

ISSN 2713-3192
DOI 10.15622/ia.2023.22.6
<http://ia.spcras.ru>

ТОМ 22 № 6

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

INFORMATICS AND AUTOMATION



СПб ФИЦ РАН

Санкт-Петербург
2023



INFORMATICS AND AUTOMATION

Volume 22 № 6, 2023

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

Founded in 2002

Founder and Publisher

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

Editor-in-Chief

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

Editorial Council

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

Editorial Board

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

Editor: A. S. Lopotova

Interpreter: Ya. N. Berezina

Art editor: N. A. Dormidontova

Editorial office address

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

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

The journal is indexed in Scopus

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

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

ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ

Том 22 № 6, 2023

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

Журнал основан в 2002 году

Учредитель и издатель

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

Главный редактор

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

Редакционный совет

А. А. Ашимов	академик Национальной академии наук Республики Казахстан, д-р техн. наук, проф., Алматы, Казахстан
Н. П. Веселкин	академик РАН, д-р мед. наук, проф., Санкт-Петербург, РФ
И. А. Калыев	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. Соифер	академик РАН, д-р техн. наук, проф., Самара, РФ

Редакционная коллегия

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

Выпускающий редактор: А. С. Лопотова

Переводчик: Я. Н. Березина

Художественный редактор: Н. А. Дормидонтова

Адрес редакции

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

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

Журнал индексируется в международной базе данных Scopus

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

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

© Федеральное государственное бюджетное учреждение науки

«Санкт-Петербургский Федеральный исследовательский центр Российской академии наук», 2023
Разрешается воспроизведение в прессе, а также сообщение в эфир или по кабелю опубликованных
в составе печатного периодического издания - журнала «ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ»
статей по текущим экономическим, политическим, социальным и религиозным вопросам
с обязательным указанием имени автора статьи и печатного периодического издания
журнала «ИНФОРМАТИКА И АВТОМАТИЗАЦИЯ»

CONTENTS

Artificial Intelligence, Knowledge and Data Engineering

M.J. Diván, D. Shchemelinin, M.E. Carranza, C.I. Martinez-Spessot, M. Buinevich REAL-TIME RELIABILITY MONITORING ON EDGE COMPUTING: A SYSTEMATIC MAPPING	1243
---	------

K. Krishna, R. Thirumuru A BALANCED INTRUSION DETECTION SYSTEM FOR WIRELESS SENSOR NETWORKS IN A BIG DATA ENVIRONMENT USING CNN-SVM MODEL	1296
--	------

A. Poletaev, I. Paramonov, E. Boychuk ALGORITHM OF CONSTITUENCY TREE FROM DEPENDENCY TREE CONSTRUCTION FOR A RUSSIAN-LANGUAGE SENTENCE	1323
--	------

Robotics, Automation and Control Systems

V. Golovin, K. Yakovlev MOTION PRIMITIVES IN THE TRAJECTORY PLANNING PROBLEM WITH KINEMATIC CONSTRAINTS	1354
---	------

B. Mayorov APPLICATION OF HARMONIC HALF-WAVES FOR AUTOMATION OF HIGH-SPEED TRAINS CONTROL	1387
---	------

Mathematical modeling and applied mathematics

K. Krotov MATHEMATICAL MODELING OF THE PROCESSES OF EXECUTING PACKAGES OF TASKS IN CONVEYOR SYSTEMS WITH INTERMEDIATE BUFFERS OF LIMITED SIZE	1415
--	------

V. Vorotnikov ON THE PARTIAL STABILITY OF NONLINEAR DISCRETE-TIME SYSTEMS WITH DELAY	1451
--	------

A. Stepanov, E. Fomina, L. Illarionova, K. Dubrovin, D. Fedoseev VEGETATION INDICES (NDVI AND EVI) TIME SERIES APPROXIMATION FOR MONITORING CROPS OF Khabarovsk Territory	1473
---	------

