internet-of-Things (IoT), network security, fuzzy anomaly detection, Naïve Bayes classification, RSA.

1. Introduction. An IoT device is an electronic device that is embedded with one or more sensors, internet connection ability, data processing ability and optional drivers or actuators [1]. Literally, they are all small embedded devices equipped with at least IEEE 802.11 wireless communication capability. IoT devices are omnipresent these days, and that takes place in almost every automation and other operations [2]. From agricultural to industrial automation, the Internet of Things (IoT) has reformed several industries by combining daily objects and devices with the Internet, enabling them to collect, share, and exchange data. IoT applications span a wide range of sectors, from smart homes that allow remote control of appliances and thermostats to healthcare where wearable devices monitor vital signs and transmit patient data for remote diagnosis. Industrial IoT optimizes manufacturing processes through real-time monitoring of machinery and predictive maintenance, while agriculture employs IoT for precision farming, optimizing irrigation and monitoring crop conditions. Smart cities utilize IoT to enhance urban services, such as intelligent traffic management and waste management systems [3]. As IoT continues to evolve, it holds the potential to enhance efficiency,

convenience, and sustainability across numerous domains, transforming the way we interact with the world.

An attack on IoT is one of the spiteful attempts to utilize weakness in internet-connected devices. Internet-connected devices can be smart home devices or industrial control systems, even medical devices [4]. If an attacker could obtain control of any of these devices, he could abuse sensitive data or even destroy a system. Since IoT devices are cost effective the computational resources are used to be very limited. This limited resource nature is prone to several attacks that cause network instability [5]. The heterogeneous property of IoT nodes is also an important gateway to malicious attacks. Following a standard static security method is not sufficient to maintain acceptable security in IoT network environments [6]. Most modern life-supporting systems are developed using IoT devices. Any possible threat to these IoT devices can cause severe physical consequences. Therefore, it is important to devise a precise lightweight dynamic security system for IoT network environments.

2. Existing Methods. A set of relevant existing methods is taken here to understand the fundamental principles, methodologies, advantages and limitations of IoT network security. Securing Wireless Sensor Networks Against Denial-of-Sleep Attacks Using RSA Cryptography Algorithm and Interlock Protocol [7], An authentication information exchange scheme in WSN for IoT applications [8], Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN [9], A secure IoT-based mutual authentication for healthcare applications in wireless sensor networks using ECC [10], and Towards an improved energy efficient and end-to-end secure protocol for IoT healthcare applications [11] are the methods examined in this work.

2.1. Securing Wireless Sensor Networks against Denial-of-Sleep Attacks Using RSA Cryptography Algorithm and Interlock Protocol. In 2019, in study [7] the authors presented Abnormal Sensor Detection Accuracy with RSA (ASDARSA) work to protect IoT networks in particular for Denial of Sleep (DoSL) type attacks. Energy and Distance-based cluster head selection methodology, and RSA Cryptography based interlock protocol are the two phases defined in ASDARSA work. The standard RSA algorithm is used in this ASDARSA work for key generation, encryption and decryption. The novelties identified in ASDARSA work are an introduction to an energy consumption model, a Cluster head selection procedure, a dedicated DoSL prevention procedure, and cluster-level node authentication. NS-2 network simulator is used to evaluate the performance parameters such as throughput, packet delivery ratio, detection range, residual energy and lifetime of the network. Preliminary assumptions such

that nodes are homogeneous and placed statically, cluster head can only communicate with the base station, and a constant communication rate is followed in ASDARSA work.

Higher security against DoSL-type attacks is the noted advantage of ASDARSA work whereas decreased network performance and lack of security against modern attacks are the identified limitation.

2.2. An authentication information exchange scheme in WSN for IoT applications (AIES). AIES work is introduced by the authors in [8] to confront node capture attacks in IoT-based wireless sensor networks. Dedicated procedures for the System setup phase, a new association scheme and a dynamic contracting mechanism are introduced in AIES. The sensor registration phase, User Registration phase, login session, Authentication, and key agreement phase are described clearly in AIES work. Along with node capture attacks, AIES also resists Replay attacks and Sensor impersonation attacks. AIES work also facilitates anonymity to protect user information. The security evaluation is done by BAN logic.

The achievement of security against node capture attacks is the advantage of AIES work. Missing assessment against essential network performance metrics such as Throughput, Communication Delay and Packet delivery ratio are identified as the limitation of AIES work. Security against several modern attacks is not included in AIES work, which is another observed limitation.

2.3. Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN. Paper [9] proposed a fog-assisted secure healthcare data aggregation scheme represented as Enhanced Healthcare Data Aggregation (EHDA). EHDA work uses the Fog server to minimize the storage, communication and energy overheads. EHDA uses peer-to-peer communication between healthcare IoT devices and wearable IoT devices to aggregate data. As per EHDA work, the fog server receives the aggregated data from IoT devices and stores it in the local repository. Thus, the cloud server can extract the required data from the fog server and store it in cloud storage. Individual algorithms are provided in EHDA work for message reception in the Aggregator node, and Aggregator node to Base station. The lightweight cryptography scheme reduces the energy overhead. NS2.35 simulator is used to measure the discussed overheads in EHDA work.

The achievement of reduction in storage, communication and energy overheads are the stated advantages of EHDA work. The lightweight cryptography diminishes security is one of the limitations of EHDA. Network performance metrics such as throughput, Packet delivery ratio, and

communication delay are not evaluated during the experiments carried out for the EHDA work.

2.4. A secure IoT-based mutual authentication for healthcare applications in wireless sensor networks using ECC (SIMA). SIMA work is proposed by the authors in paper [10] to improve security in Wireless Medical Sensor Networks (WMSN). SIMA ensures a new privacy-perceiving user authentication scheme with the help of Elliptic Curve Cryptography. A lightweight authentication scheme is introduced in SIMA work to withstand Smartcard stolen attacks, Insider attacks, User impersonation attacks, Gateway node impersonation attacks, Sensor node impersonation attacks, forward secrecy attacks, and replay attacks. BAN logic and the Random Oracle model are used to demonstrate the security characteristics of the SIMA model. AVISPA simulation tool is used to perform a clear security analysis of SIMA work. Dedicated methods are provided for standard network authentication phases such as the System initialization phase, User registration phase, Sensor node registration phase, gateway registration phase, Login phase, Authentication phase, Password revocation phase, and Dynamic node adjunct phase.

The achievement of less computational and communication overheads is the advantage of SIMA work, whereas, attack detection accuracy and precision are not measured during the evaluation process. Common network performance measurement metrics such as throughput, latency and packet delivery ratio are not discussed in SIMA.

2.5. Towards an improved energy efficient and end-to-end secure protocol for IoT healthcare applications. An Asynchronous duty cycle medium access control protocol is proposed by the author in [11] in the name Local Coordination XMAC (LCX-MAC) to reduce data transmission delay and energy. LCX-MAC method uses a set of timers to operate and to reduce power consumption by periodically operating nodes in Sleep mode, Data availability, back-off timeout, Channel status, Early acknowledgment and Data transmission status. The regular duty cycle of MAC protocol is modified with the Markov model to optimize the Sleep/Wake time cycles. Throughput, communication delay and Average energy consumption are calculated using the formula.

The improvements in Throughput, communication delay and energy are the stated advantages of LCX-MAC work. The entire parameters computations are performed using calculations. There are no simulation-based results produced in the evaluation process – which is the limitation of this work. The Security aspect of LCX-MAC work is totally absent in the evaluation process.

The used methodologies, Advantages and Limitations of the existing works are enumerated in Table 1.

Table 1. Existing works, Methodologies, advantages and limitations

No.	Work	Year	Methodology	Advantages	Limitations
[7]	Securing Wireless Sensor Networks Against Denial-of-Sleep Attacks Using RSA Cryptography Algorithm and Interlock Protocol	2019	Energy-Distance-based Clustering, RSA	Security against DoSL	Vulnerable to Probe, U2R attacks
[8]	An authentication information exchange scheme in WSN for IoT applications	2020	Legacy association scheme, Dynamic contact mechanism	Security against Node capture attack	Impact on network performance
[9]	Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN	2020	Peer-to-peer communication, Fog data aggregation	Reduced communication and Energy overheads	Diminished network performance
[10]	A secure IoT-based mutual authentication for healthcare applications in wireless sensor networks using ECC	2021	ECC-based lightweight authentication	Reduced communication and computational overheads	Low throughput, Attack detection accuracy
[11]	Towards an improved energy efficient and end-to-end secure protocol for IoT healthcare applications	2020	Timer-based node mode control	Better network performance	Compromised security

3. Background. Fuzzy logic and the Rivest-Shamir-Adleman (RSA) algorithm are acknowledged as the background methodologies of the

proposed work. A clear understanding of Fuzzy logic and the RSA algorithm is required to explicate the ASORI work, covered in this section.

3.1. Fuzzy Logic. Fuzzy logic is a mathematical concept and a computing paradigm that deals with reasoning and decision-making in situations that involve uncertainty, ambiguity, and imprecision. It was introduced by Lotfi A. Zadeh in 1965 as an extension of classical (Boolean) logic to handle situations where the boundaries between true and false are not clearly defined. In classical logic, propositions are either true or false, with no middle ground. Fuzzy logic, on the other hand, allows for degrees of truth to be represented. It introduces the concept of membership functions, which assign a degree of membership (between 0 and 1) to an element in a set [12]. These membership functions capture the gradual transition between different states rather than abrupt distinctions. Fuzzy logic is commonly used in control systems and decision-making processes where human expertise and intuition play a significant role. Fuzzy logic stands in between analog and digital signal measurements; and litigations are the best to formulate fast decision-making [13]. The fuzzy logic system is used to select an optimal path for data transmission during uncertainty [14]. The flow of the fuzzy logic process is illustrated in Figure 1.

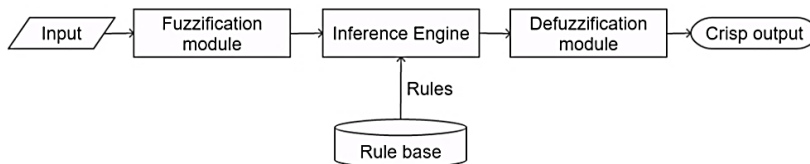


Fig. 1. Fuzzy Logic block diagram

The key idea is to represent the flow of information from input variables through fuzzification, rule processing, inference, aggregation, and finally defuzzification to obtain a crisp output value or action.

3.2. RSA. RSA, which stands for Rivest-Shamir-Adleman, is a widely used public key cryptosystem and encryption algorithm. It was introduced by Ron Rivest, Adi Shamir, and Leonard Adleman in 1977 and remains one of the most important and secure methods for secure communication and data protection. RSA is based on the mathematical properties of large prime numbers and their difficulty in being factored [15]. The security of RSA relies on the fact that it is computationally infeasible to factorize a large semiprime (the product of two large prime numbers) into its constituent primes, especially when the primes are chosen to be sufficiently large [16].

RSA key generation:

- Step 1: Generate two distinct large prime numbers, as "p" and "q".
- Step 2: Compute the product of these primes, " $n = p * q$ ", which becomes the modulus used as the public and private keys.
- Step 3: Compute the totient of n, denoted as " $\varphi(n) = (p - 1) * (q - 1)$ ".
- Step 4: Choose an encryption exponent "e" that is relatively prime to $\varphi(n)$, typically a small prime like $65537 (2^{16} + 1)$.
- Step 5: Compute the decryption exponent "d" such that $(d * e) \bmod \varphi(n) = 1$.
- Step 6: The public key is (n, e), and the private key is (n, d)

RSA Encryption:

- Step 1: To encrypt a message "M", the sender uses the recipient's public key (n, e)
- Step 2: The sender converts the message into a numerical value "m"
- Step 3: The sender calculates the ciphertext "C" using the formula:
$$C = (m^e) \bmod n.$$
- Step 4: The ciphertext "C" is sent to the recipient

RSA Decryption:

- Step 1: The recipient uses their private key (n, d) to decrypt the ciphertext "C"
- Step 2: The recipient calculates the plaintext message "m" using the formula: $m = (c^d) \bmod n$.
- Step 3: The recipient converts the numerical value "m" back into the original message format.

RSA is widely used for secure communication, digital signatures, and various other cryptographic applications. It is considered secure as long as the key sizes are chosen appropriately large (e.g., 2048 bits or more) and the algorithms are implemented correctly. However, with the advancement of computing power and algorithms, it is important to keep up with best practices and periodically update key lengths to ensure continued security.

4. Proposed method ASORI. There are three discrete modules formulated to construct the ASORI model. They are Fast Fuzzy Anomaly Detector, Legacy Naïve Bayes Attack Classifier, and Variable RSA Security Schemer. The comprehensive details about the functionalities of these modules are described in this section.

4.1. Fast Fuzzy Anomaly Detector (FFAD). FFAD is designed in a way to identify anomaly network communication attempts. The

incorporation of Fuzzy logic ensures accurate anomaly detection swiftly. FFAD uses Traffic volumes, Baseline traffic, Time of the day, Protocol outliers, Location assessment, change in User privilege, Header interpretation, Payload dissection, Port scanning, Ping sweeps, Connection duration, and Number of connections as the constituents to detect an intruder [17]. These 12 parameters are individually categorized into 4 anomaly severity labels such as Nada, Low, Medium, and High. The severity labels are aggregated for all 12 parameters to detect anomaly network activities. The label weights are assigned from 0 to 3 for Nada to High respectively. Therefore, the maximum label weight ω_{max} is 3.

Traffic Volume(v_T) refers to the amount of data transferred through a node at any particular time. v_t is computed using Equation 1.

$$v_T = I_T \times t, \tag{1}$$

where I_T is the traffic intensity, t is the time.

FFAD measures the traffic volume difference v_{T_d} between the previous network volume v_{T_p} which is measured at a time $t - 1$ and the current traffic volume v_T as follows:

$$V_{T_d} = |V_T - V_{T_p}|. \tag{2}$$

Since the possible lowest traffic volume is 0 in IoT networks, the severity label L_{tr} is calculated by Equation 3.

$$L_{tr} = \begin{cases} Nada & \text{if } \frac{v_{T_d}}{v_{T_m}} < \frac{1}{4} \\ Low & \text{if } \frac{1}{4} \leq \frac{v_{T_d}}{v_{T_m}} < \frac{1}{2} \\ Medium & \text{if } \frac{1}{2} \leq \frac{v_{T_d}}{v_{T_m}} < \frac{3}{4} \\ High & \text{otherwise} \end{cases}, \tag{3}$$

where v_{T_m} is the maximum possible traffic volume.

Baseline traffic is the typical network traffic between specific nodes to cluster heads or base stations. FFDA uses the baseline traffic between the cluster heads and base station as in the following algorithm to determine the anomaly severity label.

Algorithm 1. Baseline traffic anomaly severity labeler

Input: Baseline traffic information

Output: Anomaly severity label

Step 1: Let n_{CH} be the number of cluster heads in the network

Step 2: Let Γ be the set of cluster heads with members $\{\gamma_1, \gamma_2 \dots \gamma_{n_{CH}}\}$

Step 3: Let b_T be the set of baseline traffic set by 1:1 correspondence to Γ with members $\{b_{T_1}, b_{T_2}, \dots b_{T_{n_{CH}}}\}$

Step 4: Initialize baseline traffic anomaly severity label $L_{bl} = \text{Nada}$

Step 5: Let L_{temp} be the temporary label

Step 6: $\forall i = 1 \rightarrow n_{CH}$

Step 7: Compute L_{temp} by Equation 4

Step 8: if $L_{bl} < L_{temp}$, set $L_{bl} = L_{temp}$

Step 7: return L_{bl}

The temporary baseline traffic anomaly severity label is determined by the following equation:

$$L_{temp} = \begin{cases} \text{Nada} & \text{if } \frac{b_{T_i}}{b_{T_m}} < \frac{1}{4} \\ \text{Low} & \text{if } \frac{1}{4} \leq \frac{b_{T_i}}{b_{T_m}} < \frac{1}{2} \\ \text{Medium} & \text{if } \frac{1}{2} \leq \frac{b_{T_i}}{b_{T_m}} < \frac{3}{4} \\ \text{High} & \text{otherwise} \end{cases} \quad (4)$$

The flowchart for the algorithm 1 is given below (Figure 2).

By this way, the baseline traffic anomaly severity label will get the highest severity index of all cluster head communications with the base station. That is, the baseline traffic anomaly label will get the value of ‘‘High’’ with even a single malicious node entry in any of the clusters in the entire network.

Time of day is one of the important factors in anomaly detection in scheduled communication an IoT network environment. The Time of day anomaly severity Label L_{td} is set to have only two labels in the FFAD module for ease of processing. The assignable labels are Nada and high. For all non-scheduled IoT network nodes, L_{td} is assigned with a Nada label For prescheduled IoT nodes, let Δ_{comm} be the set of preapproved time slots $\{\delta_1, \delta_2 \dots \delta_n\}$. Then for a communication attempt made at a time t_{ca} , the time of day anomaly severity label is assigned as in Equation 5.

$$L_{td} = \begin{cases} Nada & \text{if } t_{ca} \in \Delta \\ High & \text{otherwise} \end{cases} \quad (5)$$

The protocol outlier severity label L_{po} is set to high if a node tries to communicate with another node through a new protocol. Otherwise L_{po} is assigned with the label Nada. Location assessment is used to check whether an intruder node is trying to get into the network.

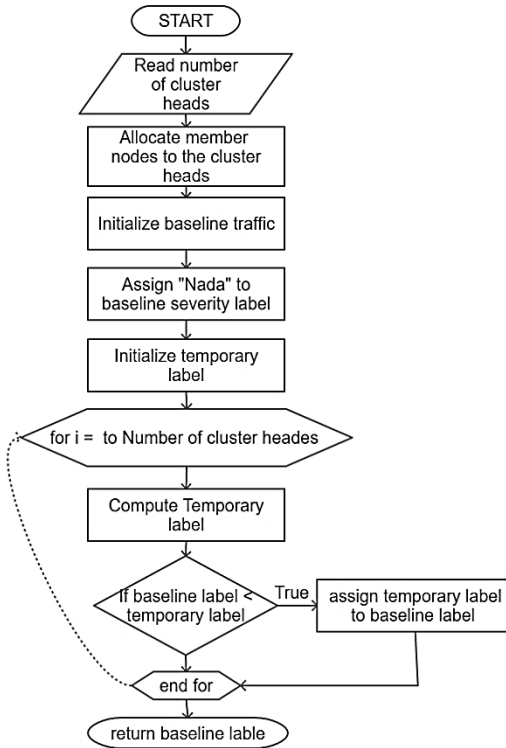


Fig. 2. Algorithm 1 flowchart

FFAD permits node mobility in IoT networks. But whenever a node location is changed to an impossible distance between two successive communications, it will be flagged as the intruder node. Let the geographical location of a node n_x at time t be $(Lat_{n_{x_t}}, Long_{n_{x_t}})$ and the previous location is $(Lat_{n_{x_{t-1}}}, Long_{n_{x_{t-1}}})$, then the displacement D_{n_x} of the node n_x is calculated using the following equation.

$$D_{n_x} = \sqrt{\left(\text{Lat}_{n_{x_t}} - \text{Lat}_{n_{x_{t-1}}}\right)^2 + \left(\text{Long}_{n_{x_t}} - \text{Long}_{n_{x_{t-1}}}\right)^2}. \quad (6)$$

The Location anomaly severity label is assigned as follows.

$$L_{lo} = \begin{cases} \text{High if } D_{n_x} > D_{max_{n_x}}, \\ \text{Nada otherwise} \end{cases}, \quad (7)$$

where $D_{max_{n_x}}$ is the maximum possible displacement for the node n_x .

In FFAD, Any change in user privilege sets the Privilege anomaly severity label to high. Let $P_{n_{x_t}}$ be the privilege of the node n_x at a time t , the privilege anomaly severity label is determined as follows:

$$L_{pr} = \begin{cases} \text{High if } P_{n_{x_t}} \neq P_{n_{x_{t-1}}}, \\ \text{Nada otherwise} \end{cases}, \quad (8)$$

where $P_{n_{x_{t-1}}}$ is the previous privilege of the node n_x .

A typical data packet header contains information such as header length, total length, source address, destination address, and time-to-live (TTL); FFAD uses a header flag set H with members $\{h_1, h_2, h_3, h_4, h_5\}$ for the corresponding header information. Any change in any of these information sets the corresponding flag to 1. The quantification $Q(H)$ of the overall set H is computed using Equation 9.

$$Q(H) = \frac{1}{5} \sum_{i=1}^5 h_i. \quad (9)$$

The header anomaly severity label L_{hdr} is determined by the equation:

$$L_{hdr} = \begin{cases} \text{Nada if } \frac{Q(H)}{5} = 0 \\ \text{Low if } 0 < \frac{Q(H)}{5} \leq 1/4 \\ \text{Medium if } 1/4 < \frac{Q(H)}{5} \leq 1/2 \\ \text{High otherwise} \end{cases}. \quad (10)$$

Every IoT node is deployed to send one or more data from the sensor that can be categorized as numerical or text data. The payload dissection allows the FFAD module to monitor the data type during the transaction. If a change in the data type means the behavior of the node is suspicious. The Payload anomaly severity label L_{pl} is set to High if there is a change detected during the communication.

Any attempt to detect the possible connectivity ports in the base station between the communications will set the Port scanning anomaly severity label L_{ps} to High. Otherwise L_{ps} will be assigned to Nada. The tantamount procedure is followed to set the Ping sweep anomaly severity label L_{pi} .

Connection duration is also an important parameter in detecting anomalies. Usually, a drastic increase in connection durations indicates a possibility of an intruder attack. Let $\lambda_{n_{x_{t-1}}}$ and $\lambda_{n_{x_t}}$ the two successive connection durations of the node n_x , the connection duration anomaly severity label is determined by the following equation:

$$L_{cd} = \begin{cases} \text{High if } (\lambda_{n_{x_t}} - \lambda_{n_{x_{t-1}}}) > 1/4 \lambda_{n_{x_{max}}}, \\ \text{Nada otherwise} \end{cases} \quad (11)$$

where $\lambda_{n_{x_{max}}}$ is the maximum permitted connection duration for the node n_x .

A change in the number of connections is also treated in the FFAD module since most of the IoT nodes are accustomed to using a single connection at a time. Let i_{n_x} and c_{n_x} be the initial and current number of connections of the node n_x , the number of connections anomaly sensitivity label L_{nc} is set to high wherever the value of c_{n_x} is greater than i_{n_x} .

The overall quantization of FFAD anomaly severity labels is computed using equation 12:

$$Q(L) = \frac{L_{tr} + L_{bl} + L_{td} + L_{po} + L_{lo} + L_{pr} + L_{hdr} + L_{pl} + L_{ps} + L_{pi} + L_{cd} + L_{nc}}{\omega_{max} \times \text{number of parameters}}. \quad (12)$$

FFAD sets an Anomaly detection flag if there is more than a 25% chance of any combination of the above-discussed anomalies as by the following equation.

$$\text{Anomaly} = \begin{cases} \text{TRUE if } Q(L) > 1/4, \\ \text{FALSE otherwise} \end{cases} \quad (13)$$

4.2. Legacy Naïve Bayes Attack Classifier (LNBAC). LNBAC is intended to classify a network anomaly into either of the following categories. They are Denial-of-Service, User-to-Root, Remote-to-Local, Probe and Normal. As per the standard Naïve Bayes algorithm is customized here to operate with 5 above-mentioned classifications and 12 features discussed in the FFAD module.

Algorithm 2. LNBAC Model Training

Input: Network Transaction Data

Output: Trained Model

Step 1: Let Attack Classification Set $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\}$

Step 2: Let the selected feature set be $f = \{f_1, f_2 \dots f_{12}\}$

Step 3: Let $P(\alpha)$ be the prior probability

Step 4: Let $P(f)$ be the Marginal Probability

Step 5: Let $P(\alpha|f)$ be the posterior probability of the class α for the feature f

Step 6: Let $P(f|\alpha)$ be the probability of the feature f given that the probability of the class α

Step 7: $\forall i = 1 \rightarrow$ Attack classifications = 5 :: $\forall j = 1 \rightarrow$ Number of features = 12 :=

Step 8: Compute $P(\alpha_i)$ for the feature f_j

Step 9: Compute the conditional probability $P(f_j|\alpha_i)$

Step 8: Compute $P(\alpha|f) = \frac{P(f|\alpha)P(\alpha)}{P(f)}$

Step 9: return $P(\alpha|f)$

The flow diagram for the LNBAC algorithm is given below (Figure 3). By this way, the attack classification is achieved by the LNBAC module.

Probe (Probing Attacks): These are attempts to gather information about a target network or system to identify potential vulnerabilities. Probing attacks are usually considered less severe because they are preliminary steps attackers take before launching more damaging attacks.

R2L (Remote-to-Local Attacks): These attacks involve unauthorized attempts to gain access to a local system from a remote machine. They are often considered more severe than probing attacks, as they can potentially lead to unauthorized access and data compromise.

U2R (User-to-Root Attacks): User-to-Root attacks are attempts by a user to gain administrative privileges on a system. These attacks are more serious because if successful, they can provide the attacker with full control over the compromised system, allowing them to install malicious software or manipulate data.

DoS (Denial-of-Service Attacks): Denial-of-Service attacks are among the most severe types of attacks. They involve overwhelming a system,

network, or service with excessive traffic or resource requests, rendering it unavailable to legitimate users. These attacks can lead to significant disruptions in services, financial losses, and reputational damage [18].

Normal network transaction refers to there is no evidence of any intruder attacks or network anomalies in the particular network communication attempt.

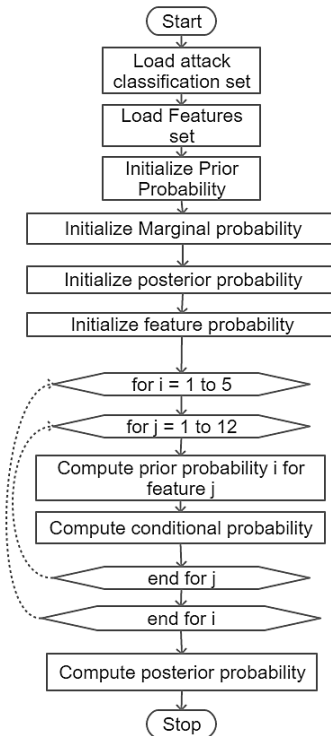


Fig. 3. LNBAC flowchart

4.3. Variable RSA Security Scheme (VRSS). The purpose of the VRSS module is to change the security strength dynamically based on the outputs of the FFAD and LNBAC modules. Instead of using a fixed security key size in RSA, the dynamic key selection VRSS will improve the security as well as the performance of the network by reducing the computational complexity when there is no or a little security anomaly detected. Whenever there is a hazard signal from FFAD and the classification of the attack from LNBAC, VRSS determines the required security key size circumspectly.

The base key size of VRSS is set to 256 bits, which can scale up to 1152 bits [19]. The key size for the RSA procedure is determined using the Anomaly severity index $Index_s$ and attack threat index $Index_t$.

The overall anomaly severity index $Index_s$ is determined as follows:

$$Index_s = \begin{cases} 1 & \text{if } Q(L) < 1/4 \\ 2 & \text{if } 1/4 \leq Q(L) < 1/2 \\ 3 & \text{if } 1/2 \leq Q(L) < 3/4 \\ 4 & \text{otherwise} \end{cases} \quad (14)$$

The attack types Probe, R2L, U2R and DoS are assigned with the weights 2,3,4 and 5 respectively. The normal transaction type is assigned with the weight 1 to ensure basic security. That is the value of $Index_t$ will be 1, 2, 3, 4 and 5 for Normal, Probe, R2L, U2R and DoS respectively. The key size k is determined by using the following equation:

$$k = \left(\frac{1}{n_{sl}} \times (2^{n_{sl}} \times 2^{n_t}) \right) \times (Index_s + Index_t), \quad (15)$$

where n_{sl} is the number of sensitivity labels, and n_t is the number of network transaction types.

Be like that the integration of FFAD, LNBAC and VRSS modules ensures fast and secure network communications in IoT network environments.

5. Experimental Setup. A computer with an i7 -8250U processor with 6MB Cache, 16GB DDR4 RAM and 1TB SSD storage is used to develop and evaluate the discussed procedures. Visual Studio IDE [20] is used to create the implementation solution, and C++ 20.0 [21] programming language is used to code the methodologies of ASORI. OPNET [22], which stands for "Optimized Network Engineering Tools", was a widely used software suite for network simulation, modeling, and performance analysis. The software allowed engineers, researchers, and network professionals to simulate and analyze various aspects of computer networks, telecommunications systems, and other communication technologies. OPNET provides several features such as Network Modeling, Simulation, Performance Analysis, Protocol Evaluation, and Resource monitoring/management. OPNET facilitated the testing of various network scenarios without the need for physical implementation, helping users identify potential issues before deployment. Users could analyze resource

utilization, bottlenecks, and optimization strategies within the network. OPNET was also used in academia to teach networking concepts and provide hands-on experience with network simulation.

The proposed IoT network Security method has extensive applications across diverse sectors, revolutionizing industries and enhancing efficiency, safety, and sustainability. From asset tracking and environmental monitoring to smart agriculture and energy management, IoT devices enable real-time data collection and analysis, empowering stakeholders to make informed decisions. In healthcare, IoT facilitates remote patient monitoring and personalized care, while in smart cities, it optimizes urban infrastructure and services for better quality of life. Retail and manufacturing benefit from IoT-driven automation and optimization, while home automation enhances convenience and energy efficiency for consumers. Industrial processes become more efficient and predictive with IoT-enabled automation, while safety and security are heightened through real-time monitoring and response capabilities. Overall, the broad spectrum of IoT applications underscores its transformative potential across various domains, shaping the future of technology-enabled innovation and connectivity.

6. Results and Analysis. Two different categories of results are obtained during the evaluation process. The first one is anomaly detection parameters such as Accuracy, Precision, Sensitivity, Specificity and F-Score. The second category is about network performance metrics such as Throughput, Latency, Jitter, End-to-End delay, Packet Delivery Ratio, Power consumption, and Security. Readings are taken for 1 real-world hour for every 6 minutes. Thus, there are 10 different timestamps used to log the parameters during the evaluation.

6.1. Accuracy. Anomaly detection accuracy is an important parameter in maintaining network stability. Since an anomaly can be an intruder attack, the anomaly detection process is a vital one in network security. It is calculated using the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The accuracy is calculated using the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The ASORI method scored 99.135% of accuracy which is 1.2% higher than the nearest follower SIMA. The average accuracy of the proposed ASORI is 98.95% which is 1.4% higher than the second best SIMA method. The performance rank based on accuracy parameters is

ASORI, SIMA, ASDARS, EHDA, AIES and LCXMAC listed as the best. The measured accuracy values are provided in Table 2.

Table 2. Accuracy (%)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	96.37	91.41	95.57	97.61	91.64	98.82
2	96.13	91.72	95.15	97.56	91.37	99.08
3	96.33	91.68	95.49	97.23	91.51	99.03
4	96.66	91.40	95.68	97.72	91.43	98.86
5	96.22	91.86	95.38	97.93	91.43	99.01
6	96.61	91.18	95.61	97.25	91.44	98.79
7	96.54	91.36	95.52	97.50	91.65	99.04
8	96.76	91.06	95.28	97.36	91.16	99.13
9	96.65	91.14	95.52	97.54	91.76	98.97
10	96.37	91.28	95.67	97.76	91.31	98.78

6.2. Precision. Precision, in the context of data science and machine learning, is a metric that measures the accuracy of positive predictions made by a model. It is a concept often used in binary classification problems, where the goal is to classify instances into one of two classes: "positive" and "negative".

Precision is calculated using the formula $\frac{TP}{TP+FP}$. The computed Precision values are given in Table 3.

Table 3. Precision (%)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	95.69	92.34	96.93	96.45	92.47	98.05
2	95.14	92.78	96.1	96.92	92.01	98.46
3	95.28	92.56	96.64	96.22	92.35	98.35
4	95.49	92.69	96.78	96.53	92	98.23
5	95.13	92.81	96.12	96.97	92.66	98.22
6	95.47	92.09	96.74	96.12	92.68	98.18
7	95.90	92.35	96.99	96.47	92.81	98.3
8	95.54	92.01	96.56	96.35	92.27	98.53
9	95.87	92.22	96.07	96.23	92.89	98.32
10	95.38	92.48	96.68	96.98	92.49	98.06

As per the observations, ASORI scores the highest precision value of 98.53%, which is 1.54% higher than 96.99% of the closest performer EHDA. The precision average of ASORI is 98.27% during the overall experiment. The performance rank based on the average precision is

ASORI, EHDA, SIMA, ASDARSA, LCXMAC, and AIES listed as the best.

6.3. Sensitivity. Sensitivity refers to a metric used to evaluate the performance of a classification model, particularly in binary classification problems. It measures the ability of the model to correctly identify positive instances from the total actual positive instances. Sensitivity is also known as recall, hit rate, or true positive rate. Sensitivity is calculated using the formula $\frac{TP}{TP+FN}$. The observed sensitivity values are given in Table 4.

Table 4. Sensitivity (%)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	97.00	90.65	94.37	98.74	90.96	99.59
2	97.05	90.86	94.32	98.18	90.84	99.71
3	97.32	90.96	94.48	98.20	90.83	99.71
4	97.79	90.36	94.69	98.88	90.97	99.47
5	97.26	91.08	94.72	98.87	90.44	99.79
6	97.70	90.44	94.60	98.34	90.44	99.39
7	97.14	90.56	94.22	98.50	90.71	99.78
8	97.92	90.29	94.15	98.33	90.27	99.74
9	97.39	90.27	95.02	98.81	90.85	99.63
10	97.31	90.31	94.77	98.51	90.37	99.49

The experimental results show that ASORI work scored 99.78% anomaly sensitivity, which is higher than other methods. The average sensitivity of ASORI is 99.63% that shows the stability in terms of anomaly sensitivity.

6.4. Specificity. Specificity is one of the important metrics used to evaluate the performance of a binary classification model, particularly in scenarios where correctly identifying negative instances is crucial. Specificity measures the ability of the model to correctly identify negative instances from the total actual negative instances.

Specificity is calculated by $\frac{TN}{TN+FP}$. The measured specificity values for the proposed and existing methods are given in Table 5.

The ASORI method gained a specificity score of 98.55% which is 1.5% higher than the nearest achievement of 97.03% of the SIMA method. The performance rank based on the average specificity is ASORI, SIMA, EHDA, ASDARS, LCXMAC, and AIES with the scores 98.29%, 96.59%, 96.49%, 95.58%, 92.31% and 92.27% in their respective order listed as the best.

Table 5. Specificity (%)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	95.75	92.19	96.84	96.53	92.34	98.08
2	95.23	92.62	96.02	96.96	91.91	98.48
3	95.38	92.43	96.56	96.29	92.22	98.37
4	95.59	92.50	96.71	96.61	91.91	98.25
5	95.23	92.67	96.06	97.03	92.47	98.25
6	95.57	91.94	96.66	96.21	92.49	98.20
7	95.95	92.20	96.90	96.54	92.64	98.32
8	95.65	91.85	96.47	96.42	92.10	98.55
9	95.93	92.05	96.03	96.33	92.73	98.34
10	95.47	92.30	96.61	97.03	92.31	98.09

6.5. F-Score. The F-score, also known as the F1-score, is a metric used in classification tasks to assess the performance of a model, particularly in scenarios where class imbalance exists. It combines precision and recall into a single value and provides a balanced measure of a model's accuracy. The formula for F-Score calculation is $2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$. The F-Score values of the compared methods are given in Table 6.

Table 6. F-Score

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	0.9634	0.9149	0.9563	0.9758	0.9171	0.9882
2	0.9609	0.9181	0.9520	0.9754	0.9142	0.9908
3	0.9629	0.9175	0.9555	0.9720	0.9159	0.9902
4	0.9663	0.9151	0.9572	0.9769	0.9148	0.9885
5	0.9618	0.9194	0.9541	0.9791	0.9153	0.9900
6	0.9657	0.9126	0.9566	0.9722	0.9154	0.9878
7	0.9652	0.9144	0.9558	0.9747	0.9175	0.9903
8	0.9672	0.9114	0.9534	0.9733	0.9126	0.9913
9	0.9662	0.9123	0.9554	0.9750	0.9186	0.9897
10	0.9633	0.9138	0.9571	0.9774	0.9142	0.9877

The performance rank in terms of F-Score is ASORI, SIMA, ASDARS, EHDA, LCXMAC, and AIES the F-Score index averages 0.9894, 0.9752, 0.9643, 0.9554, 0.9156 and 0.9150.

The collective performance of the proposed ASORI method is perpetually higher than the compared methods in terms of Accuracy, Precision, Sensitivity, Specificity and F-Score. The experimental result evidently validates the enhanced performance of the ASORI method in Anomaly detection.

6.6. Throughput. Throughput refers to the rate at which data is successfully transmitted or received over a network. It is a measure of the

network's efficiency and capacity. The OPNET measures the throughput values during the simulation and the values are logged in Table 7.

Table 7. Throughput (kbps)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	27287	24274	27837	28688	24568	29303
2	27084	24039	27704	28905	24905	29407
3	27072	24082	27743	28917	24412	29475
4	27317	24418	27903	28933	24872	29637
5	27097	24260	28247	28908	25006	29642
6	27126	24281	28189	29051	24538	29743
7	27366	24043	27612	29029	24647	29332
8	27027	24330	27910	28524	24480	29450
9	27231	24541	28246	28543	24782	29558
10	27272	24528	27930	28504	24729	29589

As per the simulation results, the highest throughput of 29743 kbps is achieved by ASORI during the 6th timestamp. The comparison graphs are given in Figure 4.

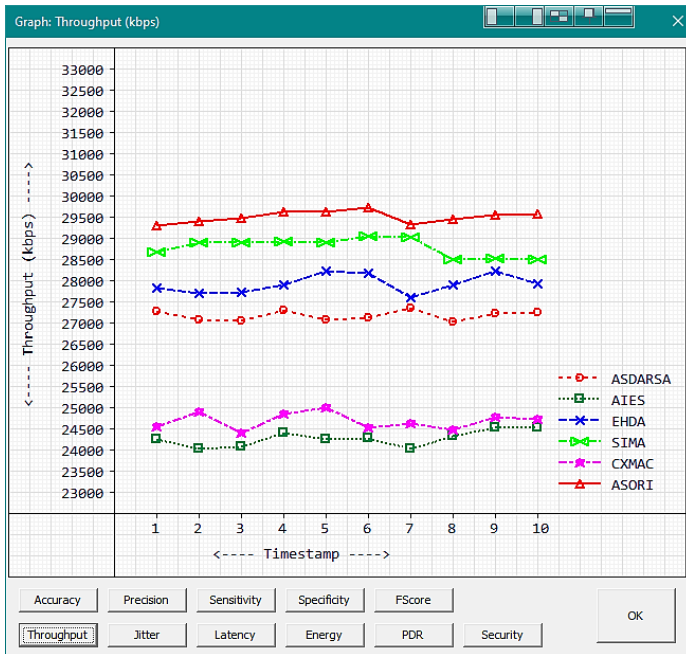


Fig. 4. Throughput graph

6.7. Latency. Latency refers to the delay or time lag that occurs when data packets travel from one point in a network to another. In general, Latency is measured in millisecond (ms) units. The observed Latency values during the simulation are given in Table 8. Latency is inversely proportional to the performance of the network architecture. Lower latency indicates the higher performance of the network.

The lowest latency 168mS is achieved by ASORI at the 6th timestamp during the simulation. The average latency of ASORI is about 180mS which is lesser than other compared methods (Figure 5).

Table 8. Latency (mS)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	299	460	269	224	444	191
2	310	473	277	212	426	186
3	310	470	275	212	453	182
4	297	452	266	211	428	173
5	309	461	248	212	421	173
6	307	460	251	205	446	168
7	295	472	282	206	440	190
8	313	457	266	233	449	183
9	302	446	248	232	433	177
10	300	446	265	234	436	176

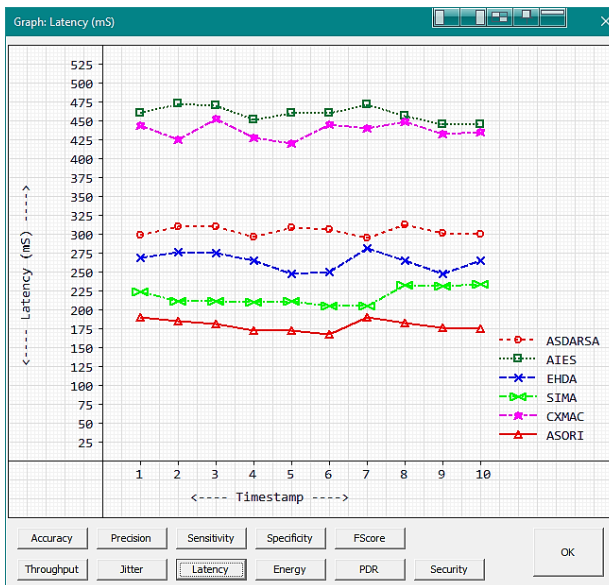


Fig. 5. Latency graph

6.8. Jitter. Jitter in the context of networking refers to the variation in the delay of packet delivery in a network. It is the irregular timing of data packets arriving at their destination. Higher jitter values lead to inconsistent network performance. The observed jitter values during the simulation are given in Table 9.

Table 9. Jitter (mS)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	102	150	93	80	145	70
2	105	154	95	76	140	68
3	105	153	95	76	148	67
4	101	148	92	76	141	64
5	105	151	87	76	139	64
6	104	150	88	74	146	63
7	101	154	97	74	144	69
8	106	149	92	82	147	67
9	103	146	87	82	142	66
10	102	146	92	83	143	65

The lowest jitter reading of 63mS is achieved by the ASORI method during the entire simulation. The jitter average of the ASORI method is 66mS which is lesser than other compared methods (Figure 6).

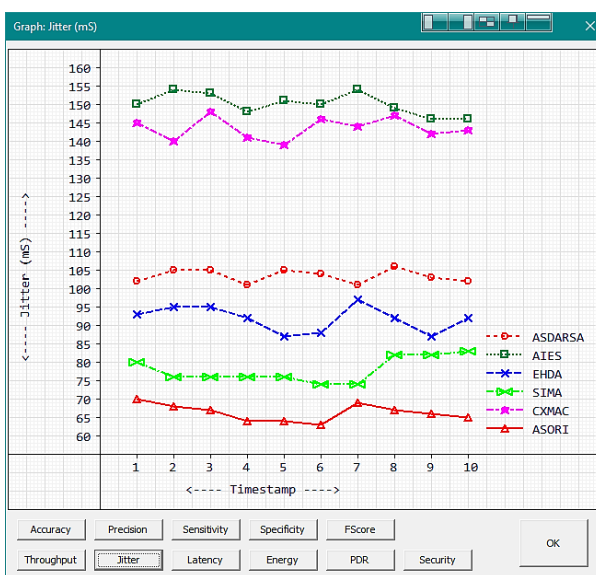


Fig. 6. Jitter graph

6.9. Energy. Energy efficiency is of paramount importance in Internet of Things (IoT) networks due to several key reasons such as power resource limitation of the nodes, Scalability of the network, and limited maintainability, Energy consumption is measured in millijoules units in a network. The energy readings during the simulation are given in Table 10.

Table 10. Energy (mJ)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	880	586	796	670	629	576
2	852	600	784	666	629	517
3	826	547	807	663	590	504
4	848	578	799	696	658	580
5	847	560	765	696	605	530
6	860	554	772	713	630	537
7	827	582	755	710	626	567
8	873	571	744	712	604	540
9	820	578	760	662	599	553
10	804	537	787	646	615	556

The experimental results show that the lowest energy consumption is achieved by the ASORI method during the complete simulation. ASORI is managed to operate with an average energy consumption of 546mJ. The comparison graph is given in Figure 7.

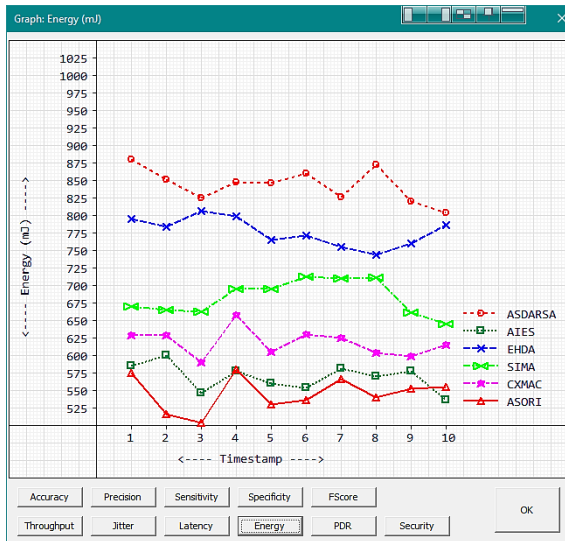


Fig. 7. Energy graph

6.10. Packet Delivery Ratio. Packet Delivery Ratio is a metric used to measure the reliability and performance of a communication network, particularly in wireless and packet-switched networks. It indicates the proportion of the successfully delivered packets compared to the total number of the sent packets. Table 11 is provided with the PDR values during the simulation.