Yu. Bibik FORECASTING IN STOCK MARKETS USING THE FORMALISM OF STATISTICAL MECHANICS	1499
---	------

СОДЕРЖАНИЕ

Искусственный интеллект, инженерия данных и знаний

- М.Х. Диван, Д.А. Щемелинин, М. Карранса, Ц.И. Мартинес-Спессот,
М.В. Буйневич
МОНИТОРИНГ НАДЕЖНОСТИ ПОЛЬЗОВАТЕЛЬСКИХ
ВЫЧИСЛИТЕЛЬНЫХ УСТРОЙСТВ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ:
СИСТЕМАТИЧЕСКОЕ ОТОБРАЖЕНИЕ 1243

- К. Кришна, Р. Тирумуру
СБАЛАНСИРОВАННАЯ СИСТЕМА ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ ДЛЯ
БЕСПРОВОДНЫХ СЕНСОРНЫХ СЕТЕЙ В СРЕДЕ БОЛЬШИХ ДАННЫХ
С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ CNN-SVM 1296

- А.Ю. Полетаев, И.В. Парамонов, Е.И. Бойчук
АЛГОРИТМ ПОСТРОЕНИЯ ДЕРЕВА СИНТАКСИЧЕСКИХ ЕДИНИЦ
РУССКОЯЗЫЧНОГО ПРЕДЛОЖЕНИЯ ПО ДЕРЕВУ СИНТАКСИЧЕСКИХ
СВЯЗЕЙ 1323

Робототехника, автоматизация и системы управления

- В.А. Головин, К.С. Яковлев
ПРИМИТИВЫ ДВИЖЕНИЯ РОБОТА В ЗАДАЧЕ ПЛАНИРОВАНИЯ
ТРАЕКТОРИИ С КИНЕМАТИЧЕСКИМИ ОГРАНИЧЕНИЯМИ 1354

- Б.Г. Майоров
ПРИМЕНЕНИЕ ГАРМОНИЧЕСКИХ ПОЛУВОЛН ДЛЯ
АВТОМАТИЗАЦИИ УПРАВЛЕНИЯ ВЫСОКОСКОРОСТНЫМИ
ПОЕЗДАМИ 1387

Математическое моделирование и прикладная математика

- К.В. Кротов
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ ВЫПОЛНЕНИЯ
ПАКЕТОВ ЗАДАНИЙ В КОНВЕЙЕРНЫХ СИСТЕМАХ С
ПРОМЕЖУТОЧНЫМИ БУФЕРАМИ ОГРАНИЧЕННЫХ РАЗМЕРОВ 1415

- В.И. Воротников
О ЧАСТИЧНОЙ УСТОЙЧИВОСТИ НЕЛИНЕЙНЫХ ДИСКРЕТНЫХ
СИСТЕМ С ЗАПАЗДЫВАНИЕМ 1451

- А.С. Степанов, Е.А. Фомина, Л.В. Илларионова, К.Н. Дубровин,
Д.В. Федосеев
АППРОКСИМАЦИЯ ВРЕМЕННЫХ РЯДОВ ИНДЕКСОВ ВЕГЕТАЦИИ
(NDVI И EVI) ДЛЯ МОНИТОРИНГА СЕЛЬХОЗКУЛЬТУР (ПОСЕВОВ)
ХАБАРОВСКОГО КРАЯ 1473

- Ю.В. Бибик
ПРОГНОЗИРОВАНИЕ НА ФОНДОВЫХ РЫНКАХ С ИСПОЛЬЗОВАНИЕМ
ФОРМАЛИЗМА СТАТИСТИЧЕСКОЙ МЕХАНИКИ 1499

M.J. DIVÁN, D. SHCHEMELININ, M.E. CARRANZA,
C.I. MARTINEZ-SPESSOT, M. BUINEVICH
**REAL-TIME RELIABILITY MONITORING ON EDGE
COMPUTING: A SYSTEMATIC MAPPING**

Diván M.J., Shchemelinin D., Carranza M.E., Martinez-Spessot C.I., Buinevich M. **Real-Time Reliability Monitoring on Edge Computing: a Systematic Mapping.**