PDR is a directly proportional network metric. Higher PDR indicates the higher performance of the network. The highest PDR 99.25% is achieved by the AOSRI method. The PDR average of the ASORI method is 99.17% which is also higher than the other compared methods (Figure 8).

Table 11. Packet Delivery Rate (%)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	98.43	97.42	98.61	98.90	97.52	99.10
2	98.36	97.35	98.57	98.97	97.64	99.14
3	98.36	97.36	98.58	98.97	97.47	99.16
4	98.44	97.47	98.63	98.98	97.62	99.21
5	98.37	97.42	98.75	98.97	97.67	99.21
6	98.38	97.43	98.73	99.02	97.51	99.25
7	98.46	97.35	98.54	99.01	97.55	99.11
8	98.34	97.44	98.64	98.84	97.49	99.15
9	98.41	97.51	98.75	98.85	97.59	99.19
10	98.42	97.51	98.64	98.83	97.58	99.20

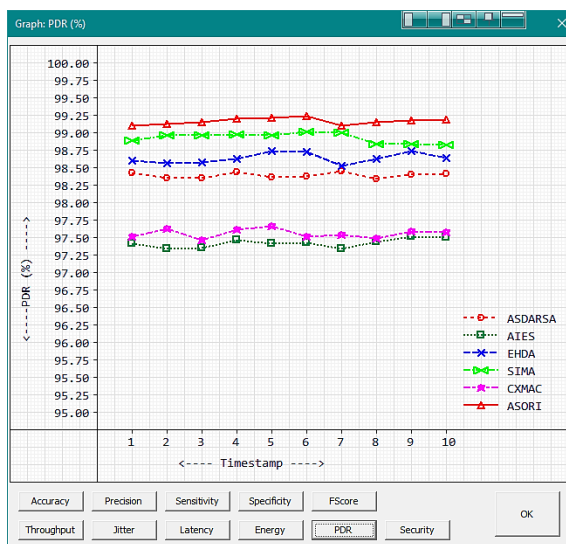


Fig. 8. Packet Delivery Ratio

6.11. Security. Security is crucial for Internet of Things (IoT) devices due to several significant reasons such as privacy protection, data integrity, device control, safety concerns, and long lifecycles. OPNET can determine the security level of the network environment under simulation by triggering a variety of intruder attacks. The measured Security scores of the examined methods are given in Table 12.

Table 12. Security Score (%)

Timestamp	ASDARSA	AIES	EHDA	SIMA	LCXMAC	ASORI
1	96.0000	91.5294	95.4118	97.3529	91.5882	99.5941
2	96.5882	91.7059	95.3529	97.2353	91.4706	99.5353
3	96.1176	91.0588	95.5882	97.2353	91.3529	99.3000
4	96.4118	91.2941	95.5882	97.1765	91.1176	99.5353
5	96.2941	91.2941	95.7647	97.7059	91.0000	99.6529
6	96.1765	91.7059	95.2941	97.4118	91.5882	99.4176
7	96.2941	91.2353	95.1765	97.7647	91.7059	99.4176
8	96.2353	91.8235	95.2941	97.8235	91.3529	99.5941
9	96.2941	91.7647	95.5294	97.7647	91.7059	99.3000
10	96.5294	91.7059	95.2941	97.7647	91.5294	99.4765

The comparison graph for security score is given in Figure 7.

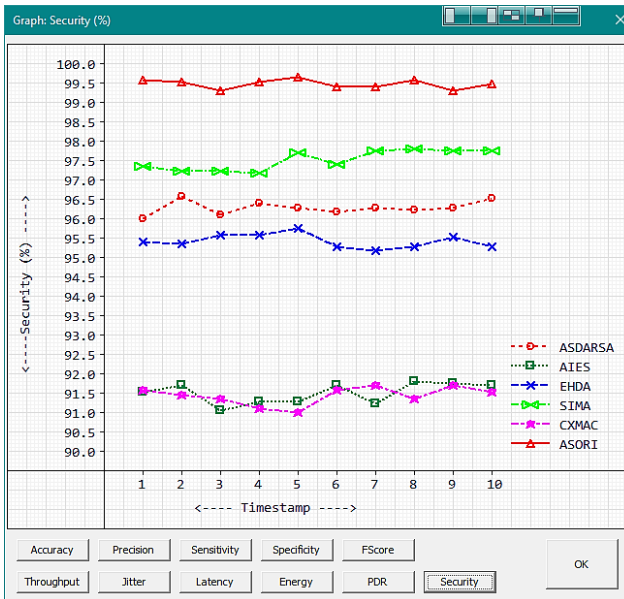


Fig. 9. Security score graph

The evaluation results show that the proposed ASORI method is the method that achieved the maximum security score of 99.66% at the 5th timestamp during the simulation. ASORI is capable of maintaining a security score of not less than 99.3%. The average security score of ASORI for the entire simulation is 99.48%. An improvement of 1.83% is achieved by the ASORI method than the security score of 97.82% of the nearest performing method SIMA.

7. Conclusion. The requirements can vary based on the specific use case, industry, and goals of the IoT deployment. A well-designed IoT architecture will carefully address these considerations to create a robust, secure, and efficient ecosystem which is attempted in this work. The experiments carried out during the evaluation process of the proposed ASORI method show that the work comes in handy with the expected advantages. The three novel modules FFAD, LNBAK and VRSS serve the purpose affirmatively. Improved IoT security is critical for safeguarding data privacy, protecting against unauthorized access, and ensuring the reliability of IoT devices and networks is addressed by the proposed method. Additionally, advancements in technologies like AI and blockchain offer innovative solutions for enhancing security and integrity. However, ongoing collaboration, innovation, and proactive measures are essential to stay ahead of evolving threats and maintain trust in IoT systems.

The advancements in terms of both network anomaly detection and overall network performance of the ASORI method put forward that the work can be encouraged for real-time IoT network environments. The dependency of RSA core is the limitation of ASORI work, and the incorporation of modern cryptography algorithms could be the future work.

Availability of Dataset: The work is intended to ensure dynamic IoT network security, thus dynamic network transactional simulation is used for evaluation.

Code Availability: The complete implementation source code is available on GitHub, the link will be provided on request.

References

1. Farrukh H., Ozmen M., Kerem Ors F. Celik Z. One Key to Rule Them All: Secure Group Pairing for Heterogeneous IoT Devices. 2023 IEEE Symposium on Security and Privacy (SP). 2023. pp. 3026–3042. DOI: 10.1109/SP46215.2023.10179369.
2. Quy V., Hau N., Anh D., Ngoc L. Smart healthcare IoT applications based on fog computing: architecture, applications and challenges. *Complex & Intelligent Systems*. 2022. vol. 8. pp. 3805–3815. DOI: 10.1007/s40747-021-00582-9.
3. Rejeb A., Rejeb K., Simske S., Treiblmaier H., Zailani S. The big picture on the internet of things and the smart city: a review of what we know and what we need to know. *Internet of Things*. 2022. vol. 19. DOI: 10.1016/j.iot.2022.100565.

4. Inayat U., Zia M., Mahmood S., Khalid H., Benbouzid M. Learning-Based Methods for Cyber Attacks Detection in IoT Systems: A Survey on Methods, Analysis, and Future Prospects. *Electronics*. 2022. vol. 11(9). DOI: 10.3390/electronics11091502.
5. Hatami M., Leinonen M., Chen Z., Pappas N., Codreanu M. On-Demand AoI Minimization in Resource-Constrained Cache-Enabled IoT Networks With Energy Harvesting Sensors. *IEEE Transactions on Communications*. 2022. vol. 70. no. 11. pp. 7446–7463. DOI: 10.1109/TCOMM.2022.3208873.
6. Nagaraju R, C V, J K, G M, Goyal SB, Verma C, Safirescu C, Mihaltan T. Secure Routing-Based Energy Optimization for IoT Application with Heterogeneous Wireless Sensor Networks. *Energies*. 2022. vol. 15(13). DOI: 10.3390/en15134777.
7. Fotohi R., Bari S., Yusefi M. Securing Wireless Sensor Networks against Denial-of-Sleep Attacks Using RSA Cryptography Algorithm and Interlock Protocol. *International Journal of Communication Systems*. 2019. vol. 33(4). DOI: 10.1002/dac.4234.
8. Yang S.-K., Shiue Y.-M., Su Z.-Y., Liu I.-H., Liu C.-G. An Authentication Information Exchange Scheme in WSN for IoT Applications. *IEEE Access*. 2020. vol. 8. pp. 9728–9738. DOI: 10.1109/ACCESS.2020.2964815.
9. Ullah A., Said G., Sher M., Ning H. Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN. *Peer-to-Peer Networking and Applications*. 2020. vol. 13. pp. 163–174. DOI: 10.1007/s12083-019-00745-z.
10. Singh D., Kumar B., Singh S., Chand S. A Secure IoT-Based Mutual Authentication for Healthcare Applications in Wireless Sensor Networks Using ECC. *International Journal of Healthcare Information Systems and Informatics*. 2021. vol. 16. no. 2. pp. 21–48. DOI: 10.4018/IJHISI.20210401.oa2.
11. Ahmad A., Ullah A., Feng C., Khan M., Ashraf S., Adnan M., Nazir S., Ullah Khan H. Towards an Improved Energy Efficient and End-to-End Secure Protocol for IoT Healthcare Applications. *Security and Communication Networks*. 2020. vol. 2020. DOI: 10.1155/2020/8867792.
12. Nada A., Bayoumi M. Development of a constraint stabilization method of multibody systems based on fuzzy logic control. *Multibody System Dynamics*. 2024. vol. 61. pp. 233–265. DOI: 10.1007/s11044-023-09921-9.
13. Liu L., Xue D., Zhang S. General type industrial temperature system control based on fuzzy fractional-order PID controller. *Complex and Intelligent Systems*. 2023. vol. 9. pp. 2585–2597. DOI: 10.1007/s40747-021-00431-9.
14. Sivapriya N., Ravi T. Efficient Fuzzy based Multi-constraint Multicast Routing with Multi-criteria Enhanced Optimal Capacity–Delay Tradeoff. *International Journal of Scientific & Technology Research*. 2019. vol. 8(8). pp. 1468–1473.
15. Jasim A., Kashmar A. An Evaluation of RSA and a Modified SHA-3 for a New Design of Blockchain Technology. *Artificial Intelligence for Smart Healthcare. EAI/Springer Innovations in Communication and Computing*. Cham: Springer, 2023. pp. 477–489. DOI: 10.1007/978-3-031-23602-0_28.
16. Abid R., Iwendi C., Javed A., Rizwan M., Jalil Z., Anajemba J., Biamba C. An optimised homomorphic CRT-RSA algorithm for secure and efficient communication. *Personal and Ubiquitous Computing*. 2023. vol. 27. pp. 1405–1418. DOI: 10.1007/s00779-021-01607-3.
17. Anushiya R., Lavanya V. A new deep-learning with swarm based feature selection for intelligent intrusion detection for the Internet of things. *Measurement: Sensors*. 2023. vol. 26. DOI: 10.1016/j.measen.2023.100700.
18. Roldan-Gomez J., Boubeta-Puig J., Carrillo-Mondejar J., Manuel Castelo Gomez J., del Rincon J. An automatic complex event processing rules generation system for the recognition of real-time IoT attack patterns. *Engineering Applications of Artificial Intelligence*. 2023. vol. 123. DOI: 10.1016/j.engappai.2023.106344.

19. Size considerations for public and private keys. Available at: <https://www.ibm.com/docs/en/zos/2.3.0?topic=certificates-size-considerations-public-private-keys> (accessed 26.01.2024).
20. Visual Studio 2022. Available at: <https://visualstudio.microsoft.com/vs/> (accessed 10.02.2024).
21. Features of C++ 20. Available at: <https://www.geeksforgeeks.org/features-of-c-20/> (accessed 04.02.2024).
22. OPNET Network Simulator. Available at: <https://opnetprojects.com/opnet-network-simulator/> (accessed 16.14.2024).

Jenifer R. Rita — Research Scholar, PG & Research Department of Computer Science, Cauvery College for Women (Autonomous), Affiliated to Bharathidasan University. Research interests: computer science. The number of publications — 1. rita.jenifer@gmail.com; Cauvery Nagar Main Rd, Annamalai Nagar, 620018, Tiruchirappalli, Tamil Nadu, India; office phone: +91(431)275-1232.

Prakash V. Sinthu — Professor, PG & Research Department of Computer Science, Cauvery College for Women (Autonomous), Affiliated to Bharathidasan University. Research interests: computer science. The number of publications — 54. sinthujanita@gmail.com; Cauvery Nagar Main Rd, Annamalai Nagar, 620018, Tiruchirappalli, Tamil Nadu, India; office phone: +91(431)275-1232.

Acknowledgements. This research has been supported by the grant obtained under the scheme of Seed Money for Research projects from Cauvery College for Women (Autonomous), Tiruchirappalli.

Р. ДЖЕНИФЕР, В.Д. ПРАКАШ
**АЛГОРИТМ RIVEST-SHAMIR-ADLEMAN,
ОПТИМИЗИРОВАННЫЙ ДЛЯ ЗАЩИТЫ УСТРОЙСТВ
ИНТЕРНЕТА ВЕЩЕЙ ОТ КОНКРЕТНЫХ АТАК**

Дженифер Р., Пракаш В.Д. Алгоритм Rivest-Shamir-Adleman, оптимизированный для защиты устройств Интернета вещей от конкретных атак.

Аннотация. Устройства Интернета вещей играют важнейшую роль в современном мире во многих отношениях, поскольку они обеспечивают поддержку для зондирования окружающей среды, автоматизации и ответственного сохранения ресурсов. В «умном» мире повсеместное присутствие устройств Интернета вещей в повседневной жизни неизбежно. Широкое использование устройств Интернета вещей привлекает к себе любопытные взгляды злонамеренных хакеров. Несмотря на то, что существует несколько систем и протоколов безопасности, доступных для обычных беспроводных сетей, наблюдается необходимость в разработке современного механизма безопасности исключительно для сетевых сред Интернета вещей. Эта работа представляет улучшения безопасности сетей Интернета вещей. В ней собраны три специализированных способа для достижения более высоких показателей безопасности в сетевых средах Интернета вещей. Fast Fuzzy Anomaly Detector, Legacy Naïve Bayes Attack Classifiers и Variable Security Schemes of Rivest-Shamir-Adleman algorithm – это новые модули, представленные в этой работе, сокращенно ASORI. Уникальные преимущества встроенного механизма сертификации Интернета вещей и выбор динамической стратегии безопасности являются новшествами, представленными в данной работе. Модель ASORI была проверена с использованием промышленного стандартного симулятора сети OPNET для обеспечения улучшенной безопасности наряду с существенными улучшениями параметров производительности сети.

Ключевые слова: интернет вещей (IoT), сетевая безопасность, нечеткое обнаружение аномалий, наивная байесовская классификация, RSA.

Литература

1. Farrukh H., Ozmen M., Kerem Ors F. Celik Z. One Key to Rule Them All: Secure Group Pairing for Heterogeneous IoT Devices. 2023 IEEE Symposium on Security and Privacy (SP). 2023. pp. 3026–3042. DOI: 10.1109/SP46215.2023.10179369.
2. Quy V., Hau N., Anh D., Ngoc L. Smart healthcare IoT applications based on fog computing: architecture, applications and challenges. Complex & Intelligent Systems. 2022. vol. 8. pp. 3805–3815. DOI: 10.1007/s40747-021-00582-9.
3. Rejeb A., Rejeb K., Simske S., Treiblmaier H., Zailani S. The big picture on the internet of things and the smart city: a review of what we know and what we need to know. Internet of Things. 2022. vol. 19. DOI: 10.1016/j.iot.2022.100565.
4. Inayat U., Zia M., Mahmood S., Khalid H., Benbouzid M. Learning-Based Methods for Cyber Attacks Detection in IoT Systems: A Survey on Methods, Analysis, and Future Prospects. Electronics. 2022. vol. 11(9). DOI: 10.3390/electronics11091502.
5. Hatami M., Leinonen M., Chen Z., Pappas N., Codreanu M. On-Demand AoI Minimization in Resource-Constrained Cache-Enabled IoT Networks With Energy Harvesting Sensors. IEEE Transactions on Communications. 2022. vol. 70. no. 11. pp. 7446–7463. DOI: 10.1109/TCOMM.2022.3208873.

6. Nagaraju R, C V, J K, G M, Goyal SB, Verma C, Safirescu C, Mihaltan T. Secure Routing-Based Energy Optimization for IoT Application with Heterogeneous Wireless Sensor Networks. *Energies*. 2022. vol. 15(13). DOI: 10.3390/en15134777.
7. Fotohi R., Bari S., Yusefi M. Securing Wireless Sensor Networks against Denial-of-Sleep Attacks Using RSA Cryptography Algorithm and Interlock Protocol. *International Journal of Communication Systems*. 2019. vol. 33(4). DOI: 10.1002/dac.4234.
8. Yang S.-K., Shiue Y.-M., Su Z.-Y., Liu I.-H., Liu C.-G. An Authentication Information Exchange Scheme in WSN for IoT Applications. *IEEE Access*. 2020. vol. 8. pp. 9728–9738. DOI: 10.1109/ACCESS.2020.2964815.
9. Ullah A., Said G., Sher M., Ning H. Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN. *Peer-to-Peer Networking and Applications*. 2020. vol. 13. pp. 163–174. DOI: 10.1007/s12083-019-00745-z.
10. Singh D., Kumar B., Singh S., Chand S. A Secure IoT-Based Mutual Authentication for Healthcare Applications in Wireless Sensor Networks Using ECC. *International Journal of Healthcare Information Systems and Informatics*. 2021. vol. 16. no. 2. pp. 21–48. DOI: 10.4018/IJHISI.20210401.0a2.
11. Ahmad A., Ullah A., Feng C., Khan M., Ashraf S., Adnan M., Nazir S., Ullah Khan H. Towards an Improved Energy Efficient and End-to-End Secure Protocol for IoT Healthcare Applications. *Security and Communication Networks*. 2020. vol. 2020. DOI: 10.1155/2020/8867792.
12. Nada A., Bayoumi M. Development of a constraint stabilization method of multibody systems based on fuzzy logic control. *Multibody System Dynamics*. 2024. vol. 61. pp. 233–265. DOI: 10.1007/s11044-023-09921-9.
13. Liu L., Xue D., Zhang S. General type industrial temperature system control based on fuzzy fractional-order PID controller. *Complex and Intelligent Systems*. 2023. vol. 9. pp. 2585–2597. DOI: 10.1007/s40747-021-00431-9.
14. Sivapriya N., Ravi T. Efficient Fuzzy based Multi-constraint Multicast Routing with Multi-criteria Enhanced Optimal Capacity–Delay Tradeoff. *International Journal of Scientific & Technology Research*. 2019. vol. 8(8). pp. 1468–1473.
15. Jasim A., Kashmar A. An Evaluation of RSA and a Modified SHA-3 for a New Design of Blockchain Technology. *Artificial Intelligence for Smart Healthcare. EAI/Springer Innovations in Communication and Computing*. Cham: Springer, 2023. pp. 477–489. DOI: 10.1007/978-3-031-23602-0_28.
16. Abid R., Iwendi C., Javed A., Rizwan M., Jalil Z., Anajemba J., Biamba C. An optimised homomorphic CRT-RSA algorithm for secure and efficient communication. *Personal and Ubiquitous Computing*. 2023. vol. 27. pp. 1405–1418. DOI: 10.1007/s00779-021-01607-3.
17. Anushiya R., Lavanya V. A new deep-learning with swarm based feature selection for intelligent intrusion detection for the Internet of things. *Measurement: Sensors*. 2023. vol. 26. DOI: 10.1016/j.measen.2023.100700.
18. Roldan-Gomez J., Boubeta-Puig J., Carrillo-Mondejar J., Manuel Castelo Gomez J., del Rincon J. An automatic complex event processing rules generation system for the recognition of real-time IoT attack patterns. *Engineering Applications of Artificial Intelligence*. 2023. vol. 123. DOI: 10.1016/j.engappai.2023.106344.
19. Size considerations for public and private keys. Available at: <https://www.ibm.com/docs/en/zos/2.3.0?topic=certificates-size-considerations-public-private-keys> (accessed 26.01.2024).
20. Visual Studio 2022. Available at: <https://visualstudio.microsoft.com/vs/> (accessed 10.02.2024).
21. Features of C++ 20. Available at: <https://www.geeksforgoeks.org/features-of-c-20/> (accessed 04.02.2024).

22. OPNET Network Simulator. Available at: <https://opnetprojects.com/opnet-network-simulator/> (accessed 16.14.2024).

Дженифер Р. Рита — научный сотрудник, научно-исследовательский отдел компьютерных наук, Женский колледж Кавери, Университет Бхаратидасан. Область научных интересов: компьютерные науки. Число научных публикаций — 1. rita.jenifer@gmail.com; Главная улица Кавери-Нагар, Аннамалай-Нагар, 620018, Тируччираппалли, Тамил Наду, Индия; р.т.: +91(431)275-1232.

Пракаш В. Синту Джанита — профессор, научно-исследовательский отдел компьютерных наук, Женский колледж Кавери, Университет Бхаратидасан. Область научных интересов: компьютерные науки. Число научных публикаций — 54. sinthujanita@gmail.com; Главная улица Кавери-Нагар, Аннамалай-Нагар, 620018, Тируччираппалли, Тамил Наду, Индия; р.т.: +91(431)275-1232.

Поддержка исследований. Исследование было поддержано грантом, полученным в рамках программы начального финансирования исследовательских проектов от Женского колледжа Кавери, Тируччираппалли.

А.А. ВОЕВОДА, Ю.П. ФИЛОШОВ, В.Ю. ФИЛОШОВ
**РАЗРАБОТКА ЛИНЕЙНОЙ СИСТЕМЫ УПРАВЛЕНИЯ ТЯГОЙ
ВИНТОМОТОРНОЙ ГРУППЫ ДЛЯ БПЛА**

Воевода А.А., Филошов Ю.П., Филошов В.Ю. Разработка линейной системы управления тягой винтомоторной группы для БПЛА.

Аннотация. Управление ориентацией и позиционированием беспилотного летательного аппарата (БПЛА) вертикального взлета и посадки мультироторного типа в пространстве неразрывно связано с формированием вектора управления движением, состоящего из комбинации тяг и аэродинамических моментов создаваемых каждой винтомоторной группой. Точность и скорость формирования вектора управления движением в значительной степени влияет на ошибки позиционирования и ориентации БПЛА. В большинстве работ, посвященных синтезу систем управления БПЛА, используется вектор управления движением без учета динамики винтомоторных групп, что в некоторых случаях вынуждает снижать быстродействие системы управления. Повысить быстродействие можно за счет повышения быстродействия формирования тяги винтомоторных групп, для чего предложена линейная система управления тягой винтомоторной группы. Винтомоторная группа в своем составе имеет нелинейную внутреннюю связь по аэродинамическому моменту и выходной сигнал – тягу, нелинейно зависящую от квадрата скорости вращения винта. Обычно, винтомоторной группой управляют как электродвигателем – внутреннюю связь по аэродинамическому моменту рассматривают как внешнее возмущение, а тягой управляют посредством изменения скорости вращения винта, которая вычисляется на основании требуемого вектора управления движением. Предлагается рассматривать тягу и аэродинамический момент как составную часть винтомоторной группы, для которой построить линейную систему управления тягой. Для этого выполнена линеаризация обратной связи системы винтомоторной группы, связывающей подаваемое на двигателях напряжение с вектором управления движением, являющимся выходной величиной. Процесс линеаризации разбит на два этапа: на первом этапе выполнена линеаризация обратной связи по состоянию для электродвигателя с внутренней нелинейной связью по аэродинамическому моменту; на втором этапе выполнена линеаризация обратной связи по выходу, полученной на первом этапе системы с нелинейным выходным сигналом – тягой. В соответствии с принципами подчиненного регулирования для линеаризованной обратной связи винтомоторной группы сформировано управление двигателем. Выполнено моделирование. Важным вопросом при применении линеаризации обратной связи является сохранение качественных характеристик системы управления при несоответствии параметров объекта и модели, параметры которой используются для вычисления линеаризующей обратной связи. В работе проведено моделирование при несоответствии некоторых параметров до 50%.

Ключевые слова: линеаризация обратной связи, квадрокоптер, винтомоторная группа, подчиненное регулирование, двигатель переменного тока, синтез, система управления.

1. Введение. Существует несколько направлений построения беспилотных летательных аппаратов, одно из них – это создание беспилотных аппаратов с вертикальным взлетом и посадкой к которым относятся мультикоптеры (*multicopter*, *multirotor*) – летательные

аппараты с более чем двумя несущими винтами, к ним относятся бикоптеры, трикоптеры, квадрокоптеры и так далее. Построение системы управления такими летательными аппаратами является сложной задачей, что привлекает внимание многих исследователей, испытывающих свои подходы к её решению. Область применения таких летательных аппаратов растет с каждым годом, предлагая все новые направления для исследований.

Мультикоптеры бывают с фиксированными и с поворотными роторами. Одной из проблем мультикоптеров с фиксированными роторами является недостаток управляющих воздействий. Такие объекты называются объектами с недостаточным количеством степеней свободы (*underactuated systems*). К ним можно отнести объекты типа перевернутого маятника на тележке [1]. Для решения этой проблемы в некоторых работах, на примере квадрокоптера, предлагается использовать поворотные роторы, выполняющие линейное перемещение центра масс квадрокоптера без его наклона [2, 3]. Помимо поворотных роторов бывают винты изменяемого шага (*variable pitch propellers*) [4], позволяющие не только повышать эффективность полета, но и изменять направление тяги для повышения маневренности.

Сравнение беспилотных летательных аппаратов осуществляется по качеству управления. Высокого качества управления можно достигнуть построением системы управления учитывающей неточности модели и устойчивую к внешним воздействиям. Для летательных аппаратов внешним воздействием является ветер. В работе [5] рассматривается система управления с компенсацией ветра, при условии, что изменение ветра происходит медленнее, чем переходные процессы в системе. В работе [6] возмущение ветром рассматривается как гармонический сигнал, амплитуда и частота которого вычисляется из уравнений собственной угловой скорости квадрокоптера, сводя задачу компенсации возмущений к задаче адаптивного управления. В работе [7] приводится широкий обзор применения наблюдателей возмущений, в том числе для управления беспилотными летательными аппаратами.

В работе [8] рассмотрена система управления при кватернионном представлении вращений квадрокоптера. Другим важным направлением является управление группой беспилотных летательных аппаратов, пример такой системы рассмотрен в работе [9].

Модель мультикоптера можно представить в виде трех зависящих друг от друга подсистем. Подсистема позиционирования

описывает линейное перемещение. Подсистема ориентации описывает повороты – крен, тангаж и рысканье. Подсистема *электромеханического преобразования* (ЭМП), состоящая из нескольких винтомоторных групп (ВМГ), выступает преобразователем подаваемого напряжения на двигатели в скорости вращения винтов, из которых далее создается тяга и аэродинамический момент, используемые для управления ориентацией и позиционированием. В большинстве работ динамику винтомоторной группы не учитывают [10 – 12], предполагая, что она значительно выше динамики подсистем ориентации и перемещения. Такое допущение не всегда верно, что вынуждает разрабатывать системы управления ориентацией и позиционированием менее быстродействующими для работоспособности используемых алгоритмов. Поэтому для построения систем управления позиционированием и ориентацией в первую очередь необходимо построить быстродействующую систему управления ЭМП.

В качестве двигателя винтомоторной группы наибольшее распространение получили синхронные двигатели с постоянными магнитами управляемые посредством контроллеров скорости (*electric speed controller, ESC*). В данной работе разрабатывается алгоритм работы контроллера скорости, основывающийся на векторном управлении с подчиненным регулированием [13, 14]. В работе [15] разрабатывается система управления для явнополюсного (*salient pole, interior mounted magnets*) синхронного двигателя с постоянными магнитами при управлении током по поперечной оси ($i_d=0$) с использованием принципа плоского регулирования (*flatness based control*).

В работе [16] в качестве объекта взята винтомоторная группа совместно с регулятором скорости. Для этого объекта предложена следующая передаточная функция со звеном запаздывания $G(s) = 9.412e^{-0.295s} / (s + 3.507)$, для которой строится регулятор дробного порядка (*fractional order controller*). В составе винтомоторной группы аэродинамический момент противоположен моменту электродвигателя и нелинейно зависит от скорости вращения винта, при этом тяга двигателя также нелинейно зависит от скорости вращения винта. В работе [17] используется модель неявнополюсного (*surface mounted magnets, non-salient pole*) синхронного двигателя с постоянными магнитами с векторным управлением двигателем током по поперечной оси, для чего ток по продольной оси делают равным нулю ($i_d=0$), что упрощает модель системы. Регулятор скорости формируется по принципу плоского регулирования с учетом

аэродинамического момента, без датчиков тока, но с известным углом поворота ротора ВМГ.

В отличие от других работ по управлению винтомоторными группами, в данной работе предлагается сформировать линейную систему управления тягой, а не скоростью, посредством применения линеаризации обратной связью [18 – 22]. Она заключается в формировании такого управляющего сигнала посредством нелинейных обратных связей, что замкнутая система получает линейный вид. Для полученной линейной системы применяются хорошо разработанные линейные методы синтеза.

2. Постановка задачи. Цель работы – построить линейную систему управления тягой винтомоторной группы, что позволит повысить быстродействие подсистемы ЭМП. Для этого предложено линеаризовать обратную связью по состоянию и по выходу [23] модель винтомоторной группы, построить управление и проверить, что система остается устойчивой при изменении параметров объекта.

Винтомоторная группа относится к классу нелинейных аффинных систем – нелинейных систем, линейных по управлению, которые можно представить следующими уравнениями:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})u \\ y = h(\mathbf{x}) \end{cases}, \quad (1)$$

где \mathbf{x} – n -мерный вектор состояния $\mathbf{x} \in R^n$, управляющее воздействие u и выходной сигнал y – скалярные величины $y, u \in R$, $\mathbf{f}(\mathbf{x})$, $\mathbf{g}(\mathbf{x})$ – гладкие векторные функции в некоторой области $\Omega \subset R^n$.

Линеаризация обратной связью по состоянию системы (1) выполняется на первом этапе, для этого запишем только первое уравнение:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})u. \quad (2)$$

В некоторых случаях, для уравнения (2) существует диффеоморфизм*.

* Гладкая векторная функция $\mathbf{z} = T(\mathbf{x}) = (T_1(\mathbf{x}) \ T_2(\mathbf{x}) \ \dots \ T_n(\mathbf{x}))^T$, где $\mathbf{z}, \mathbf{x} \in R^n$, определенная в области Ω , называется диффеоморфизмом в области Ω , если существует однозначная векторная функция $\mathbf{x} = T^{-1}(\mathbf{z})$ и эта функция гладкая.

$$\mathbf{z} = T(\mathbf{x}) = (T_1(\mathbf{x}) \ T_2(\mathbf{x}) \ \dots \ T_n(\mathbf{x}))^T, \quad (3)$$

где индекс T обозначает транспонирование, $z_1 = T_1(\mathbf{x})$, $z_2 = T_2(\mathbf{x}) = L_f T_1(\mathbf{x})$, $z_3 = L_f T_2(\mathbf{x}) = L_f^2 T_1(\mathbf{x})$, $z_n = L_f^{n-1} T_1(\mathbf{x})$, который преобразует систему (2) к виду:

$$\begin{cases} \dot{z}_1 = z_2 \\ \dot{z}_2 = z_3 \\ \vdots \\ \dot{z}_n = L_f^n T_1 + L_g L_f^{n-1} T_1 u \end{cases},$$

где $L_f^n T_1$ и $L_g L_f^{n-1} T_1$ некоторые нелинейные гладкие скалярные функции, определяемые производными Ли, такие, что $(L_g L_f^{n-1} T_1)^{-1}$ существует в некоторой области. Тогда следующее преобразование обратной связью:

$$u = \frac{1}{L_g L_f^{n-1} T_1} (v - L_f^n T_1), \quad (4)$$

приводит систему (2) к линейному виду:

$$\begin{cases} \dot{z}_1 = z_2 \\ \dot{z}_2 = z_3 \\ \vdots \\ \dot{z}_{n-1} = z_n \\ \dot{z}_n = v \end{cases}, \quad (5)$$

v – входной сигнал системы (5).

В результате применения оператора ∇ к гладкой скалярной функции $\alpha(\mathbf{x})$ получена вектор-строка $\nabla\alpha$:

$$\nabla = \left(\frac{\partial}{\partial x_1} \ \frac{\partial}{\partial x_2} \ \dots \ \frac{\partial}{\partial x_{n-1}} \ \frac{\partial}{\partial x_n} \right), \quad \nabla\alpha = \left(\frac{\partial\alpha}{\partial x_1} \ \frac{\partial\alpha}{\partial x_2} \ \dots \ \frac{\partial\alpha}{\partial x_{n-1}} \ \frac{\partial\alpha}{\partial x_n} \right),$$

а к гладкой векторной функции $\boldsymbol{\beta}(\mathbf{x}) = (\beta_1(\mathbf{x}) \dots \beta_m(\mathbf{x}))^T$ – матрица:

$$\nabla \boldsymbol{\beta} = \begin{pmatrix} \frac{\partial \beta_1}{\partial x_1} & \frac{\partial \beta_1}{\partial x_2} & \dots & \frac{\partial \beta_1}{\partial x_{n-1}} & \frac{\partial \beta_1}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial \beta_m}{\partial x_1} & \frac{\partial \beta_m}{\partial x_2} & \dots & \frac{\partial \beta_m}{\partial x_{n-1}} & \frac{\partial \beta_m}{\partial x_n} \end{pmatrix}.$$

Производная Ли L_f скалярной функции $\alpha(\mathbf{x})$ по векторной функции $\mathbf{f}(\mathbf{x})$ описывается выражением:

$$L_f \alpha = \nabla \alpha \mathbf{f}.$$

Скобка Ли ad_f векторной функции $\mathbf{g}(\mathbf{x})$ по векторной функции $\mathbf{f}(\mathbf{x})$ описывается выражением:

$$ad_f \mathbf{g} = \nabla \mathbf{g} \mathbf{f} - \nabla \mathbf{f} \mathbf{g} = L_f \mathbf{g} - L_g \mathbf{f}.$$

Для нахождения диффеоморфизма (3) и линеаризующего управления (4) воспользуемся алгоритмом, приведенным в [23]:

Шаг 1. Составить матрицу управляемости [19]:

$$\mathbf{Y} = \left(\mathbf{g} \quad ad_f \mathbf{g} \quad \dots \quad ad_f^{n-1} \mathbf{g} \right), \quad (6)$$

если матрица управляемости невырожденная – $\det \mathbf{Y} \neq 0$ и множество $\left(\mathbf{g} \quad ad_f \mathbf{g} \quad \dots \quad ad_f^{n-2} \mathbf{g} \right)$ инволютивно [23].

Шаг 2. Определить функцию $T_1(\mathbf{x})$ из следующих соотношений:

$$\nabla T_1 ad_f^i \mathbf{g} = 0, \quad \nabla T_1 ad_f^{n-1} \mathbf{g} \neq 0, \quad i = 0, 1, \dots, n-2.$$

Шаг 3. Определить преобразование состояния в соответствии с (3) и сформировать управляющее воздействие в соответствии с (4).

Результатом выполнения первого этапа линеаризации обратной связью является система (5). Эта система является линейной относительно \mathbf{z} , но нелинейной относительно выхода $y = h(\mathbf{x})$

объекта (1). На втором этапе добавим к системе (5), полученной в результате первого этапа линеаризации, выходной сигнал объекта (1) $h(\mathbf{x})$, который в новых переменных равен $\bar{h}(\mathbf{z}) = h(T^{-1}(\mathbf{z}))$ и получим:

$$\begin{cases} \dot{\mathbf{z}} = \bar{\mathbf{f}}(\mathbf{z}) + \bar{\mathbf{g}}(\mathbf{z})v, \\ y = \bar{h}(\mathbf{z}) \end{cases}, \quad (7)$$

где $\bar{h}(\mathbf{z})$ дифференцируема необходимое количество раз, $\bar{\mathbf{f}}(\mathbf{z}) = (z_2 \ z_3 \ \dots \ z_n \ 0)^T$, $\bar{\mathbf{g}}(\mathbf{z}) = (0 \ 0 \ \dots \ 0 \ 1)^T$.

Линеаризация обратной связью по выходу для системы (7) выполняется на втором этапе, который заключается в дифференцировании выходного сигнала $y = \bar{h}(\mathbf{z})$ до получения прямой зависимости от входного сигнала v :

$$\dot{y} = \frac{d\bar{h}(\mathbf{z})}{dz} \dot{\mathbf{z}} = \frac{d\bar{h}(\mathbf{z})}{dz} \bar{\mathbf{f}}(\mathbf{z}) + \frac{d\bar{h}(\mathbf{z})}{dz} \bar{\mathbf{g}}(\mathbf{z})v = L_{\bar{\mathbf{f}}}\bar{\mathbf{h}} + L_{\bar{\mathbf{g}}}\bar{\mathbf{h}}v,$$

если $L_{\bar{\mathbf{g}}}\bar{\mathbf{h}} = 0$, то \dot{y} не зависит от v и необходимо продолжить дифференцирование:

$$\ddot{y} = \frac{d(L_{\bar{\mathbf{f}}}\bar{\mathbf{h}})}{dz} \bar{\mathbf{f}}(\mathbf{z}) + \frac{d(L_{\bar{\mathbf{f}}}\bar{\mathbf{h}})}{dz} \bar{\mathbf{g}}(\mathbf{z})v = L_{\bar{\mathbf{f}}}^2\bar{\mathbf{h}} + L_{\bar{\mathbf{g}}}L_{\bar{\mathbf{f}}}\bar{\mathbf{h}}v,$$

если $L_{\bar{\mathbf{g}}}L_{\bar{\mathbf{f}}}\bar{\mathbf{h}} = 0$, то необходимо продолжить дифференцирование до тех пор, пока не появится явная зависимость y от v , таким образом получим:

$$y^{(r)} = L_{\bar{\mathbf{f}}}^r\bar{\mathbf{h}} + L_{\bar{\mathbf{g}}}L_{\bar{\mathbf{f}}}^{r-1}\bar{\mathbf{h}}v, \quad (8)$$

где r – относительная степень объекта и $r \leq n$, $L_{\bar{\mathbf{g}}}L_{\bar{\mathbf{f}}}^{r-1}\bar{\mathbf{h}} \neq 0$. При формировании входного сигнала v системы (7) следующим образом:

$$v = (L_{\bar{\mathbf{g}}}L_{\bar{\mathbf{f}}}^{r-1}\bar{\mathbf{h}})^{-1} (\bar{v} - L_{\bar{\mathbf{f}}}^r\bar{\mathbf{h}}), \quad (9)$$

будет получена система вида последовательного соединения нескольких интеграторов:

$$y^{(n)} = \bar{v}, \quad (10)$$

где \bar{v} – управляющий сигнал линеаризованной системы (10), а производные выходного сигнала, необходимые для обратной связи регуляторов, вычисляются следующим образом:

$$y^{(i)} = G_{i+1}(\mathbf{z}) = L_{\mathbf{f}}^i \bar{\mathbf{h}}, \quad i = 0, \dots, n-1.$$

В результате двухэтапной линеаризации получена система (10), для которой применимы различные линейные методы синтеза.

3. Описание модели. Модель мультикоптера состоит из трех подсистем. Подсистему ориентации и подсистему позиционирования объединим в одну подсистему – *подсистему движения*. Подсистема ЭМП, состоящая из нескольких ВМГ, формирует вектор скоростей вращения винтов, из которых создается управляющий сигнал для подсистемы движения, состоящий из комбинации аэродинамических моментов и тяг входящий в состав ВМГ. В данной работе рассмотрен квадрокоптер, ЭМП которого состоит из четырех одинаковых ВМГ, перекрестные связи между которыми не учитываются из-за достаточного удаления друг от друга. Так как ВМГ одинаковы и не имеют перекрестных связей между друг другом, то системы управления для них также одинаковы. В данном разделе рассмотрим, как формируется управляющий сигнал для подсистемы движения и составим модель отдельного ВМГ.

3.1. Составление векторного управляющего сигнала для подсистемы движения квадрокоптера. Винтомоторная группа представляет собой соединение электродвигателя и винта. При работе электродвигателей винтомоторной группы создаётся два аэродинамических эффекта – сила тяги и аэродинамический момент, обеспечивающих перемещение квадрокоптера. Полет квадрокоптера осуществляется на высотах с одинаковой плотностью воздушной среды, поэтому силу тяги F_i и аэродинамический момент M_{ai} можно представить следующим образом [24, 25]:

$$F_i = k_{F_i} \omega_i^2, \quad M_{ai} = k_{M_{ai}} \omega_i^2, \quad (11)$$

где k_{F_i} и $k_{M_{ai}}$ константы i -го винта, ω_i – скорость i -ой ВМГ, i – порядковый номер ВМГ. Действие сил тяги и моментов ВМГ квадрокоптера, показаны на рисунке 1.

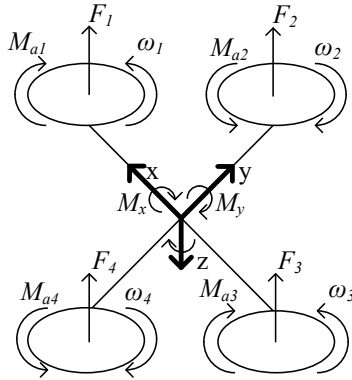


Рис. 1. Схема расположения ВМГ квадрокоптера

Для поворота квадрокоптера вокруг осей x , y , z , как показано на рисунке 1, необходимо создать соответствующие моменты M_x , M_y , M_z , которые получаются из комбинации тяг и аэродинамических моментов следующим образом:

$$M_x = b(F_4 - F_2), \quad M_y = b(F_1 - F_3), \quad M_z = M_{a1} + M_{a3} - M_{a2} - M_{a4},$$

где b – плечо, в рамках данной работы плечо одинаково для всех двигателей. Суммарная тяга двигателей $F_\Sigma = F_1 + F_2 + F_3 + F_4$ обеспечивает изменение высоты квадрокоптера и его движение вдоль осей x , y , z . При равенстве для каждого двигателя констант силы тяги $k_{F_i} = k_F$ и аэродинамического момента $k_{M_i} = k_M$, моменты M_x , M_y , M_z и тягу двигателей F_Σ , можно выразить через скорости вращения двигателей следующим образом:

$$\mathbf{u}_d = P\mathbf{y} = \begin{pmatrix} k_F & k_F & k_F & k_F \\ 0 & -k_F & 0 & k_F \\ k_F & 0 & -k_F & 0 \\ k_M & -k_M & k_M & -k_M \end{pmatrix} \mathbf{y}, \quad (12)$$

где $\mathbf{y} = (\omega_1^2 \ \omega_2^2 \ \omega_3^2 \ \omega_4^2)^T$ – вектор выходных величин подсистемы ЭМП, $\mathbf{u}_d = (F_\Sigma \ M_x \ M_y \ M_z)^T$ – вектор управления подсистемой движения. Электродвигатели ВМГ вращаются за счёт формирования вектора напряжений \mathbf{u}_{Si} , подводимого к их обмоткам. Квадрокоптер можно разделить на две, связанные уравнением (12) подсистемы – подсистема ЭМП и движения. Выходной величиной подсистемы движения является вектор \mathbf{q} , состоящий из координат квадрокоптера и его углов ориентации. Структурные связи подсистем квадрокоптера представлены на рисунке 2.