Abstract. Scenario: System reliability monitoring focuses on determining the level at which the system works as expected (under certain conditions and over time) based on requirements. The edge computing environment is heterogeneous and distributed. It may lack central control due to the scope, number, and volume of stakeholders. Objective: To identify and characterize the Real-time System Reliability Monitoring strategies that have considered Artificial Intelligence models for supporting decision-making processes. Methodology: An analysis based on the Systematic Mapping Study was performed on December 14, 2022. The IEEE and Scopus databases were considered in the exploration. Results: 50 articles addressing the subject between 2013 and 2022 with growing interest. The core use of this technology is related to networking and health areas, articulating Body sensor networks or data policies management (collecting, routing, transmission, and workload management) with edge computing. Conclusions: Real-time Reliability Monitoring in edge computing is ongoing and still nascent. It lacks standards but has taken importance and interest in the last two years. Most articles focused on Push-based data collection methods for supporting centralized decision-making strategies. Additionally, to networking and health, it concentrated and deployed on industrial and environmental monitoring. However, there are multiple opportunities and paths to walk to improve it. E.g., data interoperability, federated and collaborative decision-making models, formalization of the experimental design for measurement process, data sovereignty, organizational memory to capitalize previous knowledge (and experiences), calibration and recalibration strategies for data sources.

Keywords: real-time, system reliability, monitoring, edge computing, systematic mapping study.

1. Introduction. The continuously evolving technological landscape has driven the need for real-time data processing to meet emerging requirements. Consequently, there is a renewed interest in reevaluating real-time concepts and their alignment with advancements and new contexts [1]. These reasons vary, ranging from leveraging up-to-date information about entities to making critical decisions [2–4] and mitigating natural disasters [5]. These advancements have paved the way for various monitoring applications in sectors such as retail [6], industry [7, 8], and natural disaster management [9–11], among others. In applications directly associated with people (e.g., natural disasters and rescue), real-time represents the difference between whether or not keeping a person alive [10]. To achieve this, the Internet of Things (IoT) has played a crucial role by enabling data processing, sensor integration, storage, and communication systems to be deployed as close to the monitored

environment as possible [12, 13]. Simultaneously, Edge Computing (EC) has emerged to enhance resource management and workload processing [14, 15]. This approach has increased computing power, local resource availability, and communication capabilities while enabling effective resource allocation and management within specific events. Various networks, including body sensor networks and road sensor networks, have been developed and deployed as part of a common and heterogeneous infrastructure, each serving different purposes [16].

Real-Time Reliability Monitoring (RTRM), which relies on IoT devices and EC, is based on a crucial assumption – the system is reliable. However, it is important to establish a clear and shared definition of reliability to ensure a common understanding. Reliability refers to the level of achievement at which a system, product, or component meets its functional specifications within specified conditions and time intervals [17]. A reliable system should exhibit maturity, availability, fault tolerance, and recoverability.

Maturity entails meeting the reliability requirements during typical system operation. Availability ensures that the system is operational and accessible whenever it is required. Fault tolerance focuses on maintaining the system's expected functionality even in the presence of hardware or software faults. Recoverability deals with the system's ability to recover affected data and restore the desired state of a component following an interruption or failure event. Therefore, while reliability represents the desired objective, resiliency encompasses the various elements and procedures employed to attain such a status.

In critical areas such as healthcare, real-time analysis relies on the utmost reliability and monitoring. The advent of IoT has enabled the proliferation of sensing applications, leading to increased density and coverage of data collection related to various aspects, such as the impact of particulate matter on people's health. However, it is crucial to acknowledge and address concerns associated with sensor measurements, such as miscalibrations, before making decisions based on such data [18]. The assumptions underlying the measurement process must be precise and treated with the same level of importance as any decision-making process, whether automated or not.

The collection of measures should adhere to certain principles: repeatability (consistent data collection methods), extensibility (ability to accommodate new requirements), clear definition, and alignment with the intended purpose. The measures obtained must be comparable and consistent, reflecting a cohesive and coherent representation of the property being quantified. They should also possess a known degree of accuracy and precision, achieved using calibrated devices (instruments or sensors). When

introducing data-driven decision-making [19], decision-makers expect the data to be interesting, consistent, trustworthy, and timely. They seek information, rather than raw data, to inform their decisions across various aspects. This concern becomes even more critical in the context of automated sensor networks, where decision-making processes are automated. Therefore, ensuring system reliability monitoring, along with a robust measurement process, becomes essential.

Given the critical areas (e.g., health) where real-time analysis is applied, reliability and monitoring are mandatory. On the one hand, IoT fostered increasing the density and coverage of sensing applications (e.g., particulate matter and its impact on people's health). On the other hand, sensors suffer from different concerns that need to be considered (e.g., miscalibrations) before deciding based on such data [18]. The underlying assumptions around the measurement process are sharp and must be regarded as in any decision-making process (automated or not). How the measures are collected must be repeatable (i.e., collects data in a compatible way), extensible (e.g., new requirements), defined, and aim guided. The obtained measures must be comparable and consistent (i.e., cohesive and coherent based on the property trying to quantify), with a known degree of accuracy and precision using a calibrated device (instrument or sensor). Always Data-driven decision-making [19] gets introduced, and decision-makers believe that data are interesting, consistent, trust, and timely. Thus, decision-makers expect to use information (and not data) before deciding on any aspect. Now, it extrapolates this concern in automated sensor networks where the decision-making is automated. System reliability monitoring (and the underlying measurement process) becomes essential.

The Systematic Mapping Study (SMS) describes a set of guides to perform a study focused on tracing the evidence in a broad domain, such as RTRM on EC. This approach allows for a preliminary exploration of the topic before delving into specific aspects [20 – 22].

This work contributes in several ways: a) an SMS was conducted, providing a detailed identification and characterization of strategies related to RTRM on EC; b) different approaches focusing on reliability monitoring were identified and characterized; c) a comparative analysis was performed to elucidate the relationship between artificial intelligence, reliability, and monitoring on edge devices; d) the reliability monitoring needs and their impact on automated decision-making processes were analyzed within the context of edge computing.

The study is divided into seven sections. Section 2 provides a summary of related articles that share similar objectives. Section 3 outlines the methodological strategy, including the selected data sources, criteria for

article selection, and retention criteria used to compile the final list of articles. Section 4 analyzes the collected documents using the SMS methodology, describing the application of SMS and the analysis of the subjects, along with a scoring model for prioritizing further reading. Section 5 presents a comparative analysis of the main strategies based on the scoring model. Section 6 discusses the results obtained in relation to the research questions. Finally, the study concludes with a section dedicated to presenting the overall findings and conclusions.

2. Related Works. There is a vast number of reviews available on the topics of IoT and EC, particularly related to sensor networks and data gathering for monitoring purposes. However, this section specifically focuses on the most relevant works pertaining to RTRM on EC, which examines system reliability monitoring at an abstract level. The objective is to identify reviews that emphasize reliability as a core concept.

Additionally, the monitoring schema should prioritize the degree to which a system achieves functional specifications under defined conditions within a specific time interval. This means that the system can monitor various entities such as vehicles, trains, roads, outpatients, etc., but the crucial aspect lies in ensuring that the primary monitoring system operates as intended.

In a review by [23], the authors delve into edge computing applications in smart grids and compare their efficiency to traditional systems. They discuss the advantages of employing edge devices in smart meters, which results in improved accuracy, reduced latency, and optimized bandwidth consumption. The article explores various EC-based applications, including the monitoring of transmission lines using Unmanned Aerial Vehicles, video surveillance systems, microgrid systems, and power supply management in charging points. It also discusses RTRM of edge frameworks, the interplay between IoT, EC, and cloud computing, as well as the correlation between productivity and security in smart grids based on EC.

While this review shares similarities in addressing the impact of EC devices on latency, bandwidth, and accuracy, it differs in its primary focus on RTRM as a core concern, regardless of the specific application domain.

In [24], the authors introduce a review of mobile and wearable sensors for data-driven health monitoring systems. They discuss strengths and weaknesses associated with wearable sensors, available alternatives, and monitoring systems. Also, the work addresses data types and issues related to disease diagnosis domains, critical challenges inhibiting sensor-based health monitoring systems, and proposed solutions. As a similarity, it addresses a systematic mapping study on Sensor-based applications and solutions. As a difference, this systematic mapping study focuses on the reliability of EC-based systems and how different works try to achieve defined system requirements.