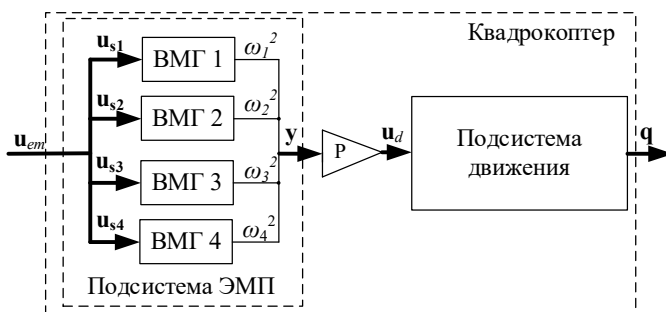


Рис. 2. Структурная схема квадрокоптера

Управляющим сигналом для подсистемы ЭМП является вектор управляющих напряжений \mathbf{u}_{em} , а управляющим сигналом для подсистемы движения – вектор управления движением \mathbf{u}_d . Как видно из рисунка 2, вектор управления движением \mathbf{u}_d формируется за счет изменения вектора напряжений питания \mathbf{u}_{em} , посредством подсистемы ЭМП, то есть управляющий сигнал подсистемы движения \mathbf{u}_d не может быть сформирован самостоятельно, а только за счет подачи соответствующего управляющего напряжения \mathbf{u}_{em} на подсистему ЭМП.

3.2. Описание винтомоторной группы. Исходя из того, что все винтомоторные группы одинаковы, для лаконичной записи уберем индексы i из описания ВМГ. В состав винтомоторной группы входит синхронный двигатель с постоянными магнитами (*permanent magnet synchronous machine*), математическое описание которого представлено уравнениями [26, 27]:

$$\begin{cases} u_{Sd} = L_{Sd} \frac{di_{Sd}}{dt} + r_S i_{Sd} - \omega n_p L_{Sq} i_{Sq}, \\ u_{Sq} = L_{Sq} \frac{di_{Sq}}{dt} + r_S i_{Sq} + \omega n_p L_{Sd} i_{Sd} + \omega n_p \psi_m, \\ M = 1.5 n_p (i_{Sq} (i_{Sd} L_{Sd} + \psi_m) - i_{Sd} i_{Sq} L_{Sq}), \\ M - M_c = J_r \frac{d\omega}{dt}, \end{cases} \quad (13)$$

где u_{Sd} , u_{Sq} – статорные напряжения (stator's voltage), $\mathbf{u}_S = (u_{Sd} \ u_{Sq})^T$; i_{Sd} , i_{Sq} – статорные токи (stator's current); L_{Sd} , L_{Sq} – индуктивности статора (stator's inductance), r_S – статорное сопротивление (stator's resistance), J_r – момент инерции ротора, ω – скорость вращения двигателя, n_p – число пар полюсов; ψ_m – потокосцепление с постоянными магнитами, M – момент, M_c – момент сопротивления, индекс d – проекция на продольную ось, индекс q – проекция на поперечную ось.

Напряжение \mathbf{u}_S формируется таким образом, чтобы выполнялась развязка перекрестных связей и ток статора по продольной оси был равен нулю $i_{Sd} = 0$, обеспечивая ориентацию тока по поперечной оси $i_S = i_{Sq}$.

Рассматривается синхронный двигатель с неявнополюсным ротором $L = L_{Sd} = L_{Sq}$, момент двигателя зависит только от тока по поперечной оси $M = 1.5 n_p \psi_m i_S$, или по аналогии с двигателем постоянного тока, запишем $M = \phi i_S$.

При $i_{Sd} = 0$ и $L = L_{Sd} = L_{Sq}$ уравнение (13) можно записать по аналогии с двигателем постоянного тока следующим образом:

$$\frac{i_S}{u} = \frac{r_S^{-1}}{T_s p + 1}, \quad u = u_{Sq} - u_b, \quad u_b = \psi_m n_p \omega, \quad M - M_c = J_r \frac{d\omega}{dt}, \quad (14)$$

где $T_s = L / r_S$ – постоянная времени характеризующая темп изменения тока статора, u_b – обратное ЭДС двигателя.

В соответствии с уравнением (12), выходной величиной подсистемы ЭМП является вектор скорости, состоящий из скоростей вращения каждого двигателя. Допустим, что двигатели в подсистеме ЭМП расположены на достаточном удалении друг от друга и не влияют друг на друга, поэтому их можно рассматривать по отдельности. Аэродинамический момент M_a (11) создается в результате вращения винта и направлен в противоположную сторону от момента двигателя, являясь для него нагрузкой $M_c = M_a$, как показано на рисунке 1. Подставив в систему уравнений (14) выражение $M_c = k_M \omega^2$, получено математическое описание ВМГ в пространстве состояний:

$$\begin{cases} \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})u_{Sq} \\ y = h(\mathbf{x}) \end{cases}, \quad (15)$$

где $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} i_s \\ \omega \end{pmatrix}$, $\mathbf{f}(\mathbf{x}) = \begin{pmatrix} -\frac{1}{T_S}x_1 - \frac{n_p\Psi_m r_s^{-1}}{T_S}x_2 \\ \frac{\varphi}{J_r}x_1 - \frac{k_M}{J_r}x_2^2 \end{pmatrix}$, $\mathbf{g}(\mathbf{x}) = \begin{pmatrix} \frac{r_s^{-1}}{T_S} \\ 0 \end{pmatrix}$,
 $h(\mathbf{x}) = x_2^2$.

Вектор состояния \mathbf{x} полностью доступен. Представим уравнение (15) в виде структурной схемы на рисунке 3.

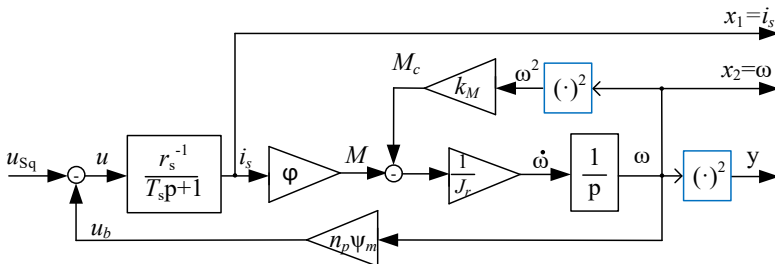


Рис. 3. Структурная схема ВМГ

Как видно из рисунка 3, аэродинамическое сопротивление винта вносит к уравнениям электродвигателя нелинейную внутреннюю

обратную связь по скорости и выход ВМГ y зависит от скорости также нелинейно.

4. Линеаризация обратной связью модели ВМГ.

Первый этап – линеаризация обратной связью по состоянию. Воспользуемся алгоритмом в разделе 2.

Шаг 1. Порядок системы (15) второй $n = 2$, тогда матрица управляемости (6), следующая:

$$\mathbf{Y} = (\mathbf{g} \text{ ad}_f \mathbf{g}),$$

$$\text{ad}_f \mathbf{g} = \nabla \mathbf{g} \mathbf{f} - \nabla \mathbf{f} \mathbf{g} = \begin{pmatrix} \frac{r_s^{-1}}{T_s^2} \\ -\frac{\varphi r_s^{-1}}{T_s J_r} \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} \frac{r_s^{-1}}{T_s} & \frac{r_s^{-1}}{T_s^2} \\ 0 & -\frac{\varphi r_s^{-1}}{T_s J_r} \end{pmatrix}, \text{rank}(\mathbf{Y}) = 2.$$

Матрица управляемости невырожденная, а множество, составленное из одного вектора \mathbf{g} – инволютивно.

Шаг 2. Определим функцию $z_1 = T_1(\mathbf{x})$. Так как $n = 2$, то:

$$\nabla T_1 \text{ad}_f^i \mathbf{g} = \nabla T_1 \mathbf{g} = \begin{pmatrix} \frac{\partial T_1}{\partial x_1} & \frac{\partial T_1}{\partial x_2} \end{pmatrix} \begin{pmatrix} \frac{r_s^{-1}}{T_s} \\ 0 \end{pmatrix} = \frac{r_s^{-1}}{T_s} \frac{\partial z_1}{\partial x_1} = 0,$$

$$\nabla T_1 \text{ad}_f \mathbf{g} = \begin{pmatrix} \frac{\partial T_1}{\partial x_1} & \frac{\partial T_1}{\partial x_2} \end{pmatrix} \text{ad}_f \mathbf{g} = \frac{r_s^{-1}}{T_s^2} \frac{\partial T_1}{\partial x_1} - \frac{\varphi r_s^{-1}}{T_s J_r} \frac{\partial T_1}{\partial x_2} \neq 0,$$

откуда следует, что $T_1(\mathbf{x})$ зависит только от x_2 , примем $T_1(\mathbf{x}) = x_2$.

Шаг 3. Определим преобразование состояния (3) и управляющее воздействие (4). На шаге 2 найдена функция $T_1(\mathbf{x}) = x_2$, тогда найдем $T_2(\mathbf{x})$:

$$T_2(\mathbf{x}) = L_f T_1 = \nabla T_1 \mathbf{f} = (0 \ 1) \mathbf{f} = \frac{1}{J_r} (\varphi x_1 - k_M x_2^2).$$

Преобразование состояния (3) запишется следующим образом:

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = T(\mathbf{x}) = \begin{pmatrix} x_2 \\ \frac{1}{J_r}(\varphi x_1 - k_M x_2^2) \end{pmatrix}. \quad (16)$$

Для определения линеаризующей обратной связи (4) найдем $L_{\mathbf{g}}z_2$ и $L_{\mathbf{f}}z_2$:

$$L_{\mathbf{g}}z_2 = \nabla z_2 \mathbf{g} = \frac{1}{J_r}(\varphi, -2k_M x_2) \begin{pmatrix} r_S^{-1} \\ T_S \\ 0 \end{pmatrix} = \frac{\varphi r_S^{-1}}{J_r T_S},$$

$$L_{\mathbf{f}}z_2 = \nabla z_2 \mathbf{f} = \frac{1}{J_r}(\varphi, -2k_M x_2) \begin{pmatrix} -\frac{1}{T_S}x_1 - \frac{n_p \Psi_m r_S^{-1}}{T_S}x_2 \\ \frac{\varphi}{J_r}x_1 - \frac{k_M}{J_r}x_2^2 \end{pmatrix},$$

$$L_{\mathbf{f}}z_2 = \frac{2k_M^2 x_2^3 - 2k_M \varphi x_1 x_2}{J_r^2} - \frac{\varphi x_1 + \varphi n_p \Psi_m r_S^{-1} x_2}{T_S J_r}.$$

Управляющий сигнал по обратной связи (4), приводящий систему (15) к виду (5) следующий:

$$u_{Sq} = \frac{T_S J_r}{\varphi r_S^{-1}} \left(\frac{2k_M \varphi x_1 x_2 - 2k_M^2 x_2^3}{J_r^2} + \frac{\varphi x_1 + \varphi n_p \Psi_m r_S^{-1} x_2}{T_S J_r} + v \right). \quad (17)$$

Таким образом выполнен первый этап линеаризации – изменение скорости вращения ВМГ по линейному закону.

Второй этап – линеаризация обратной связью по выходу. На первом этапе была получена линеаризованная система ВМГ вида (5). На втором этапе, примем в качестве выхода системы (5) квадрат скорости $y = \omega^2$, запишем полученную систему, которая имеет вид (7):

$$\begin{cases} \dot{\mathbf{z}} = \bar{\mathbf{f}}(\mathbf{z}) + \bar{\mathbf{g}}(\mathbf{z})v \\ y = \bar{h}(\mathbf{z}) = z_1^2 = \omega^2 \end{cases}, \quad (18)$$

где $\bar{\mathbf{f}}(\mathbf{z}) = (z_2 \ 0)^T$, $\bar{\mathbf{g}}(\mathbf{z}) = (0 \ 1)^T$. Найдем зависимость выхода y от входа v в системе (18). Продифференцируем выход необходимое количество раз:

$$\dot{y} = L_{\bar{\mathbf{f}}}\bar{\mathbf{h}} + L_{\bar{\mathbf{g}}}\bar{\mathbf{h}}v = (2z_1 \ 0) \begin{pmatrix} z_2 \\ 0 \end{pmatrix} + (2z_1 \ 0) \begin{pmatrix} 0 \\ 1 \end{pmatrix} v = 2z_1 z_2.$$

Как видно $L_{\bar{\mathbf{g}}}\bar{\mathbf{h}} = 0$, продолжаем дифференцировать:

$$\ddot{y} = L_{\bar{\mathbf{f}}}^2\bar{\mathbf{h}} + L_{\bar{\mathbf{g}}}L_{\bar{\mathbf{f}}}\bar{\mathbf{h}}v = (2z_2 \ 2z_1) \begin{pmatrix} z_2 \\ 0 \end{pmatrix} + (2z_2 \ 2z_1) \begin{pmatrix} 0 \\ 1 \end{pmatrix} v = 2(z_2^2 + z_1v),$$

получено выражение вида (8), где $n = r$, $L_{\bar{\mathbf{f}}}^2\bar{\mathbf{h}} = 2z_2^2$, $L_{\bar{\mathbf{g}}}L_{\bar{\mathbf{f}}}\bar{\mathbf{h}} = 2z_1$. По формуле (9) найдем линеаризующее обратной связью по выходу воздействие:

$$v = \frac{1}{2z_1}(\bar{v} - 2z_2^2),$$

преобразующее систему (18) к виду (10):

$$\begin{cases} y = \bar{h}(\mathbf{z}) = z_1^2 \\ \dot{y} = L_{\bar{\mathbf{f}}}\bar{h} = 2z_1 z_2, \\ \ddot{y} = \bar{v} \end{cases} \quad (19)$$

где y и \dot{y} используются для построения регуляторов. Как видно управляющий сигнал имеет разрыв в точке $z_1 = \omega = 0$, который проявляется только при остановленном двигателе.

На рисунке 4 представлена структурная схема линеаризованной системы, где новыми состояниями являются величины y , \dot{y} , которые выражаются из исходного вектора состояния \mathbf{x} в соответствии с (16).

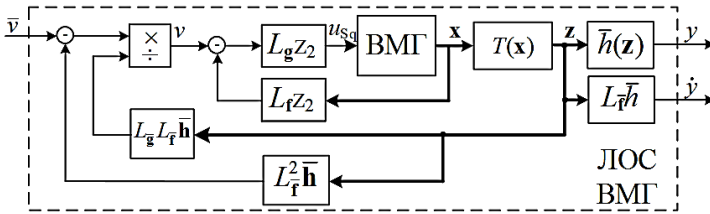


Рис. 4. Структурная схема линейризованной обратной связью винтомоторной группы

На рисунке 4 представлена линейризованная обратной связью винтомоторная группа (ЛОС ВМГ), которая эквивалентна последовательному соединению двух интеграторов, как показано в уравнении (10). Сам выходной сигнал y и его производная \dot{y} вычисляется из системы уравнений (19), где значения вектора z вычисляются из вектора состояния по (16). Управление такой линейризованной системой можно строить разными способами, например добавить П-регуляторы по состоянию в обратную связь, чем обеспечить заданные полюса в замкнутой системе. Также можно добавить полиномиальный матричный регулятор [28] в прямую или обратную связь, что позволит повысить порядок регулятора, тем самым добавит дополнительные качественные свойства замкнутой системы – увеличится порядок астатизма или улучшится точность за счет лучшей компенсации возмущений [29].

Предполагая, что все четыре двигателя одинаковы, получим линейную многоканальную систему преобразования напряжения в тягу двигателей:

$$\begin{pmatrix} F_{\Sigma} \\ M_x \\ M_y \\ M_z \end{pmatrix} = \frac{1}{s^2} \begin{pmatrix} k_F & k_F & k_F & k_F \\ 0 & -k_F & 0 & k_F \\ k_F & 0 & -k_F & 0 \\ k_M & -k_M & k_M & -k_M \end{pmatrix} \begin{pmatrix} \bar{v}_1 \\ \bar{v}_2 \\ \bar{v}_3 \\ \bar{v}_4 \end{pmatrix}. \quad (20)$$

Уравнение (20) представляет собой матричную передаточную функцию, связывающую входной сигнал \bar{v} , полученной в результате линейризации обратной связью системы (19) и вектор управления движением \mathbf{u}_d .

5. Синтез системы управления. Структуры управления, сформированные по принципу подчиненного регулирования, имеют в своем составе несколько последовательно соединенных контуров.

Особенностью систем подчиненного регулирования (СПР) является последовательное расположение регуляторов, образующих замкнутые контуры по своей выходной переменной. Внутренний контур $F_{i-1}(p)$ подчинен внешнему контуру $F_i(p)$. Каждый последующий регулятор вырабатывает задание для предыдущего, как показано на рисунке 5. Рассмотрим трехконтурную СПР [30]:

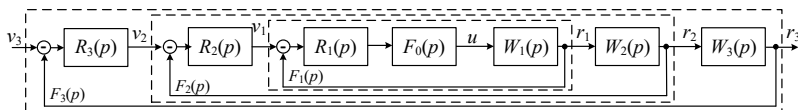


Рис. 5. Структурная схема трехконтурной СПР

На рисунке 5, передаточные функции объекта обозначены как $W_i(p)$, а передаточные функции регулятора как $R_i(p)$, где i – номер звена объекта и регулятора (контура). Корректирующее устройство необходимо для компенсации инерционности объекта и является обратной передаточной функцией звена $R_{Ki}(p) = W_i^{-1}(p)$. Астатизм достигается введением интегратора последовательно со звеном объекта $R_{Ui}(p) = (T_i p)^{-1}$, тогда регулятор звена равен:

$$R_i(p) = R_{Ki}(p)R_{Ui}(p) = W_i^{-1}(p)(T_i p)^{-1}, \quad (21)$$

где T_i удобно выражать в долях от некомпенсируемой постоянной времени T_μ , используя коэффициент a^i , тогда $T_i = a^i T_\mu$. В данной работе примем $a = 2$. Фильтр описывается передаточной функцией $F_0(p) = (T_\mu p + 1)^{-1}$, быстродействие которого характеризуется некомпенсируемой постоянной времени T_μ .

Система подчиненного регулирования для линеаризованной модели ВМГ. Регуляторы для линеаризованной обратной связью системы (19) строятся аналогичным образом. Представим систему (19) в виде многоконтурной системы, как показано на рисунке 5, тогда:

$$r_1 = W_1(p)u, \quad r_2 = W_2(p)r_1, \quad r_3 = W_3(p)r_2, \quad (22)$$

где $W_1(p) = p^{-1}$, $W_2(p) = p^{-1}$, $W_3(p) = 1$, $r_1 = \dot{y}$, $r_2 = r_3 = y$, $u = \bar{v}$. Выходы y , \dot{y} вычисляются из вектора состояния ВМГ x в соответствии с (16) и (19), сигнал \bar{v} является входным воздействием в системе (22). Построенные по принципу подчиненного регулирования (21) регуляторы имеют следующий вид:

$$R_1(p) = \frac{1}{2T_\mu}, \quad R_2(p) = \frac{1}{4T_\mu}, \quad R_3(p) = \frac{1}{8T_\mu p}. \quad (23)$$

На рисунке 6 представлена структурная схема системы подчиненного регулирования для линеаризованной обратной связью винтомоторной группы (ЛОС ВМГ).

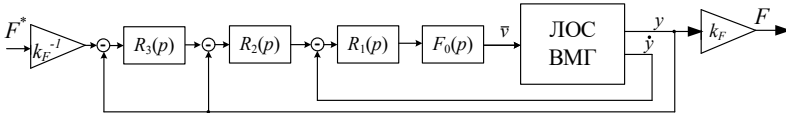


Рис. 6. Структурная схема СПР линеаризованной обратной связью ВМГ

Система управления на рисунке 6 линейна относительно квадрата скорости $y = \omega^2$ и как следствие – линейна относительно тяги. Блок ЛОС ВМГ представлен на рисунке 4, в котором блок ВМГ описывается системой (15). Система управления на рисунке 6 имеет следующую передаточную функцию:

$$F(p) = \frac{1}{64T_\mu^4 p^4 + 64T_\mu^3 p^3 + 32T_\mu^2 p^2 + 8T_\mu p + 1},$$

которая устойчива и соответствует СПР при $a = 2$.

6. Численное моделирование системы. Параметры моделируемого ВМГ представлены в таблице 1.

Таблица 1. Параметры ВМГ

Наименование	r_s [МОм]	T_S [с]	n_p	ψ_m [Вб]
Значение	288	0.001	8	0.0104
Наименование	J_r [кг·м ²]	ω_{\max} [об/мин]	k_M	k_F
Значение	0.005	4000	$8 \cdot 10^{-6}$	$25 \cdot 10^{-6}$

Из таблицы 1 можно найти, что $\varphi = 1.5n_p \psi_m = 0.1246$. Зададим ненулевые начальные условия по скорости, чтобы избежать разрыва функции управления во время линеаризации $x_2(0) = \omega(0) = 0.001$. В данной работе не рассмотрено ограничение по напряжению источника питания [31], предельные значения по тяге определяются максимальной скоростью вращения ω_{\max} , кратковременное превышение которой допустимо.

Подставим значения параметров объекта из таблицы 1 в выражение управляющего сигнала (17), выполняющего первый этап линеаризации обратной связи:

$$u_{Sq} = 1.15 \cdot 10^{-5} v + (-5.91 \cdot 10^{-11} x_2^3 + 9.216 \cdot 10^{-7} x_1 x_2 + 0.288 x_1 + 0.0831 x_2).$$

Второй этап линеаризации заключается в формировании следующего сигнала: $v = 0.5 z_1^{-1} (\bar{v} - 2 z_2^2)$, который подставляется в выражение для u_{Sq} из первого этапа. Подставим значения вектора z , тогда:

$$v = 0.5 x_2^{-1} \bar{v} - 621.2 x_1^2 x_2^{-1} - 0.0399 x_1 x_2 + 1.28 \cdot 10^{-6} x_2^3.$$

Приведенные выше выражения u_{Sq} и v приводят исходную модель на рисунке 3 к виду (19):

$$\begin{cases} y = x_2^2, \\ \dot{y} = 49.85 x_2 x_1 - 0.0032 x_2^3, \\ \dot{y} = \bar{v}. \end{cases} \quad (24)$$

Система уравнений (24) описывает подсистему ЛОС ВМГ на рисунках 4 и 6. Система (24) с регуляторами (23) и соответствующими коэффициентами тяги образует следующую передаточную функцию по тяге:

$$F = \frac{1}{64T_\mu^4 p^4 + 64T_\mu^3 p^3 + 32T_\mu^2 p^2 + 8T_\mu p + 1} F^*. \quad (25)$$

Выполним моделирование системы (24) с регулятором (23) при $T_\mu = 0.02$, которое представлено на рисунке 7.

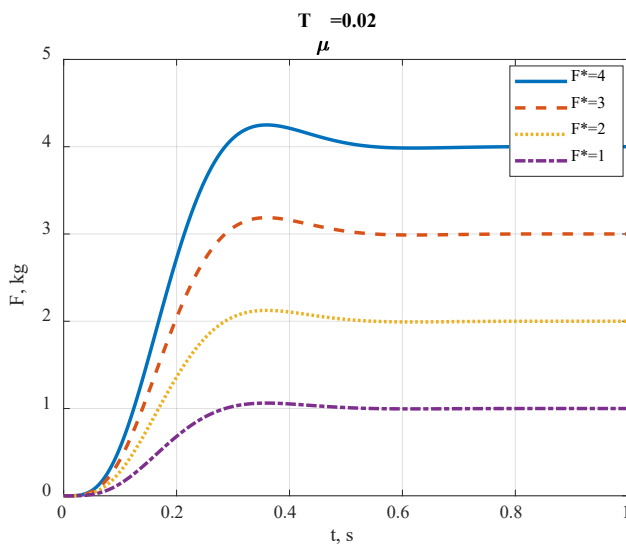


Рис. 7. Графики переходных процессов при различных значениях требуемой тяги F^*

По рисунку 7 видно, что разработанная система управления линейна и соответствует системе подчиненного регулирования (25). Линейное поведение системы управления относительно тяги позволяет реализовывать более сложные алгоритмы управления движением квадрокоптера. Приведенные в таблице 1 параметры ВМГ, в изделии никогда не соответствуют реальным значениям, что приводит к неточной линеаризации модели объекта и отклонении переходных процессов от приведенных на рисунке 7. На рисунке 8 приведены графики переходных процессов при несоответствии некоторых параметров объекта и модели.

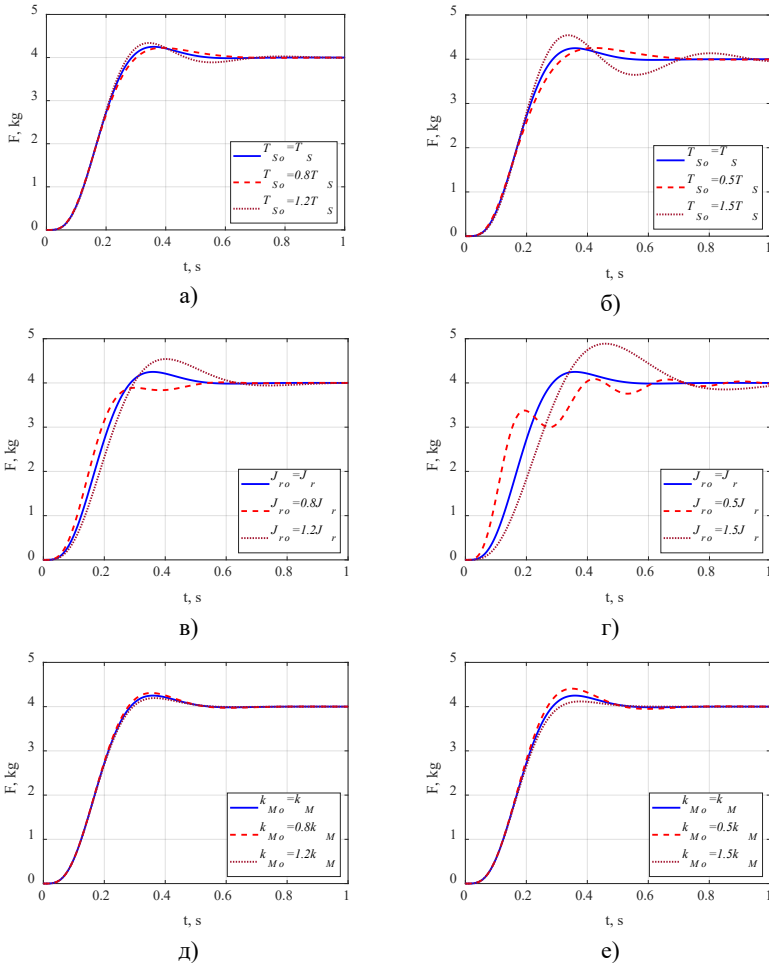


Рис. 8. Переходные процессы при несоответствии параметров объекта и модели: а) электромагнитная постоянная времени отличается на 20%; б) электромагнитная постоянная времени отличается на 50%; в) момент инерции отличается на 20%; г) момент инерции отличается на 50%; д) коэффициент аэродинамического момента отличается на 20%; е) коэффициент аэродинамического момента отличается на 50%

Как видно из рисунка 8, где T_{So} , J_{ro} , k_{Mo} – параметры объекта, а T_S , J_r , k_M – параметры модели, система управления ЛОС ВМГ не теряет устойчивости при несоответствии выбранных параметров на

20% и на 50%. При этом заметно, что отличие k_{Mo} объекта от k_M модели влияет на переходной процесс меньше, чем отличие T_{So} и J_{ro} от T_S и J_r .

Сравним *быстродействие* разработанной системы управления тягой ВМГ с системой подчиненного регулирования скорости ВМГ, где аэродинамический момент является возмущением, а уставка по скорости вычисляется из уставки по тяге. На рисунке 9 приведена эквивалентная схема двукратной системы автоматического регулирования (САР) скорости, взятой из [14] с учетом (11). Она получена для уравнения двигателя (14) при применении принципа подчиненного регулирования (21).

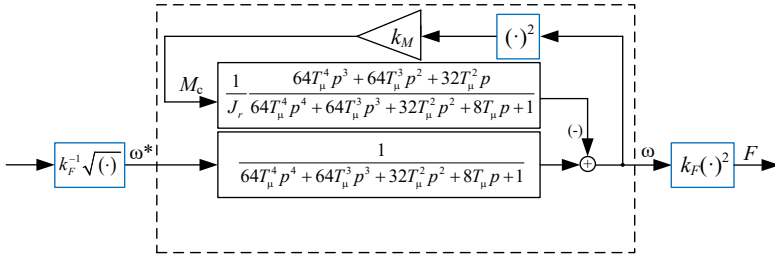


Рис. 9. Эквивалентная структурная схема двукратной системы автоматического регулирования скорости ВМГ

Динамика разработанной линейной системы управления тягой ВМГ описывается передаточной функцией (25), которая совпадает с передаточной функцией на рисунке 9, описывающей систему управления скоростью без возмущения, от входа ω^* к выходу ω_1 . Сравним графики переходных процессов разработанной системы управления тягой (25) с двукратной САР скорости ВМГ на рисунке 8.

На рисунке 10 показаны графики переходных процессов по тяге (а), напряжению (б) и току (в) для разработанной линейной системы управления тягой (*Linearized*) и двукратной САР скорости (*Non-Linearized*), показанной на рисунке 9. Линеаризованная система управления тягой (*Linearized*) описывается передаточной функцией (25) куда подставляется значение некомпенсируемой постоянной времени $T_{\mu L} = 0.05$ (сплошная линия). В показанную на рисунке 9 двукратную САР скорости (*Non-Linearized*), подставляются значения некомпенсируемой постоянной времени $T_{\mu nL} = 0.05$ (штрихованная

линия), $T_{\mu nL} = 0.04$ (штрих-пунктирная линия), $T_{\mu nL} = 0.03$ (пунктирная линия).

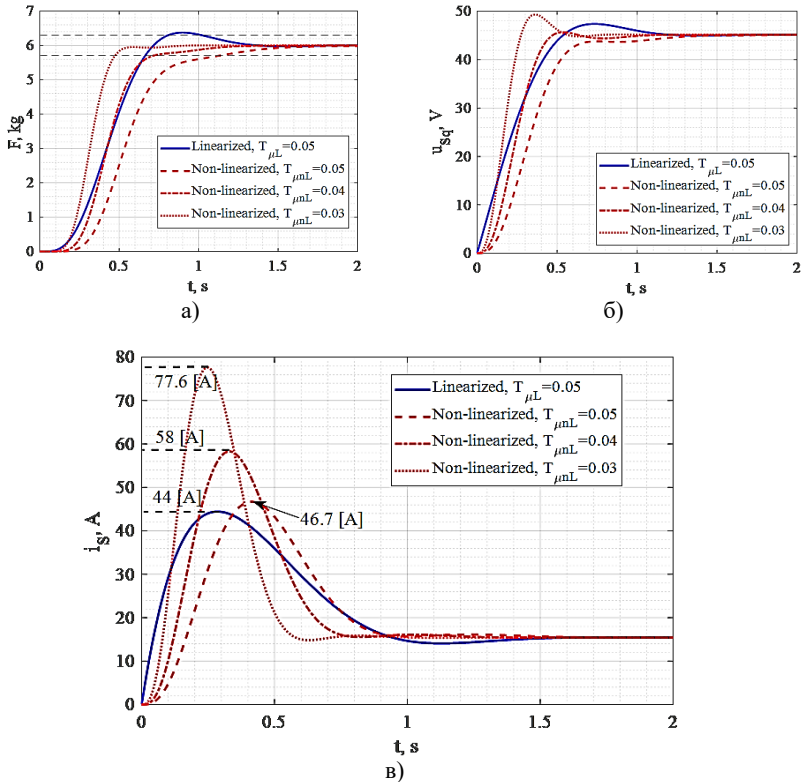


Рис. 10. Графики переходных процессов разработанной линейной системы управления тягой (Linearized) и двукратной САР скорости (Non-linearized) при различных некомпенсируемых постоянных времени:

а) по тяге; б) по напряжению; в) по току

В таблице 2 введены следующие обозначения, t_1 – время достижения зоны 5%, $\max(q)$ – максимальное значение q , Δ – разница между столбцами *Non-linearized* и *Linearized*, например, $\Delta t_1 = t_1^{Non-linearized} - t_1^{Linearized}$, δ – отношение между столбцами *Non-linearized* и *Linearized*, например, $\delta t_1 = \frac{t_1^{Non-linearized} - t_1^{Linearized}}{t_1^{Linearized}} \cdot 100\%$.

Таблица 2. Параметры переходных процессов на рисунке 10

	<i>Linearized</i> $T_{\mu L} = 0.05$	<i>Non-linearized</i> $T_{\mu nL} = 0.05$	<i>Non-linearized</i> $T_{\mu nL} = 0.04$	<i>Non-linearized</i> $T_{\mu nL} = 0.03$
t_1 [с]	0.67	1.13	0.71	0.47
Δt_1 [с]		0.46	0.04	-0.2
δt_1 [%]		68.6	5.9	-30
$\max(u_{sq})$ [В]	47.3	45	45.6	49.3
$\Delta(\max(u_{sq}))$ [В]		-2.3	-1.7	2
$\delta(\max(u_{sq}))$ [%]		-5	-3.6	4.2
$\max(i_s)$ [А]	44	46.7	58	77.6
$\Delta(\max(i_s))$ [А]		2.7	14	33.6
$\delta(\max(i_s))$ [%]		6.1	31.8	76.4

Увеличение быстродействия системы по тяге можно достичь уменьшением некомпенсируемой постоянной времени, но при этом возрастают максимальные значения тока, как показано на рисунке 10(в), что вынудит использовать электрооборудование с большей перегрузочной способностью. Поэтому под повышением быстродействия понимаем уменьшение времени t_1 при схожих максимальных значениях тока и напряжения.

В соответствии с таблицей 2, при одинаковых некомпенсируемых постоянных времени $T_{\mu L} = T_{\mu nL} = 0.05$ время t_1 уменьшено на 68.6% или 0.44 [с], а максимальные значения тока и напряжения близки к друг другу $\Delta(\max(u_{sq})) = -2.3$ [В], $\Delta(\max(i_s)) = 2.7$ [А].

При уменьшении некомпенсируемой постоянной времени двукратной САР скорости $T_{\mu nL} = 0.04$, время достижения зоны 5% становится близким, двукратная САР скорости обрабатывает ступенчатый сигнал задания медленнее на 5.9%, но при этом её максимальное значение тока увеличено на 14 [А] или 31.8 [%].

При $T_{\mu nL} = 0.03$ время достижения зоны 5% двукратной САР скорости быстрее на 0.2 [с] или на 30%, чем у разработанной системы, но при этом максимальное значение тока возрастает на 33.6 [А] или 76.4%.

7. Заключение. Математическая модель винтомоторной группы отличается от математической модели электродвигателя наличием двух нелинейностей как показано на рисунке 3. Обе эти нелинейности зависят от квадрата скорости ω^2 . В отличие от других работ в данной работе выполнена линеаризация обратной связью модели ВМГ (15), которая содержит эти две нелинейности. В результате получена линеаризованная обратной связью система (24), что позволило линейно связать сигнал управления движением \mathbf{u}_d со входным сигналом системы $\bar{\mathbf{v}}$ (20).

Для линеаризованной обратной связью системы (24) построена линейная система управления тягой ВМГ (25) по принципу подчиненного регулирования, которая сравнена с двукратной САР скорости, показанной на рисунке 9. В результате у разработанной системы, относительно двукратной САР скорости увеличено быстродействие (уменьшено время достижения 5% зоны) на 68% при схожих максимальных значениях тока, как показано на рисунке 10 и в таблице 2. При схожем быстродействии линейной системы управления тягой и двукратной САР скорости, максимальное значение тока уменьшилось на 14 [А]. Проведена проверка сохранения устойчивости системы при отличии параметров объекта и модели, как показано на рисунке 8. Таким образом разработанная линейная система управления тягой позволяет повысить быстродействие подсистемы ЭМП, тем самым улучшить маневренность БПЛА.

Литература

1. Ткачев С.Б. Стабилизация неминимально фазовых аффинных систем с использованием линеаризации по части переменных // Наука и образование. изд. МГТУ им. Н.Э. Баумана. 2011. № 10. С. 1–25.
2. Шавин М.Ю. Управляемая динамика квадрокоптера с поворотными роторами // Инженерный журнал: наука и инновации. 2018. № 4(76). С. 1–16.
3. Shavin M., Pritykin D. Tilt-Rotor Quadrotor Control System Design and Mobile Object Tracking // Mekhatronika, Avtomatizatsiya, Upravlenie. 2019. vol. 20. no. 10. pp. 629–639. <https://doi.org/10.17587/mau.20.629-639>.
4. Cutler M., How J.P. Analysis and Control of a Variable-Pitch Quadrotor for Agile Flight. // Journal of Dynamic Systems, Measurement, and Control. 2015. vol. 137(10). DOI: 10.1115/1.4030676.
5. Pyrkin A., Bobtsov A., Kolyubin S., Borisov O., Gromov V., Aranovskiy S. Output Controller for Quadcopters with Wind Disturbance Cancellation // IEEE Conference on Control Applications (CCA). 2014. pp. 166–170. DOI: 10.1109/CCA.2014.6981346.
6. Demircioglu H., Basturk H. Adaptive Attitude and Altitude Control of a Quadrotor Despite Unknown Wind Disturbances // IEEE 56th Annual Conference on Decision and Control. 2017. pp. 274–279. DOI: 10.1109/CDC.2017.8263678.

7. Andrievsky B., Furtat I. Disturbance observers: methods and applications. II. Applications // Automation and Remote Control. 2020. vol. 81. pp. 1775–1818. DOI: 10.1134/S0005117920100021.
8. Kusaka T., Tanaka R. Stateful Rotor for Continuity of Quaternion and Fast Sensor Fusion Algorithm Using 9-Axis Sensors // Sensors. 2022. vol. 22(20). DOI: 10.3390/s22207989.
9. Пшихопов В.Х., Медведев М.Ю. Групповое управление движением мобильных роботов в неопределенной среде с использованием неустойчивых режимов // Труды СПИИРАН. 2018. Т. 60. № 5. С. 39–63. DOI: 10.15622/sp.60.2.
10. Zulu A., John S. A review of control algorithms for autonomous quadrotors // Open Journal of Applied Sciences. 2014. no. 4. pp. 547–556. DOI: 10.4236/ojapps.2014.414053.
11. Gasparyan O., Darbinyan H., Asatryan A., Simonyan T. On the control of quadcopters based on the feedback linearization method // Proceedings of National Polytechnic University of Armenia. Information Technologies, Electronics, Radio Engineering. 2020. vol. 2. pp. 44–54.
12. Itaketo U., Inyang H. Dynamic Modeling and Performance Analysis of an Autonomous Quadrotor Using Linear and Nonlinear Control Techniques // International Journal of Advances in Engineering and Management. 2021. vol. 3. no. 12. pp. 1629–1641.
13. Воевода А.А., Филюшов В.Ю. Многоканальная система подчиненного регулирования в многоканальном неквадратном представлении // Вестник Рязанского государственного радиотехнического университета. 2021. Т. 76. С. 90–100. DOI: 10.21667/1995-4565-2021-76-90-100.
14. Шрейнер Р.Т. Системы подчиненного регулирования электроприводов: учеб. пособие для вузов. Ч. 1. Электроприводы постоянного тока с подчиненным регулированием координат // Урал. гос. проф.-пед. ун-т. Екатеринбург: Издательство УГПТУ. 1997. 277 с.
15. Fezzani A., Drid S., Makouf A., Chrifi L. Speed sensorless flatness-based control of PMSM using a second order sliding mode observer // 2013 Eighth International Conference and Exhibition on Ecological Vehicles and Renewable Energies (EVER). 2013. pp. 1–9. DOI: 10.1109/EVER.2013.6521553.
16. Копесну Л., Hnidka J., Bajer J. Drone Motor Control using Fractional-Order PID Controller // International Conference on Military Technologies (ICMT). 2023. pp. 1–5. DOI: 10.1109/ICMT58149.2023.10171276.
17. Herrmann L., Bruckmann T., Bröcker M., Schramm D. Development of a Dynamic Electronic Speed Controller for Multicopters // 18th European Control Conference (ECC). Naples, Italy. 2019. pp. 4010–4015. DOI: 10.23919/ECC.2019.8795711.
18. Krener A., Isidori A. Linearization by output injection and nonlinear observers // Systems & Control Letters. 1983. vol. 3. pp. 47–52.
19. Жевнин А.А., Крищенко А.П. Управляемость нелинейных систем и синтез алгоритмов управления // Докл. АН СССР. 1981. Т. 258. № 4. С. 805–809.
20. Fetisov D. Linearization of affine systems based on control-dependent changes of independent variable // Diff Equat. 2017. vol. 53. pp. 1483–1494. DOI: 10.1134/S0012266117110106.
21. Поляк Б.Т., Хлебников М.В., Рапопорт Л.Б. Математическая теория автоматического управления: учебное пособие // М.: ЛЕНАНД. 2019. 500 с.
22. Филюшов В.Ю. Линеаризация нелинейного трехканального динамического объекта обратной связью // Вестник научных трудов НГТУ. 2017. Т. 66. № 1. С. 74–85.
23. Ким Д.П. Теория автоматического управления. Т. 2. // М.: Физматлит. 2004. 464 с.

24. Арзамасцев А.А., Крючков А.А. Математические модели для инженерных расчетов летательных аппаратов мультироторного типа // Вестник ТГУ. 2014. Т. 19. № 6. С. 1821–1828.
25. Kato Y. Performance Evaluation of a Gain-scheduled Propeller Thrust Controller Using Wind Velocity and Rotor Angular Velocity under Fluctuating Wind // IEEE 17th International conference on advanced motion control. 2022. pp. 12–17. DOI: 10.1109/AMC51637.2022.9729317.
26. Виноградов А. Векторное управление электроприводами переменного тока // ГОУ ВПО Ивановский государственный энергетический университет имени В.И. Ленина. 2008. 298 с.
27. Анучин А.С. Системы управления электроприводов. Учебник для ВУЗов // М.: Изд. МЭИ. 2015. 373 с.
28. Гайдук А.Р. Теория и методы аналитического синтеза систем автоматического управления (полиномиальный подход). // М.: ФИЗМАТЛИТ. 2012. 360 с.
29. Филошов В.Ю. Полиномиальный метод синтеза регуляторов по задающему и возмущающим воздействиям // Системы анализа и обработки данных. 2022. Т. 85. № 1. С. 93–108.
30. Филошов В.Ю. Полиномиальный матричный метод синтеза для многоканальных объектов с неквадратной матричной передаточной функцией: дис. канд. техн. наук: 2.3.1. 2022. 177 с.
31. Andrievsky B., Kuznetsov N.V., Leonov G.A., Pogromsky A.Yu. Hidden oscillations in aircraft flight control system with input saturation // International Federation of Automatic Control proceedings. 2013. vol. 46. no. 12. pp. 75–79. DOI: 10.3182/20130703-3-FR-4039.00026.

Воевода Александр Александрович — д-р техн. наук, профессор кафедры, кафедра автоматике, Новосибирский государственный технический университет. Область научных интересов: многоканальные системы управления, нелинейные системы управления. Число научных публикаций — более 350. voevoda@corp.nstu.ru; проспект Карла Маркса, 20, 630073, Новосибирск, Россия; р.т.: +7(383)319-6110.

Филошов Юрий Петрович — д-р техн. наук, профессор, Сибирский государственный университет водного транспорта. Область научных интересов: быстродействующий и энергоэффективный электропривод. Число научных публикаций — 70. filushov@mail.ru; улица Щетинкина, 33, 630099, Новосибирск, Россия; р.т.: +7(905)945-3923.

Филошов Владислав Юрьевич — канд. техн. наук, старший преподаватель кафедры, кафедра автоматике, Новосибирский государственный технический университет. Область научных интересов: многоканальные системы управления, нелинейные системы управления. Число научных публикаций — 40. filushov.vladislav@gmail.com; проспект Карла Маркса, 20, 630073, Новосибирск, Россия; р.т.: +7(952)935-6734.

A. VOEVODA, YU. FILIUSHOV, V. FILIUSHOV
**DEVELOPMENT OF A LINEAR CONTROL SYSTEM FOR A
THROTTLE OF A UAV PROPELLER-MOTOR GROUP**

Voevoda A., Filiushov Yu., Filiushov V. Development of a Linear Control System for a Throttle of a UAV Propeller-Motor Group.