It discusses the articulation between cloud and EC for cooperating on meteorological radar applications and data interoperability [25]. The authors discuss the main problems of weather radar data quality control and different scenarios where edge and cloud computing are complimented for data quality control. The article introduces and contrasts various techniques for data quality control jointly with the challenges and association with edge-cloud cooperation. As a similarity, the review focuses on real-time alternatives to deal with data quality control and the associated implementation (and not the data gathering itself). As a difference, this SMS addresses the system reliability considering different data collection strategies (e.g., push-based, pull-based, hierarchical variations, etc.) jointly with the decision-making process and how to implement actions to keep aligned reliability requirements.

In [26], the authors discuss the impact of information and communication technologies on Digital Twins as a Service (DTaaS). They substantiate the increasing interest in the industrial field due to technologies related to EC, network function virtualization, and 5G. As a similarity, it analyses the latency, data rate, reliability, and scalability requirements in different functionalities related to digital twins (i.e., monitoring, simulation, and operation). However, it limits the level of scalability outlined based on use cases. As a difference, this work discusses the SMS of the RTRM on EC as a central concept, analyzing research volume, types of applications, data collection strategies, and decision-making approaches (e.g., centralized, distributed, etc.).

3. Methodology. SMS proposes a set of directives to specify an action protocol to analyze the shell of a topic. It is beneficial in those areas where a topic is extensive and needs to be reduced based on determined criteria [20 – 22]. This way, after the topic's shell is known and described, it is feasible to explain it before focusing on further detailed analysis.

Figure 1 outlines a Business Process Model and Notation (BPMN) diagram depicting the crucial steps and activities related to the SMS. First, the aim that guided the study was defined jointly with the Research Questions (RQ). An essential aspect was to cover the purpose through RQs before advancing; In other words, the research questions should add different approaches to explore the topic based on the objective. After that, it chooses the databases where the inquiries will be executed. In this sense, every RQ is translated as a part of the search string and suited to the particularities of each chosen database, incorporating the filtering criteria.

This way, the final search strings (suited to each chosen database, including the filtering criteria) were run on each database. Then, the results were processed and reduced through established criteria.

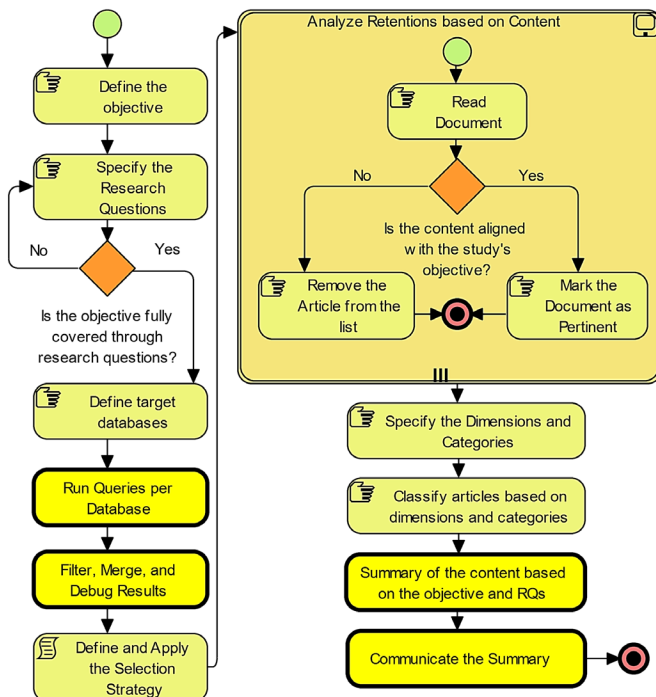


Fig. 1. Crucial Steps of the Systematic Mapping Study, using the BPMN notation

It is worth mentioning that database results were unified in a single list. Then, a quick analysis removed the duplicate records from the list while the selection criteria were defined. The retention criteria focused on the wished properties in articles to sustain the analysis based on coverage and the study's objective. At the same time, those that did not meet the requirements were separated, which helped debug the results and obtain a debugged list.