Abstract. Orientation and positioning control of an unmanned aerial vehicle (UAV) vertical take-off and landing multi-rotor type in space is inextricably linked with the formation of a motion control vector, consisting of a combination of thrusts and aerodynamic moments of each propeller-motor group. The accuracy and speed of formation of the motion control vector greatly affect the positioning and orientation errors of the UAV. Most works devoted to the synthesis of UAV control systems use a motion control vector without taking into account the dynamics of the rotor-motor groups, which in some cases forces the control system to reduce its performance. The performance of the UAV control system can be increased by increasing the speed of generation of the thrust of the propeller-motor groups, for which a system for controlling the thrust of the propeller-motor group has been proposed. The propeller-motor group in its composition has a nonlinear internal connection in the aerodynamic torque and an output signal – thrust, that nonlinearly depends on the square of the propeller rotation speed. Typically, the propeller group is controlled like an electric motor – the internal coupling of the aerodynamic torque is considered an external disturbance, and the thrust is controlled by changing the speed of rotation of the propeller which is calculated based on the required motion control vector. It is proposed to consider thrust and aerodynamic torque an integral part of the propeller-engine group, for which to build a linear thrust control system. For this purpose, we carried out feedback linearization of the rotor-motor group system, connecting the voltage supplied to the motors with the motion control vector, which is the output value. The linearization process is divided into two stages: at the first stage feedback linearization is performed for an electric motor with internal nonlinear coupling by aerodynamic torque; at the second stage, linearization is performed with feedback on the output obtained at the first stage of the system with a nonlinear output signal – thrust. In accordance with the principles of subordinate control, motor control is formed for linearized feedback of the propeller group. Simulation was completed. An important issue when using feedback linearization is the preservation of the quality characteristics of the control system in the event of a mismatch between the parameters of the object and the model, the parameters of which are used to calculate the linearizing feedback. In this work, modeling was carried out with a discrepancy of some parameters up to 50%.

Keywords: feedback linearization, quadcopter, propeller group, slave control, AC motor, synthesis, control system.

References

1. Tkachev S. [Stabilization of non-minimal phase affine systems with linearization by some variables]. *Nauka i obrazovanie izd. MG TU imeni Baumana – Science and education*. ed. MSTU im. N.E. Bauman. 2011. no. 10. pp. 1–25. (In Russ.).
2. Shavin M.Yu. [Controllable dynamics of a quadcopter with rotating rotors] *Inzhenernyj zhurnal: nauka i innovacii – Engineering Journal: Science and Innovation*. 2018. no. 4(76). pp. 1–16. (In Russ.).
3. Shavin M., Pritykin D. Tilt-Rotor Quadrotor Control System Design and Mobile Object Tracking. *Mekhatronika, Avtomatizatsiya, Upravlenie*. 2019. vol. 20. no. 10. pp. 629–639. DOI: 10.17587/mau.20.629-639.

4. Cutler M., How J. Analysis and Control of a Variable-Pitch Quadrotor for Agile Flight. *Journal of Dynamic Systems, Measurement, and Control*. 2015. vol. 137(10). DOI: 10.1115/1.4030676.
5. Pyrkin A., Bobtsov A., Kolyubin S., Borisov O., Gromov V., Aranovskiy S. Output Controller for Quadcopters with Wind Disturbance Cancellation. *IEEE Conference on Control Applications (CCA)*. 2014. pp. 166–170. DOI: 10.1109/CCA.2014.6981346.
6. Demircioglu H., Basturk H. Adaptive Attitude and Altitude Control of a Quadrotor Despite Unknown Wind Disturbances. *IEEE 56th Annual Conference on Decision and Control*. 2017. pp. 274–279. DOI: 10.1109/CDC.2017.8263678.
7. Andrievsky B., Furtat I. Disturbance observers: methods and applications. II. Applications. *Automation and Remote Control*. 2020. vol. 81. pp. 1775–1818. DOI: 10.1134/S0005117920100021.
8. Kusaka T., Tanaka R. Stateful Rotor for Continuity of Quaternion and Fast Sensor Fusion Algorithm Using 9-Axis Sensors. *Sensors*. 2022. vol. 22(20). DOI: 10.3390/s22207989.
9. Pshihopov V., Medvedev M. [Group control of autonomous robots motion in uncertain environment via unstable modes]. *Trudi SPIIRAN – SPIIRAS Proceedings*. 2018. vol. 60. no. 5. pp. 39–63. (In Russ.).
10. Zulu A., John S. A review of control algorithms for autonomous quadrotors. *Open Journal of Applied Sciences*. 2014. no. 4. pp. 547–556. DOI: 10.4236/ojapps.2014.414053.
11. Gasparyan O., Darbinyan H., Asatryan A., Simonyan T. On the control of quadcopters based on the feedback linearization method. *Proceedings of National Polytechnic University of Armenia. Information Technologies, Electronics, Radio Engineering*. 2020. vol. 2. pp. 44–54.
12. Itaketo U., Inyang H. Dynamic Modeling and Performance Analysis of an Autonomous Quadrotor Using Linear and Nonlinear Control Techniques. *International Journal of Advances in Engineering and Management*. 2021. vol. 3. no. 12. pp. 1629–1641.
13. Voevoda A., Filiushov V. [Multi-loop slave control system in multi-channel non-square representation]. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta – Bulletin of the Ryazan State Radio Engineering University*. 2021. vol. 76. no. 2. pp. 90–101. (In Russ.).
14. Shrejner R. *Sistemy podchinennogo regulirovaniya jelektroprivodov: ucheb. posobie dlja vuzov. Ch. 1. Jelektroprivody postojannogo toka s podchinennym regulirovanijem koordinat* [Systems of subordinate control of electric drives: textbook. manual for universities. Part 1. DC electric drives with subordinate coordinate control.]. Ural. state prof.-ped. univ. Ekaterinburg: Publishing house UGPPU. 1997. 277 p. (In Russ.).
15. Fezzani A., Drid S., Makouf A., Chrifi L. Speed sensorless flatness-based control of PMSM using a second order sliding mode observer. *2013 Eighth International Conference and Exhibition on Ecological Vehicles and Renewable Energies (EVER)*. 2013. pp. 1–9. DOI: 10.1109/EVER.2013.6521553.
16. Kopečný L., Hnidka J., Bajer J. Drone Motor Control using Fractional-Order PID Controller. *International Conference on Military Technologies (ICMT)*. 2023. pp. 1–5. DOI: 10.1109/ICMT58149.2023.10171276.
17. Herrmann L., Bruckmann T., Bröcker M., Schramm D. Development of a Dynamic Electronic Speed Controller for Multicopters. *18th European Control Conference (ECC)*. Naples. Italy. 2019. pp. 4010–4015. DOI: 10.23919/ECC.2019.8795711.
18. Krener A., Isidori A. Linearization by output injection and nonlinear observers. *Systems & Control Letters*. 1983. vol. 3. pp. 47–52.

19. Zhevnin A.A., Krishchenko A.P. [Controllability of nonlinear systems and synthesis of control algorithms]. Dokl. AN SSSR – Reports of the Academy of Sciences of the USSR. 1981. vol. 258. no. 4. pp. 805–809. (In Russ.)
20. Fetisov D. Linearization of affine systems based on control-dependent changes of independent variable. Diff. Equat. 2017. vol. 53. pp. 1483–1494. DOI: 10.1134/S0012266117110106.
21. Polyak B., Hlebnikov M., Rapoport L. Matematicheskaja teorija avtomaticheskogo upravlenija: uchebnoe posobie [Mathematical theory of automatic control: textbook]. Moscow: LENAND, 2019. 500 p. (In Russ.).
22. Filiushov V. [Linearization of a nonlinear three-channel dynamic object with feedback]. Vestnik nauchnyh trudov NGTU – Bulletin of scientific works of NSTU. 2017. vol. 66. no. 1. pp. 74–85. (In Russ.).
23. Kim D. Teorija avtomaticheskogo upravlenija. T. 2 [Automation control theory Volume 2]. Moscow: Fizmatlit, 2004. 464 p. (In Russ.).
24. Arzamascev A., Kryuchkov A. [Mathematical models for engineering calculations of multi-rotor aircraft]. Vestnik TGU – Bulletin of TSU. 2014. vol. 19. no. 6. pp. 1821–1828. (In Russ.).
25. Kato Y. Performance Evaluation of a Gain-scheduled Propeller Thrust Controller Using Wind Velocity and Rotor Angular Velocity under Fluctuating Wind. IEEE 17th International conference on advanced motion control. 2022. pp. 12–17. DOI: 10.1109/AMC51637.2022.9729317.
26. Vinogradov A. Vektornoe upravlenie jelektrivodami peremennogo toka [Vector control of AC electric drives] // GOU VPO Ivanovskij gosudarstvennyj energeticheskij universitet. 2008. 298 p. (In Russ.).
27. Anuchin A. Sistemy upravlenija jelektrivodov [Electric drive control systems]. Moscow: Ed. MPEI, 2015. 373 p. (In Russ.).
28. Gaiduk A. Teorija i metody analiticheskogo sinteza sistem avtomaticheskogo upravlenija (polinomial'nyj podhod) [Theory and methods of analytical synthesis of automatic control systems (polynomial approach)]. Moscow: Fizmatlit, 2012. 360 p. (In Russ.).
29. Filiushov V. [Polynomial method for synthesizing controllers based on master and disturbing influences]. Sistemy analiza i obrabotki dannyh – Data analysis and processing systems. 2022. vol. 85. no. 1. pp. 93–108. (In Russ.).
30. Filiushov V. [Polynomial matrix synthesis method for multichannel objects with non-square matrix transfer function]: diss. Ph.d.: 2.3.1. St-Petersburg, 2022. 177 p. (In Russ.).
31. Andrievsky B., Kuznetsov N., Leonov G., Pogromsky A. Hidden oscillations in aircraft flight control system with input saturation. International Federation of Automatic Control proceedings. 2013. vol. 46. no. 12. pp. 75–79. DOI: 10.3182/20130703-3-FR-4039.00026.

Voevoda Alexander — Ph.D., Dr.Sci., Professor of the department, Automation department, Novosibirsk State Technical University (NSTU). Research interests: multichannel control systems, nonlinear control systems. The number of publications — более 350. voevoda@corp.nstu.ru; 20, Karl Marx Ave., 630073, Novosibirsk, Russia; office phone: +7(383)319-6110.

Filiushov Yuriy — Ph.D., Dr.Sci., Professor, Siberian State University of Water Transport. Research interests: high-speed and energy-efficient electric drive. The number of publications — 70. filushov@mail.ru; 33, Shchetinkina St., 630099, Novosibirsk, Russia; office phone: +7(905)945-3923.

Filiushov Vladislav — Ph.D., Senior lecturer of the department, Automation department, Novosibirsk State Technical University (NSTU). Research interests: multichannel control systems, nonlinear control systems. The number of publications — 40. filiushov.vladislav@gmail.com; 20, Karl Marx Ave., 630073, Novosibirsk, Russia; office phone: +7(952)935-6734.

A. ISKANDAR, A. HAMMOUD, B. KOVÁCS
**IMPLICIT UNDERSTANDING: DECODING SWARM BEHAVIORS IN
ROBOTS THROUGH DEEP INVERSE REINFORCEMENT
LEARNING**

Iskandar A., Hammoud A., Kovács B. **Implicit Understanding: Decoding Swarm Behaviors in Robots through Deep Inverse Reinforcement Learning.**

Abstract. Using reinforcement learning to generate the collective behavior of swarm robots is a common approach. Yet, formulating an appropriate reward function that aligns with specific objectives remains a significant challenge, particularly as the complexity of tasks increases. In this paper, we develop a deep inverse reinforcement learning model to uncover the reward structures that guide autonomous robots in achieving tasks by demonstrations. Deep inverse reinforcement learning models are particularly well-suited for complex and dynamic environments where predefined reward functions may be difficult to specify. Our model can generate different collective behaviors according to the required objectives and effectively copes with continuous state and action spaces, ensuring a nuanced recovery of reward structures. We tested the model using E-puck robots in the Webots simulator to solve two tasks: searching for dispersed boxes and navigation to a predefined position. Receiving rewards depends on demonstrations collected by an intelligent pre-trained swarm using reinforcement learning act as an expert. The results show successful recovery of rewards in both segmented and continuous demonstrations for two behaviors – searching and navigation. By observing the learned behaviors of the swarm by the expert and proposed model, it is noticeable that the model does not merely clone the expert behavior but generates its own strategies to achieve the system’s objectives.

Keywords: deep inverse reinforcement learning, reward function, demonstrations, searching behavior, navigation behavior.

1. Introduction. Swarm robotics (SR) is a self-organized system with a decentralized control architecture. Robots in SR interact with each other and their environment to achieve the desired behavior collectively. The key features of SR systems rely on autonomy and the local perception of individuals. These features lead to SR with robustness to individual failures, flexibility with changes in environmental conditions, and scalability for different sizes of swarm [1 – 4]. Methods of generating the collective behavior of the SR typically depend on the required task. The methods used to generate collective behavior in swarm robotics (SR) are closely tailored to the specific tasks the robots are designed to perform. For example, in search and rescue missions, SR systems utilize algorithms that optimize area coverage and ensure rapid localization of targets, such as finding survivors in a collapsed building. In navigation tasks, SR might employ pathfinding algorithms that enable the swarm to efficiently move through complex environments, avoiding obstacles and minimizing travel time [5 – 11]. This concept addresses the task-oriented nature of SR. It is challenging because of the unpredictable interactions

within a swarm and the diverse potential applications and environments. There is no generalized, applicable method for designing desired collective behavior, making this a key area of research. Many directions addressed this challenge, like bio-inspired algorithms: Drawing inspiration from nature, these algorithms emulate behaviors observed in biological systems, such as the flocking of birds or the foraging patterns of ants. Such models help in designing decentralized control systems where each robot in the swarm behaves according to simple rules based on local information and interactions [12 – 15], modular design: This approach focuses on creating robots with interchangeable modules, which can dynamically reconfigure based on the task at hand. Modular design enhances the flexibility and scalability of swarm robotics systems, allowing for adaptability to different environments and tasks by rearranging the modules to fit specific needs [16, 17], evolutionary robotics: This method uses evolutionary algorithms to develop the control systems for robots, effectively allowing the robots' behavior to evolve and optimize over time. It mimics natural selection processes to automatically generate solutions that are well-adapted to their environment and task, continually improving as the system encounters new scenarios [18], and machine learning, where reinforcement learning (RL) provides a robust framework for developing SR systems with diverse tasks [19]. RL supplies robots with autonomy and the ability to learn from others and the environment. It can manage the complexity of designing collective behavior by breaking down the learning process into simpler, manageable parts, with dynamic adaptation. The RL approach in SR is represented as a Markov decision process (MDP) as (S, A, R, T, γ) , where the robot moves from state S to a new state S_{t+1} by executing an action A . The essential function in RL is the reward R where robots learn to perform the actions that maximize the cumulative received rewards during period T by weighting them by factor γ . Thus, the problem of designing R reflects the main objectives of the given task that correspond to generating the collective behavior of SR [20]. To generate the collective behavior of SR by RL. Firstly, define the environment in which multiple agents can coexist. Secondly, representing the states and actions spaces for policy π representation. Then define the reward function by deciding whether the agents receive rewards based on individual performance, collective outcomes, or a combination of both. Finally, choose an appropriate RL algorithm. Choosing R is critical because the reward function directly shapes the agent's behavior, guiding it towards desired objectives and away from undesirable actions. Inverse reinforcement learning (IRL) is a sophisticated approach that involves learning the underlying reward function based on the observed behavior of experts. Unlike traditional RL, which directly learns a policy based on a pre-defined reward formula, IRL provides a deeper

understanding of complex behaviors by demonstrations instead of explicitly tuning the reward formula that describes these behaviors. This methodology is particularly advantageous in swarm robotics, where explicit reward functions are challenging to formulate due to the interactions and collective dynamics of robots.

2. Related works. Defining an appropriate reward function requires mathematical knowledge and a deep understanding of the operating conditions of the system. Formulating the R equation becomes more complex by increasing the objectives of the given task. Many methods have been used to formulate R which corresponds to generating the collective behavior of SR, like sparse rewards, which are infrequent or only given sparsely throughout the training phase, where rewards or punishment are given for a specific action like when each robot near to its fellow or when the swarm reaches the target, and negatives values for colliding with obstacles [21, 22]. Shaping rewards is another method that depends on providing additional rewards to guide agents toward the desired behavior more efficiently. This involves rewarding each action through each time step during the episode, as opposed to sparse rewards where specific actions are rewarded. Both of sparse and shaping methods were used to solve the foraging SR problem. It demonstrates that the shaping method was able to solve the problem while the sparse one failed. RL was modified with a hierarchal structure to solve the problem with sparse rewards [23].

To eliminate the need for manually crafting rewards and provide more structured learning guidance than traditional methods, IRL introduces a solution to infer the reward function from demonstrations collected by an expert. So, the agent learns a policy that is similar to or better than the expert policy based on the inferred reward function. The obtained policy does not require mathematical experience or a full understanding of the conditions and operations needed to formulate R , thus avoiding human bias and potential suboptimality [24].

The main idea behind IRL is to understand what motivates these behaviors by analyzing the decisions that experts make in various situations. In IRL, the expert is typically an agent (human or robotic) who performs a task with high proficiency. The expert's behavior serves as a benchmark or model that the IRL algorithm attempts to emulate. By observing the expert, the IRL aims to deduce the reward structure that guides the expert's decisions, assuming that the expert's actions are optimized to maximize some form of cumulative reward. Developing an IRL model to be deployed in swarm robotics is an interesting research area due to its ability to facilitate autonomous decision-making in complex, dynamic environments. Many researchers have carried out IRL to generate collective behavior for many missions like [25],

where maximum entropy IRL is used for each agent to infer the birds' reward functions from observed GPS data of pigeon flocks. This approach allowed them to not only simulate flocking behavior but also infer potential leader-follower dynamics within the flock. IRL was also used in SR for area coverage problems, particularly focusing on improving efficiency in unstructured search and rescue scenarios [26]. The solution involved humans in the loop with IRL. Human expert demonstrations are used to train SR, allowing them to learn optimal area coverage strategies. The author in [27] combines IRL with automatic modular design to generate control software for robot swarms based on only the demonstrations, without the need for explicitly defined reward and objective functions. Many methodologies in cited IRL can not handle high-dimensional, continuous state-action spaces and are capable of generalizing across different tasks and dynamic environments. This is crucial for developing adaptive and robust swarm robotic systems that can operate effectively in a wide range of scenarios. Our paper investigates the ability of IRL to generalize across different scenarios and automate reward design, making it robust and efficient, particularly in complex and continuous environments. It introduces an IRL model able to deal with continuous state and action spaces with simplified segmented or continuous demonstrations. This model can be generalized to produce different collective behaviors such as navigation and searching tasks.

3. System Description. This section describes the framework of the SR by testing two tasks: searching for the boxes represented as light sources and navigating from initial positions to a predefined position illustrated as a circular yellow area called (P), as shown in Figure 1.

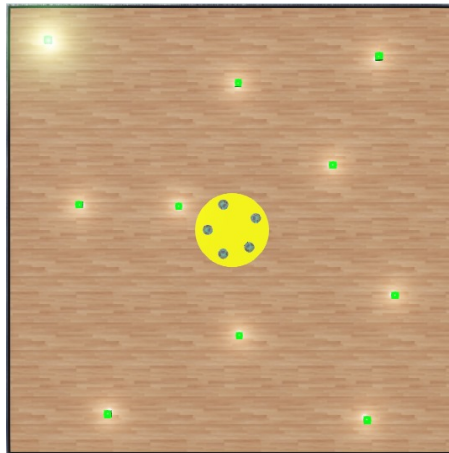


Fig. 1. SR environment

The swarm system was implemented in a 3D robot simulator called Webots where the E-Puck mobile robot was selected to build the swarm. The dimensions of the workspace were defined as $3 \times 3m^2$, forming a square area surrounded by four walls. The parameters of E-Puck robots were set as follows: linear velocity $[0, 0.25]m/s$, angular velocity $[-3.14, 3.14]rad/s$, and light sensors' readings corresponded to the light intensity $[0, 4095]$.

3.1. RL architecture. The proximal policy optimization algorithm (PPO) is used for both searching and navigation tasks. Most of the studies mentioned in the related work section used PPO. It is favored in robotics for its balance between sample efficiency and computational simplicity, avoiding the need for complex calculations like those in Trust Region Policy Optimization (TRPO). PPO's stability is enhanced via a policy gradient method that maximizes an objective function by using a clipped surrogate objective to keep updates stable [28], thus maintaining steady training progress. This makes PPO an adaptable and robust choice for a variety of applications, particularly those involving continuous action spaces and environments with complex dynamics.

The problem is formulated as a MDP represented by the tuple (S, A, T, R, γ) . The state space S has two frames, one for the searching task contains light sensor readings, and the other frame for the navigation task includes the distance D , besides the angle θ between the robot and P . The action space A includes the velocities of both the left and right motors. The transition function T describes the dynamics of the system. In continuous states and actions spaces, the transition dynamics function typically cannot be explicitly defined for every possible state and action due to the infinite possibilities. PPO optimizes a policy function that outputs a probability distribution over actions given the current state. The policy is typically parameterized by a neural network where the weights are adjusted to maximize the cumulative reward. The reward function R provides feedback based on the system's behavior. The main architecture of the PPO has two neural networks, actor and critic with fully connected layers, as shown in Figure 2. Table 1 presents the parameters of PPO.

3.1.1. Reward formulating. In the searching task, As we mentioned the PPO receives the light sensor readings as inputs and produces wheel velocities as outputs, Figure 2. The learned velocities attempt to adjust the robot's trajectory toward the light source (boxes). To formulate the equation of the reward function to motivate robots to steer their directions to the light sources where the boxes are located, it is better to measure the intensity of the light between two states at times $t - 1$ and t and be rewarded if it is increased.

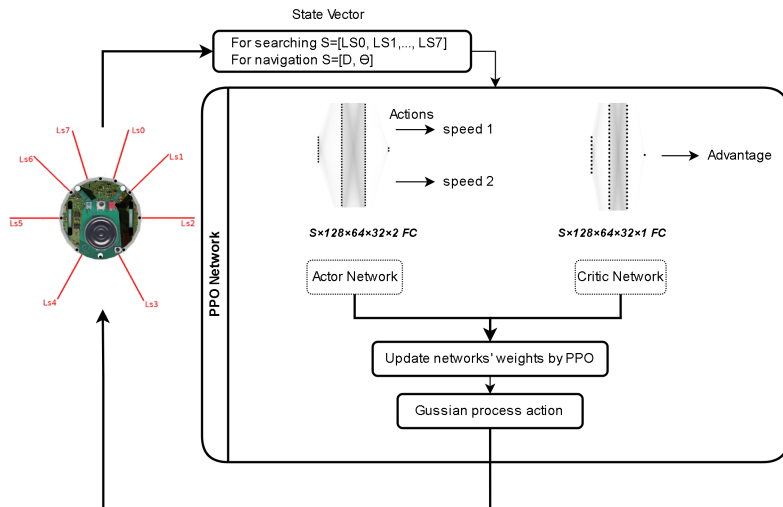


Fig. 2. PPO architecture

Table 1. PPO hyperparameters

Parameter	Value
max training timesteps	RL:1000000, IRL-RL:250000
max timesteps per episode	800
state space dimension	Searching:8, Navigation:2
action space dimension	2
discount factor (γ)	0.99
PPO epsilon clip	0.2
PPO K epochs	80
optimizer learning rate actor	0.0003
optimizer learning rate critic	0.001
Layers size	input,128,64,32,output

An additional value 1.1 is given when the robot finds the box, as in Equations 1 and 2. Notably, both shaping and sparse methods were used.

$$r_{\text{box}} = \begin{cases} 1.1 & \text{if } LS_0^{(t)} > \text{FindThreshold}_{\text{searching}} \\ 1.1 & \text{if } LS_7^{(t)} > \text{FindThreshold}_{\text{searching}} \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

$$R(t)_{searching} = \frac{(LS_0^{(t-1)} - LS_0^{(t)}) + (LS_7^{(t-1)} - LS_7^{(t)})}{2} + r_{box}, \quad (2)$$

where:

$R(t)_{searching}$ – Reward value at each time step t .

$LS_0^{(t)}, LS_7^{(t)}$ – The current readings of light sensors 0 and 7, respectively, at time t .

$LS_0^{(t-1)}, LS_7^{(t-1)}$ – The previous readings of light sensors 0 and 7, respectively, at time $t - 1$.

$FindThreshold_{searching}$ – The threshold value for the light sensor where the box is found.

r_{box} – The additional reward when the robot finds the box.

For the navigation task, the inputs of the PPO network are the robot's current distance and angle relative to P , where the outputs modify the wheel velocities to navigate P . Additional reward is sparse for successfully reaching P , as in Equation 3, incorporating the shaping method to speed up the learning process, as in Equation 4.

$$r_P = \begin{cases} 0.1 & \text{if } D_t < FindThreshold_{navigation} \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

$$R(t)_{navigation} = (D_{t-1} - D_t) + r_P + \frac{\cos(\theta_t)}{1000}, \quad (4)$$

r_P – The additional reward when the robot reaches P .

$FindThreshold_{navigation}$ – The robot is inside P .

D_t – The distance between the robot and P at time t .

D_{t-1} – The distance between the robot and P at time $t-1$.

θ_t – The angle between the robot and P .

3.2. IRL-RL model. In this approach, instead of formulating the reward mathematically train RL to find the policy. IRL is implemented to infer the reward by demonstrations collected via a pre-trained swarm. RL used them to generate the policy to obtain the searching and navigation behaviors, as in Figure 3.

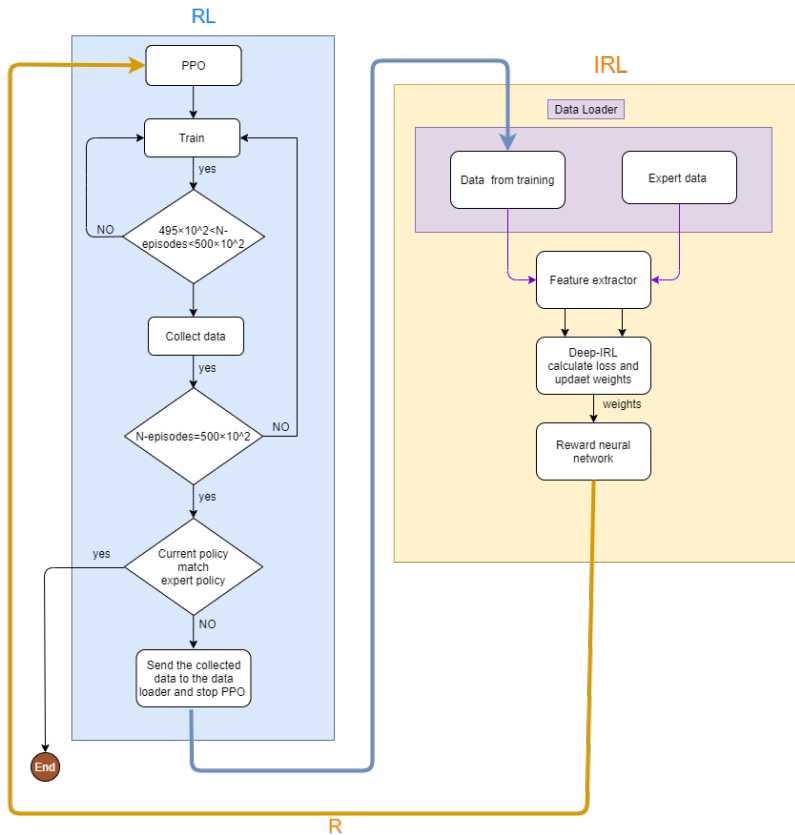


Fig. 3. IRL-RL model

IRL components:

Data loader: it is a container for the data that flows from the expert and training process. Expert data is collected by an expert pre-trained model while training data is collected during the PPO training process. Both of them contain only state frames with flags without actions. The flags are indicators of completing the task like finding a box in the searching task or reaching P in the navigation. The model is able to cope with two types segmented and continuous states. Both types of states were tested in the results section. In segmented mode, the sensors' readings after normalization are divided into five ranges between 0 and 1, each corresponding to a value.

Features extractor: the given Table 2 shows the types of data and functions that are used on the received data from the data loader. The received data is raw, where the values of the light sensors are measured in a range of $[0, 4095]$. While the distance D is in the range of $[0, 3]m$ and the angle θ is $[-\pi, \pi]rad$, the purpose of $\phi(s)$ in Equations 5, and 6 is to convert raw states S into a feature vector which is more suitable as input for the model. The shift function is applied to the states after normalization to obtain the values in $t - 1$. These values are used to produce the correlation between states to encourage the R network to perceive the directions of changes in states.

Table 2. Features Extractor Input and Output for Searching and Navigation Tasks

Task	Input of features extractor (from the data loader)	Output of features extractor
Searching	$LS_0^{(t)}, LS_7^{(t)}$, flag (Finding a box)	Normalized $[LS_0^{(t-1)}, LS_0^{(t)}, LS_7^{(t-1)}, LS_7^{(t)}]$, flag (Finding a box)
Navigation	$D^{(t)}, \theta^{(t)}$, flag (Reaching P)	Normalized $[D^{(t-1)}, D^{(t)}, \theta^{(t-1)}, \theta^{(t)}]$, flag (Reaching P)

$$\phi(s) : S \rightarrow [0, 1], \quad (5)$$

$$\phi(s) = \frac{\text{Max}_{\text{Output}} - \text{Min}_{\text{Output}}}{\text{Max}_{\text{Value}} - \text{Min}_{\text{Value}}} \cdot (s - \text{Max}_{\text{Value}}) + \text{Max}_{\text{Output}}, \quad (6)$$

$\text{Max}_{\text{Value}}$ – the upper value in the raw range of states, for example, in the searching task equals 4095 according to the light sensor reading

$\text{Min}_{\text{Value}}$ – the lower value in the raw range of states, for example, in the searching task equals 0.

$\text{Max}_{\text{Output}}$ – the upper value in the output range of $\phi(s)$, equals 1 as in Equation 5.

$\text{Min}_{\text{Output}}$ – the lower value in the output range of $\phi(s)$, equals 0 as in Equation 5.

Deep IRL – the backpropagation process of the reward network is performed by calculating the losses according to Equation 7, which guarantees updating the weights of the reward neural network. This objective function is the binary cross-entropy loss function applied for distinguishing between expert and training rewards. This loss function is designed to penalize the deviation of the predicted rewards from the "true" rewards indicated by the expert's behavior.

$$\text{loss} = -\log(\text{sigmoid}(R_{\text{expert}})) - (1 - \log(\text{sigmoid}(R_{\text{training}}))), \quad (7)$$

R_{expert} – the output of the reward neural network for states from the expert.

R_{training} – the output of the reward neural network for collected states from the training process.

Reward network – The purpose of the reward neural network is to approximate the reward function. This is done by passing the feature vector through the neural network. Then, producing a scalar reward value as an output. It is constructed as fully connected layers of $\text{length}(\text{feature} - \text{vectors}) \times 15 \times 1FC$, where the length of feature vectors in the proposed tasks is 5, as in Table 2.

4. Results and discussion. We have focused on examining the reward that affects swarm behavior in a simulated environment, demonstrating how IRL can approximate the reward function without the need for mathematical formalities. Our implementation involved two tasks: searching for boxes in continuous RL and segmented features, and navigation task to a pre-defined position known as P also in continuous RL but in continuous features. The swarm's performance was evaluated by comparing the rewards between the IRL-based model and an expert-pre-trained RL model, demonstrating the ability to generate behavior to achieve the required tasks. Finally, we analyze the generated behavior of the swarm under both models. Choosing the features plays a major role in recovering a correct reward function. They differ based on the defined problem and objective function of the swarm system as in Table 2. This table illustrates the differences in the chosen features for searching and navigation behavior. In addition, consider the readings in the time step t and $t - 1$ to make the R network recognize the difference in the light intensity for the searching task or change in the distance in the navigation task. Using a deep neural network to represent R with binary cross-entropy loss function makes the model able to handle continuous environments. So, in our models, we recovered reward in the continuous and segmented mode of the IRL model to generate the policy in RL with continuous states and action spaces.

4.1. Searching task. In this task, the reward structure is related to the change in the intensity of light detected by sensors. The reward increases as the robots move towards stronger light intensity, collecting a higher reward once the boxes are located.

The training process required three rounds to recover the reward which generates a successful behavior as follows.

In the first round, the reward neural network was initially configured with arbitrary weights, denoted as ω_0 . So, RL learned a stochastic policy π_0 . Based on this policy, data was collected and forwarded to the data loader. By harnessing both expert and collected data, the IRL part conducted training on the reward network, resulting in an update to the weights, yielding ω_1 . Subsequently, the RL part in Figure 2 trained again to generate its policy π_1 by the new weights of reward. The iterative process finished with the weights ω_2 , so the robots trained to learn policy π_2 achieved the required task as shown in Figure 4. The reward function in the expert-RL model, indicated by a red line, maximizes the reward as the robot approaches the light source (box) and maintains peak values upon reaching it. In contrast, the inferred reward by the IRL-RL model, represented by a blue line, captures the increasing light intensity in a segmented fashion, echoing the RL model's behavior but with discrete transitions due to the segmented mode of features. Thus, the data is segmented into specific ranges, such as states from 0 to 0.2 representing darker areas rewarded uniformly.

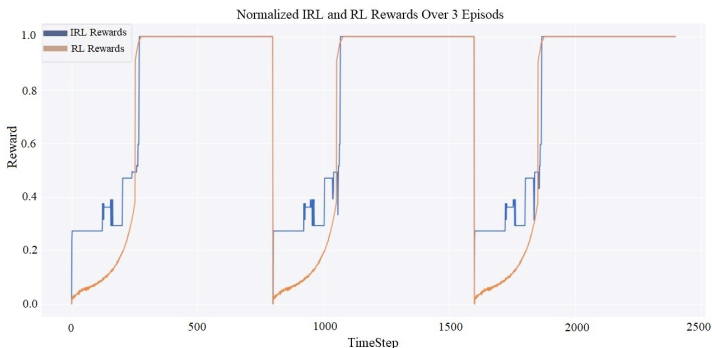


Fig. 4. Normalized IRL and RL rewards over 3 episodes for ω_2 -searching task

As shown in Figure 5, the IRL-RL model effectively reconstructs the reward function, when sensor readings LS0 and LS7 in the 0 to 0.2 range result in minimal rewards, increasing as the robot transitions to the 0.2-0.4 range – the reward spikes in the 0.8-1 range, indicating the robot's proximity to the box. The comparative visualization of the reward functions from the expert RL model (pre-trained) and the IRL-RL model highlights differences, with the darker blue associated with the RL model. Nonetheless, a crucial observation is the shared gradient pattern between the models, suggesting a direct correlation of rewards with incremental light intensity, affirming that the robots have learned to search and find boxes, thereby validating the designed

behavior. The IRL-RL model’s ability to mimic the decision-making strategy of the pre-trained robot.

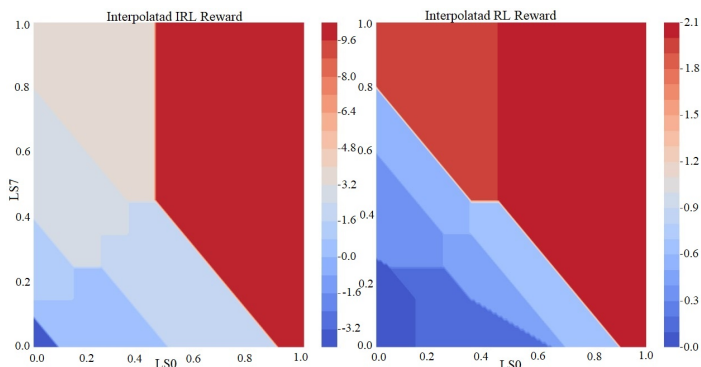


Fig. 5. Heat map of the true reward (right) and the recovered reward (left) for the searching task

For the robot’s behavior, the bar chart in Figure 6 highlights a successful behavior of the IRL-RL model compared to pre-trained expert RL across ten incremental collected boxes. The IRL-RL model generates a behavior that enables the robots to collect the boxes in a round-trip manner. However, it is notable that there are differences in the collecting box’s times which reflect different behaviors. This means the IRL-RL model does not clone the behavior or actions of RL instead, it learns how to achieve the task with its own generated behavior. This suggests that IRL successfully learns the complex reward structures governing task completion rather than merely copying actions.

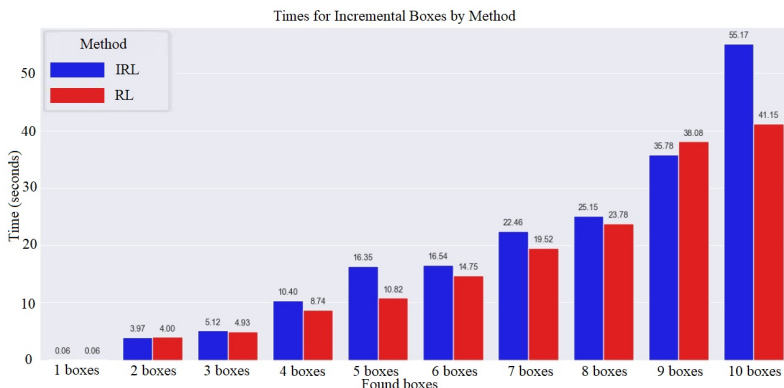


Fig. 6. Swarm searching behavior by IRL and RL

4.2. Navigation task. The same process with the same number of rounds was applied to the navigation task. Initially, the reward neural network was configured with random weights, labeled as ω_0 , leading to the production of a stochastic policy π_0 from which the RL model began its learning process. After the first round, the collected data was used to train the reward network, updating the weights to ω_1 , which in turn allowed the RL model to refine its policy to π_1 . The process concluded after a second round of adjustments, resulting in final weights ω_2 , enabling the model to successfully execute the required navigation tasks as depicted as in Figure 7.

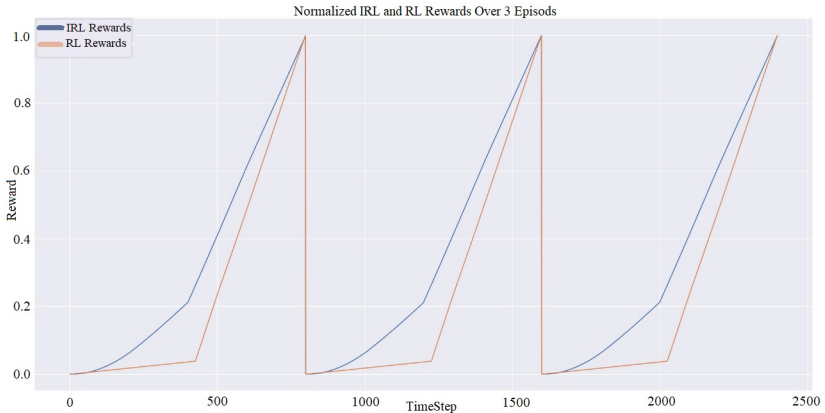


Fig. 7. Normalized IRL and RL rewards over 3 episodes for ω_2 - Navigation task

Figure 8 illustrates both the interpolated IRL reward and interpolated RL reward models under these conditions. It shows a gradient of colors from red to blue, indicating varying reward intensities based on the robot's angle and distance to the target.

High rewards are shown in red, corresponding to smaller angles and distances – indicative of the robot directly facing and being close to the target. As the angle increases or the distance increases, the reward diminishes, as shown by the gradient transitioning to blue. Unlike the search, there is no segmentation into discrete state ranges. The continuous nature of the data allows for a smoother gradient in the visual representation and a more nuanced adjustment of the reward based on the robot's proximity and alignment with the target.

The similarities between the paths in both graphs in Figure 9 indicate that the IRL has effectively learned from the RL data, closely replicating the expert RL's behavior. This suggests the successful application of IRL where

the algorithm has inferred the strategies and decisions that the RL considered optimal.

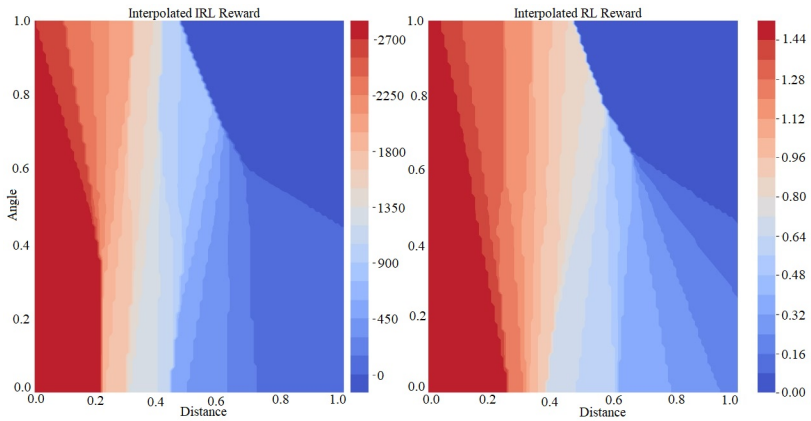


Fig. 8. Heat map of the true reward (right) and the recovered reward (left) for the Navigation task

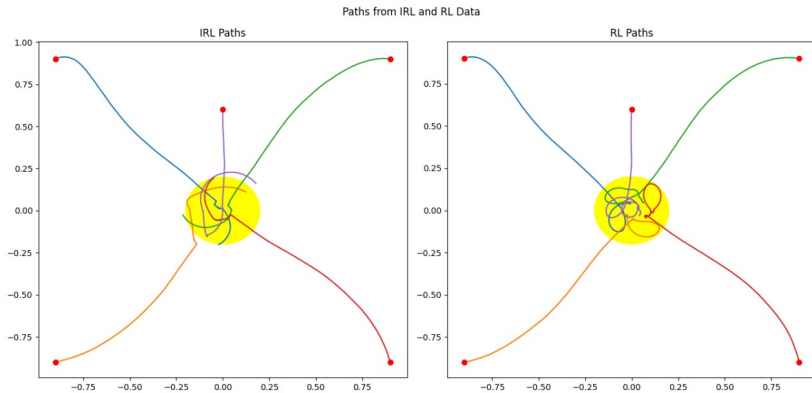


Fig. 9. Robots navigation paths

5. Conclusion. This paper presented an advanced model employing IRL to effectively recover the reward function by demonstrations of expert behaviors. Instead of directly learning the behavior, IRL seeks to understand the reasons behind specific actions or strategies by inferring the reward to solve the task by generating its own behavior. It eliminates the need for extensive manual tuning of reward functions and facilitates a more intuitive

setup via demonstration-based learning. The proposed IRL-RL model has the ability to handle continuous state spaces and dynamic environments to deal with continuous RL problems due to employing a deep neural network for representing R, in addition, to recovering reward function based on two types of data that flow from data loader: segmented features and continuous features for naive strategies. This model was tested across two tasks, navigating towards a predefined position and searching for specific objects within a simulated swarm robot environment. It demonstrated its robust capability to infer and adapt the reward structures essential for guiding autonomous robotic swarms to accomplish tasks. Moreover, our findings highlight the potential of the proposed model to generalize across different scenarios. For future directions, this model will be developed to achieve composed and more complex tasks like generating foraging collective behaviors and aggregation behavior.

References

1. Shahzad M., Saeed Z., Akhtar A., Munawar H., Yousaf M., Baloach N., Hussain F. A review of swarm robotics in a nutshell. *Drones*. 2023. vol. 7. no. 4.
2. Schranz M., Umlauf M., Sende M., Elmenreich W. Swarm robotic behaviors and current applications. *Frontiers in Robotics and AI*. 2020. vol. 7.
3. Cheraghi A., Shahzad S., Graffi K. Past, present, and future of swarm robotics. In *Intelligent Systems and Applications: Proceedings of the 2021 Intelligent Systems Conference (IntelliSys)*. 2022. vol. 3. pp. 190–233.
4. Brambilla M., Ferrante E., Birattari M., Dorigo M. Swarm robotics: a review from the swarm engineering perspective. *Swarm Intelligence*. 2013. vol. 7. pp. 1–41.
5. Nauta J., Van Havermaet S., Simoens P., Khaluf Y. Enhanced foraging in robot swarms using collective Lévy walks. In *24th European Conference on Artificial Intelligence (ECAI)*. 2020. pp. 171–178.
6. Misir O., Gokrem I. Flocking-based self-organized aggregation behavior method for swarm robotics. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*. 2021. vol. 45. no. 4. pp. 1427–1444.
7. Sadeghi A., Raoufi M., Turgut A. A self-adaptive landmark-based aggregation method for robot swarms. *Adaptive Behavior*. 2022. vol. 30. no. 3. pp. 223–236.
8. Lu Q., Hecker J., Moses M. The MPFA: A multiple-place foraging algorithm for biologically-inspired robot swarms. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016. pp. 3815–3821.
9. Lopes Y., Trenkwalder S., Leal A., Dodd T., Groß R. Supervisory control theory applied to swarm robotics. *Swarm Intelligence*. 2016. vol. 10. pp. 65–97.
10. Hamann H. *Swarm robotics: A formal approach*. Cham: Springer, 2018. 210 p.
11. Berlinger F., Gauci M., Nagpal R. Implicit coordination for 3D underwater collective behaviors in a fish-inspired robot swarm. *Science Robotics*. 2021. vol. 6. no. 50. DOI: 10.1126/scirobotics.abd8668.
12. Zhang J., Lu Y., Che L., Zhou M. Moving-distance-minimized PSO for mobile robot swarm. *IEEE Transactions on Cybernetics*. 2021. vol. 52. no. 9. pp. 9871–9881.
13. Parhi D., Sahu C., Kumar P. Navigation of multiple humanoid robots using hybrid adaptive swarm-adaptive ant colony optimisation technique. *Computer Animation and Virtual Worlds*. 2018. vol. 29. no. 2. DOI: 10.1002/cav.1802.

14. Jiang L., Mo H., Tian P. An adaptive decentralized control strategy for deployment and aggregation of swarm robots based on bacterial chemotaxis. *Applied Intelligence*. 2023. vol. 53. no. 10. pp. 13018–13036.
15. Hu C., Arvin F., Bellotto N., Yue S., Li H. Swarm neuro-robots with the bio-inspired environmental perception. *Frontiers in Neurorobotics*. 2024. vol. 18.
16. Hasselmann K., Ligot A., Birattari M. Automatic modular design of robot swarms based on repertoires of behaviors generated via novelty search. *Swarm and Evolutionary Computation*. 2023. vol. 83.
17. Birattari M., Ligot A., Francesca G. AutoMoDe: a modular approach to the automatic off-line design and fine-tuning of control software for robot swarms. *Automated Design of Machine Learning and Search Algorithms*. 2021. pp. 73–90.
18. Stolfi D., Danoy G. Evolutionary swarm formation: From simulations to real world robots. *Engineering Applications of Artificial Intelligence*. 2024. vol. 128. DOI: 10.1016/j.engappai.2023.107501.
19. Blais M., Akhloufi M. Reinforcement learning for swarm robotics: An overview of applications, algorithms and simulators. *Cognitive Robotics*. 2023. vol. 3. pp. 226–256. DOI: 10.1016/j.cogr.2023.07.004.
20. Sutton R., Barto A. *Reinforcement learning: An introduction*. The MIT press, 2018. 552 p.
21. Iskandar A., Rostum H., Kovacs B. Using Deep Reinforcement Learning to Solve a Navigation Problem for a Swarm Robotics System. *Proceedings of the 24th International Carpathian Control Conference (ICCC)*. IEEE, 2023. pp. 185–189.
22. Wei Y., Nie X., Hiraga M., Ohkura K., Car Z. Developing end to end control policies for robotic swarms using deep Q-learning. *Journal of Advanced Computational Intelligence and Intelligent Informatics*. 2019. vol. 23. no. 5. pp. 920–927.
23. Jin B., Liang Y., Han Z., Hiraga M., Ohkura K. A hierarchical training method of generating collective foraging behavior for a robotic swarm. *Artificial Life and Robotics*. 2022. vol. 27. pp. 137–141.
24. Arora S., Doshi P. A survey of inverse reinforcement learning: Challenges, methods and progress. *Artificial Intelligence*. 2021. vol. 297. DOI: 10.1016/j.artint.2021.103500.
25. Pinsler R., Maag M., Arenz O., Neumann G. Inverse reinforcement learning of bird flocking behavior. *ICRA Swarms Workshop*. 2018.
26. Chen M., Zhang P. Area Coverage for Swarm Robots Via Inverse Reinforcement Learning. 2023. 9 p.
27. Gharbi I., Kuckling J., Ramos D., Birattari M. Show me what you want: Inverse reinforcement learning to automatically design robot swarms by demonstration. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023. pp. 5063–5070.
28. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*. 2017.

Alaa Iskandar — Ph.D. student, Faculty of mechanical engineering and informatics (istvan salyi doctoral school of mechanical engineering sciences – mathematic institute), University of Miskolc. Research interests: reinforcement learning for swarm robotics, Navigation, and foraging behaviors. The number of publications — 3. iskandar.alaa@student.uni-miskolc.hu; Egyetemvaros, 3515, Miskolc city, Hungary; office phone: +(36)46-565-111.

Ali Hammoud — Ph.D. student, Faculty of applied informatics, Federal State Budgetary Educational Institution of Higher Education “Kuban State Agrarian University named after I.T. Trubilin”. Research interests: multi-agent systems and decision-making. The number of publications — 2. ali-hammoud@mail.ru; 13, Kalinina St., 350044, Krasnodar, Russia; office phone: +7(861)221-5942.

Béla Kovács — Ph.D., Dr.Sci., Associate professor, Faculty of mechanical engineering and information, institute of mathematics department of analysis, University of Miskolc. Research interests: mechanical engineering and differential equations. The number of publications — 99. matmn@uni-miskolc.hu; Egyetemvaros, 3515, Miskolc city, Hungary; office phone: +(36)46-565-111.

А. ИСКАНДАР, А. ХАММУД, Б. КОВАЧ
**СКРЫТЫЙ СМЫСЛ: ДЕКОДИРОВКА РОЕВОГО ПОВЕДЕНИЯ
РОБОТОВ С ПОМОЩЬЮ ГЛУБОКОГО ОБРАТНОГО
ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ**

Искандар А., Хаммуд А., Ковач Б. **Скрытый смысл: декодировка роевого поведения роботов с помощью глубокого обратного обучения с подкреплением.**

Аннотация. Использование обучения с подкреплением для создания коллективного поведения роевых роботов является распространенным подходом. Тем не менее, формулирование соответствующей функции вознаграждения, которая соответствовала бы конкретным целям, остается серьезной проблемой, особенно по мере увеличения сложности задач. В этой статье мы разрабатываем модель глубокого обратного обучения с подкреплением, чтобы раскрыть структуры вознаграждения, которые помогают автономным роботам выполнять задачи посредством демонстраций. Модели глубокого обратного обучения с подкреплением особенно хорошо подходят для сложных и динамичных сред, где может быть сложно указать заранее определенные функции вознаграждения. Наша модель может генерировать различное коллективное поведение в соответствии с требуемыми целями и эффективно справляется с непрерывными пространствами состояний и действий, обеспечивая детальное восстановление структур вознаграждения. Мы протестировали модель с помощью роботов E-puck в симуляторе Webots для решения двух задач: поиска рассредоточенных коробок и навигации к заданной позиции. Получение вознаграждения зависит от демонстраций, собранных интеллектуальным предварительно обученным роем, использующим обучение с подкреплением в качестве эксперта. Результаты показывают успешное получение вознаграждения как в сегментированной, так и в непрерывной демонстрации двух типов поведения — поиска и навигации. Наблюдая за изученным поведением роя экспертом и предложенной моделью, можно заметить, что модель не просто копирует поведение эксперта, но генерирует свои собственные стратегии для достижения целей системы.

Ключевые слова: обратное обучение с подкреплением, функция вознаграждения, демонстрации, поисковое поведение, навигационное поведение.

Литература

1. Shahzad M., Saeed Z., Akhtar A., Munawar H., Yousaf M., Baloach N., Hussain F. A review of swarm robotics in a nutshell. *Drones*. 2023. vol. 7. no. 4.
2. Schranz M., Umlauf M., Sende M., Elmenreich W. Swarm robotic behaviors and current applications. *Frontiers in Robotics and AI*. 2020. vol. 7.
3. Cheraghi A., Shahzad S., Graffi K. Past, present, and future of swarm robotics. In *Intelligent Systems and Applications: Proceedings of the 2021 Intelligent Systems Conference (IntelliSys)*. 2022. vol. 3. pp. 190–233.
4. Brambilla M., Ferrante E., Birattari M., Dorigo M. Swarm robotics: a review from the swarm engineering perspective. *Swarm Intelligence*. 2013. vol. 7. pp. 1–41.
5. Nauta J., Van Havermaet S., Simoens P., Khaluf Y. Enhanced foraging in robot swarms using collective lévy walks. In *24th European Conference on Artificial Intelligence (ECAI)*. 2020. pp. 171–178.

6. Misir O., Gokrem L. Flocking-based self-organized aggregation behavior method for swarm robotics. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*. 2021. vol. 45. no. 4. pp. 1427–1444.
7. Sadeghi A., Raoufi M., Turgut A. A self-adaptive landmark-based aggregation method for robot swarms. *Adaptive Behavior*. 2022. vol. 30. no. 3. pp. 223–236.
8. Lu Q., Hecker J., Moses M. The MPFA: A multiple-place foraging algorithm for biologically-inspired robot swarms. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016. pp. 3815–3821.
9. Lopes Y., Trenkwalder S., Leal A., Dodd T., Groß R. Supervisory control theory applied to swarm robotics. *Swarm Intelligence*. 2016. vol. 10. pp. 65–97.
10. Hamann H. *Swarm robotics: A formal approach*. Cham: Springer, 2018. 210 p.
11. Berlinger F., Gauci M., Nagpal R. Implicit coordination for 3D underwater collective behaviors in a fish-inspired robot swarm. *Science Robotics*. 2021. vol. 6. no. 50. DOI: 10.1126/scirobotics.abd8668.
12. Zhang J., Lu Y., Che L., Zhou M. Moving-distance-minimized PSO for mobile robot swarm. *IEEE Transactions on Cybernetics*. 2021. vol. 52. no. 9. pp. 9871–9881.
13. Parhi D., Sahu C., Kumar P. Navigation of multiple humanoid robots using hybrid adaptive swarm-adaptive ant colony optimisation technique. *Computer Animation and Virtual Worlds*. 2018. vol. 29. no. 2. DOI: 10.1002/cav.1802.
14. Jiang L., Mo H., Tian P. An adaptive decentralized control strategy for deployment and aggregation of swarm robots based on bacterial chemotaxis. *Applied Intelligence*. 2023. vol. 53. no. 10. pp. 13018–13036.
15. Hu C., Arvin F., Bellotto N., Yue S., Li H. Swarm neuro-robots with the bio-inspired environmental perception. *Frontiers in Neurorobotics*. 2024. vol. 18.
16. Hasselmann K., Ligot A., Birattari M. Automatic modular design of robot swarms based on repertoires of behaviors generated via novelty search. *Swarm and Evolutionary Computation*. 2023. vol. 83.
17. Birattari M., Ligot A., Francesca G. AutoMoDe: a modular approach to the automatic off-line design and fine-tuning of control software for robot swarms. *Automated Design of Machine Learning and Search Algorithms*. 2021. pp. 73–90.
18. Stolfi D., Danoy G. Evolutionary swarm formation: From simulations to real world robots. *Engineering Applications of Artificial Intelligence*. 2024. vol. 128. DOI: 10.1016/j.engappai.2023.107501.
19. Blais M., Akhloufi M. Reinforcement learning for swarm robotics: An overview of applications, algorithms and simulators. *Cognitive Robotics*. 2023. vol. 3. pp. 226–256. DOI: 10.1016/j.cogr.2023.07.004.
20. Sutton R., Barto A. *Reinforcement learning: An introduction*. The MIT press, 2018. 552 p.
21. Iskandar A., Rostum H., Kovacs B. Using Deep Reinforcement Learning to Solve a Navigation Problem for a Swarm Robotics System. *Proceedings of the 24th International Carpathian Control Conference (ICCC)*. IEEE, 2023. pp. 185–189.
22. Wei Y., Nie X., Hiraga M., Ohkura K., Car Z. Developing end to end control policies for robotic swarms using deep Q-learning. *Journal of Advanced Computational Intelligence and Intelligent Informatics*. 2019. vol. 23. no. 5. pp. 920–927.
23. Jin B., Liang Y., Han Z., Hiraga M., Ohkura K. A hierarchical training method of generating collective foraging behavior for a robotic swarm. *Artificial Life and Robotics*. 2022. vol. 27. pp. 137–141.
24. Arora S., Doshi P. A survey of inverse reinforcement learning: Challenges, methods and progress. *Artificial Intelligence*. 2021. vol. 297. DOI: 10.1016/j.artint.2021.103500.

25. Pinsler R., Maag M., Arenz O., Neumann G. Inverse reinforcement learning of bird flocking behavior. ICRA Swarms Workshop. 2018.
26. Chen M., Zhang P. Area Coverage for Swarm Robots Via Inverse Reinforcement Learning. 2023. 9 p.
27. Gharbi I., Kuckling J., Ramos D., Birattari M. Show me what you want: Inverse reinforcement learning to automatically design robot swarms by demonstration. In 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023. pp. 5063–5070.
28. Schulman J., Wolski F., Dhariwal P., Radford A., Klimov O. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. 2017.

Искандар Алаа — аспирант, факультет машиностроения и информатики (докторантура машиностроительных наук иштвана сали – математический институт), Университет Мишкольца. Область научных интересов: обучение с подкреплением для роевой робототехники, навигации и поиска пищи. Число научных публикаций — 3. iskandar.alaa@student.uni-miskolc.hu; Эгьетемварош, 3515, Мишкольц, Венгрия; р.т.: +(36)46-565-111.

Хаммуд Али — аспирант, факультет прикладной информатики, Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный аграрный университет имени И.Т. Трубилина». Область научных интересов: мультиагентные системы и принятие решений. Число научных публикаций — 2. ali-hammoud@mail.ru; улица Калинина, 13, 350044, Краснодар, Россия; р.т.: +7(861)221-5942.

Ковач Бела — Ph.D., Dr.Sci., доцент, факультет машиностроения и информатики института математики, кафедра анализа, Университет Мишкольца. Область научных интересов: машиностроение и дифференциальные уравнения. Число научных публикаций — 99. matmn@uni-miskolc.hu; Эгьетемварош, 3515, Мишкольц, Венгрия; р.т.: +(36)46-565-111.

А.А. Шилин, Х. ФАМ ЧОНГ, В. НГУЕН ВОНГ
**СИНТЕЗ FUZZY-РЕГУЛЯТОРА ОБЪЕКТОМ ВТОРОГО
ПОРЯДКА С ЗАПАЗДЫВАНИЕМ**

Шилин А.А., Фам Чонг Х., Нгуен Вонг В. Синтез Fuzzy-регулятора объектом второго порядка с запаздыванием.

Аннотация. В работе предлагается метод реализации синтеза оптимального управления динамическим объектом второго порядка с запаздыванием на базе Fuzzy-контроллера. Применена идея построения фазовой поверхности, совмещающей оптимальное релейное управление в удалении от области равновесного состояния и линейное управление в самой области. Такой подход позволяет избежать автоколебаний в установившемся режиме, при этом сохранив свойства оптимального управления по быстродействию. Траектория переключения в фазовом пространстве, соответствующая решению задачи оптимального управления согласно принципу Максимума, определяется методом обратного по времени вычисления разностного уравнения объекта второго порядка. Для определения области вокруг точки равновесного состояния, где применяется линейный регулятор, предложено использовать результаты моделирования движения точки в фазовом пространстве при оптимальном управлении для объекта с запаздыванием в режиме автоколебаний. Данная область представлена эллипсом, описывающим движение в фазовом пространстве при автоколебательном режиме. Для дальнейшего исключения автоколебаний согласно известным методам субоптимального управления в этой области применён линейный регулятор, настроенный средствами решения вариационной задачи оптимального управления. Предложено использовать инструментарий для синтеза Fuzzy-регулятора, где поверхность переключения и вычисления значения управления может задаваться произвольно. В результате получена переменная структура регулятора для совмещения этих двух подходов. Сформированная модель Fuzzy-регулятора представлена стандартной FLS-структурой, которая была реализована на языке Python во встраиваемом компьютере Orange Pi. Для подключения к действующему объекту управления использован промышленный контроллер FX3U-24MR, связанный с компьютером по сети ModBus. Приведены испытания на эксплуатируемом объекте управления температурой горячего водоснабжения, который максимально близко соответствует исследуемой модели объекта. Метод, идея и результаты, полученные в работе, можно применять и исследовать в синтезе управления динамическими объектами в скользящем режиме для решения актуальных задач, связанных с исключением нежелательного chattering-эффекта.

Ключевые слова: оптимальное управление, Fuzzy-регулятор, фазовая траектория, поверхность переключения, ПД-регулятор.

1. Введение. В течение последних двух десятилетий теория нечёткого управления приобрела отдельное научное направление, которому посвящены целые научные издания, имеющие большую популярность в научной среде [1 – 3]. Одним из важных преимуществ направления является возможность использования экспертных оценок на лингвистическом уровне, понятных специалистам-технологам. При этом технологам необязательно глубоко знать теорию автоматического управления, математическое моделирование и другие наукоемкие

направления. В результате сам подход к созданию систем управления [4, 5] позволяет использовать интуитивные знания эксперта, представленные на достаточно простом и понятном языке.

Процедура фазификации и дефазификации также может быть реализована несколькими потоками с использованием конвейерных вычислений в Fuzzy-контроллерах. Поэтому, несмотря на большое количество вычислений, сохраняется возможность использования полученных методов в быстродействующих системах с временем принятия решения меньшим миллисекунды. Примером может послужить система управления летательным аппаратом, где время дискретизации в контурах управления скоростью двигателей вращения составляет единицы миллисекунд. Данное преимущество чаще отмечается в задачах распознавания [6, 7 – 9] изображений, где требуется большое количество вычислений.

Спектр задач, где используется теория нечёткого управления, постоянно расширяется благодаря простоте использования методов нечёткой логики. Этому способствует ряд причин:

- простое и понятное представление решений в виде доступных восприятию специалистом признаков и выводов;
- единый и типовой подход к синтезу системы управления;
- полученное решение реализуется в рамках единого стандарта реализации систем нечёткого вывода [10];
- имеются библиотеки с открытым исходным кодом для использования метода в программируемых логических контроллерах (ПЛК) [11];
- результат синтеза нечёткой модели имеет строго формализованную структуру для реализации на различных платформах, включая ПЛК и математические пакеты программного обеспечения MatLab или SciLab.

Для задач, выполняемых аналитическими методами, теория нечёткого управления может рассматриваться как один из очень эффективных инструментов аппроксимации нелинейных динамических моделей в управлении и принятии решений. Возможность построения аппроксимаций методами теории нечётких множеств уместно, поскольку имеются работы, в которых доказана возможность [12] и пригодность к практическому применению [13]. Также в качестве способов оценки модели, построенной на нечёткой логике, в математических пакетах предлагается графическое представление в двух и трёх мерных пространствах нечёткого вывода по входным параметрам.

Для управления динамическими объектами также имеются примеры модернизации классических ПИД регуляторов путём придания необходимых нелинейных свойств [14] для достижения лучших показателей качества работы замкнутых систем. В роли выходных переменных нечёткого вывода могут быть как переменные состояния ПИД регулятора [15, 16], так и его параметры [17]. Выявленные преимущества, как правило, определены для заданных объектов в определённых режимах. Так например, в работе [18] в качестве преимущества приведён показатель для системы управления объектом поддержания температуры с элементом Пельте. Сложность практического использования для других систем ограничивается объектом, к которому имеются качественные экспертные оценки, что затрудняет распространение метода на другие объекты.

Касательно методов управления динамическими объектами второго порядка сложилась практика формирования нелинейной поверхности, определяющей сигнал управления в фазовом пространстве [19]. Для этих целей в программном обеспечении MatLab и SciLab имеются инструменты для оценки соответствия поверхностей выбранных множеств входных переменных состояния выходным [20]. Следует отметить, что подобными средствами оценки пользуются в работах, связанных с развитием оптимального управления по быстродействию [21], а также робастного управления в скользящем режиме [22]. Исследование в этом направлении позволяет получить методы, обладающие полезными свойствами относительно быстродействия и робастности [23]. Эти методы, в отличие от построения экспертных оценок для Fuzzy-регулятора, основаны на аналитических методах управления использующих принцип Максимума [24] и теории скользящего управления [25].

Вопросам синтеза управления, где результатом являются семейства фазовых траекторий, посвящено достаточно много работ. В частности в работе [26] используется хорошо изученный подход анализа поведения замкнутого контура на фазовых портретах движения динамической системы. Проанализировав траектории переключения в скользящих режимах на фазовой плоскости, можно достаточно наглядно представить семейства траекторий переключения с лучшим быстродействием или с широкими областями изменения параметров объекта робастных режимов. У авторов данной работы имеются исследования по синтезу оптимальных регуляторов, где на основании анализа фазовых поверхностей синтезируется регулятор, формирующий программное движение [27].

Стандартная процедура синтеза оптимального управления по быстродействию согласно принципа Максимиума включает в себя вычисления функций Гамильтона, где результатом синтеза является переходный процесс с заданными моментами переключения управляющего воздействия [28]. Однако это решение является частным, поскольку рассматриваются начальные условия не в общем виде. Для формирования алгоритма управления для произвольных начальных условий требуется исследование нескольких фазовых траекторий с целью создания обобщённого алгоритма [29]. Сам алгоритм вычисления управления может включать как нелинейные функции, так и условия, меняющие структуру регулятора.

С развитием теории оптимального управления, использующий принцип максимума [30], возникают нестандартные численные решения оптимального управления, где эффективность и показатели быстродействия также анализируются на фазовых траекториях.

Принцип Максимиума и методы синтеза управления в скользящем режиме имеют общий инструмент анализа фазовых траекторий, где можно показать, как подбор траектории переключения [31] позволяет получить несколько режимов управления, включая оптимальный по быстродействию с конечным числом переключений.

Однако большинство удачных с точки зрения реализации решений связаны с динамическими объектами второго порядка. Применение этих решений для объектов более высокого порядка связано с проблемой возникновения нежелательного chattering-эффекта. Этой проблеме посвящено много исследований, о которой можно прочесть в работе [32].

В прикладных задачах часто результатом является семейство выбранных фазовых траекторий [33] с целью их дальнейшего применения в алгоритме, программном коде или математической модели. Для реализации эксперимента требуется обеспечить работу ПЛК согласно полученным траекториям на штатном языке программирования выбранного контроллера.

Если воспользоваться способностью Fuzzy-регулятора реализовывать аппроксимацию произвольных многомерных поверхностей, то можно применить результаты нескольких исследований, где алгоритм вычисления содержит обобщённую поверхность переключения в фазовой плоскости. Такая реализация управления будет представлена стандартной FLS-структурой, которую можно использовать в промышленных ПЛК.

Данная работа посвящена исследованию метода построения экспертных оценок Fuzzy-регулятора, построенных на известных и хорошо зарекомендовавших себя решениях в области оптимального

управления по быстродействию для стационарного динамического объекта второго порядка с астатизмом и запаздыванием. Линейная или линеаризованная математическая модель такого объекта может быть представлена выражением

$$W_o(s) = \frac{1}{T_a s} \cdot \frac{1}{(1 + T_{ob} s)} e^{-\tau s}, \quad (1)$$

где: T_a – время интегрирования части объекта, обладающей астатизмом; T_{ob} – время инерции объекта; τ – обобщённое время запаздывания, включающая в себя инерционные свойства датчиков, преобразователей и фильтров. На сигнал управление наложено ограничение

$$|u| \leq 1, 0. \quad (2)$$

Поскольку интегральная часть контура ПИД-регулятора включена в состав объекта управления и определена временем T_a , синтезу подлежит изменяемая часть регулятора, представленная пропорциональной и дифференциальной составляющей (ПД) регулятора.

2. Синтез. Поскольку объект (1) не является объектом второго порядка, для которых задача оптимального управления по быстродействию с ограничением на значение управления решена, то это несоответствие, как правило, приводит к автоколебаниям в области равновесного состояния [34]. Для систем управления в скользящем режиме наличие обобщённого запаздывания приводит к нежелательному chattering-эффекту [35], который также возникает в области равновесного состояния замкнутой системы управления. При этом линейный ПД-регулятор возможно настроить известными методами так, чтобы в области равновесного состояния обеспечивалась асимптотическая устойчивость в цепи управления (1). Так как свойства линейного регулятора обеспечивают асимптотическую устойчивость в области равновесного состояния, можно применить известный метод использования переменной структуры регулятора [36, 37], где вне области равновесного состояния можно исследовать работу быстродействующего регулятора на основе кривой переключений в фазовом пространстве рассогласования и его производной. В упомянутых работах подробно представлены хорошо зарекомендовавшие себя методы совмещения нескольких принципов управления для достижения быстродействия и асимптотической устойчивости в области равновесного состояния.

Прикладная часть, обеспечивающая реализацию метода управления, ориентирована на доступные 30 лет назад способы, ограниченные технологическими возможностями того времени. Учитывая современные возможности реализации сложных методов управления, можно выделить актуальное направление исследований и разработки более эффективных и простых в реализации методов. Анализ возможностей инструментов в синтезе Fuzzy-управления в рамках математического пакета SciLab для решения подобных задач позволяет получить эффективные методы и алгоритмы реализации управления. В работе предлагается процедура синтеза метода управления, построенная на основе двух поверхностей в фазовом пространстве, где граничная область определяется из результата моделирования автоколебательного режима.

Процедура синтеза заключается в следующих этапах:

- решение задачи оптимального управления объектом (1) с учётом ограничений (2) LQR-методом минимизации функционала;
- построение кривой переключения для области, не входящей в окрестность равновесного состояния, численным методом обратного движения по времени, предложенным А.А. Фельдбаумом;
- определение области равновесного состояния, где обеспечивается (2);
- обобщение полученных решений для формирования совмещенной поверхности средствами аппроксимации в среде Fuzzy-tools;
- формирование экспертных оценок фазификации и дефазификации Fuzzy-регулятора;
- реализация полученной FLS-структуры на базе ПЛК.

3. Оптимальный LQR. Для решения задачи поиска оптимальных параметров воспользуемся численными методами доступными в математических пакетах SciLab или MatLab. Перечислим основные используемые функции на примере среды SciLab.

– Функция $[f, x_{opt}] = optim(cst, p_0)$ выполняет вариационную задачу поиска минимума функционала f , подбирая вектор параметров p . Начальные значения этого вектора представлены переменной p_0 . Функция возвращает оптимальный вектор параметров x_{opt} и значение соответствующего ему функционала. Описание объекта включено в функцию cst .

– Функция $[f, g, ind] = cst(p, ind)$ с заданной структурой определяется пользователем и включает процедуру вычисления функционала f для текущего вектора параметров p . Кроме функционала,

необходимо возвращать градиент от текущего изменения параметров, который вычисляется средствами математического пакета с помощью встроеной функции $g = \text{numderivative}(fkn, p)$.

– Функция $J = fkn(p)$ вычисления значения функционала для текущего вектора параметров p определяется пользователем. В данном случае значение функционала формируется в виде суммы среднеквадратических отклонений переменных состояния со своими весовыми коэффициентами согласно LQR-метода оптимизации. Сумма определяется по всему переходному процессу в решении однородного дифференциального уравнения. По этой причине начальный вектор параметров должен соответствовать устойчивой замкнутой системе.

– Функция $y = \text{ode}(\text{"rk"}, x_0, t_0, t, \text{syst1})$ решения однородного дифференциального уравнения для которой следует указать метод численного решения уравнения ("rk" – Рунге-Кутта), исходные вектор переменных состояния и время, вектор значений времени переходного процесса, функция syst1 , описывающая замкнутую динамическую систему в матричной форме. Функция возвращает набор векторов переменных состояния в виде матрицы.

Для обеспечения вычислений математическую модель замкнутой системы управления для объекта (1) и ПД-регулятора представим в матричной форме. Для этого объект представим в виде

$$W_o(s) = \frac{1}{T_a s} \frac{1}{(1 + s(\tau/3))^3} \frac{1}{(1 + T_{obs})}, \quad (3)$$

где звено запаздывания представим в виде аппроксимации тремя аperiодическими звеньями. Выражение (3) разделим на переменные состояния

$$X_5 = \frac{U}{(1 + \frac{\tau}{3}s)}, \underbrace{X_{3+i}}_{(i \in 0,1)} = \frac{X_{4+i}}{(1 + \frac{\tau}{3}s)}, X_2 = \frac{X_3}{T_a(1 + T_{obs})}, X_1 = \frac{X_2}{s}, \quad (4)$$

где изображения переменных состояния X_1 и X_2 соответствуют значению рассогласования и его производной в точке равновесного состояния. В этом случае значение выхода линейного регулятора представлено выражением

$$U = -(K_p X_1 + T_d X_2). \quad (5)$$

В результате после обратного преобразования Лапласа (4) и (5) замкнутая система управления будет представлена в матричной форме

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} \\ a_{51} & a_{52} & 0 & 0 & a_{55} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}, \quad (6)$$

где ненулевые значения матрицы A вычисляются выражениями

$$\begin{aligned} a_{12} = 1; a_{33} = a_{44} = a_{55} = -3\tau^{-1}; a_{34} = a_{45} = 3\tau^{-1}; \\ a_{51} = -3K_p\tau^{-1}; a_{52} = -3T_d\tau^{-1}; a_{22} = -T_{ob}^{-1}; a_{23} = (T_{ob}T_a)^{-1}. \end{aligned} \quad (7)$$

Для нахождения оптимальных параметров регулятора K_p и T_d определим вариационную задачу минимизации квадратичного функционала J_q и функционала J_{qt} , соответствующего оптимальному управлению по быстродействию, которые представлены выражениями

$$\begin{aligned} J_q(K_p, T_d) &= \int_{t=0}^{t_{max}} (x_1^2(t) + K_u x_5^2(t)) dt, \\ J_{qt}(K_p, T_d) &= \int_{t=0}^{t_{max}} (x_1^2(t) + K_u x_5^2(t)) \cdot t dt, \end{aligned} \quad (8)$$

где выбраны для минимизации значения выхода объекта x_1 и сигнала управления x_5 с весовым коэффициентом K_u . Этот коэффициент подбирается для выполнения условия (2). Максимальное время интегрирования t_{max} должно быть на столько большим, чтобы все динамические процессы в замкнутой системе закончились. Чрезмерное увеличение этого времени увеличивает время вычисления. Следует отметить, что x_5 соответствует управлению (5) по амплитуде и форме, поскольку между ними имеется незначительная задержка равная значению $(\tau / 3)$. Однако на решение вариационной задачи это практически не влияет. Скрипт реализации вариационной задачи опубликован в [38, 39]. Результаты вычисления для объекта с двумя наборами параметров

$$\begin{aligned} n_\tau = 3, \quad T_a = 0,3713 \text{ с}, T_{ob} = 43,10 \text{ с}, \tau = 7,880 \text{ с}; \\ n_\tau = 1, \quad T_a = 0,3729 \text{ с}, T_{ob} = 41,24 \text{ с}, \tau = 8,706 \text{ с}; \end{aligned} \quad (9)$$

приведены в таблице 1. Параметры объекта взяты из результатов эксперимента, описанного в разделе – Реализация, на действующем объекте управления температурой воды горячего водоснабжения, где звено запаздывание аппроксимировано тремя звеньями ($n_\tau = 3$) и одним звеном ($n_\tau = 1$). В результате, рассматривая варианты с двумя функционалами $J_q()$ и $J_{qt}()$, с порядком аппроксимации $n_\tau \in (1; 3)$ и начальными условиями ($x_1 = -10$) и ($x_2 = 0, 5$) получены результаты вычислений для восьми случаев. На рисунке 1 приведены переходные процессы для этих случаев.

Таблица 1. Оптимальные параметры линейного регулятора

N	Располагание	n_τ	$J(8)$	K_p	T_d, c	K_u	t_{calc}, c
1	$x_1 = -10; x_2 = 0$	3	$J_q = 2686$	0.0610	1.35	120	30.7
2	$x_1 = -10; x_2 = 0$	1	$J_q = 2989$	0.0486	1.15	200	37.3
3	$x_1 = 0; x_2 = 0.5$	3	$J_q = 635$	0.0604	1.84	30	12.3
4	$x_1 = 0; x_2 = 0.5$	1	$J_q = 770$	0.0593	1.68	50	20.0
5	$x_1 = -10; x_2 = 0$	3	$J_{qt} = 34656$	0.0664	1.47	150	26.9
6	$x_1 = -10; x_2 = 0$	1	$J_{qt} = 42222$	0.0538	1.27	250	36.1
7	$x_1 = 0; x_2 = 0.5$	3	$J_{qt} = 12877$	0.0633	1.77	70	40.2
8	$x_1 = 0; x_2 = 0.5$	1	$J_{qt} = 14234$	0.0565	1.65	70	32.9

Начальные условия представлены двумя вариантами:

– изменение задающего значения когда целевая переменная не равна нулю $x_1 = -10$;

– изменение объёма потребляемой воды приводит к смещению производной от целевой переменной $x_2 = 0.5$.

Весовой коэффициент K_u переменной состояния, соответствующая сигналу управления, подбирается итерационным способом для обеспечения условия (2). В таблице также приведено время вычисления вариационной задачи t_{calc} для одного ядра процессора с частотой 1,7 ГГц.

Как видно из таблицы, решение вариационной задачи поиска оптимальных параметров незначительно зависит от начальных условий в решении однородного дифференциального уравнения. Во всех случаях значение функционала меньше при использовании аппроксимации звена запаздывания тремя звеньями, а время вычисления меньше при аппроксимации одним инерционным звеном. Разницу в быстродействии по значению функционалов $J_q()$ и $J_{qt}()$ оценивать некорректно, поэтому на рисунке 1 приведены переходные процессы. Можно отметить, что они не симметричны на втором и третьем колебании для случаев, когда начальные условия соответствуют $x_2 = 0.5$, но при этом амплитуда

отклонения рассогласования меньше чем для настройки при начальных условиях $x_1 = -10$.

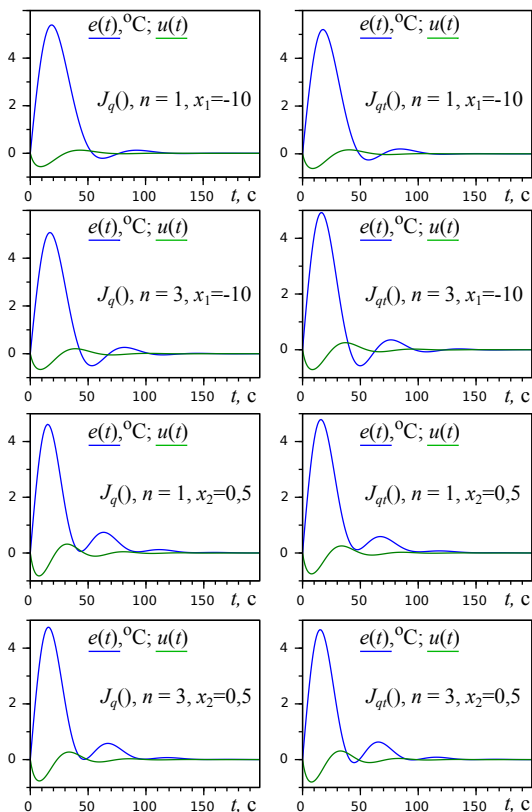


Рис. 1. Переходные процессы настроенной системы при изменении $x_2 = 0.5$

Колебательность соответствующих процессов для функционала $J_q()$ незначительно меньше, чем при использовании функционала $J_{qt}()$. В дальнейшем при моделировании работы замкнутой системы и при реализации экспериментов на объекте для различных наборов параметров ПД-регулятора в рамках таблицы 1 существенную разницу определить достаточно сложно из-за схожести процессов. Поскольку все наборы параметров обеспечивают приемлемое качество регулирования в области точки равновесного состояния и можно выбрать любой. Для дальнейшей настройки выбран случай, где переходный процесс наиболее симметричен

с минимальной калембательностью. При таком выборе используется обычный квадратичный функционал для объекта, где звено запаздывания аппроксимирована одним инерционным звеном для вариационной задачи с начальными условиями, соответствующими $x_1 = -10$.

4. Траектория. Процедура применения принципа Максимиума для объектов с заданным конечным порядком дифференциального уравнения объекта решена и обобщена. Поэтому в данном случае достаточно воспользоваться теоремой и числе переключений в оптимальном процессе по быстройдействию, которая доказана Фельдбаумом А.А. в работе [40].

Сама теорема утверждает следующее. Если объект управление представлен дифференциальным уравнением в форме Коши или в матричной форме, которое имеет n порядок, то при ограничении (2) управление представлена кусочно-постоянной функций в заданном интервале времени ($0 < t < t_1$) с точками переключения, число которых не превышает $(n - 1)$. Для частного случая, когда объект представлен дифференциальным уравнением второго порядка, для произвольно заданного состояния объекта момент переключения будет один и оптимальное управление в интервале ($0 < t < t_1$) будет соответствовать двум максимально возможным значениям из условия (2).

Момент времени переключения сигнала управления определяется начальными значениями переменных состояния объекта управления. Для объекта второго порядка эти переменные состояния соответствуют значению выхода объекта и его производной. Если обобщить результаты решения для разных начальных условий, то точки переключения можно представить в виде траектории в фазовом пространстве. Как отмечает в своей работе Фельдбаум А.А., для объектов второго порядка данную траекторию можно получить методом вычисления дифференциального уравнения в обратном по времени направлении. Для этого можно применить дискретное преобразования, где, в отличие от стандартного разностного уравнения, за знак равенства выносятся предыдущее значение выхода, а не последующее.

Таким образом, траекторию переключения в фазовом пространстве построим для объекта второго порядка с переменными состояния x_1 и x_2 , где $u = x_3$. Предыдущее значение переменной состояния на основании (4) можно получить из разностных уравнений

$$\begin{aligned} x_2(t - \Delta t) &= x_2(t) - (u(t) \frac{1}{T_a} - x_2(t)) \cdot \frac{\Delta t}{T_{obz}}, \\ x_1(t - \Delta t) &= x_1(t) - x_2(t) \Delta t, \end{aligned} \quad (10)$$

где Δt – время дискретизации. Для учёта времени запаздывания обобщённое время объекта второго порядка рассчитывается с помощью функции

$$T_{obz} = T_{ob} \cdot \left(1 + \frac{k_z T}{T_{ob}}\right). \quad (11)$$

Происхождение коэффициента k_z рассмотрим в процессе анализа фазового портрета. Программа вычисления траектории (рисунок 2(а)) написана на языке Си [38, 41].

После того как получена траектория переключения оптимальное управление по быстродействию для рассматриваемого объекта можно представить в виде поверхности в координатах x_1 и x_2 (рисунок 2(б)). Далее достаточно написать программу вычисления этой поверхности для ПЛК на встроенных языках программирования. Другим, не требующим программирования ПЛК, способом может стать использование Fuzzy-регулятора.

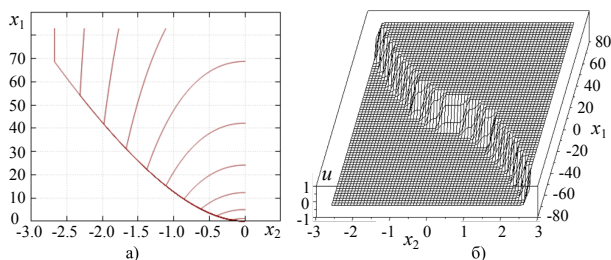


Рис. 2. Траектория переключения и поверхность Fuzzy управления

Настройки Fuzzy-регулятора получим встроенными в SciLab-средствами аппроксимации поверхности управления (рисунок 2(б)) используя скрипт [38, 42]. Для этого в пакете Fuzzy-Tools имеется функция

$$fls = fuzzapp([x_{min} \ x_{max}; y_{min} \ y_{max}], [-11], [n_{part}; n_{part}], F),$$

где указываются диапазон изменения: значений переменных состояния объекта x_1 и x_2 в матрице $[x_{min} \ x_{max}; y_{min} \ y_{max}]$, значений управления и в векторе $[-11]$. В векторе $[n_{part}; n_{part}]$ указывается количество частей в сетке изменения x_1 и x_2 . Функция возвращает готовую FLS-структуру,

которую можно применить в любом контроллере, имеющем поддержку Fuzzy-регулятора. Функции принадлежности по переменным x_1 и x_2 представлены треугольными, количество которых соответствует числу ($n_{part} - 1$) и по краям z и s функции. Все они согласно функции *fuzzapp* равномерно распределены и не несут в себе особой информации, поэтому в статье не приводятся. Также и правила дефазификации не имеет смысла подробно описывать, поскольку они сформированы автоматически для задач аппроксимации. На данном этапе важно, что получена FLS-структура, по которой можно получить трёхмерное изображение (рисунок 1(б)) средствами Fuzzy-Tools.

Работа Fuzzy-регулятора с полученной FLS-структурой исследована на математической модели, построенной в среде X-Cos SciLab. Функциональная схема показана на рисунке 3. Как уже отмечалось, что в модели (1) звено запаздывания может быть заменено как на апериодическое первого порядка ($n = 1$), так и представлена тремя звеньями ($n = 3$). В цепь оценки производной в условиях шумов канала АЦП и ошибок дискретизации включён фильтр с постоянной времени T_f .

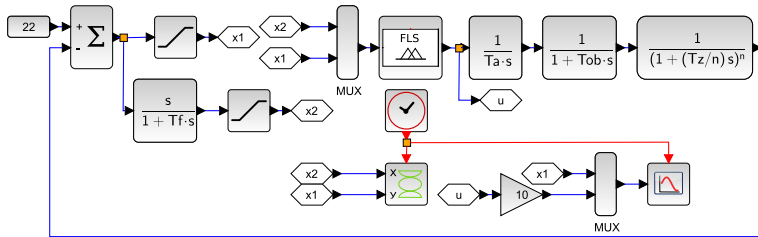


Рис. 3. Функциональная схема математической модели

Диапазон входных переменных Fuzzy-регулятора ограничен соответствующими блоками областью, где определена траектория (рисунок 2). Настройка Fuzzy-регулятора (FLS) производится выбором файла, который сформирован и сохранен встроенными в пакет «Fuzzy Toolbox» функцией $fsl = fuzzapp(\dots, F)$ и функцией $savefsl(fsl, "Tr")$.

Результаты моделирования представлены в виде фазового портрета (рисунок 4(а)) и переходного процесса (рисунок 4(б)). Коэффициент ($k_z = 6,6$) подобран таким образом, чтобы траектория движения переменных x_1 и x_2 при первом приближении к эллипсу входила в него, стремясь к точке равновесного состояния, как показано на рисунке 4(а). Далее наблюдается автоколебательный процесс (рисунок 4(б)) по

траектории эллипса. В процессе исследования фазовый портрет получен для обоих случаев ($n = 1$) и ($n = 3$), при этом результаты (форма, размер эллипса, точка входа в него) полностью совпадали. По этой причине приведён только один набор фазового портрета и переходного процесса.

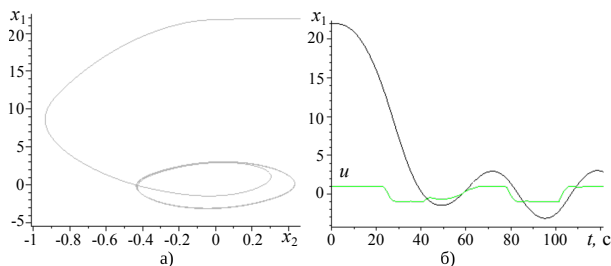


Рис. 4. Фазовый портрет и переходный процесс

В области фазового пространства, представленной эллипсом, невозможно обеспечить используя полученную траекторию переключения как оптимальное управление по быстродействию, так и асимптотическое приближение к точке равновесного состояния объекта. При этом настроенный ПД-регулятор способен обеспечить асимптотическое приближение.

Так как поверхность Fuzzy-регулятора можно строить произвольно, то в окрестности эллипса можно организовать работу ПД-регулятора, который обеспечивает асимптотическое приближение к точке равновесного состояния. Поверхность переключения, обладающая быстродействием от траектории переключения и асимптотической устойчивостью ПД-регулятора представлена на листинге 1. Со всем кодом программы можно ознакомиться в [38, 43].

```

Kp = 0.0486; Td = 1.15;
x2l = 4; x1l = 0.5;
function y = F(x)
    if ( ( x(:,2) .* x(:,2)) / (x2l * x2l) +
        ( x(:,1) .* x(:,1)) / (x1l * x1l) ) < 1.4
    then
        y = Kp * x(:,2) + Td * x(:,1);
        if y > 1 then y = 1; end
        if y < -1 then y = -1; end
    else
        y = sign( x(:,2) - interp ( x(:,1), Xt, Yt, dk ) );
    end
endfunction
    
```

Листинг 1. Функция совмещения поверхностей

Поверхность ПД-регулятора при условии (2) представлена на рисунке 5(а), а совмещённая с траекторией переключения ее поверхность показана на рисунке 5(б).