The next step was to read each article from the debugged list. Records were kept depending upon whether the contents were related to the study objective and the associated points of view (e.g., perhaps an article figured in the list because keywords were present, but the content could mention them as a related example – or potential application – without specific details on the topic). After analyzing every article, a unique list of articles aligned with the objective and proposed questions was produced.

The next step aimed to define dimensions and categories for analyzing those articles from the list. On the one hand, dimensions describe a broad perspective from research questions, while the categories focus on specific content in the dimension. For instance, a dimension could analyze

the approaches used for monitoring the system's reliability; however, categories could indicate different typologies for implementing the monitoring, such as "Push-based, Pull-based, distributed, or centralized." Once the dimensions and categories were specified, they helped to classify the articles in the list. Then, articles were summarized, organized, and communicated through the conclusion in alignment with the study aim.

This section is organized into four sections. The first section specifies the study aim and research questions. The second focuses on the search strategy, target databases, and how to suit the search string in each case. The third one discusses the selection process. The last part of this section makes explicit the limitations and potential biases of this study.

3.1. Research Questions. RQ emerged naturally from the defined aim to define the analysis perspectives and describe the broadness of the search strategy. As it was introduced, the objective of this SMS resided in identifying and characterizing the IoT-based RTRM strategies, which have implemented a measurement process to study the effects of the particulate matter on human health.

The first one was *What is the volume of research on Artificial Intelligence (AI) applied to RTRM on edge devices?* (i.e., RQ1). It is focused on exploring the main concentration of AI-related techniques on current reliability concerns and challenges. It is interesting to review and characterize the current trends. Also, it is worth analyzing the reasons behind the topics with the closest attention in contrast to those with the lowest popularity.

The second question was *What kind of AI approaches are used for reliability monitoring on edge devices?* (i.e., RQ2) The inquiry aims to dive into the different approaches used for reliability monitoring. It could be associated with a holistic approach or dive into some reliability sub-attributes such as maturity, availability, fault-tolerance, or recoverability. Thus, methods could refer to the data processing mode (i.e., online or offline), how metrics are collected (i.e., push or pull), etc.

The third question analyzed the relationship between AI approaches and Data-driven decision-making. It was *"Are the AI approaches related to some real-time decision-making strategy?"* (i.e., RQ3). In this perspective, the underlying assumption implies there is an AI approach guided by data and applied to some situations. Thus, the research question focused on the articulation between the AI model and the Decision-making strategy.

The fourth question advocated for the data collection methods and the articulation with Real-time Reliability Monitoring. This question was *What data collection methods are used for implementing RTRM on eEC?"* (i.e., RQ4). These aspects are worth considering due to the edge challenges such as device heterogeneity, device distribution, data ownership, data sovereignty, etc.

The last research question was *What are the main applications of AI-based RTRM on EC?* (i.e., RQ5). In other words, the idea is to dive into kinds of applications, fields, areas, or subjects where RTRM gets required, and the AI models represent a positive plus. Thus, the characterization of such applications would allow extracting features under interest.

From the research questions, the idea was to identify the concepts involved in each question to delineate the query strings. Thus, once the query strings were identified, they were adapted following the syntax of each queried database [21].

Table 1 describes the keywords surged from each research question, defining the alternative terms for known concepts such as Reliability. While the research questions indicate each complementary perspective under analysis based on the aim of the view, keywords allowed for addressing such views converting them into conceptual search strings to locate articles satisfying each view (acting as decision criteria) [22]. Keywords related to AI, RTRM, Strategy, Reliability, and EC emerged from RQ1 to analyze the articulations between the devices, collecting systems, and those AI approaches to monitor them [27 – 29]. Keywords associated with the types and uses emerged from RQ2, intending to analyze the alternative approach for reliability monitoring [30 – 31]. Keywords about AI, Decision-Making, and Relationship get derived from RQ3 to explore the articulation between knowledge (or previous experiences) and decision [32 – 34]. Keywords related to the data collection use and reliability used from RQ4 to analyze the common collection methods and the reliability monitoring [35 – 37]. Keywords associated with the applications surged from RQ5 [38 – 40].