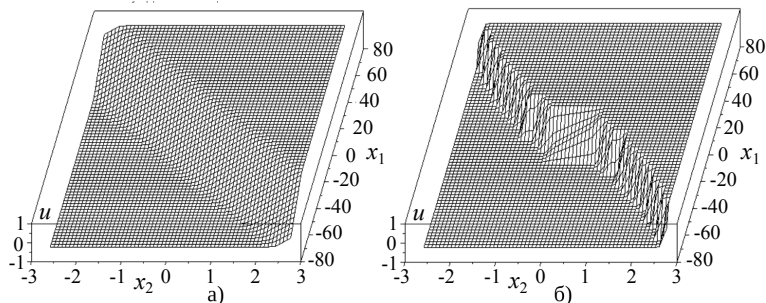


Рис. 5. Поверхности ПД регулятора и Fuzzy-регулятора

Чтобы определить преимущества Fuzzy-регулятора рассмотрим исходное состояние рассогласования ($x_2 = 22; x_1 = 0$) на рисунке 6 соответствует точке А. Траектория движения (A, B_f) и (C_f, D_f) соответствуют оптимальному по быстродействию управлению в условиях ограничения на управление (2). Траектория движения (A, B_{pd}, E_{pd}) проходит в линейной области значений управления, где предполагается более длительный переходный процесс замкнутой системы. В этом можно убедиться, выполнив моделирование замкнутой системы (рисунок 3), используя поверхности управления на рисунке 5 в Fuzzy-регуляторе.

5. Реализация. Для экспериментального подтверждения результатов использования метода выбран объект управления температурой горячего водоснабжения [44, 45], который имеет астатизм 1-го порядка, инерционные свойства и запаздывание. Линеаризованная модель объекта [46] соответствует (1), поэтому для получения переходного процесса используем максимально возможное воздействие на привод клапана в течение 5 секунд. В результате получим переходный процесс для части объекта, где астатизм отсутствует. На рисунке 7 он представлен красной линией.

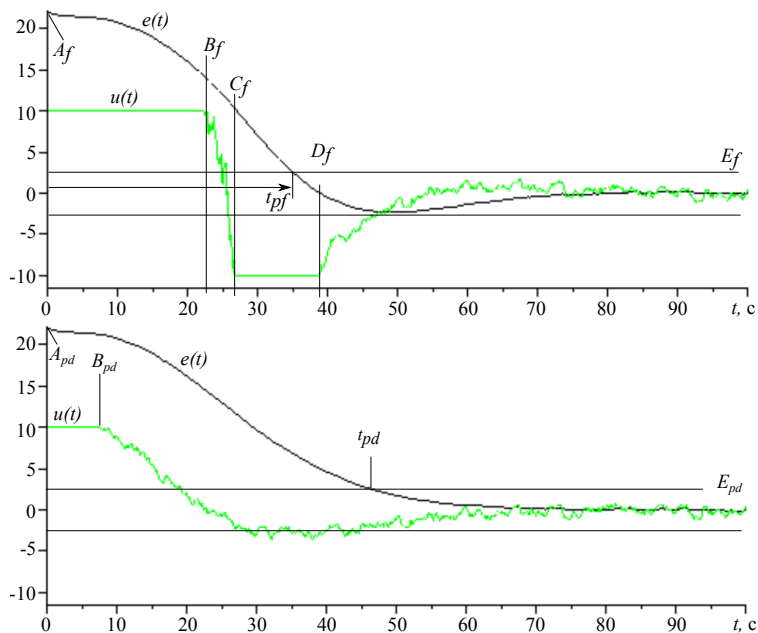


Рис. 6. Переходные процессы ПД-регулятора и Fuzzy-регулятора

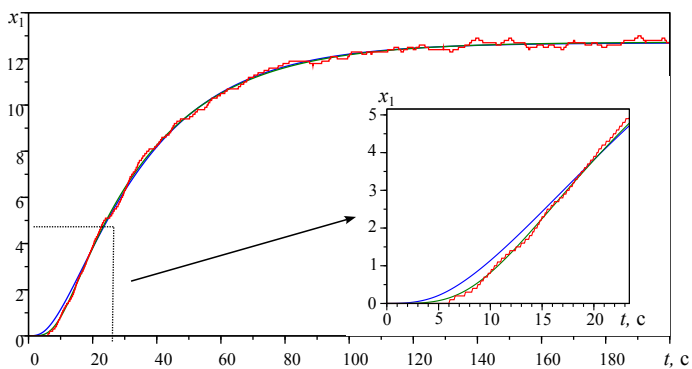


Рис. 7. Переходные процессы ПД-регулятора и Fuzzy-контроллера

Аппроксимацию математической модели будем выполнять для передаточной функции

$$W_m(s) = \frac{1}{T_{am}s} \frac{1}{(1 + T_{ob}s)} \frac{1}{(1 + (\tau/n)s)^n}, \quad (12)$$

используя алгоритм поиска минимума функционала по параметрам (T_{am} , T_{om} , T_{zm}), построенного на разности значений переходных процессов. Чтобы избежать появления нескольких локальных экстремумов, звено запаздывания заменено обычным апериодическим или тремя одинаковыми апериодическими звеньями. Скрипт на языке SciLab доступен в [38, 47], результаты вычисления параметров объекта приведены в выражении (9), на примере которых выполнен синтез Fuzzy-регулятора и уже получены поверхности на рисунок 2 и 5 вместе с его настройками. Синяя линия соответствует аппроксимации одним звеном, а зеленая – аппроксимации тремя звеньями. Поскольку переходный процесс, представленный зеленой линией, лучшим образом повторяет результат эксперимента, то можно сделать вывод, что объект имеет запаздывание. Однако, как показано выше, можно использовать любой из этих вариантов для вычисления оптимальных параметров ПД-регулятора.

Для реализации Fuzzy-регулятора использован одноплатный компьютер Orange Pi 3 с операционной системой Debian, где реализован интерпретатор языка Python библиотеками PyIT2FLS и MinimalModbus. Структурная схема экспериментальной установки показана на рисунке 8.

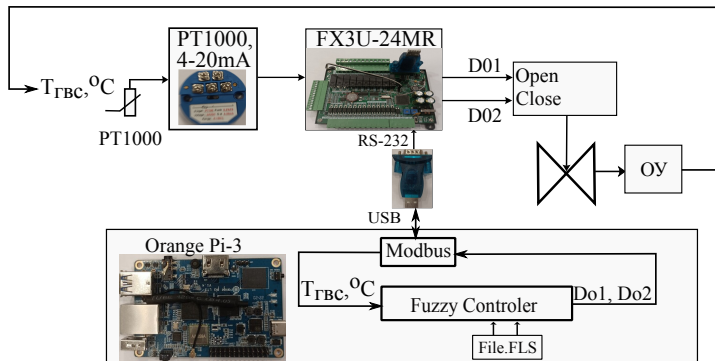


Рис. 8. Структурная схема экспериментальной установки

Для сопряжения с датчиком температуры применён преобразователь сопротивления в ток (pt1000 – 4-20 mA), который подключён к контроллеру FX3U-24MR, где имеется аналоговый вход (4-20 mA) и реализован ModBus-протокол. Результаты эксперимента представлены на рисунке 6, где моменты времени, соответствующие фазовой траектории движения (A , B_f , C_f , D_f) с экстремальными значениями управления отмечены соответствующими символами на оси

времени. Также на рисунке отмечено время переходного процесса из значений которых очевидно преимущества предложенного метода по быстродействию на классическим ПД-регулятором и асимптотическая устойчивость в окрестности точки равновесного состояния.

6. Обсуждение. Свойства Fuzzy-регулятора с использованием оптимальной по быстродействию траектории переключения позволяет несколько улучшить показатели переходного процесса перед Fuzzy-регулятором, где не учитываются ограничения на сигнал управления, поскольку отношение времени ПП по сравнению с ПД-регулятором меньше чем достигнутая в работе [48]. Однако у авторов статьи есть исследования для объекта второго порядка без астатизма, где учитываются ограничения на управляющий сигнал [49] при создании Fuzzy-поверхности. Достаточно полно и адекватно оценить полученные результаты и сравнить с их результатами оказалось достаточно сложно, поскольку там не приведены переходные процессы сигнала управления. Тем не менее, важно отметить, что в отличие от работы, где в области оптимальной траектории используется линейная экстраполяция с более быстродействующими параметрами ПД (Fig. 11 в статье [48]), в нашей работе эта часть представлена траекторией переключения. В результате, в областях, отмеченных на рисунке 5 используется максимально возможный сигнал управления (рисунок 6, $t = 0-23$ с. и $27-40$ с.), что соответствует более быстрому управлению.

Полученный Fuzzy-контроллер включает в себя 32 правила фазификаций по двум переменным состояния и более тысячи логических выводов, полученных в автоматическом режиме, что ограничивает применение метода для недорогих ПЛК и микроконтроллеров. В таблице 2 приведены результаты практического использования некоторых известных подходов, где указаны следующие показатели:

- N_{calc} – количество пар вычислений (умножение, сложение);
- t_{calc} – время вычисления в секундах на один шаг дискретизации;
- σ_d – относительная дисперсия шума и ошибок дискретизации по отношению к ширине допустимого диапазона отклонения целевой величины в установившемся режиме;
- название применённого принципа оптимальности.

Следует заметить, что большое количество вычислений может выполнено за столь малое время при условии использования специализированных Fuzzy – контроллеров с возможностью параллельных вычислений. И если предложенная Fuzzy-поверхность будет создана путём выбора оптимальных правил фазификации и логических выводов, используя различные методы и подходы построения логики нечёткого

вывода, то время вычисления может быть уменьшено до нескольких микросекунд. Оптимизация нечёткого вывода, также может позволить исключить нежелательный сглаженный переход в области переключения исключив плавный переход сигнала управления (рисунок 6, $t = 23-27$ с.). Однако описание этих методов выходит за рамки статьи, и должны рассматриваться в отдельной работе.

Таблица 2. Сравнительные характеристики

Регулятор	N_{calc}	t_{calc}, c	σ_d	Принцип оптимальности
PID	10 - 20	< 0.001	< 0.2	LQR
Slide	10 - 20	< 0.001	< 0.05	Робастные свойства
λ – Slide	20 - 40	< 0.001	< 0.5	Быстродействие
Fuzzy	> 4000	< 0.001	< 0.05	Быстродействие
Optimal Fuzzy	> 40	< 0.00001	< 0.05	Быстродействие

Согласно таблицы более предпочтительным кажется λ – Slide – регулятор, однако при своей устойчивости к шумам в канале измерения [50] обеспечивает оптимальный процесс по быстродействию только для заданных начальных условий переходного процесса.

Реализовать предложенный метод управления можно не используя инструменты нечёткой логики. Более того, возможно написать программу управления, которая достаточно лаконично будет использовать ресурсы микроконтроллера. Однако для такой реализации требуются навыки программирования и знания теории оптимального управления, что может стать препятствием для использования метода в инженерной деятельности. Использование инструментальных средств Fuzzy Toolbox SciLab совместно с разработанными в рамках этой работы скриптами позволяет автоматизировать процесс формирования Fuzzy-контроллера вместе с его структурой и параметрами. Это даёт возможность использования метода в технологических процессах для объектов с линеаризованной математической моделью (1).

7. Заключение. Предложен новый подход для реализации методов построения оптимальных или робастных регуляторов, использующих принцип Максимума или скользящих режимов. В таких методах управление формируется в виде фазовых траекторий. Показано, что реализация управления может быть выполнена Fuzzy-инструментами и средствами теории нечёткой логики. Возможность построения произвольной поверхности в фазовом пространстве сигнала управления, позволила реализовать преимущества линейного регулятора в области равновесного состояния системы управления с преимуществами

быстродействующих или робастных регуляторов с релейным управлением в удалённых от области равновесного состояния точках. Таким образом, подход позволяет реализовать более гибкий метод формирования фазовой траектории управления объединяющий ПД-регулятор и оптимальный с релейным управлением. Для объективного и качественного сравнения работы линейного регулятора и оптимального по быстродействию использован метод минимизации квадратичного функционала. При этом показано, что метод оптимальной настройки ПД-регулятора несущественно влияет на качество управления в области равновесного состояния. Поэтому в предложенном методе реализации оптимального управления допустимо использовать некурсивные процедуры вычисления параметров настройки, что позволит сократить время вычислений.

В работе не рассматривался режим скользящего управления, но сам подход может применён и для исследования регуляторов в скользящем режиме. Для этого достаточно сформировать фазовую траекторию согласно известным методам построения и представить её средствами фазификации и дефазификации в средах SciLab или MatLab.

Предложенный метод реализации управления также может быть полезен для исследователей методов, связанных с управлением динамическими объектами, где результатом является семейство фазовых траекторий управления, составляющее сложную поверхность. Реализация натурального эксперимента с использованием предложенного подхода не составит особого труда при наличии навыков построения модели Fuzzy-поверхностей для встраиваемых компьютеров серии Raspberry, Orange или других аналогов. На практике данный подход целесообразно использовать после оптимизации FLS-модели Fuzzy-контроллера.

Литература

1. Pu Q., Zhu X., Liu J., Cai D., Fu G., Wei D., Sun J., Zhang R. Integrated optimal design of speed profile and fuzzy PID controller for train with multifactor consideration // *IEEE Access*. 2020. vol. 8. pp. 152146–152160.
2. Aruna R., Christa S.T.J. Modeling, system identification and design of fuzzy PID controller for discharge dynamics of metal hydride hydrogen storage bed // *International Journal of Hydrogen Energy*. 2020. vol. 45. no. 7. pp. 4703–4719.
3. Kandiban R., Arulmozhiyal R. Speed control of BLDC motor using adaptive fuzzy PID controller // *Procedia Engineering*. 2012. vol. 38. pp. 306–313.
4. Shi J.Z. A fractional order general type-2 fuzzy PID controller design algorithm // *IEEE Access*. 2020. vol. 8. pp. 52151–52172.
5. Osinski C., Leandro G.V., da Costa Oliveira G.H. Fuzzy PID controller design for LFC in electric power systems // *IEEE Latin America Transactions*. 2019. vol. 17. no. 01. pp. 147–154.

6. Daradkeh Y.I. et al. Development of effective methods for structural image recognition using the principles of data granulation and apparatus of fuzzy logic // *IEEE Access*. 2021. vol. 9. pp. 13417–13428.
7. Bobyr M.V., Yakushev A.S., Kulabukhov S.A., Arkhipov A.E. System of stereovision based on fuzzy-logical method of constructing depth map // 2018 International Russian Automation Conference (RusAutoCon). IEEE, 2018. pp. 1–5.
8. Osinski C., Leandro G.V., da Costa Oliveira G.H. Fuzzy PID controller design for LFC in electric power systems // *IEEE Latin America Transactions*. 2019. vol. 17. no. 01. pp. 147–154.
9. Wang Y., Jin Q., Zhang R. Improved fuzzy PID controller design using predictive functional control structure // *ISA transactions*. 2017. vol. 71. pp. 354–363.
10. Oudah M.K., Sulttan M.Q., Shneen S.W. Fuzzy type 1 PID controllers design for TCP/AQM wireless networks // *Indonesian Journal of Electrical Engineering and Computer Science*. 2021. vol. 21. no. 1. pp. 118–127.
11. Hao Z., Li C., Zhang H., Ren W. Control and realization of intelligent fishing system based on PLC and fuzzy PID algorithm // *Journal of Physics: Conference Series*. IOP Publishing, 2022. vol. 2246. no. 1. DOI: 10.1088/1742-6596/2246/1/012042.
12. Zhou J., Pedrycz W., Gao C., Lai Z., Yue X. Principles for constructing three-way approximations of fuzzy sets: A comparative evaluation based on unsupervised learning // *Fuzzy Sets and Systems*. 2021. vol. 413. pp. 74–98.
13. Lv H., Wang G. Approximations of Fuzzy Numbers by Using r -s Piecewise Linear Fuzzy Numbers Based on Weighted Metric // *Mathematics*. 2022. vol. 10. no. 1.
14. Tang K., Man K.F., Chen G., Kwong S. An optimal fuzzy PID controller // *IEEE transactions on industrial electronics*. 2001. vol. 48. no. 4. pp. 757–765.
15. Bobyr M.V., Emelyanov S.G. A nonlinear method of learning neuro-fuzzy models for dynamic control systems // *Applied soft computing*. 2020. vol. 88. DOI: 10.1016/j.asoc.2019.106030.
16. Hu Y., Yang Y., Li S., Zhou Y. Fuzzy controller design of micro-unmanned helicopter relying on improved genetic optimization algorithm // *Aerospace Science and Technology*. 2020. vol. 98. DOI: 10.1016/j.ast.2020.105685.
17. Mehedi I.M., Shah H.S., Al-Saggaf U.M., Mansouri R., Bettayeb M. Fuzzy PID control for respiratory systems // *Journal of Healthcare Engineering*. 2021. vol. 2021. no. 1. DOI: 10.1155/2021/7118711.
18. Perevoshchikov F.P., Vuong N.V., Shabo A.K., Shilin A.A. Identification of Mathematical Model of Drying Unit // *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2020. vol. 795. no. 1. DOI: 10.1088/1757-899X/795/1/012021.
19. Zhao F., May J.A., Loh S.C. Controller Synthesis and Verification for Nonlinear Systems: A computational approach using phase-space geometric models // *IEEE Control Systems Magazine*, 1998. pp. 198–224.
20. Shakin V.N., Semyonova T.I., Kudryashova A.Y., Frisk V.V. Comparison of computer modeling of rc filter in matlab and scilab environments // 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). IEEE, 2020. pp. 1–5.
21. Phu N.D., Hung N.N. Minimum stability control problem and time-optimal control problem for fuzzy linear control systems // *Fuzzy Sets and Systems*. 2019. vol. 371. pp. 1–24.
22. Zhang J., Shi P., Xia Y. Robust adaptive sliding-mode control for fuzzy systems with mismatched uncertainties // *IEEE Transactions on Fuzzy Systems*. 2010. vol. 18. no. 4. pp. 700–711.

23. Somwanshi D. et al. Comparison of fuzzy-PID and PID controller for speed control of DC motor using LabVIEW // *Procedia Computer Science*. 2019. vol. 152. pp. 252–260.
24. Pesch H.J., Plail M. The maximum principle of optimal control: a history of ingenious ideas and missed opportunities // *Control and Cybernetics*. 2009. vol. 38. no. 4A. pp. 973–995.
25. Utkin V. et al. Conventional and high order sliding mode control // *Journal of the Franklin Institute*. 2020. vol. 357. no. 15. pp. 10244–10261.
26. Хорошавин В.С., Грудинин В.С. Примеры построения квазиоптимальной по быстродействию и энергозатратам замкнутой системы управления электроустановкой // *Известия Тульского государственного университета. Технические науки*. 2023. № 2. С. 449–453.
27. Хорошавин В.С., Грудинин В.С. Оптимальное программное движение с изменяемым временем регулирования // *Радиопромышленность*. 2020. Т. 30. № 3. С. 40–49.
28. Долятовский В.А., Рауль Б. Оптимальное управление динамическим объектом на основе принципа максимума // *Системный анализ в проектировании и управлении*. 2020. Т. 24. № 2. С. 37–45.
29. Костоготов А.А., Костоготов А.И., Лазаренко С.В., Шевцова Л.А. Синтез оптимального управления на основе объединенного принципа максимума // *Известия высших учебных заведений. Северо-Кавказский регион. Технические науки*. 2010. № 2. С. 31–37.
30. Bourdin L., Dhar G. Optimal sampled-data controls with running inequality state constraints: Pontryagin maximum principle and bouncing trajectory phenomenon // *Mathematical Programming*. 2022. vol. 191. no. 2. pp. 907–951.
31. Рустамов Г.А., Намазов М.Б., Гасымов А.Ю., Рустамов Р.Г. Управление динамическими объектами в условиях неопределенности в точечном скользящем режиме // *Мехатроника, автоматизация, управление*. 2019. Т. 20. № 12. С. 714–722.
32. Utkin V., Poznyak A., Orlov Y.V., Polyakov A. Chattering Problem. Road Map for Sliding Mode Control Design. *SpringerBriefs in Mathematics*. 2020. pp. 73–82.
33. Балабанов А. О синтезе желаемых фазовых траекторий скользящего режима в релейных системах управления // 8-я конференция «Микроэлектроника и информатика» (Кишинев, Молдова, 22-25 октября 2014 г.). 2014. С. 396–403.
34. Píkina G.A., Pashchenko F.F., Pashchenko A.F. Some Problems of Implementing Optimal Control Theory in Automated Control Systems // *Automation and Remote Control*. 2022. vol. 83. no. 10. pp. 1630–1638.
35. Utkin V., Lee H. Chattering problem in sliding mode control systems // *International Workshop on Variable Structure Systems, VSS'06. IEEE*, 2006. pp. 346–350.
36. Нейдорф Р.А., Чан Н.Н. Композиционный синтез квазиоптимальных по быстродействию систем управления высокого порядка // *Вестник Донского государственного технического университета*. 2007. Т. 7. № 4(35). С. 353–359.
37. Клюев А.С. Оптимизация автоматических систем управления по быстродействию / А.С. Клюев, А.А. Колесников. Москва: Энергоиздат, 1982. 240 с.
38. Shilin A.A., Pham H.T., Nguen V.V. Research data for this article. URL: <https://cloud.mail.ru/public/vS3k/sLULhtwKk> (дата обращения: 03.01.2024).
39. Verma M., Chen C.Y., Kılıcman A., Mat Hasim R.A. Systematic Review on the Advancement in the Study of Fuzzy Variational Problems // *Journal of Function Spaces*. 2022. vol. 2022. no. 1. DOI: 10.1155/2022/8037562.
40. Фельдбаум А.А. Основы теории оптимальных автоматических систем. 2-е изд., испр. и доп. Москва: Наука, 1966. 623 с.
41. Ritchie D., Johnson S., Lesk M., Kernighan B. UNIX Time-Sharing System: The C programming language // *Bell Sys. Tech. J*. 1978. vol. 57. no. 6. pp. 1991–2019.

42. Campbell S., Chancelier J., Nikoukhah R. Modeling and Simulation in SCILAB // Modeling and Simulation in Scilab/Scicos with ScicosLab 4.4. Springer, New York, 2010. pp. 73–106.
43. Mikac M., Horvatic M., Mikac V. Using vectorized calculations in Scilab to improve performances of interpreted environment // INTED2020 Proceedings. IATED, 2020. pp. 2127–2136.
44. Prokhorov S., Krivoshein Y., Shilin A. Automatic Control of Hot Water Supply System on Solar Collectors. 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). IEEE, 2019. pp. 1–6.
45. Tsvetkov N., Boldyryev S., Shilin A., Krivoshein Y., Tolstykh A. Hardware and Software Implementation for Solar Hot Water System in Northern Regions of Russia. *Energies* 2022. vol. 15. no. 4. DOI: 10.3390/en15041446.
46. Shilin A., Bukreev V. The reduction of the multidimensional model of the nonlinear heat exchange system with delay // Proceedings of the Information Technologies and Mathematical Modelling: 13th International Scientific Conference, ITMM 2014, named after A.F. Terpigov. Springer International Publishing, 2014. pp. 387–396.
47. Godlewski M., Rogowski K. Data Acquisition and Control System Based on Scilab Software Environment // *Pomiary Automatyka Robotyka*. 2022. vol. 26. no. 3. pp. 49–55.
48. Chao C.T., Sutarna N., Chiou J.S., Wang C.J. Equivalence between fuzzy PID controllers and conventional PID controllers // *Applied Sciences*. 2017. vol. 7. no. 6. DOI: 10.3390/app7060513.
49. on conventional PID control and nonlinear factors // *Applied Sciences*. 2019. vol. 9. no. 6. DOI: 10.3390/app9061224.
50. Shilin A., Bukreev V., Perevoshchikov F. Synthesis and implementation of σ -approach of slide control in heat-consumption system // *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. 2022. no. 3. pp. 501–508.

Шилин Александр Анатольевич — д-р техн. наук, доцент, профессор отделения, отделение электроэнергетики инженерной школы энергетики, Национальный исследовательский Томский политехнический университет (ТПУ). Область научных интересов: системы управления технологическими процессами, оптимальное управление. Число научных публикаций — 200. shilin@trn.ru; улица Усова, Учебный корпус № 8, 7, 634050, Томск, Россия; р.т.: +7(3822)70-1777,,3425.

Фам Чонг Хай — аспирант, отделение электроэнергетики инженерной школы энергетики, Национальный исследовательский Томский политехнический университет (ТПУ). Область научных интересов: системы управления технологическими процессами, оптимальное управление. Число научных публикаций — 2. tronghai180598@gmail.com; улица Усова, Учебный корпус № 8, 7, 634050, Томск, Россия; р.т.: +7(923)430-3989.

Нгуен Ван Вьюнг — канд. техн. наук, исследователь информации, Военно-морской технический институт. Область научных интересов: элементы и устройства вычислительной техники и систем управления. Число научных публикаций — 14. vuongoicolen@gmail.com; Деревня Нгокбинг, община Куангфук, округ Куангсьонг, 1, 42624, Тханьхоа, Вьетнам; р.т.: +84(934)221-092.

A.SHILIN, H. PHAM TRONG, V. NGUEN VUONG
**SYNTHESIS OF A FUZZY CONTROLLER BY A SECOND-ORDER
OBJECT WITH DELAY**

Shilin A., Pham Trong H., Nguen Vuong V. Synthesis of a Fuzzy Controller by a Second-Order Object with Delay.

Abstract. The paper proposes a method for using Fuzzy controller tools to synthesize optimal control of a second-order dynamic object with a delay. The idea is to construct a phase surface that combines optimal relay control away from the equilibrium state region and linear control in the region itself. This approach made it possible to avoid self-oscillations in a steady state, while maintaining the properties of optimal control in terms of speed. The switching trajectory in phase space, corresponding to the solution of the optimal control problem according to the Maximum principle, is determined by the method of inverse time calculation of the second-order difference equation of the plant. The region near the equilibrium state, where the linear controller is used, is determined from the results of modeling the motion of a point in phase space with optimal control for an object with a delay. This region is represented by an ellipse that describes motion in phase space in a self-oscillatory mode. To further eliminate self-oscillations in this area, a linear controller is used, tuned by means of solving the variational optimal control problem. It is proposed to use a tool for the synthesis of Fuzzy controllers, where the switching surface and calculation of the control value can be set arbitrarily. As a result, a variable controller structure is proposed to combine these two approaches. The resulting Fuzzy controller model is represented by a standard FLS structure, which was implemented in Python on the Orange Pi embedded computer. To connect to the current control object, an industrial controller FX3U-24MR is used, connected to a computer via a Modbus network. Full-scale tests are presented on a hot water supply temperature control object, which corresponds as closely as possible to the object model under study. The method, idea and results obtained in the work can be applied and investigated in the synthesis of control of dynamic objects in sliding mode to solve current problems related to eliminating the unwanted Chattering effect.

Keywords: optimal control, Fuzzy controller, phase trajectory, PID controller.

References

1. Pu Q., Zhu X., Liu J., Cai D., Fu G., Wei D., Sun J., Zhang R. Integrated optimal design of speed profile and fuzzy PID controller for train with multifactor consideration. *IEEE Access*. 2020. vol. 8. pp. 152146–152160.
2. Aruna R., Christa S. Modeling, system identification and design of fuzzy PID controller for discharge dynamics of metal hydride hydrogen storage bed. *International Journal of Hydrogen Energy*. 2020. vol. 45. no. 7. pp. 4703–4719.
3. Kandiban R., Arulmozhiyal R. Speed control of BLDC motor using adaptive fuzzy PID controller. *Procedia Engineering*. 2012. vol. 38. pp. 306–313.
4. Shi J. A fractional order general type-2 fuzzy PID controller design algorithm. *IEEE Access*. 2020. vol. 8. pp. 52151–52172.
5. Osinski C., Leandro G., da Costa Oliveira G. Fuzzy PID controller design for LFC in electric power systems. *IEEE Latin America Transactions*. 2019. vol. 17. no. 01. pp. 147–154.

6. Daradkeh Y. et al. Development of effective methods for structural image recognition using the principles of data granulation and apparatus of fuzzy logic. *IEEE Access*. 2021. vol. 9. pp. 13417–13428.
7. Bobyr M., Yakushev A., Kulabukhov S., Arkhipov A. System of stereovision based on fuzzy-logical method of constructing depth map. 2018 International Russian Automation Conference (RusAutoCon). IEEE, 2018. pp. 1–5.
8. Osinski C., Leandro G., da Costa Oliveira G. Fuzzy PID controller design for LFC in electric power systems. *IEEE Latin America Transactions*. 2019. vol. 17. no. 01. pp. 147–154.
9. Wang Y., Jin Q., Zhang R. Improved fuzzy PID controller design using predictive functional control structure. *ISA transactions*. 2017. vol. 71. pp. 354–363.
10. Oudah M., Sulttan M., Shneen S. Fuzzy type 1 PID controllers design for TCP/AQM wireless networks. *Indonesian Journal of Electrical Engineering and Computer Science*. 2021. vol. 21. no. 1. pp. 118–127.
11. Hao Z., Li C., Zhang H., Ren W. Control and realization of intelligent fishing system based on PLC and fuzzy PID algorithm. *Journal of Physics: Conference Series*. IOP Publishing, 2022. vol. 2246. no. 1. DOI: 10.1088/1742-6596/2246/1/012042.
12. Zhou J., Pedrycz W., Gao C., Lai Z., Yue X. Principles for constructing three-way approximations of fuzzy sets: A comparative evaluation based on unsupervised learning. *Fuzzy Sets and Systems*. 2021. vol. 413. pp. 74–98.
13. Lv H., Wang G. Approximations of Fuzzy Numbers by Using r-s Piecewise Linear Fuzzy Numbers Based on Weighted Metric. *Mathematics*. 2022. vol. 10. no. 1.
14. Tang K., Man K., Chen G., Kwong S. An optimal fuzzy PID controller. *IEEE transactions on industrial electronics*. 2001. vol. 48. no. 4. pp. 757–765.
15. Bobyr M., Emelyanov S. A nonlinear method of learning neuro-fuzzy models for dynamic control systems. *Applied soft computing*. 2020. vol. 88. DOI: 10.1016/j.asoc.2019.106030.
16. Hu Y., Yang Y., Li S., Zhou Y. Fuzzy controller design of micro-unmanned helicopter relying on improved genetic optimization algorithm. *Aerospace Science and Technology*. 2020. vol. 98. DOI: 10.1016/j.ast.2020.105685.
17. Mehedi I., Shah H., Al-Saggaf U., Mansouri R., Bettayeb M. Fuzzy PID control for respiratory systems. *Journal of Healthcare Engineering*. 2021. vol. 2021. no. 1. DOI: 10.1155/2021/7118711.
18. Perevoshchikov F., Vuong N., Shabo A., Shilin A. Identification of Mathematical Model of Drying Unit. *IOP Conference Series: Materials Science and Engineering*. IOP Publishing, 2020. vol. 795. no. 1. DOI: 10.1088/1757-899X/795/1/012021.
19. Zhao F., May J., Loh S. Controller Synthesis and Verification for Nonlinear Systems: A computational approach using phase-space geometric models. *IEEE Control Systems Magazine*, 1998. pp. 198–224.
20. Shakin V., Semyonova T., Kudryashova A., Frisk V. Comparison of computer modeling of rc filter in matlab and scilab environments. 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). IEEE, 2020. pp. 1–5.
21. Phu N.D., Hung N.N. Minimum stability control problem and time-optimal control problem for fuzzy linear control systems. *Fuzzy Sets and Systems*. 2019. vol. 371. pp. 1–24.
22. Zhang J., Shi P., Xia Y. Robust adaptive sliding-mode control for fuzzy systems with mismatched uncertainties. *IEEE Transactions on Fuzzy Systems*. 2010. vol. 18. no. 4. pp. 700–711.
23. Somwanshi D. et al. Comparison of fuzzy-PID and PID controller for speed control of DC motor using LabVIEW. *Procedia Computer Science*. 2019. vol. 152. pp. 252–260.

24. Pesch H.J., Plail M. The maximum principle of optimal control: a history of ingenious ideas and missed opportunities. *Control and Cybernetics*. 2009. vol. 38. no. 4A. pp. 973–995.
25. Utkin V. et al. Conventional and high order sliding mode control. *Journal of the Franklin Institute*. 2020. vol. 357. no. 15. pp. 10244–10261.
26. Horoshavin V., Grudin V. [Examples of constructing a closed-loop control system for an electrical installation that is quasi-optimal in terms of speed and energy consumption]. *Izvestija Tul'skogo gosudarstvennogo universiteta. Tehnicheskie nauki – Proceedings of Tula State University. Technical sciences*. 2023. no. 2. pp. 449–453. (In Russ.).
27. Horoshavin V., Grudin V. [Optimal program motion with variable control time]. *Radiopromyshlennost' – Radio industry*. 2020. vol. 30. no. 3. pp. 40–49. (In Russ.).
28. Doljatovskij V., Raul B. [Optimal control of a dynamic object based on the maximum principle]. *Sistemnyj analiz v proektirovanii i upravlenii – System analysis in design and management*. 2020. vol. 24. no. 2. pp. 37–45. (In Russ.).
29. Kostoglotov A., Kostoglotov A.I., Lazarenko S., Shevcova L. [Synthesis of optimal control based on the combined maximum principle]. *Izvestija vysshih uchebnyh zavedenij. Severo-Kavkazskij region. Tehnicheskie nauki – News of higher educational institutions. The North Caucasus region. Technical sciences*. 2010. no. 2. pp. 31–37. (In Russ.).
30. Bourdin L., Dhar G. Optimal sampled-data controls with running inequality state constraints: Pontryagin maximum principle and bouncing trajectory phenomenon. *Mathematical Programming*. 2022. vol. 191. no. 2. pp. 907–951.
31. Rustamov G., Namazov M., Gasymov A., Rustamov R. [Control of dynamic objects under conditions of uncertainty in point sliding mode]. *Mehatronika, avtomatizacija, upravlenie – Mechatronics, automation, control*. 2019. vol. 20. no. 12. pp. 714–722. (In Russ.).
32. Utkin V., Poznyak A., Orlov Y.V., Polyakov A. Chattering Problem. *Road Map for Sliding Mode Control Design. SpringerBriefs in Mathematics*. 2020. pp. 73–82.
33. Balabanov A. [On the synthesis of desired phase trajectories of the sliding mode in relay control systems]. 8-ya konferenciya «Mikroelektronika i informatika» [8th Conference "Microelectronics and Informatics"]. 2014. pp. 396–403. (In Russ.).
34. Pikina G.A., Pashchenko F.F., Pashchenko A.F. Some Problems of Implementing Optimal Control Theory in Automated Control Systems. *Automation and Remote Control*. 2022. vol. 83. no. 10. pp. 1630–1638.
35. Utkin V., Lee H. Chattering problem in sliding mode control systems. *International Workshop on Variable Structure Systems, VSS'06. IEEE*, 2006. pp. 346–350.
36. Nejdorf R., Chan N. [Kompozicionnyj sintez kvaziopimal'nyh po bystrodejstviju sistem upravlenija vysokogo porjadka]. *Vestnik Donskogo gosudarstvennogo tehničeskogo universiteta – Bulletin of the Don State Technical University*. 2007. vol. 7. no. 4(35). pp. 353–359. (In Russ.).
37. Kljuev A. Optimization of automatic control systems for performance [Optimizacija avtomaticheskijh sistem upravlenija po bystrodejstviju] A.S. Kljuev, A.A. Kolesnikov (eds.). Moscow: Jenergoizdat, 1982. 240 p. (In Russ.).
38. Shilin A.A., Pham H.T., Nguen V.V. Reasearh data for this article. Available at: <https://cloud.mail.ru/public/vS3k/sLULhtwkK> (accessed: 03.01.2024).
39. Verma M., Chen C.Y., Kilicman A., Mat Hasim R.A Systematic Review on the Advancement in the Study of Fuzzy Variational Problems. *Journal of Function Spaces*. 2022. vol. 2022. no. 1. DOI: 10.1155/2022/8037562.
40. Fel'dbaum A. Osnovy teorii optimalnyh avtomaticheskijh sistem [Fundamentals of the theory of optimal automatic systems]. Izd. 2-oe, izd-vo Nauki, 1966. 623 p. (In Russ.).
41. Ritchie D., Johnson S., Lesk M., Kernighan B. UNIX Time-Sharing System: The C programming language. *Bell Sys. Tech. J*. 1978. vol. 57. no. 6. pp. 1991–2019.

42. Campbell S., Chancelier J., Nikoukhah R. Modeling and Simulation in SCILAB. Modeling and Simulation in Scilab/Scicos with ScicosLab 4.4. Springer, New York, 2010. pp. 73–106.
43. Mikac M., Horvatic M., Mikac V. Using vectorized calculations in Scilab to improve performances of interpreted environment. INTED2020 Proceedings. IATED, 2020. pp. 2127–2136.
44. Prokhorov S., Krivoshein Y., Shilin A. Automatic Control of Hot Water Supply System on Solar Collectors. 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon). IEEE, 2019. pp. 1–6.
45. Tsvetkov N., Boldyryev S., Shilin A., Krivoshein Y., Tolstykh A. Hardware and Software Implementation for Solar Hot Water System in Northern Regions of Russia. Energies 2022. vol. 15. no. 4. DOI: 10.3390/en15041446.
46. Shilin A., Bukreev V. The reduction of the multidimensional model of the nonlinear heat exchange system with delay. Proceedings of the Information Technologies and Mathematical Modelling: 13th International Scientific Conference, ITMM 2014, named after A.F. Terpugov. Springer International Publishing, 2014. pp. 387–396.
47. Godlewski M., Rogowski K. Data Acquisition and Control System Based on Scilab Software Environment. Pomiary Automatyka Robotyka. 2022. vol. 26. no. 3. pp. 49–55.
48. Chao C., Sutarna N., Chiou J., Wang C. Equivalence between fuzzy PID controllers and conventional PID controllers. Applied Sciences. 2017. vol. 7. no. 6. DOI: 10.3390/app7060513.
49. Chao C., Sutarna N., Chiou J., Wang C. An optimal fuzzy PID controller design based on conventional PID control and nonlinear factors. Applied Sciences. 2019. vol. 9. no. 6. DOI: 10.3390/app9061224.
50. Shilin A., Bukreev V., Perevoshchikov F. Synthesis and implementation of -approach of slide control in heat-consumption system. Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2022. no. 3. pp. 501–508.

Shilin Alexander — Ph.D., Dr.Sci., Associate Professor, Professor of the department, Electrical power engineering department of the school of energy engineering, Tomsk Polytechnic University. Research interests: process control systems, optimal control. The number of publications — 200. shilin@tpu.ru; 7, Usova St., Educational building no. 8, 634050, Tomsk, Russia; office phone: +7(3822)70-1777,,3425.

Pham Trong Hai — Ph.D. student, Electrical power engineering department of the school of energy engineering, Tomsk Polytechnic University. Research interests: process control systems, optimal control. The number of publications — 2. tronghai180598@gmail.com; 7, Usova St., Educational building no. 8, 634050, Tomsk, Russia; office phone: +7(923)430-3989.

Nguyen Van Vuong — Ph.D., information researcher, Naval Technical Institute. Research interests: elements and devices of computer technology and control systems. The number of publications — 14. vuongoicolen@gmail.com; 1, Ngoc Bing Village, Quang Phuc Commune, Quang Suong District, 42624, Thanh Hoa City, Viet Nam; office phone: +84(934)221-092.

Т.А. ФЕДОРОВА, В.А. РЫЖОВ, К.С. САФРОНОВ
**ИСПОЛЬЗОВАНИЕ ГИБРИДНОЙ КОММУНИКАЦИОННОЙ
АРХИТЕКТУРЫ ПОДВОДНОЙ БЕСПРОВОДНОЙ
СЕНСОРНОЙ СЕТИ ДЛЯ ПОВЫШЕНИЯ ЕЕ ВРЕМЕНИ
ЖИЗНИ И ЭФФЕКТИВНОСТИ**

Федорова Т.А., Рыжов В.А., Сафронов К.С. **Использование гибридной коммуникационной архитектуры подводной беспроводной сенсорной сети для повышения ее времени жизни и эффективности.**

Аннотация. В работе выполнен сравнительный анализ основных функциональных характеристик подводных беспроводных сенсорных сетей (ПБСС) со стационарной и гибридной коммуникационными архитектурами. Указанные ПБСС состоят из сенсорных узлов, расположенных на морском дне и надводных межсредных шлюзов, обеспечивающих передачу информационных пакетов между подводным и надводным сегментами сети. В стационарной ПБСС роль шлюзов выполняют заякоренные буи, в гибридной – мобильные транспортные платформы. С использованием математического аппарата, основанного на вероятностном подходе, проведена оценка функциональных характеристик альтернативных коммуникационных архитектур ПБСС с энергетической точки зрения - определены общие энергетические затраты сети на пересылку сообщений и время жизни датчиков сети. Для численного анализа функциональных характеристик ПБСС рассмотрены достаточно широкие диапазоны изменения проектных параметров сети, таких как: размер акватории, требуемое количество и варианты размещения сенсорных узлов, вероятность доставки пакета в акватории (физические параметры среды), в которых осуществлялся поиск «оптимального» с энергетической точки зрения решения. Выполненное авторами исследование показывает, что мобильность играет важную роль в повышении качества функционирования подводной сети в аспектах покрытия (обеспечения связности), энергоэффективности и времени жизни. Мобильный элемент в виде волнового глайдера, выполняющий роль межсредного шлюза, способен функционировать в акватории в течение продолжительного времени, что говорит о перспективности его использования для прикладных задач сбора, накопления и ретрансляции информации в рамках интернета подводных вещей.

Ключевые слова: подводные беспроводные сенсорные сети, гибридные сети, мобильный межсредный шлюз, гидроакустическая связь, критерий эффективности сети, время жизни сети.

1. Введение. В последние два десятилетия активно развивается концепция интернета подводных вещей [1]. Интернет подводных вещей представляет собой сеть интеллектуальных взаимосвязанных подводных объектов, оснащенных исполнительными устройствами и сенсорами, позволяющими выполнять различные действия, связанные с мониторингом, контролем и управлением. Эта сеть обладает возможностями самообучения и интеллектуальных вычислений.

Интернет подводных вещей призван стать мощной технологией для поддержки различных приложений, таких как сбор информации о водной среде в реальном времени, военные приложения, морская

навигация и безопасность, прогнозирование и контроль стихийных бедствий, контроль экологической и техногенной безопасности, разведка полезных ископаемых, контроль биоресурсов и др. Таким образом, интернет подводных вещей имеет большие перспективы стать серьезной основой «умного океана» [2 – 3].

Для успешного развертывания приложений интернета подводных вещей необходимы сети связи, обеспечивающие оптимальную производительность, эффективность и надежность передачи данных.

Существующие на сегодня решения подводного интернета вещей объединяют морские технические объекты, такие как буи, надводные и подводные суда и аппараты, системы наблюдения за океаном, а также береговые центры обработки данных в гетерогенные сети различного масштаба и назначения [4 – 6].

В два последних десятилетия активно создавались кабельные сети наблюдения за океаном (например, Australian Integrated Marine Observing System, Ocean Networks Canada, Dense Oceanfloor Network System for Earthquakes and Tsunamis), предназначенные для долгосрочного и непрерывного мониторинга подводных данных и их передачи в центры обработки [7 – 10].

Однако ввиду того, что развертывание кабельных сетей в морских условиях – весьма дорогостоящее мероприятие, то наряду с их внедрением в то же самое время активно разрабатывалась концепция беспроводных сетей. Именно разработка беспроводной технологии инициировала переход подводной связи на качественно новый уровень. Примером практической реализации таких сетей является, например, Persistent Littoral Undersea Surveillance Network (PLUSNet) [11].

Исследования радиочастотной, оптической, магнитоиндуктивной и акустической технологий беспроводной передачи информации под водой показали преимущества и недостатки каждой из них. Радиочастотная и магнитоиндуктивная связь могут обеспечивать высокую скорость передачи данных, но только на коротких расстояниях. Оптическая связь обеспечивает хорошую производительность, когда она реализуется в пределах прямой видимости и при условии незначительной мутности воды. Акустическая связь может обеспечить относительно большую дальность передачи, однако ее использование существенно зависит от свойств морской среды.

Несмотря на наличие положительных и отрицательных качеств у каждой из указанных технологий подводной связи, можно

констатировать, что на сегодняшний день гидроакустические технологии ПБСС являются наиболее применимыми в морской практике [12, 13].

При проектировании гидроакустических подводных беспроводных сенсорных сетей различного масштаба под конкретные задачи на первый план встает проблема создания такой архитектуры и компонентов сети, которые будут обеспечивать высокую надежность доставки собранной информации конечному пользователю и длительную автономную работу сети в сложных условиях морской среды [14].

Существует большое количество работ посвященных моделированию ПБСС [15 – 24]: разработке алгоритмов развертывания сетей, протоколов различных уровней OSI, обеспечению связности сетей, энергетической балансировке сетей и др. Задачи решались в 2D и 3D постановках.

Для передачи информации, собранной подводными сенсорами, в надводный сегмент сети используются межсредные шлюзы, роль которых часто выполняют стационарные буи [5]. В последнее время в связи с активным развитием морских робототехнических платформ, в качестве межсредных шлюзов начали рассматриваться такие перспективные транспортные платформы, как волновые и подводные глайдеры [25 – 29], безэкипажные катера и др. В связи с указанным ПБСС, в зависимости от типа межсредного шлюза, можно классифицировать как сети со стационарной или гибридной коммуникационными архитектурами.

Необходимо отметить, что существующие работы по использованию мобильных элементов в беспроводных сенсорных сетях в основном связаны с наземными приложениями. Работ, посвященных функционированию гибридных подводных беспроводных сенсорных сетей (включающих в себя стационарные и мобильные узлы), учитывающих специфику распространения сигнала в водной среде в рамках различных практических приложений, существенно меньше [30].

В работах по подводным акустическим беспроводным сенсорным сетям с мобильными агентами рассматривались задачи моделирования функциональных характеристик в зависимости от типа мобильного агента и природы мобильности. Эти работы были посвящены: учету мобильности агентов (узлов сети), связанной с физическими процессами океана – течениями, внешними возмущениями [31, 32]; использованию мобильных узлов, выполняющих роль референсных узлов сети и роль межсредных

шлюзов передачи сигнала на границе вода-воздух [33, 34]. В последнем случае роль мобильного шлюза выполняли различные робототехнические транспортные платформы – как подводные – автономные необитаемые подводные аппараты (АНПА), подводные глайдеры (ПГ), так и надводные – волновые глайдеры (ВГ), безэкипажные катера (БЭК).

Применение АНПА, как мобильных агентов, в беспроводных сетях обсуждалось в работах [35 – 37].

Рассматривалось использование АНПА, выполняющего функции мобильного элемента ПБСС в рамках следующих сценариев. (1) АНПА использовался как мобильный референсный узел-накопитель, собирающий информацию со стационарных узлов сети (с головных узлов кластеров) по так называемому «rendezvous-based» сценарию – последовательный обход всех головных узлов и шлюзов [38, 39]. Общий объем передачи данных при таком сценарии сводится к минимуму. Кроме того, управляемая мобильность АНПА позволяет использовать подводный канал ближнего действия с высокой скоростью передачи данных для передачи больших объемов информации. (2) АНПА использовался как мобильный агент, выполняющий роль «замещающего» элемента какого-либо стационарного узла в случае его «смерти» в целях сохранения связности сети [40].

В перечисленных работах рассматривались задачи локализации узлов сети с использованием АНПА [41], предлагались различные методы планирования пути подводного аппарата и специальные протоколы маршрутизации [35, 36, 42 – 46]. Решение рассматривалось в рамках задачи коммивояжера, с использованием генетических алгоритмов, модели обучения с подкреплением и др.

В случае использования в качестве мобильного шлюза надводной транспортной платформы, сбор информации осуществляется от референсных узлов, расположенных либо на дне, либо в объеме акватории [2, 27, 47]. При этом надводный мобильный шлюз осуществляет накопление и передачу собранной информации в центр обработки. Безусловно, использование конкретного типа надводной транспортной платформы зависит от особенностей решаемой прикладной задачи, в частности, от требований по периодичности съема сенсорами информации и режиму доставки ее в центр обработки – близком к реальному времени или с возможными задержками, а также от масштабов и глубин акватории. Несмотря на активное практическое использование волновых глайдеров, число работ, посвященных надводным мобильным транспортным

платформам, как агентам ПБСС невелико. В этих работах рассматривались задачи, в основном связанные с использованием волновых глайдеров для локализации узлов сети.

Таким образом, обзор литературы по сбору данных с помощью мобильных платформ показывает, что существующие работы в основном посвящены использованию для указанных целей АНПА. К сожалению, энергетический ресурс таких транспортных платформ весьма ограничен и для реальных долгосрочных миссий по сбору данных с сенсоров ПБСС они мало применимы. Альтернативным решением является использование в качестве управляемой мобильной платформы волнового глайдера, обладающего значительными энергетическими ресурсами за счет использования энергии волн и солнца, и способного выполнять долгосрочные миссии [28]. Существует большой класс практических приложений, для которых время доставки сообщения не является критическим параметром, а большее значение имеет надежность сети. Для решения этих задач использование волнового глайдера, имеющего невысокую скорость перемещения, является весьма актуальным.

В связи с указанным, в данной работе рассматривается задача исследования коммуникационной инфраструктуры гибридной ПБСС, в состав которой входит волновой глайдер, выполняющий функции мобильного шлюза. Авторами предлагается аналитический метод планирования оптимальной траектории движения волнового глайдера для эффективного сбора данных по запросу со стационарных сенсоров сети. Для реализации указанного метода используется оригинальная стохастическая модель функционирования гибридной ПБСС, учитывающая особенности подводного канала.

Предлагаемая работа имеет своей целью дать сравнительную качественную и количественную оценку основных функциональных характеристик гидроакустических ПБСС со стационарной и гибридной коммуникационными архитектурами с точки зрения их энергетической эффективности и времени жизни сетей.

С использованием предложенной авторами вероятностной модели [48] в работе проводится анализ влияния таких проектных параметров как топологические характеристики сети и физические характеристики среды (вероятность доставки пакета) на полные энергетические затраты ПБСС и на время жизни сенсоров.

Полученные теоретические зависимости и результаты моделирования позволили дать энергетическую оценку различных моделей размещения обыкновенных и референсных агентов для стационарной и гибридной сетей, сделать выводы относительно

предпочтительности использования той или иной коммуникационной архитектуры.

2. Постановка задачи

2.1. Коммуникационная архитектура подводной беспроводной сенсорной сети. Рассматривается коммуникационная архитектура подводной беспроводной сенсорной сети (ПБСС), включающая в себя группу стационарных агентов (сенсоров), размещенных на глубоководном плоском дне, и межсредные шлюзы, роль которых могут выполнять либо стационарные буи, либо мобильные транспортные робототехнические платформы (волновой глайдер, безэкипажный катер и др.) (рисунк 1).

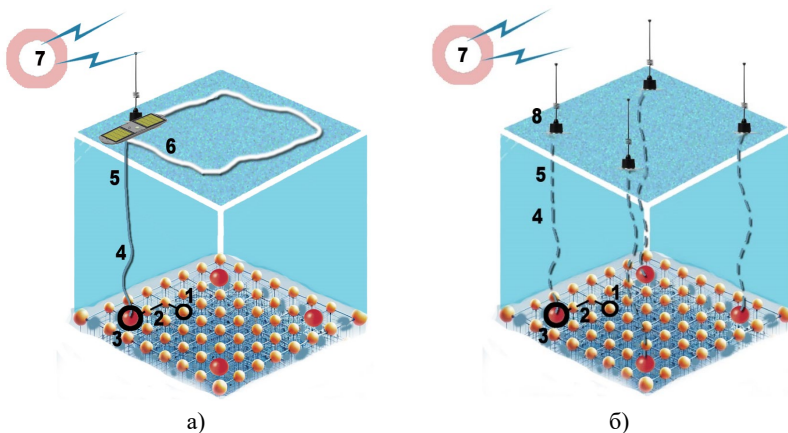


Рис. 1. Структура подводной беспроводной сенсорной сети: а) с мобильным шлюзом, б) со стационарными шлюзами. 1 – обыкновенный сенсор, 2 – путь hop-by-hop до референсного агента, 3 – референсный агент, 4 – гидроакустический канал связи, 5 – мобильный шлюз, 6 – траектория движения мобильного шлюза, 7 – центр обработки данных, 8 – Стационарный буй

Сенсоры выполняют периодические замеры некоторых характеристик среды и передают их по гидроакустическому каналу.

Существует определенное количество референсных агентов, к которым стекается информация от всех остальных – обыкновенных агентов сети. Передача пакета от обычного к референсному агенту осуществляется через ближайшего соседа (hop-by-hop). Далее референсные агенты пересылают полученные (за цикл работы сети) пакеты в шлюз.

Очевидно, что при этом вероятность доставки пакета зависит от расстояния между передающим и принимающим агентами (узлами) и от физических характеристик среды, таких как коэффициент затухания сигнала β , уровень шума в акватории SNR .

Ввиду использования однопутевой многошаговой маршрутизации через обыкновенные узлы, находящиеся рядом с референсными, и через сами референсные, проходит существенно больший информационный трафик.

Очевидно, чем больше данных проходит через узел сети, тем больше его расход энергии. Как следствие, в сети возникает проблема дисбаланса энергопотребления, приводящая к тому, что референсные узлы раньше других выходят из строя из-за разряда аккумуляторов, что влечет за собой уменьшение времени жизни сенсорной сети в целом. Для выравнивания расхода энергии всех узлов сети используют различные методы энергетической балансировки:

- индивидуальный подбор емкости батарей в зависимости от положения устройств в сети и выполняемых ими функций. В этом случае референсные агенты могут оснащаться большими по емкости аккумуляторами;

- разная плотность размещения узлов сети в акватории в зависимости от предполагаемой интенсивности трафика в конкретной зоне;

- использование протоколов маршрутизации, основанных, например, на таких принципах как, чередование дальней и ближней передачи, оценка величины остаточной энергии узлов на пути до шлюза;

- кластеризация;

- использование мобильности отдельных компонентов сети (в том числе, мобильных шлюзов). В качестве мобильного шлюза могут использоваться автономные необитаемые подводные аппараты (АНПА), подводные глайдеры (ПГ), волновые глайдеры (ВГ).

В данной работе в целях энергетической балансировки сети используются идеи кластеризации и мобильности. Роль мобильного агента выполняет надводная робототехническая платформа (например, волновой глайдер).

Качество разбиения сенсорной сети на кластеры определяется прежде всего оптимальным расположением референсных узлов, так как именно к ним стекается вся информация от обыкновенных узлов сети. Очевидно, что пропускная способность референсных узлов ограничена и при большой размерности сети и их неудачном расположении возможно появление «узких мест», в которых резко

возрастают энергозатраты, приводящие к быстрой разрядке аккумуляторов. Для моделирования подобных ситуаций, в данной работе исследуются два варианта размещения референсных агентов, влияющих на пропускную способность сети, и как следствие, на ее надежность и эффективность.

Вариант 1. Стационарная модель ПБСС. Разбиение ортогональной сетки на конечное число квадратных «подрешеток» меньшей размерности (рисунок 2(а)), в центре которых расположены стационарные шлюзы, роль которых выполняют буи. На каждой «подрешетке» обыкновенные агенты передают сообщения в центр – референсному агенту, который в свою очередь передает информацию в вертикальном направлении на шлюз. При этом шлюз передает сообщения дальше в береговой центр обработки данных.

Вариант 2. Гибридная модель ПБСС. Расположение референсных агентов на квадратном контуре (рисунок 2(б)). В этом случае используется мобильный шлюз, которому необходимо обходить этот контур.

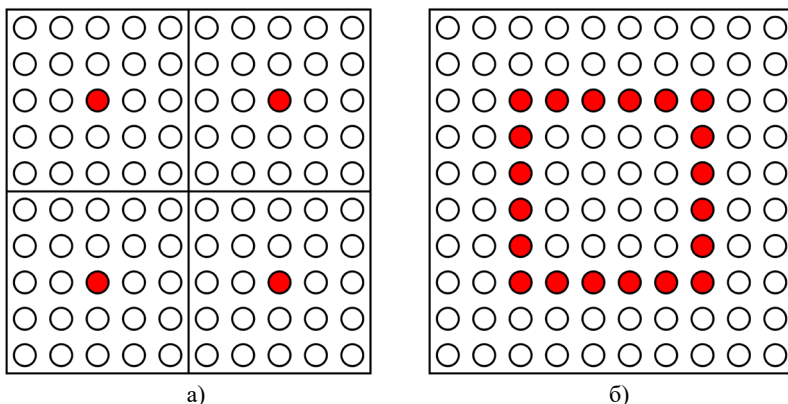


Рис. 2. Модель ПБСС при разных вариантах размещения референсных узлов: а) разбиение сети на квадратные «подрешетки», б) расположение референсных узлов на квадратном контуре. Красные кружки – референсные агенты, передающие сообщения на шлюз

Рассмотрим функционирование сети для указанных вариантов размещения референсных агентов.

2.2. Постановка модельной задачи. В работе рассматривается задача в следующей модельной постановке. Требуется выполнить измерение некоторой характеристики среды и передать эту информацию от каждого обыкновенного сенсора (через цепочку

остальных сенсоров сети) референсным сенсорам, которые в свою очередь должны передать пакеты в шлюз.

Полагается, что:

- сенсорные узлы (обыкновенные и референсные) оборудованы гидроакустическими модемами с одинаковыми характеристиками – имеют несущую частоту f , полосу пропускания B , антенна является ненаправленной.

- все пакеты S (содержащие полезную и служебную информацию) имеют одинаковую длину N_{bit} и передаются с битовой скоростью f_{bit} .

- время работы сети между двумя измерениями характеристик среды (цикл) T ограничено. За это время все обыкновенные сенсоры должны передать результаты измерений через референсных агентов в шлюз и приступить к следующему измерению.

2.3. Физическая модель. Известно, что одним из критериев производительности цифровых систем связи является зависимость вероятности появления ошибочного бита от отношения энергии сигнала, приходящейся на один бит – e_b , к спектральной плотности мощности аддитивного белого гауссовского шума – N_0 . Это отношение e_b / N_0 определяется через среднюю мощность сигнала P_s и среднюю мощность шума P_n следующим образом [49]:

$$\frac{e_b}{N_0} = \frac{P_s}{P_n} \frac{B}{f_{bit}}, \quad (1)$$

где B – ширина полосы пропускания, f_{bit} – битовая скорость передачи информации.

Потери при распространении акустического сигнала частотой f на расстояние r от источника в простейшем случае могут быть определены как

$$A(r, f) = r^s a(f)^r,$$

где степень $s = 2$ для сферической волны и $s = 1$ для цилиндрической. Считая, что распространение происходит на глубокой воде и расширение фронта волны является сферическим, возьмем $s = 2$,

$a(f)$ – связан с коэффициентом затухания сигнала выраженном в дБ/км по формуле

$$10 \log_{10} a(f) = \beta(f).$$

В пределах заданной акватории будем считать, что температура, соленость и кислотность постоянны, тогда коэффициент затухания $\beta(f)$ при заданной частоте звуковых колебаний f можно оценить по обобщенной формуле Торпа [50]:

$$\beta(f) = \frac{0.1 \cdot f^2}{1 + f^2} + \frac{40 \cdot f^2}{4100 + f^2} + 2.75 \cdot 10^{-4} \cdot f^2 + 0.0003. \quad (2)$$

Отношение сигнал/шум на расстоянии r от передатчика обозначим $\Gamma(r)$ и определим через нормированное отношение сигнал/шум и потери при распространении сигнала как

$$\Gamma(r) = \frac{e_b}{N_0 A(r, f)} = \frac{e_b}{N_0 r^2 a(f)^r}.$$

В подводном канале с рэлеевским замиранием скорость возникновения битовой ошибки (*BER*) для двоичной фазовой модуляции *BPSK* может быть вычислена как [31]:

$$q_e(r) = \frac{1}{2} \left(1 - \sqrt{\frac{\Gamma(r)}{1 + \Gamma(r)}} \right) = BER. \quad (3)$$

Тогда для пакета длиной N_{bit} вероятность успешной передачи на расстояние r будет равна

$$p(r) = (1 - q_e(r))^{N_{bit}}. \quad (4)$$

2.4. Протоколы сетевого и канального уровня модели OSI. В качестве протокола маршрутизации в работе рассматривается динамический протокол передачи информации через ближайшего соседа (hop-by-hop), который сводится к тому, что обыкновенные

сенсоры должны передавать сообщения только ближайшим соседям. Если после фиксированного числа передач всем соседям сообщение не передано, оно считается потерянным и сенсор переходит к пересылке следующего стоящего в очереди сообщения. При этом в таблице маршрутизации для каждого сенсора определено, какой сосед имеет для него более высокий приоритет для пересылки сообщения, какие соседи имеют более низкий приоритет, а какие – нулевой.

Протокол управления доступом к среде (МАС-протокол), рассматриваемый в работе – несинхронизированный доступ к гидроакустическому каналу, основанный на конкуренции. Примером такого протокола может служить хорошо известный классический протокол АЛОНА [51].

3. Вероятностная модель. Пусть ПБСС расположена в физически однородной среде и имеет вид квадратной ортогональной сетки размерности $n \times n$, в узлах которой расположены сенсоры. Расположение сенсорных узлов на ортогональной сетке удобно тем, что ее геометрические характеристики легко вычислить. Будем говорить, что обыкновенный агент (сообщение) находится на k -ом уровне удаленности от референсного агента, если для доставки сообщения требуется k пересылок. Пусть вероятность успешной передачи сообщения между соседними сенсорами постоянна в данной акватории и равна p .

Введем основные физические характеристики исследуемой системы (таблица 1).

Таблица 1. Основные физические характеристики исследуемой системы

Обозначение	Описание
$L_p(n)$	Общее ожидаемое число пересылок в сети размерности.
$L_p(k)$	Общее ожидаемое число пересылок, которое совершает один обыкновенный сенсор, находящийся на k -ом уровне удаленности от референсного.
T_{inf}	Время, затрачиваемое сенсором на сбор информации.
$T_s = N_{bit} / f_{bit}$	Время передачи сообщения и получения подтверждения о доставке.
T_w	Время ожидания и приема сообщений (прослушивания эфира).
T_c	Время от начала выполнения измерения (съема информации) до завершения доставки всех сообщений в шлюз (цикл работы сети).
T_0	Время, которое сеть находится в режиме сна (время между двумя циклами съема информации, во время которого сенсор не потребляет энергию).
$T = T_c + T_0$	Интервал, состоящий из цикла работы сети T_c и времени сна T_0 до следующего измерения.

3.1. Расчет времени жизни сенсоров. Зная начальную энергию батареи E_0 и мощность P_k , потребляемую сенсором на k -ом уровне удаленности, можно приблизительно оценить время жизни сенсора (узла сети) как

$$t_k = \frac{E_0}{P_k}. \quad (5)$$

Интервал работы сенсора T состоит из времени T_{inf} , во время которого он тратит энергию E_k^{inf} на сбор информации, общего времени $L_p(k)T_s$ на $L_p(k)$ пересылок сообщений, при которых тратится энергия E_k^s , времени прослушивания эфира T_w при котором тратится энергия E_k^w на прием сообщений и ожидание, и время T_0 , которое сеть находится в режиме сна (время между двумя циклами съема информации, во время которого сенсор не потребляет энергию), так что

$$T = T_c + T_0 = T_{inf} + L_p(k)T_s + T_w + T_0,$$

то есть общая затраченная за цикл энергия E_k определяется как

$$E_k = E_k^{inf} + E_k^s + E_k^w. \quad (6)$$

В режиме сна энергия сенсором не тратится, поэтому это слагаемое в формуле (6) равно нулю.

Запишем выражение для общей энергии, затрачиваемой сенсором за цикл работы T_c :

$$E_k = P_k T_c = P_{inf} T_{inf} + L_p(k) P_s T_s + P_w T_w. \quad (7)$$

Учтем, что $P_{inf} \approx P_w$, а время съема информации и прослушивания эфира $T_{inf} + T_w$ может быть вычислено как время, не занятое пересылками сообщений

$$T_{inf} + T_w = T_c - L_p(k)T_s. \quad (8)$$

Из формул (5), (7) и (8) несложно получить время, которое сенсор может провести в рабочих режимах съема информации, пересылок сообщений и прослушивания эфира на k -ом уровне удаленности:

$$t_p(k) = \frac{E_0 T_c}{L_p(k) P_s T_s + P_w (T_c - L_p(k) T_s)}. \quad (9)$$

Это время работы сенсора (без учета времени сна) будем называть временем жизни сенсора на k -ом уровне удаленности.

3.2. Расчет энергетических затрат. Пусть полное число пересылок на сети размерности $n \times n$ равно $L(n)$. При заданной вероятности доставки сообщения от сенсора ближайшему соседу равной p и возможности отправлять сообщения неограниченное число раз, математическое ожидание количества попыток передачи сообщения между соседними сенсорами, необходимое для успешной доставки, равно $1/p$. Общее ожидаемое число пересылок в сети размерности $n \times n$ определится выражением

$$L_p(n) = \frac{L(n)}{p}. \quad (10)$$

Если число референсных узлов равно m , то число обыкновенных сенсоров, а следовательно, и число сообщений, требующих отправки $S(n, m) = n^2 - m$.

Будем считать, что все сенсоры в сети равномерно распределены по кластерам, т.е. в каждом кластере есть один референсный и одинаковое количество обыкновенных сенсоров и уровней удаленности k_{max} . На каждом уровне сенсоры также разбиты по кластерам в среднем равномерно.

Пусть $l(p, k)$ – среднее число пересылок, необходимых для доставки одного сообщения с k -го уровня удаленности на референсный узел. Очевидно

$$l(p, k) = \frac{k}{p}.$$

Зная количество уровней k_{max} и количество сообщений, передаваемых через каждый уровень $S_k(n)$, $1 \leq k \leq k_{max}$, можно получить общую формулу для общего ожидаемого числа пересылок в сети:

$$L_p(n) = \sum_{k=1}^{k_{max}} S_k(n) \cdot \frac{k}{p}. \quad (11)$$

Сделаем еще одно допущение. Предположим, что в режиме ожидания и приема сообщения мощность пренебрежимо мала $P_w \approx 0$. Такое предположение позволяет не оценивать энергетические затраты на ожидание и прослушивание эфира. Полные энергетические затраты такой «идеальной» сети связаны только с пересылками сообщений

$$E_p(n) = L_p(n) P_s T_s. \quad (12)$$

Необходимо отметить, что чем выше нагрузка сети, тем ближе рассматриваемая сеть к «идеальной», так как большая часть времени действительно тратится на передачу сообщений, а не на ожидание. Это снова приводит к необходимости рассматривать только высокую нагрузку сети.

3.3. Критерии оптимальности работы сети. Оценим предложенные выше расположения сенсоров в акватории – разбиение на подрешетки и обход периметра – с точки зрения двух критериев:

- *Энергоэффективности сети* – общих энергетических затрат на пересылку всех сообщений в шлюз за один цикл (чем ниже эти затраты, тем более эффективна сеть);
- *Времени жизни* референсных сенсоров t_{ref} и всей сети в целом.

4. Сравнение работы оптимизированных стационарной и гибридной ПБСС в акватории произвольного размера. Предполагается, что размер акватории может быть выбран любым. При этом имеется возможность разместить в акватории любое число сенсоров на любом расстоянии друг от друга, а, следовательно, обеспечить любую вероятность доставки сообщений в шлюз.

Для сравнения эффективности работы стационарной и гибридной ПБСС будем считать, что они расположены в одной и той же акватории (с одинаковыми свойствами среды) и имеют одинаковые характеристики:

- общее число сенсоров в акватории n^2 ;
- число референсных узлов m ;
- технические характеристики модемов.

4.1. Разбиение сети на равные подрешетки. Рассмотрим случай стационарной ПБСС, когда в центрах подрешеток расположены стационарные буи, собирающие сообщения с сенсоров своего кластера (своей подрешетки) и передающие их в центр обработки данных. Пусть сеть разбита на равные подрешетки одинаковой размерности (рисунок 2(а)). Будем считать, что каждая подрешетка имеет вид квадратной сетки $l \times l$ (где l – нечётное число) с референсным узлом, расположенным в центре. Предположим, что n^2 кратно l^2 , т.е. $n^2 = m \cdot l^2$ (m – число референсных узлов, будем считать, что оно четное для удобства сравнения с гибридной сетью).

Учтем, что из энергетических соображений, количество референсных узлов m предпочтительно выбирать порядка 5% от общего числа агентов в сети [52]

$$m \approx 0,05n^2.$$

С учетом сделанного предположения, число агентов на стороне подрешетки оказывается не зависящим от масштаба сети

$$l = \frac{n}{\sqrt{m}} = \frac{1}{\sqrt{0,05}} \approx 4,54$$

и может быть выбрано либо равным 4, либо 5. В данной работе, учитывая, что l – нечётное число, будет использоваться параметр разбиения на подрешетки $l = 5$.

Зависимость числа пересылок S_k через уровень k от длины стороны подрешетки l в общем случае может быть описана формулой:

$$S_k = 4m \min(k, l - k), \quad 1 \leq k \leq l - 1, \quad k_{\max} = l - 1, \quad l = 5.$$

Общее ожидаемое число пересылок в сети $L_p^{sub-latt}(n)$ при вероятности доставки p для оптимального параметра разбиения на подрешетки $l = 5$ определяется по формуле (11) как

$$L_p^{sub-latt}(n) = \frac{4n^2}{pl^2} \sum_{k=1}^{l-1} k \cdot \min(k, l-k) \Big|_{l=5} = \frac{12n^2}{5p}. \quad (13)$$

Подстановка полученного выражения (13) в формулу (12) дает возможность определить полные энергетические затраты на пересылку сообщений в сети.

Важно отметить тот факт, что общее ожидаемое число пересылок зависит в стационарной ПБСС только от двух параметров – от однократной вероятности доставки сообщения p и от размера сети n^2 .

Общее ожидаемое число пересылок, которое совершает один обыкновенный сенсор, находящийся на k -ом уровне удаленности от референсного определится выражением

$$L_p^{sub-latt}(k) = \frac{1}{S_k} \sum_{i=k}^{l-1} \frac{S_i}{p} = \frac{1}{\min(k, l-k)} \sum_{i=k}^{l-1} \frac{\min(i, l-i)}{p}. \quad (14)$$

Общее ожидаемое число пересылок через референсные узлы равно

$$L_p^{sub-latt}(k=0) = \frac{l^2}{p} \Big|_{l=5} = \frac{25}{p}. \quad (15)$$

Зная загруженность узлов, можно определить время их жизни и оценить время жизни всей сети подставив формулы (14) и (15) в формулу (9).

Важно отметить тот факт, что общее ожидаемое число пересылок на референсные узлы зависит в стационарной ПБСС только от одного параметра – от однократной вероятности доставки сообщения p .

Небольшое количество проектных параметров в стационарной сети уменьшает возможности для ее оптимизации.

4.2. Обход внутреннего контура. Рассмотрим случай гибридной ПБСС, когда референсные узлы расположены на некотором контуре, форма которого совпадает с формой периметра акватории. Мобильный шлюз, роль которого может выполнять волновой глайдер, обходит все референсные узлы по непрерывной замкнутой траектории. Пусть эта траектория имеет форму квадрата, причем центр квадрата совпадает с центром акватории.

В предположении, что n – чётное, определим количество контуров $\nu = n/2$. Пронумеруем эти контуры, начиная от центра акватории к ее периметру, и предположим, что референсные агенты находятся на μ -ом контуре, $1 \leq \mu \leq \nu$. Число референсных узлов будет равно

$$m = 4(2\mu - 1).$$

Рассматривая сети, в которых линейный размер n кратен 10 ($n = 10q$, $q \in \mathbb{N}$), найдем номер контура гибридной сети, который содержит такое же количество референсных узлов, как и рассмотренная выше стационарная модель разбиения на подрешетки, т.е.

$$4(2\mu - 1) = \frac{n^2}{25} = 4q^2, \quad q \in \mathbb{N}. \quad (16)$$

Решая уравнение (16) и округляя результат в большую сторону, получим номер контура μ :

$$\mu = \frac{q^2}{2} + \frac{1}{2}, \quad q \in \mathbb{N}.$$

Приведем таблицу соответствия стационарной и гибридной ПБСС с примерно равным числом референсных узлов (таблица 2).

По отношению к референсным узлам обычные сенсоры сети можно разделить на внешние (лежащие в области между контуром референсных узлов и внешним периметром акватории) и внутренние (лежащие внутри контура референсных узлов). Обозначим число уровней удалённости узлов от референсных k_{max} и количество

сообщений, передаваемых через каждый уровень S_k для внешних и внутренних сенсоров как k_{max}^{in} , S_k^{in} , k_{max}^{out} , S_k^{out} .

Таблица 2. Основные физические характеристики исследуемой системы

q	n	n^2	$m^{sub-latt}$	Номер контура μ	m^{count}
1	10	100	4	1	4
2	20	400	16	[2,5]=3	20
3	30	900	36	5	36
4	40	1600	64	[8,5]=9	68
5	50	2500	200	13	200

Определение k_{max}^{in} , S_k^{in} дает:

$$k_{max}^{in} = \mu - 1,$$

$$S_k^{in} = 4(2\mu - 2k - 1), 1 \leq k \leq k_{max}^{in}.$$

Величины же k_{max}^{out} , S_k^{out} определяются следующими формулами:

$$k_{max}^{out} = 2(v - \mu),$$

$$S_k^{out} = 4 \cdot \begin{cases} 2\mu + k - 1 & 1 \leq k \leq (v - \mu) \\ 2v - (2\mu + k - 1) & (v - \mu) + 1 \leq k \leq 2(v - \mu) \end{cases}.$$

Общее ожидаемое число пересылок в сети при чётном n в сети определим по формуле (11). Оно будет состоять из двух слагаемых (соответствующих внешней и внутренней областям акватории относительно контура референсных узлов):

$$L_p^{count}(n) = \sum_{k=1}^{\mu-1} S_k^{in} \cdot \frac{k}{p} + \sum_{k=1}^{2(v-\mu)} S_k^{out} \cdot \frac{k}{p}.$$

Упрощая, получим, что число пересылок зависит от трех параметров – от вероятности доставки p , от размерности сети n и от

контура μ , который определяет траекторию движения мобильного шлюза:

$$L_p^{count}(n, \mu) = \frac{4}{6p} \left((\mu - 1)\mu(2\mu - 1) + 6\frac{n}{2} \left(\frac{n}{2} - \mu \right) \left(1 + \frac{n}{2} - \mu \right) \right). \quad (17)$$

Подстановка полученной формулы (17) в формулу (12) дает полные энергетические затраты на пересылку сообщений.

Общее ожидаемое число пересылок, которое совершает один обыкновенный сенсор, находящийся на k -ом уровне удаленности от референсного определится выражением

$$L_p^{count}(k) = \frac{1}{S_k^{in} + S_k^{out}} \left(\sum_{i=k}^{\mu-1} \frac{S_i^{in}}{p} + \sum_{i=k}^{2(v-\mu)} \frac{S_i^{out}}{p} \right). \quad (18)$$

Общее ожидаемое число пересылок через референсные узлы равно

$$L_p^{count}(k=0) = \frac{n^2}{4(2\mu-1)p}. \quad (19)$$

Зная загруженность узлов, можно определить время их жизни и оценить время жизни всей сети подставив формулы (17) и (19) в формулу (9).

Получили, что общее ожидаемое число пересылок на референсные узлы зависит от трех параметров – от вероятности доставки p , от размерности сети n и от контура μ , который определяет траекторию движения мобильного шлюза.

Таким образом, в гибридной ПБСС появляется еще один параметр оптимизации – длина контура μ , который является траекторией движения мобильного шлюза.

Если мы будем искать оптимальный контур исходя из энергетических соображений, не принимая во внимание время съема информации в сети, то это легко сделать, найдя производную от $L_p^{count}(n, \mu)$ в формуле (17).

$$\frac{dL_p^{count}}{d\mu}(n, \mu) = 0.$$

Корень получающегося при этом квадратного уравнения и есть оптимальный контур $\mu = \mu_{opt}$:

$$\mu_{opt} \approx \frac{(\sqrt{3}-1)n+1}{2} \approx 0,366n + 0,5.$$

Полученные вероятностные оценки были использованы для проведения вычислительных экспериментов с целью сравнения функциональных характеристик стационарной и гибридной ПБСС. Результаты приведены ниже.

5. Результаты моделирования. Для вычислительных экспериментов были использованы следующие значения проектных параметров (таблица 3).

Таблица 3. Значения проектных параметров

Наименование проектного параметра	Обозначение, единица измерения	Значение
Длина сообщения	N_{bit} , бит	256
Битовая скорость	f_{bit} , кбит/с	12,8
Несущая частота модема	f , кГц	60
Полоса пропускания	B , кГц	30
Мощность модема в режиме отправки сообщения	P_s , Вт	25
Мощность модема в режиме приема сообщения	P_r , Вт	0,3
Мощность модема в режиме ожидания	P_w , Вт	0,3
Мощность модема в режиме сбора информации (замера характеристик среды)	P_{inf} , Вт $P_{inf} \approx P_w$	0,3
Энергетическая емкость батареи напряжением 12 В	E_0 , кДж	864

Указанные в таблице параметры соответствуют параметрам используемых на практике модемов [53].

Подставляя указанные выше значения параметров в выражение (9) и переводя время в часы, получим формулу, позволяющую оценить

время жизни сенсоров $t_p(k)$ на любом уровне удаленности k от референсного

$$t_p(k) = \frac{240}{0,49L_p(k)/T_c + 0,3}. \quad (20)$$

Из формулы (20) видно, что при $0,49L_p(k)/T \ll 0,3$ время жизни сенсора на данном уровне удаленности перестает зависеть от числа передаваемых пакетов, и стремится к 800 часам. Назовем такой случай режимом низкой загрузки сенсора, а режим, при котором выполнено $4,94L_p(k)/T \approx 0,3$, режимом высокой загрузки. Если референсный агент работает в режиме низкой загрузки, то время его жизни будет очень мало зависеть от его расположения в акватории. По этой причине для исследования времени жизни рассматривается только высокая загрузка сети. В модельном исследовании, при сравнении вариантов расположения сенсорных узлов рассматривается цикл продолжительностью $T_c = 120$ с.

Будем считать, что вероятность лежит в широком диапазоне изменения $0,5 < p < 1$.

В рамках сделанных допущений ниже приводятся результаты вычислительных экспериментов, позволяющих определить оптимальный тип размещения сенсоров для различных размеров сети.

5.1. Полные ожидаемые энергетические затраты в стационарной ПБСС. Использование вероятностной оценки (12) для стационарной ПБСС позволили определить полные энергетические затраты на пересылку сообщений. Эти результаты показаны на рисунке 3. Видно, что энергетические затраты сети монотонно растут в зависимости от ее размера при любых заданных вероятностях доставки сообщения p .

При заданном размере сети n^2 и выбранном количестве референсных узлов 5% от общего количества, дальнейшая оптимизация энергоэффективности в такой модели оказывается невозможной. Чем больше размер сети n^2 , тем больше оказываются энергозатраты.

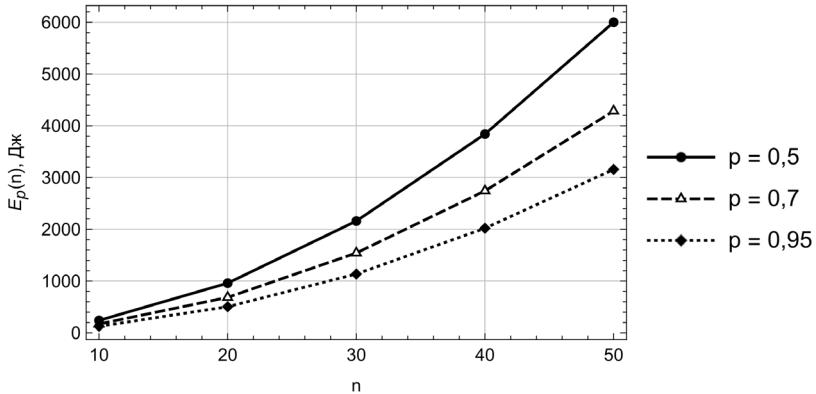


Рис. 3. Зависимость полных энергетических затрат $E_p(n)$ от линейного размера сети n в случае размещения референсных узлов в центрах подрешеток при различных вероятностях доставки сообщения p

5.2. Время жизни референсных узлов в стационарной ПБСС.

Зная загруженность узлов, можно определить время их жизни и оценить время жизни всей сети используя формулу (20). Результаты численных оценок показаны на рисунке 4.

Размер сети n^2 в данном случае не влияет на время жизни сенсоров. Достаточно рассмотреть одну произвольную подрешетку с заданной длиной стороны $l = 5$.

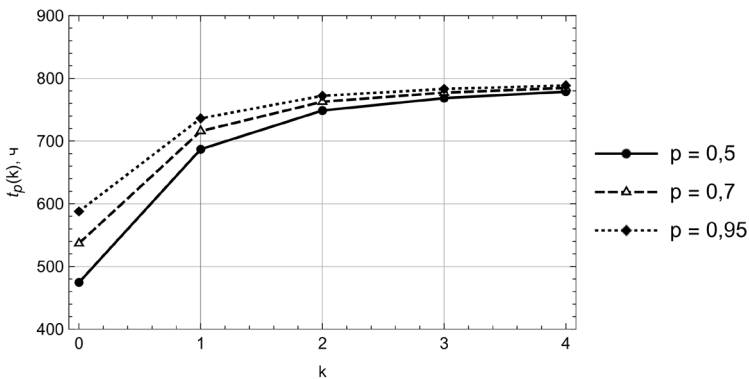


Рис. 4. Время жизни сенсоров $t_p(k)$ на разных уровнях удаленности k от референсного агента, расположенного в центре подрешетки из 25 сенсоров при длительности цикла $T_c = 120$ с

На рисунке 4 показано как изменяется время жизни сенсоров в зависимости от их удаленности от референсных узлов при длительности цикла $T_c = 120$ с.

Отметим, что для подрешетки со стороны $l = 5$ существует всего 4 уровня удаленности от референсного агента. Из графиков видно, что для уровней $k = 3$ и $k = 4$ время жизни оказывается близким к максимально возможному, равному 800 часам, что говорит о низкой загрузке этих сенсоров. Однако, для уровней 0, 1 и 2 время жизни заметно меньше, что говорит о том, что эти сенсоры выйдут из строя значительно раньше. Так, при вероятности доставки $p = 0,95$ время жизни референсного агента оказывается немного меньше 600 часов.

Эту несбалансированность стационарной сети сложно улучшить. Поскольку стационарные буи жестко связаны с расположением референсных узлов (которые находятся под ними), то после выхода референсных узлов из строя, связность сети будет нарушена. Необходимо физически переместить заякоренные буи и затем выполнить перестройку маршрутов передачи сообщений на новые выбранные референсные узлы. Процедура перемещения заякоренных буев на практике мало реализуема.

Перейдем к анализу гибридной сети.

5.3. Полные ожидаемые энергетические затраты в гибридной ПБСС. Использование вероятностной оценки (12) позволяет оценить полные энергетические затраты на пересылку сообщений в случае обхода внутреннего контура мобильным шлюзом. Для сетей с линейной размерностью $n = 2\nu$ можно рассмотреть движение шлюза по μ различным контурам, $1 \leq \mu \leq \nu$. Каждый контур характеризуется своими энергозатратами. Для наглядности на рисунке 5 представлена зависимость полных энергетических затрат за один цикл от номера контура μ при разных размерах сети для заданной вероятности доставки сообщения $p = 0,95$.

Из графиков (рисунок 5) видно, что кривые энергозатрат не монотонны, они имеют минимум, то есть для каждой сети существует контур оптимальной длины, при выборе которого в качестве референсного, энергозатраты будут минимальны. Однако, решаемая задача является многокритериальной и нельзя выбрать параметр обхода акватории не учитывая фактор времени. Исследование времени съема информации как параметра оптимизации будет исследовано в следующей работе.

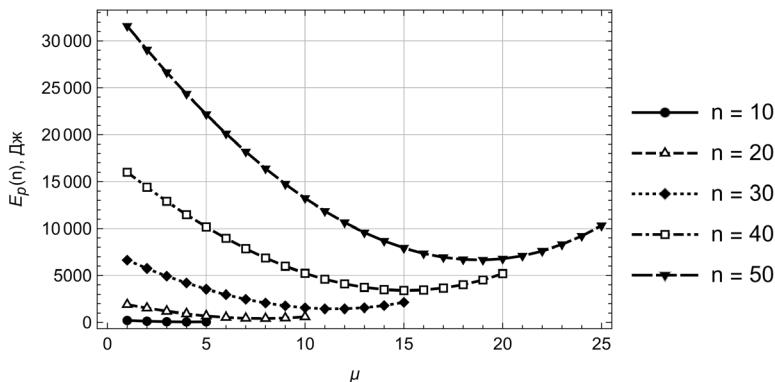


Рис. 5. Зависимость полных энергетических затрат $E_p(n)$ от номера контура μ в случае расстановки референсных узлов на внутреннем контуре при различных размерах сети n при вероятности доставки сообщения $p = 0,95$

В отличие от стационарного случая, появляется дополнительный параметр оптимизации μ , управление которым, совместно с учетом времени съема информации может улучшить энергоэффективность сети.

5.4. Время жизни референсных узлов в гибридной ПБСС.

Зная загруженность узлов, можно определить время их жизни и оценить время жизни всей сети используя формулу (20). Результаты численных оценок дали результат, который показан на графиках (рисунок 6 и рисунок 7). В данном случае размер сети n^2 оказывает существенное влияние на время жизни сенсоров. Поэтому, в отличие от стационарной сети, для гибридной было рассмотрено два случая – малая ($n = 10$) и большая ($n = 50$) сети.

На рисунке 6 для сети из 100 сенсоров показано как растет время жизни в зависимости от их удаленности от референсных узлов при длительности цикла $T_c = 120$ с. Из таблицы соответствия следует, что стационарную ПБСС для случая $n=10$ следует сравнивать с гибридной для $\mu=1$. Сравнение этих случаев на рисунке 5 и рисунке 6 показывает, что они дают приблизительно одинаковый результат около 600 часов. Однако, расстановка референсных узлов на других контурах с $\mu > 1$ может существенно увеличить время жизни. Так, при вероятности доставки $p = 0,95$ время жизни самого загруженного референсного агента на контуре $\mu = 3$ оказывается более

750 часов. Однако, необходимо иметь в виду, что это одновременно удлинит время съема информации мобильным шлюзом, что не всегда допустимо.

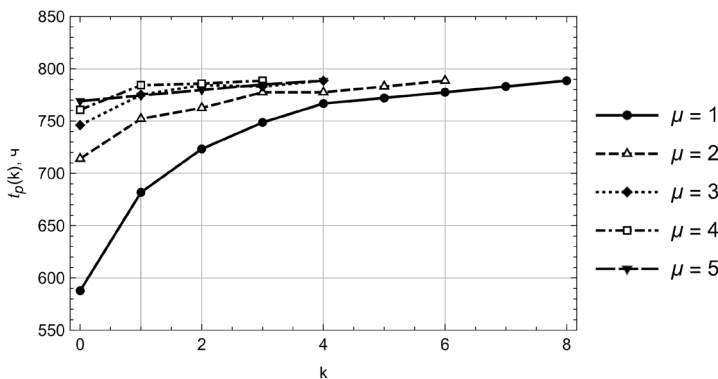


Рис. 6. Время жизни сенсоров $t_p(k)$ на разных уровнях удаленности от референсного агента в случае расстановки референсных узлов на замкнутых контурах μ акватории с вероятностью доставки сообщений $p = 0,95$ для сети из 100 сенсоров при длительности цикла $T_c = 120$ с

Теперь рассмотрим случай большой сети $n = 50$ (рисунок 7).

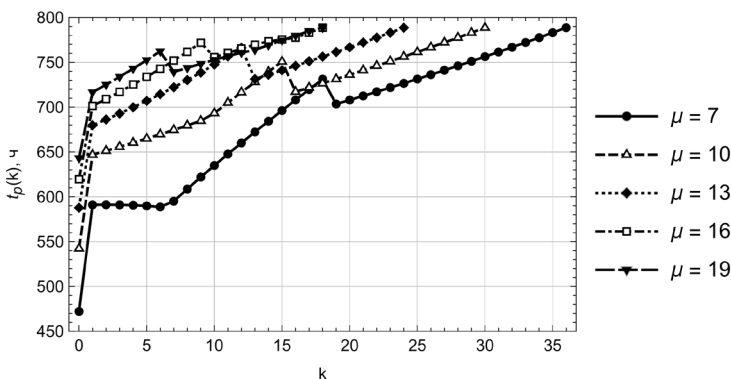


Рисунок 7. Время жизни сенсоров $t_p(k)$ на разных уровнях удаленности от референсного агента в случае расстановки референсных узлов на замкнутых контурах μ акватории с вероятностью доставки сообщений $p = 0,95$ для сети из 2500 сенсоров при длительности цикла $T_c = 120$ с

На рисунке 7 для сети из 2500 сенсоров показано как растет время жизни в зависимости от их удаленности от референсных узлов при длительности цикла $T_c = 120$ с.