Table 1. Keywords related to Research Questions

RQ	Keywords
1	AI (alternatively “Artificial Intelligence,” “Machine Learning,” “Expert Systems,” “Agents,” “Case-based Reasoning,” “Linked Systems,” “Optimization,” “Automated Planning and Scheduling,” or “Computer Vision”), Real-time , Reliability (alternatively “Maturity,” “Availability,” “Fault-Tolerance,” or “Recoverability”), Monitoring (alternatively “Observation” or “Monitor” or “Observability”), Strategy (alternatively, “Approach,” “Method,” “Procedure,” “Scheme”), and “Edge Computing” .
2	<i>Kind</i> (alternatively, “Class,” or “Category,” or “Type”), <i>AI¹</i> , <i>Approach¹</i> , <i>“Used for”</i> (alternatively “Applied to”), <i>Reliability¹</i> , <i>Monitoring¹</i> , and <i>“Edge Computing”</i> .
3	<i>AI¹</i> , <i>Approach¹</i> , <i>Relationship</i> (alternatively “Association”), <i>Real-time</i> , <i>Decision-making</i> .
4	<i>“Data collection”</i> (alternatively “data gathering”), <i>Strategy¹</i> , <i>“used for”¹</i> , <i>Real-time</i> , <i>Reliability¹</i> , <i>Monitor</i> , <i>“Edge Computing”</i> .
5	<i>Applications</i> (alternatively “Use Case”), <i>AI¹</i> , <i>Real-time</i> , <i>Reliability¹</i> , <i>“Edge Computing”</i> .

¹ As defined in previous research questions

Table 2 describes the conceptual search string derived from keywords in Table 1.

Table 2. Conceptual Search String

Terms	Structure
Common	("Artificial Intelligence" or "AI" or "Machine Learning" or "Expert Systems" or "Agents" or "Case-based Reasoning" or "Linked Systems" or "Optimization" or "Automated Planning and Scheduling" or "Computer Vision") and "Real-time" and ("Reliability" or "Maturity" or "Availability" or "Fault-Tolerance" or "Recoverability") and ("Monitoring" or "Monitor" or "Observation" or Observability) and ("Strategy" or "Approach" or "Method" or "Procedure" or "Schema") and ("Edge Computing" or "Edge" or "IoT" or "Internet-of-Thing")
Alternative	("Kind" or "Class" or "Category" or "Type") or ("Used for" or "Applied to") or ("Relationship" or "Association") or ("Decision-making") or ("Data collection" or "data gathering") or ("Applications" or "Use Case")

Table 2 discriminates between common and alternative restrictions for logical expression. On the one hand, the common terms must be simultaneously present in the abstract, title, or keywords. On the other hand, the alternative terms represent different concepts eventually associated. However, at least one of them must be present in the abstract, title, or keywords jointly with the familiar words.

Thus, an article was retained in the initial result if it contained all the common terms plus at least one alternative term. It is essential to mention that Table 2 describes the conceptual search string, which was fitted to each database's syntaxis before applying it.

3.2. Search Strategy. The conceptual search string was fitted to each database where the query would be performed. For this study, the digital libraries of IEEE and Scopus were considered. The idea was to provide broad coverage around the published articles related to the subject.

Tables 3 and 4 describe the adaptation of the conceptual search string to the Scopus and IEEE database syntaxes. As it is possible to appreciate, even when the meaning is similar, some syntactic aspects were fitted before running the query.

Table 3. Search String fitted to Scopus Database

Search String
TITLE-ABS-KEY ("Artificial Intelligence" OR "AI" OR "Machine Learning" OR "Expert Systems" OR "Agents" OR "Case-based Reasoning" OR "Linked Systems" OR "Optimization" OR "Automated Planning and Scheduling" OR "Computer Vision") AND TITLE-ABS-KEY ("Real-time") AND TITLE-ABS-KEY ("Reliability" OR "Maturity" OR "Availability" OR "Fault-Tolerance" OR "Recoverability") AND TITLE-ABS-KEY ("Monitoring" OR "Monitor" OR "Observation" OR "Observability") AND TITLE-ABS-KEY ("Strategy" OR "Approach" OR "Method" OR "Procedure" OR "Schema") AND TITLE-ABS-KEY ("Edge Computing" OR "Edge" OR "IoT" OR "Internet of Thing") AND TITLE-ABS-KEY ("Kind" OR "Class" OR "Category" OR "Type" OR "Used for" OR "Applied to" OR "Relationship" OR "Association" OR "Decision-making" OR "Data collection" OR "data gathering" OR "Applications" OR "Use Case")