Из таблицы соответствия (таблица 2) следует, что стационарную ПБСС для случая сети большой размерности – $n = 50$ следует сравнивать с гибридной ПБСС для $\mu = 13$.

Сравнение «точечной» линии в точке $k = 0$ на рисунке 4 и «штрихованной» линии с $\mu = 13$ в точке $k = 0$ на рисунке 7 показывает, что они дают приблизительно одинаковый результат по времени жизни референсного узла около 600 часов.

Причем расстановка референсных узлов на контурах с $\mu < 13$ ухудшает результат, а на контурах с $\mu > 13$ может существенно его улучшить. Из этого можно сделать вывод, что появление дополнительного параметра гибридной ПБСС, такого, как траектория обхода акватории позволяет дополнительно оптимизировать работу сети.

Кроме того, необходимо отметить, что в отличие от стационарной сети, в гибридной возможна динамическая перестройка маршрутов пересылки пакетов при уменьшении уровня заряда батарей референсных узлов до определенной критической величины. Т.е. назначение других узлов с достаточной энергией – референсными. При этом траектория движения мобильного шлюза будет согласована с положением новых референсных узлов.

6. Обсуждение результатов. Сравним между собой предложенные варианты размещения референсных узлов в акватории на стационарных и гибридных сетях различного размера при длительности цикла $T_c = 120$ с. Для моделирования работы сети и последующего сравнения была взята вероятность доставки $p = 0,95$, которая обеспечивает хорошую надежность работы сети.

Общее число сенсоров в акватории – n^2 , число референсных узлов составляет примерно 5% от общего числа сенсоров, выполнено условие $m^{sub-latt} \approx m^{count}$ (число референсных узлов в центрах подрешеток \approx числу референсных узлов на траектории движения мобильного шлюза). Для достижения этого соответствия найден соответствующий контур μ с равным числом референсных узлов.

Все выбранные технические характеристики модемов в стационарной и гибридной ПБСС одинаковы, также как и физические условия в акватории. Подстановка выбранных проектных значений в

полученные ранее формулы приводит к компактным аналитическим выражениям для энергетических затрат и времени жизни сети.

Для стационарной сети, для общих энергетических затрат при выбранных характеристиках получаем квадратичную зависимость от линейного размера сети n :

$$E_p^{sub-latt}(n) = 1,263n^2.$$

Время жизни референсного узла не зависит от размера сети, а определяется только количеством референсных узлов / подрешеток / кластеров / бுவ в акватории (у нас 5% от общего количества):

$$t_{ref}^{sub-latt} = 589 \text{ часов.}$$

Иначе дело обстоит с гибридной сетью. Тут важным параметров оптимизации оказывается траектория обхода мобильным шлюзом:

$$E_p^{count}(n) = 0,35 \left(3n \left(\frac{n}{2} - \mu + 1 \right) \left(\frac{n}{2} - \mu \right) + (2\mu - 1)(\mu - 1)\mu \right).$$

Энергозатраты зависят от размера сети как n^3 и, при неудачном выборе контура размещения референсных узлов будут расти очень быстро при росте n . Однако, несмотря на кубическую зависимость энергетических затрат от линейного размера сети n , рост энергозатрат можно компенсировать выбором оптимальной траектории обхода.

Время жизни референсного узла зависит как от размера сети, так и от траектории обхода, однако, учитывая оценку, которая следует из формулы (16) (с точностью до величины округления) $n^2 / (2\mu - 1) \approx 100$, получаем оценку для времени жизни референсного узла гибридной сети, которое находится в согласии с временем жизни в стационарной сети:

$$t_{ref}^{count} = \frac{2400}{0,01 \frac{n^2}{(2\mu - 1)} + 3} \approx 600 \text{ часов.}$$

При выбранных условиях в таблице 4 приведено сравнение полных энергетических затрат сети, связанных с пересылками

сообщений $E = E_p(n)$ и времени жизни референсных узлов $t_{ref} = t_p(0)$.

Таблица 4. Значения проектных параметров

Размер сети	Характеристики работы стационарной сети			Характеристики работы гибридной сети			
	число референсных узлов $m^{sub-latt}$	$E^{sub-latt}$	$t_{ref}^{sub-latt}$	μ	число референсных узлов m^{count}	E^{count}	t_{ref}^{count}
100	4	127	589	1	4	210	600
400	16	505	589	3	20	1186	631
900	36	1137	589	5	36	3528	600
1600	64	2021	589	9	68	5972	609
2500	200	3158	589	13	200	9555	600

Из анализа результатов, приведенных в таблице 4, следует, что время жизни сети является величиной, зависящей в основном от размера кластера, передающего сообщения на один референсный узел, а не от размера сети и количества пересылок, требующихся для доставки сообщения. Поэтому, при одинаковом соотношении обычных и референсных узлов и стационарная, и гибридная сеть дают приблизительно одинаковые результаты.

Энергоэффективность существенно зависит от размера сети и количества пересылок по ней. В стационарной ПБСС, разбитой на подрешетки, энергозатраты растут по квадратичному закону с ростом линейного размера сети, а в гибридной – по кубическому.

7. Заключение. В работе выполнен сравнительный анализ основных функциональных характеристик стационарной (с закоренными буями) и гибридной (с мобильным шлюзом) ПБСС для различных вариантов размещения сенсорных узлов в акватории.

Математический аппарат, основанный на вероятностном подходе, позволил оценить характеристики рассмотренных коммуникационных архитектур ПБСС с энергетической точки зрения – определить общие энергетические затраты сети на пересылку сообщений и время жизни сенсоров сети.

Результаты моделирования функциональных характеристик сетей для рассмотренных коммуникационных архитектур позволили сделать несколько важных выводов.

Во-первых, время жизни сети является величиной, зависящей в основном от размера кластера. Поэтому, при одинаковом соотношении обычных и референсных узлов и стационарная, и гибридная сеть дают приблизительно одинаковые результаты времени жизни сенсоров на одинаковом расстоянии от референсных.

Во-вторых, расположение сенсоров, соответствующее стационарной ПБСС, обладает плохой сбалансированностью с точки зрения времени жизни сенсоров на разных уровнях удаленности от референсного. Поскольку стационарные буи жестко связаны с расположением референсных узлов (которые находятся под ними), то после выхода референсов из строя, связность сети будет нарушена. Необходимо физически переместить заякоренные буи и затем выполнить перестройку маршрутов передачи сообщений на новые выбранные референсные узлы.

Другим способом сохранения связности в стационарной ПБСС является увеличение числа подрешеток, т.е. числа заякоренных буешлюзов, однако это приведет к существенному увеличению стоимости сети.

Эти проблемы решаются проще в гибридной сети. В ней возможна динамическая перестройка маршрутов при уменьшении уровня заряда батарей референсных узлов, не дожидаясь полного нарушения связности сети. Для этого достаточно изменить траекторию движения мобильного шлюза.

Таким образом, коммуникационная архитектура гибридной ПБСС оказывается более гибкой и настраиваемой с точки зрения сохранения связности и увеличения времени жизни сети. Кроме того, не обладая явными преимуществами с точки зрения энергозатрат, она, тем не менее, допускает увеличение энергоэффективности при использовании дополнительного параметра оптимизации – траектории мобильного шлюза. И, наконец, с экономической точки зрения – гибридная сеть является более дешевым решением. Перечисленные факты говорят о преимуществах использования гибридной ПБСС, перед стационарной.

При этом мобильный элемент – волновой глайдер, выполняющий роль межсервного шлюза, по сравнению с другими робототехническими платформами, способен функционировать в акватории в течение продолжительного времени, зачастую превосходящего время жизни сети. Это говорит о перспективности его

использования для прикладных задач сбора, накопления и ретрансляции информации при реализации различных политик обслуживания ПБСС.

В рамках предложенной стохастической модели авторами были исследованы другие важные функциональные характеристики гибридной ПБСС – надежность, пропускная способность сети, задержки передачи пакетов и коэффициент их доставки, показывающие потенциальные возможности такой сети. Эти результаты не вошли в настоящую работу и будут представлены в отдельной статье.

Литература

1. Ryabinin V., Barbiere J., Haugan P., Kullenberg, G., Smith, N., McLean, C., Trois A., Fischer A., Arico S., Aarup T., Pissierssens P., Visbeck M., Enevoldsen H., Rigaud J. The UN Decade of Ocean Science for Sustainable Development // *Oceanobs'19: An Ocean of Opportunity*. 2019. vol. 6. DOI: 10.3389/fmars.2019.00470.
2. Mohsan S., Li Y., Sadiq M., Liang J., Khan M. Recent Advances, Future Trends, Applications and Challenges of Internet of Underwater Things (IoUT): A Comprehensive Review // *Journal of Marine Science and Engineering*. 2023. no. 11. DOI: 10.3390/jmse11010124.
3. Nkenyereye L., Nkenyereye L., Ndibanje B. Internet of Underwater Things: A Survey on Simulation Tools and 5G-Based Underwater Networks // *Electronics*. 2024. vol. 13(3). DOI: 10.3390/electronics13030474.
4. Riser S., Freeland H., Roemmich D., et al. Fifteen years of ocean observations with the global Argo array // *Nature Climate Change*. 2016. vol. 6. no. 2. pp. 145–153. DOI: 10.1038/NCLIMATE2872.
5. Crout R., Conlee D., Bernard L. National Data Buoy Center (NDBC) National Backbone Contributions to the Integrated Ocean Observation System (IOOS) // *OCEANS*. 2006. pp. 1–3. DOI: 10.1109/OCEANS.2006.307073.
6. Ho T., Hagaseth M., Rialland A., et al. Internet of Things at Sea: Using AIS and VHF over Satellite in Remote Areas // *Proceedings of 7th Transport Research Arena (TRA 2018)*. Vienna, Austria. 2018. pp. 1–10. DOI: 10.5281/zenodo.1473565.
7. Oke P., Sakov P. Assessing the footprint of a regional ocean observing system // *Journal of Marine Systems*. 2012. vol. 105. pp. 30–51. DOI: 10.1016/j.jmarsys.2012.05.009.
8. Barnes C., NEPTUNE Canada Team. Building the world's first regional cabled ocean observatory (NEPTUNE): Realities, challenges and opportunities // *OCEANS*. 2007. pp. 1–8. DOI: 10.1109/OCEANS.2007.4449319.
9. Wallace D., et al. A Canadian contribution to an integrated Atlantic ocean observing system (IAOOS) // *Oceans – St. John's*. 2014. pp. 1–10. DOI: 10.1109/OCEANS.2014.7003244.
10. Kaneda Y., Kawaguchi K., Araki E., Matsumoto H., Nakamura T., Kamiya S., Ariyoshi K., Hori T. Dense Ocean Floor Network for Earthquakes and Tsunamis (DONET) Around the Nankai trough Mega Thrust Earthquake Seismogenic Zone in Southwestern Japan –Part 2: Real Time Monitoring of the Seismogenic Zone // *Proceedings of International Conference on Offshore Mechanics and Arctic Engineering*. Honolulu, Hawaii, USA. 2009. vol. 43468. pp. 715–720. DOI: 10.1115/OMAE2009-79599.

11. Grund M., Freitag L., Preisig J., Ball K. The PLUSNet Underwater Communications System: Acoustic Telemetry for Undersea Surveillance // OCEANS. 2006. pp. 1–5. DOI: 10.1109/OCEANS.2006.307036.
12. Felemban E., Shaikh F., Qureshi U.M., Sheikh A., Qaisar S. Underwater Sensor Network Applications: A Comprehensive Survey // International Journal of Distributed Sensor Networks. 2015. vol. 11(11). DOI: 10.1155/2015/896832.
13. Fattah S., Gani A., Ahmedy I., Idris M., Targio Hashem I. A Survey on Underwater Wireless Sensor Networks: Requirements, Taxonomy, Recent Advances, and Open Research Challenges // Sensors. 2020. vol. 20(18). DOI: 10.3390/s20185393.
14. Brekhovskikh L., Lysanov Yu. Fundamentals of Ocean Acoustics, 3rd Ed. // NY: Springer New York. 2003. 279 p.
15. Sandhiyaa S., Gomathy C. A Survey on underwater wireless sensor networks: challenges, requirements, and opportunities // Proceedings of Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). Palladam, India. 2021. pp. 1417–1427. DOI: 10.1109/I-SMAC52330.2021.9640791.
16. Mahmood T., Akhtar F., Ur Rehman K., Ali S., Mokbal F. A Comprehensive Survey on the Performance Analysis of Underwater Wireless Sensor Networks (UWSN) Routing Protocols // International Journal of Advanced Computer Science and Applications. 2009. vol. 10. no. 5. pp. 590–600. DOI: 10.14569/IJACSA.2019.0100576.
17. Khan A., Ali I., Ghani A., Khan N., Alsaqer M., Rahman A., Mahmood H. Routing Protocols for Underwater Wireless Sensor Networks: Taxonomy, Research Challenges, Routing Strategies and Future Directions // Sensors. 2018. vol. 18(5). DOI: 10.3390/s18051619.
18. Su X., Ullah I., Liu X., Choi D. A Review of Underwater Localization Techniques, Algorithms, and Challenges // Journal of Sensors. 2020. vol. 2020(1). DOI: 10.1155/2020/6403161.
19. Xing G., Chen Y., He L., Su W., Hou R., Li W., Zhang C., Chen X. Energy Consumption in Relay Underwater Acoustic Sensor Networks for NDN // IEEE Access. 2009. vol. 7. pp. 42694–42702. DOI: 10.1109/ACCESS.2019.2907693.
20. Gupta S., Singh N. Underwater wireless sensor networks: a review of routing protocols, taxonomy, and future directions // Journal of Supercomputing. 2024. vol. 80. pp. 5163–5196. DOI: 10.1007/s11227-023-05646-w.
21. Gola K., Arya S. Underwater acoustic sensor networks: Taxonomy on applications, architectures, localization methods, deployment techniques, routing techniques, and threats: A systematic review // Concurrency and Computational: Practice and Experience. 2023. vol. 35. DOI: 10.1002/cpe.7815.
22. Shovon I., Shin S. Survey on Multi-Path Routing Protocols of Underwater Wireless Sensor Networks: Advancement and Applications // Electronics. 2022. vol. 11(21). DOI: 10.3390/electronics11213467.
23. Pu W. A Survey of Localization Techniques for Underwater Wireless Sensor Networks // Journal of Computing and Electronic Information Management. 2023. vol. 11. no. 1. pp. 10–15. DOI: 10.54097/jceim.v11i1.9184.
24. Rani S., Anju, Sangwan A., Kumar K., Nisar K., Soomro T., Ibrahim A., Gupta M., Chand M., Khan S. A Review and Analysis of Localization Techniques in Underwater Wireless Sensor Networks // Computers, Materials and Continua. 2023. vol. 75. no. 3. pp. 5697–5715. DOI: 10.32604/cmc.2023.033007.
25. Li X., Xu X., Yan L., Zhao H. Zhang T. Energy-Efficient Data Collection Using Autonomous Underwater Glider: A Reinforcement Learning Formulation // Sensors. 2020. vol. 20. no. 13. DOI: 10.3390/s20133758.

26. Su Y., Zhang L., Li Y., Yao X. A Glider-Assist Routing Protocol for Underwater Acoustic Networks with Trajectory Prediction Methods // IEEE Access. 2020. vol. 8. pp. 154560–154572. DOI: 10.1109/ACCESS.2020.3015856.
27. Lan H., Lv Y., Jin J., Li J., Sun D., Yang Z. Acoustical Observation with Multiple Wave Gliders for Internet of Underwater Things // IEEE Internet of Things Journal. 2021. vol. 8. no. 4. pp. 2814–2825. DOI: 10.1109/JIOT.2020.3020862.
28. Никушенко Д.В., Рыжов В.А., Тряскин Н.В. Моделирование гидродинамических характеристик волнового глайдера // XII Всероссийский съезд по фундаментальным проблемам теоретической и прикладной механики: Сборник трудов (г. Уфа, 19-24 августа 2019 г.). Уфа, ИПСМ РАН, 2019. С. 71–73.
29. Овчинников К.Д., Рыжов В.А., Сенишин А.А., Кожемякин И.В. Экспериментальное исследование ходовых характеристик волнового глайдера // XV Всероссийская научно-практическая конференция «Перспективные системы и задачи управления»: Сборник материалов (Ростов-на-Дону – Таганрог, 5-9 октября 2020). Ростов-на-Дону – Таганрог: Изд-во Южного федерального университета. 2020. С. 91–97.
30. Ghanem M., Mansoor A., Ahmad R. A systematic literature review on mobility in terrestrial and underwater wireless sensor networks // International Journal of Communication Systems. 2021. vol. 34. no. 10. DOI: 10.1002/dac.4799.
31. Cui J.-H., Kong J., Gerla M., Zhou S. The challenges of building scalable mobile underwater wireless sensor networks for aquatic applications // IEEE Network. 2006. vol. 20. no. 3. pp. 12–18. DOI: 10.1109/MNET.2006.1637927.
32. Alkindi Z., Alzeidi N., Arafehand B., Touzene A. Performance evolution of grid-based routing protocol for underwater wireless sensor networks under different mobile models // International Journal of Wireless and Mobile Networks (IJWMN). 2018. vol. 10. no. 1. pp. 13–25. DOI: 10.5121/ijwmn.2018.10102.
33. Janardanan Kartha J., Jacob L. Delay and Lifetime Performance of Underwater Wireless Sensor Networks with Mobile Element Based Data Collection // International Journal of Distributed Sensor Networks. 2015. vol. 11. no. 5. DOI: 10.1155/2015/128757.
34. Jalaja M., Jacob L. On-demand data collection in sparse underwater acoustic sensor networks using mobile elements // Proceedings of the 10th International Conference on Wireless and Mobile Communication (ICWMC 2014). Seville, Spain. 2014. pp. 105–111.
35. Yoon S., Azad A., Oh H., Kim S. AURP: an AUV-aided underwater routing protocol for underwater acoustic sensor networks // Sensors. 2012. vol. 12. no. 2. pp. 1827–1845. DOI: 10.3390/s120201827.
36. Hollinger G., Choudhary S., Qarabaqi P., Murphy C., Mitra U., Sukhatme G., Stojanovic M., Singh H., Hover F. Underwater data collection using robotic sensor networks // IEEE Journal on Selected Areas in Communications. 2012. vol. 30. no. 5. pp. 899–911. DOI: 10.1109/jsac.2012.120606.
37. Rahim M., Casari P., Guerra F., Zorzi M. On the performance of delay—tolerant routing protocols in underwater networks // Proceedings of OCEANS 2011 IEEE. Santander, Spain. 2011. pp. 1–7. DOI: 10.1109/Oceans-Spain.2011.6003388.
38. Choudhary M., Goyal N. A rendezvous point-based data gathering in underwater wireless sensor networks for monitoring applications // International Journal of Communication System. 2022. vol. 35. no. 6. DOI: 10.1002/dac.5078.
39. Sharma S., Puthal D., Jena S., Zomaya A., Ranjan R. Rendezvous based routing protocol for wireless sensor networks with mobile sink // The Journal of Supercomputing. 2017. vol. 73. pp. 1168–1188. DOI: 10.1007/s11227-016-1801-0.

40. Cheng C., Li L. Data gathering problem with the data importance consideration in Underwater Wireless Sensor Networks // *Journal of Network and Computer Applications*. 2017. vol. 78. pp. 300–312. DOI: 10.1016/j.jnca.2016.10.010.
41. Maqsood H., Javaid N., Yahya A., Ali B., Khan Z., Qasim U. MobiL-AUV: AUV-aided localization scheme for underwater wireless sensor networks // *Proceedings of the 2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS)*. Fukuoka, Japan. 2016. pp. 170–175. DOI: 10.1109/IMIS.2016.140.
42. Favaro F., Casari P., Guerra F., Zorzi M. Data upload from a static underwater network to an AUV: Polling or random access? // *Proceedings of the IEEE OCEANS 2012*. Yeosu, Republic of Korea. 2012. pp. 1–6. DOI: 10.1109/OCEANS-Yeosu.2012.6263499.
43. Favaro F., Brolo L., Toso G., Casari P., Zorzi M. A study on remote data retrieval strategies in underwater acoustic networks // *Proceedings of the MTS/IEEE OCEANS Conference*. San Diego, USA. 2013. pp. 1–8.
44. Cai S., Zhu Y., Wang T., Xu G., Liu A., Liu X. Data Collection in Underwater Sensor Networks based on Mobile Edge Computing // *IEEE Access*. 2019. vol. 7. pp. 6535–65367. DOI: 10.1109/ACCESS.2019.2918213.
45. Yan J., Yang X., Luo X., Chen C. Energy-efficient data collection over AUV-assisted underwater acoustic sensor network // *IEEE Systems Journal*. 2018. vol. 12. no. 4. pp. 3519–3530. DOI: 10.1109/JSYST.2017.2789283.
46. Khan M., Ahmed S., Jembre Y., Kim D. An energy-efficient data collection protocol with AUV path planning in the Internet of Underwater Thing // *Journal of Network and Computer Applications*. 2019. vol. 135. pp. 20–31. DOI: 10.1016/j.jnca.2019.02.025.
47. Li H., Wu X., Zhang Z., Tan X., Pan J., Dai C., Luo D., Ahmed A., Xu Y. An extended-range wave-powered autonomous underwater vehicle applied to underwater wireless sensor networks // *iScience*, 2022. vol. 25. no. 8. DOI: 10.1016/j.isci.2022.104738.
48. Fedorova T., Ryzhov V., Semenov N, Sulaiman S. (2022) Optimization of an Underwater Wireless Sensor Network Architecture with Wave Glider as a Mobile Gateway // *J. Marine. Sci. Appl*. 2022. vol. 21. pp. 179–196. DOI: 10.1007/s11804-022-00268-9.
49. Rappaport T. *Wireless Communications: Principles and Practice* // Upper Saddle River, NJ: Prentice Hall. 1996. 656 p.
50. Thorp W. Deep Sound Attenuation in the Sub and Low Kilocycle per-second Range // *The Journal of the Acoustical Society of America*. 1965. vol. 38. no. 4. p. 648–654.
51. Ahn J., Syed A., Krishnamachari B., Heidemann J. Design and analysis of a propagation delay tolerant ALOHA protocol for underwater networks // *Ad Hoc Networks*. 2011. vol. 9. pp. 752–766. DOI: 10.1016/j.adhoc.2010.09.007.
52. Lindsey S., Raghavendra C. PEGASIS: Power-efficient gathering in sensor information systems // *Proceedings of IEEE Aerospace Conference*. USA: Big Sky, 2002. vol. 3. DOI: 10.1109/AERO.2002.1035242.
53. Шарафудинова Т.К. Сравнительный анализ промышленных гидроакустических модемов // *Вестник Государственного университета морского и речного флота имени адмирала С.О. Макарова*. 2021. № 13(6). С. 832–841. DOI: 10.21821/2309-5180-2021-13-6-832-841.

Федорова Татьяна Александровна — канд. физ.-мат. наук, доцент кафедры, кафедра прикладной математики и математического моделирования факультета цифровых и промышленных технологий, Санкт-Петербургский государственный морской технический университет. Область научных интересов: теоретическая физика,

мультиагентные системы, моделирование сложных динамических систем, подводные беспроводные сенсорные сети. Число научных публикаций — 46. fedorova.tan@gmail.com; улица Лоцманская, 3, 190121, Санкт-Петербург, Россия; р.т.: +7(812)495-2648.

Рыжов Владимир Александрович — д-р техн. наук, профессор, заведующий кафедрой, кафедра прикладной математики и математического моделирования факультета цифровых и промышленных технологий, Санкт-Петербургский государственный морской технический университет. Область научных интересов: морская робототехника, мультиагентные системы, гидродинамика. Число научных публикаций — 120. varyzhov@smtu.ru; улица Лоцманская, 3, 190121, Санкт-Петербург, Россия; р.т.: +7(812)495-2648.

Сафронов Кирилл Сергеевич — старший преподаватель кафедры, кафедра прикладной математики и математического моделирования факультета цифровых и промышленных технологий, Санкт-Петербургский государственный морской технический университет. Область научных интересов: теория вероятности, математическая статистика, машинное обучение, математическая акустика, мультиагентные системы. Число научных публикаций — 12. safronov.kirill.pm@gmail.com; улица Лоцманская, 3, 190121, Санкт-Петербург, Россия; р.т.: +7(812)495-2648.

Поддержка исследований. Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках программы научного центра мирового уровня «Передовые цифровые технологии» (договор № 075-15-2022-312 от 20.04.2022).

T. FEDOROVA, V. RYZHOV, K. SAFRONOV

THE USE OF HYBRID COMMUNICATION ARCHITECTURE IN UNDERWATER WIRELESS SENSOR NETWORKS TO ENHANCE THEIR LIFETIME AND EFFICIENCY

Fedorova T., Ryzhov V., Safronov K. The Use of Hybrid Communication Architecture in Underwater Wireless Sensor Networks to Enhance Their Lifetime and Efficiency.

Abstract. The paper presents a comparative analysis of the main functional characteristics of underwater wireless sensor networks (UWSNs) with stationary and hybrid communication architectures. The UWSNs consist of sensor nodes located on the seabed and intermodal gateways facilitating the transmission of information packets between the underwater and above-water segments of the network. In the stationary UWSNs, anchored buoys serve as gateways, while in the hybrid UWSNs, mobile transport platforms fulfill this role. Using a mathematical framework based on a probabilistic approach, an evaluation of the functional characteristics of alternative communication architectures for UWSNs is performed from an energy perspective. The overall energy consumption of the network for message transmission and the sensor network's lifespan are determined. or the numerical analysis of the functional characteristics of UWSNs, a wide range of design parameters is considered, such as the size of the water area, the required number and placement options of sensor nodes, and the packet delivery probability in the water area (physical parameters of the environment). The search for "optimal" solutions from an energy standpoint is conducted within these parameter ranges. The conducted research demonstrates that mobility plays a crucial role in improving the functioning of underwater networks in terms of coverage (ensuring connectivity), energy efficiency, and lifetime. The mobile element, represented by the wave glider acting as an intermodal gateway, is capable of sustained operation in the water area for an extended period, indicating its potential for practical tasks such as data collection, storage, and information relay within the context of the Internet of Underwater Things.

Keywords: underwater wireless sensor networks, hybrid networks, mobile intermodal gateway, hydroacoustic communication, network efficiency criterion, network lifetime.

References

1. Ryabinin V., Barbieri J., Haugan P., Kullenberg, G., Smith, N., McLean, C., Troisi A., Fischer A., Arico S., Aarup T., Pissierssens P., Visbeck M., Enevoldsen H., Rigaud J. The UN Decade of Ocean Science for Sustainable Development. Oceanobs'19: An Ocean of Opportunity. 2019. vol. 6. DOI: 10.3389/fmars.2019.00470.
2. Mohsan S., Li Y., Sadiq M., Liang J., Khan M. Recent Advances, Future Trends, Applications and Challenges of Internet of Underwater Things (IoUT): A Comprehensive Review. Journal of Marine Science and Engineering. 2023. no. 11. DOI: 10.3390/jmse11010124.
3. Nkenyereye L., Nkenyereye L., Ndibanje B. Internet of Underwater Things: A Survey on Simulation Tools and 5G-Based Underwater Networks. Electronics. 2024. vol. 13(3). DOI: 10.3390/electronics13030474.
4. Riser S., Freeland H., Roemmich D., et al. Fifteen years of ocean observations with the global Argo array. Nature Climate Change. 2016. vol. 6. no. 2. pp. 145–153. DOI: 10.1038/NCLIMATE2872.

5. Crout R., Conlee D., Bernard L. National Data Buoy Center (NDBC) National Backbone Contributions to the Integrated Ocean Observation System (IOOS). OCEANS. 2006. pp. 1–3. DOI: 10.1109/OCEANS.2006.307073.
6. Ho T., Hagaseth M., Rialland A., et al. Internet of Things at Sea: Using AIS and VHF over Satellite in Remote Areas. Proceedings of 7th Transport Research Arena (TRA 2018). Vienna, Austria. 2018. pp. 1–10. DOI: 10.5281/zenodo.1473565.
7. Oke P., Sakov P. Assessing the footprint of a regional ocean observing system. Journal of Marine Systems. 2012. vol. 105. pp. 30–51. DOI: 10.1016/j.jmarsys.2012.05.009.
8. Barnes C., NEPTUNE Canada Team. Building the world’s first regional cabled ocean observatory (NEPTUNE): Realities, challenges and opportunities. OCEANS. 2007. pp. 1–8. DOI: 10.1109/OCEANS.2007.4449319.
9. Wallace D., et al. A Canadian contribution to an integrated Atlantic ocean observing system (IAOOS). Oceans – St. John’s. 2014. pp. 1–10. DOI: 10.1109/OCEANS.2014.7003244.
10. Kaneda Y., Kawaguchi K., Araki E., Matsumoto H., Nakamura T., Kamiya S., Ariyoshi K., Hori T. Dense Ocean Floor Network for Earthquakes and Tsunamis (DONET) Around the Nankai trough Mega Thrust Earthquake Seismogenic Zone in Southwestern Japan –Part 2: Real Time Monitoring of the Seismogenic Zone. Proceedings of International Conference on Offshore Mechanics and Arctic Engineering. Honolulu, Hawaii, USA. 2009. vol. 43468. pp. 715–720. DOI: 10.1115/OMAE2009-79599.
11. Grund M., Freitag L., Preisig J., Ball K. The PLUSNet Underwater Communications System: Acoustic Telemetry for Undersea Surveillance. OCEANS. 2006. pp. 1–5. DOI: 10.1109/OCEANS.2006.307036.
12. Felemban E., Shaikh F., Qureshi U.M., Sheikh A., Qaisar S. Underwater Sensor Network Applications: A Comprehensive Survey. International Journal of Distributed Sensor Networks. 2015. vol. 11(11). DOI: 10.1155/2015/896832.
13. Fattah S., Gani A., Ahmedy I., Idris M., Targio Hashem I. A Survey on Underwater Wireless Sensor Networks: Requirements, Taxonomy, Recent Advances, and Open Research Challenges. Sensors. 2020. vol. 20(18). DOI: 10.3390/s20185393.
14. Brekhovskikh L., Lysanov Yu. Fundamentals of Ocean Acoustics, 3rd Ed. NY: Springer New York. 2003. 279 p.
15. Sandhiyaa S., Gomathy C. A Survey on underwater wireless sensor networks: challenges, requirements, and opportunities. Proceedings of Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). Palladam, India. 2021. pp. 1417–1427. DOI: 10.1109/I-SMAC52330.2021.9640791.
16. Mahmood T., Akhtar F., Ur Rehman K., Ali S., Mokbal F. A Comprehensive Survey on the Performance Analysis of Underwater Wireless Sensor Networks (UWSN) Routing Protocols. International Journal of Advanced Computer Science and Applications. 2009. vol. 10. no. 5. pp. 590–600. DOI: 10.14569/IJACSA.2019.0100576.
17. Khan A., Ali I., Ghani A., Khan N., Alsaqer M., Rahman A., Mahmood H. Routing Protocols for Underwater Wireless Sensor Networks: Taxonomy, Research Challenges, Routing Strategies and Future Directions. Sensors. 2018. vol. 18(5). DOI: 10.3390/s18051619.
18. Su X., Ullah I., Liu X., Choi D. A Review of Underwater Localization Techniques, Algorithms, and Challenges. Journal of Sensors. 2020. vol. 2020(1). DOI: 10.1155/2020/6403161.
19. Xing G., Chen Y., He L., Su W., Hou R., Li W., Zhang C., Chen X. Energy Consumption in Relay Underwater Acoustic Sensor Networks for NDN. IEEE Access. 2009. vol. 7. pp. 42694–42702. DOI: 10.1109/ACCESS.2019.2907693.

20. Gupta S., Singh N. Underwater wireless sensor networks: a review of routing protocols, taxonomy, and future directions. *Journal of Supercomputing*. 2024. vol. 80. pp. 5163–5196. DOI: 10.1007/s11227-023-05646-w.
21. Gola K., Arya S. Underwater acoustic sensor networks: Taxonomy on applications, architectures, localization methods, deployment techniques, routing techniques, and threats: A systematic review. *Concurrency and Computational: Practice and Experience*. 2023. vol. 35. DOI: 10.1002/cpe.7815.
22. Shovon I., Shin S. Survey on Multi-Path Routing Protocols of Underwater Wireless Sensor Networks: Advancement and Applications. *Electronics*. 2022. vol. 11(21). DOI: 10.3390/electronics11213467.
23. Pu W. A Survey of Localization Techniques for Underwater Wireless Sensor Networks. *Journal of Computing and Electronic Information Management*. 2023. vol. 11. no. 1. pp. 10–15. DOI: 10.54097/jceim.v11i1.9184.
24. Rani S., Anju, Sangwan A., Kumar K., Nisar K., Soomro T., Ibrahim A., Gupta M., Chand M., Khan S. A Review and Analysis of Localization Techniques in Underwater Wireless Sensor Networks. *Computers, Materials and Continua*. 2023. vol. 75. no. 3. pp. 5697–5715. DOI: 10.32604/cmc.2023.033007.
25. Li X., Xu X., Yan L., Zhao H. Zhang T. Energy-Efficient Data Collection Using Autonomous Underwater Glider: A Reinforcement Learning Formulation. *Sensors*. 2020. vol. 20. no. 13. DOI: 10.3390/s20133758.
26. Su Y., Zhang L., Li Y., Yao X. A Glider-Assist Routing Protocol for Underwater Acoustic Networks with Trajectory Prediction Methods. *IEEE Access*. 2020. vol. 8. pp. 154560–154572. DOI: 10.1109/ACCESS.2020.3015856.
27. Lan H., Lv Y., Jin J., Li J., Sun D., Yang Z. Acoustical Observation with Multiple Wave Gliders for Internet of Underwater Things. *IEEE Internet of Things Journal*. 2021. vol. 8. no. 4. pp. 2814–2825. DOI: 10.1109/JIOT.2020.3020862.
28. Nikushchenko D., Ryzhov V., Triaskin N. [Modeling the hydrodynamic characteristics of a wave glider] XII Vserossijskij s'ezd po fundamental'nym problemam teoreticheskoj i prikladnoj mehaniki: Sbornik trudov [XII All-Russian Congress on Fundamental Problems of Theoretical and Applied Mechanics: Collected papers]. Ufa., IPSM RAN, 2019. pp. 71–73. (In Russ.).
29. Ovchinnikov K., Ryzhov V., Sinishin A., Kozhemiakin I. [Experimental study of the performance characteristics of a wave glider] XV Vserossijskaja nauchno-prakticheskaja konferencija «Perspektivnye sistemy i zadachi upravlenija»: Sbornik materialov [XV All-Russian Scientific and Practical Conference "Promising Systems and Control Problems": Collected materials]. Rostov-na-Donu – Taganrog: Izd-vo Juzhnogo federal'nogo universiteta, 2020. pp. 91–97. (In Russ.).
30. Ghanem M., Mansoor A., Ahmad R. A systematic literature review on mobility in terrestrial and underwater wireless sensor networks. *International Journal of Communication Systems*. 2021. vol. 34. no. 10. DOI: 10.1002/dac.4799.
31. Cui J.-H., Kong J., Gerla M., Zhou S. The challenges of building scalable mobile underwater wireless sensor networks for aquatic applications. *IEEE Network*. 2006. vol. 20. no. 3. pp. 12–18. DOI: 10.1109/MNET.2006.1637927.
32. Alkindi Z., Alzeidi N., Arafehband B., Touzene A. Performance evolution of grid-based routing protocol for underwater wireless sensor networks under different mobile models. *International Journal of Wireless and Mobile Networks (IJWMN)*. 2018. vol. 10. no. 1. pp. 13–25. DOI: 10.5121/ijwmn.2018.10102.
33. Janardanan Kartha J., Jacob L. Delay and Lifetime Performance of Underwater Wireless Sensor Networks with Mobile Element Based Data Collection. *International Journal of Distributed Sensor Networks*. 2015. vol. 11. no. 5. DOI: 10.1155/2015/128757.

34. Jalaja M., Jacob L. On-demand data collection in sparse underwater acoustic sensor networks using mobile elements. Proceedings of the 10th International Conference on Wireless and Mobile Communication (ICWMC 2014). Seville, Spain. 2014. pp. 105–111.
35. Yoon S., Azad A., Oh H., Kim S. AURP: an AUV-aided underwater routing protocol for underwater acoustic sensor networks. *Sensors*. 2012. vol. 12. no. 2. pp. 1827–1845. DOI: 10.3390/s120201827.
36. Hollinger G., Choudhary S., Qarabaqi P., Murphy C., Mitra U., Sukhatme G., Stojanovic M., Singh H., Hover F. Underwater data collection using robotic sensor networks. *IEEE Journal on Selected Areas in Communications*. 2012. vol. 30. no. 5. pp. 899–911. DOI: 10.1109/jsac.2012.120606.
37. Rahim M., Casari P., Guerra F., Zorzi M. On the performance of delay-tolerant routing protocols in underwater networks. Proceedings of OCEANS 2011 IEEE. Santander, Spain. 2011. pp. 1–7. DOI: 10.1109/Oceans-Spain.2011.6003388.
38. Choudhary M., Goyal N. A rendezvous point-based data gathering in underwater wireless sensor networks for monitoring applications. *International Journal of Communication System*. 2022. vol. 35. no. 6. DOI: 10.1002/dac.5078.
39. Sharma S., Puthal D., Jena S., Zomaya A., Ranjan R. Rendezvous based routing protocol for wireless sensor networks with mobile sink. *The Journal of Supercomputing*. 2017. vol. 73. pp. 1168–1188. DOI: 10.1007/s11227-016-1801-0.
40. Cheng C., Li L. Data gathering problem with the data importance consideration in Underwater Wireless Sensor Networks. *Journal of Network and Computer Applications*. 2017. vol. 78. pp. 300–312. DOI: 10.1016/j.jnca.2016.10.010.
41. Maqsood H., Javaid N., Yahya A., Ali B., Khan Z., Qasim U. MobiL-AUV: AUV-aided localization scheme for underwater wireless sensor networks. Proceedings of the 2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS). Fukuoka, Japan. 2016. pp. 170–175. DOI: 10.1109/IMIS.2016.140.
42. Favaro F., Casari P., Guerra F., Zorzi M. Data upload from a static underwater network to an AUV: Polling or random access? Proceedings of the IEEE OCEANS 2012. Yeosu, Republic of Korea. 2012. pp. 1–6. DOI: 10.1109/OCEANS-Yeosu.2012.6263499.
43. Favaro F., Brolo L., Toso G., Casari P., Zorzi M. A study on remote data retrieval strategies in underwater acoustic networks. Proceedings of the MTS/IEEE OCEANS Conference. San Diego, USA. 2013. pp. 1–8.
44. Cai S., Zhu Y., Wang T., Xu G., Liu A., Liu X. Data Collection in Underwater Sensor Networks based on Mobile Edge Computing. *IEEE Access*. 2019. vol. 7. pp. 6535–65367. DOI: 10.1109/ACCESS.2019.2918213.
45. Yan J., Yang X., Luo X., Chen C. Energy-efficient data collection over AUV-assisted underwater acoustic sensor network. *IEEE Systems Journal*. 2018. vol. 12. no. 4. pp. 3519–3530. DOI: 10.1109/JSYST.2017.2789283.
46. Khan M., Ahmed S., Jembre Y., Kim D. An energy-efficient data collection protocol with AUV path planning in the Internet of Underwater Thing. *Journal of Network and Computer Applications*. 2019. vol. 135. pp. 20–31. DOI: 10.1016/j.jnca.2019.02.025.
47. Li H., Wu X., Zhang Z., Tan X., Pan J., Dai C., Luo D., Ahmed A., Xu Y. An extended-range wave-powered autonomous underwater vehicle applied to underwater wireless sensor networks. *iScience*, 2022. vol. 25. no. 8. DOI: 10.1016/j.isci.2022.104738.
48. Fedorova T., Ryzhov V., Semenov N, Sulaiman S. (2022) Optimization of an Underwater Wireless Sensor Network Architecture with Wave Glider as a Mobile Gateway. *J. Marine. Sci. Appl.* 2022. vol. 21. pp. 179–196. DOI: 10.1007/s11804-022-00268-9.

49. Rappaport T. *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ: Prentice Hall. 1996. 656 p.
50. Thorp W. Deep Sound Attenuation in the Sub and Low Kilocycle per-second Range. *The Journal of the Acoustical Society of America*. 1965. vol. 38. no. 4. p. 648–654.
51. Ahn J., Syed A., Krishnamachari B., Heidemann J. Design and analysis of a propagation delay tolerant ALOHA protocol for underwater networks. *Ad Hoc Networks*. 2011. vol. 9. pp. 752–766. DOI: 10.1016/j.adhoc.2010.09.007.
52. Lindsey S., Raghavendra C. PEGASIS: Power-efficient gathering in sensor information systems. *Proceedings of IEEE Aerospace Conference*. USA: Big Sky, 2002. vol. 3. DOI: 10.1109/AERO.2002.1035242.
53. Sharafutdinova T. [Comparative analysis of industrial hydroacoustic modems]. *Vestnik Gosudarstvennogo universiteta morskogo i rechnogo flota imeni admirala S.O. Makarova – Bulletin of the Admiral S.O. Makarov State University of Maritime and Inland Shipping*. 2021. no. 13(6). pp. 832–841. (In Russ.). DOI: 10.21821/2309-5180-2021-13-6-832-841.

Fedorova Tatiana — Ph.D., Associated professor of the department, Department of applied mathematics and mathematical modeling, faculty of digital industrial technologies, St. Petersburg State Marine Technical University. Research interests: theoretical physics, multiagent systems, complex dynamic systems modeling, underwater wireless sensor network. The number of publications — 46. fedorova.tan@gmail.com; 3, Lotsmanskaya St., 190121, St. Petersburg, Russia; office phone: +7(812)495-2648.

Ryzhov Vladimir — Ph.D., Dr.Sci., Professor, Head of the department, Department of applied mathematics and mathematical modeling, faculty of digital industrial technologies, St. Petersburg State Marine Technical University. Research interests: marine robotics, multiagent systems, hydrodynamics. The number of publications — 120. varyzhov@smtu.ru; 3, Lotsmanskaya St., 190121, St. Petersburg, Russia; office phone: +7(812)495-2648.

Safronov Kirill — Assistant professor of the department, Department of applied mathematics and mathematical modeling, faculty of digital industrial technologies, St. Petersburg State Marine Technical University. Research interests: probability theory, mathematical statistics, machine learning, mathematical acoustics, multiagent systems. The number of publications — 12. safronov.kirill.pm@gmail.com; 3, Lotsmanskaya St., 190121, St. Petersburg, Russia; office phone: +7(812)495-2648.

Acknowledgements. The research is partially funded by the Ministry of Science and Higher Education of the Russian Federation as a part of the World-class Research Center Program: Advanced Digital Technologies (contract No. 075-15-2022-312 dated 20 April 2022).

Руководство для авторов

Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Информатика и автоматизация» <http://ia.spcras.ru/>. При регистрации авторам рекомендуется заполнить все предложенные поля данных. Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше или LaTeX. Объем основного текста (до раздела Литература) - от 20 до 30 страниц включительно. Переносы запрещены. Номера страниц не проставляются. Основная часть текста статьи разбивается на разделы, среди которых являются обязательными: введение, хотя бы один «содержательный» раздел и заключение. Допускается также мотивированное содержанием и структурой материал а выделение подразделов. В основную часть опускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

Author guidelines

Interaction between each potential author and the Editorial board is realized through the pesoal account on the website of the journal "Informatics and Automation" <http://ia.spcras.ru/>. At the registration the authors are requested to fill out all data fields in the proposed form. The submissions should be prepared using MS Word 2007, LaTeX. The text of the paper in the main part should not exceed 30 pages. Pages are not numbered; hyphenations are not allowed. Certain figures, tables, listings and formulas are allowed in the main section, and their typography is considered in more detail at the journal web.

Signed to print 20.09.2024. Passed for print 01.10.2024.

Printed in Publishing center GUAP.

Address: 67 litera A, B. Morskaya, St. Petersburg, 190000, Russia

Founder and Publisher: SPC RAS.

Address: 39 litera A, 14th Line V.O., St. Peterburg, 199178, Russia.

The journal is registered in the Federal Service for Supervision of Communications, Information Technology, and Mass Media, Registration Certificate (registration number) ПИ № ФС77-79228 dated September 25, 2020 Subscription Index П5513, Russian Post Catalog

Подписано к печати 20.09.2024. Дата выхода в свет 01.10.2024.

Формат 60×90 1/16. Усл. печ. л. 17,1. Заказ № 330. 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»