Table 4. Search String Fitted To IEEE Database

Search String
("All Metadata": "Machine Learning" OR "All Metadata": "Artificial Intelligence" OR "All Metadata": "AI" OR "All Metadata": "Expert Systems" OR "All Metadata": "Agents" OR "All Metadata": "Case-based reasoning" OR "All Metadata": "Linked Systems" OR "All Metadata": "Optimization" OR "All Metadata": "Automated Planning and Scheduling" OR "All Metadata": "Computer Vision") AND ("All Metadata": Reliability OR "All Metadata": Maturity OR "All Metadata": Availability OR "All Metadata": "Fault-Tolerance" OR "All Metadata": Recoverability) AND ("All Metadata": Monitoring OR "All Metadata": Monitor OR "All Metadata": Observation OR "All Metadata": Observability) AND ("All Metadata": Real-time) AND ("All Metadata": "Edge Computing" OR "All Metadata": "Edge" OR "All Metadata": "IoT" OR "All Metadata": "Internet of Thing") AND ("All Metadata": Strategy OR "All Metadata": Approach OR "All Metadata": Method OR "All Metadata": Procedure OR "All Metadata": Schema) AND ("All Metadata": Kind OR "All Metadata": Class OR "All Metadata": Category OR "All Metadata": Type OR "All Metadata": "Used for" OR "All Metadata": "Applied to" OR "All Metadata": Relationship OR "All Metadata": Association OR "All Metadata": "Decision-making" OR "All Metadata": "Data collection" OR "All Metadata": "Data gathering" OR "All Metadata": "Applications" OR "All Metadata": "Use case")

3.3. The Selection Process. This SMS focused on consolidated research written in English and published in journals. It did not consider those articles related to conferences (except extended versions), abstracts, posters, general surveys, reviews, or similar. Thus, it assumes the consolidated research is published in journals and written in English, pursuing international visibility. In this sense, the idea focused on getting an updated perspective about consolidated research strongly associated with Edge-aware systems for RTRM. For sure, this decision implies a bias around the review.

Thus, the initial result list was filtered, retaining articles written in English and coming from journals. Also, surveys and reviews were kept out, fostering access to the source.

Once results from each database were obtained and filtered, a unique list was generated from the fusion of the previous ones. Over this new list merging results from each database, only one copy of duplicate records was kept.

After that, each article's abstract was read in the list to confirm whether it was retained or not according to the search string. That is to say, the searched terms (i.e., keywords) could be present in the abstract just as a mention but not as a core subject. Just articles addressing as a core subject those aspects indicated in the keywords introduced by research questions (Table 1) were retained. In this way, a debugged list for its analysis was reached. This aspect incorporated subjectivity derived from the author's criteria in this work, which was understood as a potential bias.

3.4. Biases and Limitations. The primary underlying assumption throughout this work was that consolidated results from research lines were published in journals. For this reason, no conference papers and other kinds of publishing were considered. Also, previous reviews were supposed to

only partially analyze the core aspect of real-time system reliability, avoiding intermediaries.

English is assumed to be the common language to internationalize the research results.

The limitation of the systematic review is associated with biases. The inclusion/exclusion criteria and the selection of databases compound a set of elements that could be naturally subjective due to the authors' perception. This work was limited to articles published in journals and available on IEEE and Scopus databases. Even when the work tried to get a broad scope, it is not a warranty that the universe of articles is considered.

4. Description of the Approaches. This section is broken down into three sections to describe the implementation of SMS, jointly with the analysis of topics and the proposed order to introduce each obtained work from the queries.

4.1. Implementing the SMS. The queries ran in their respective databases on December 14, 2022. After the initial results, the indicated filters were applied (i.e., limiting to journal articles written in English, etc.), obtaining a partial list of a) *IEEE*: 22 records and b) *Scopus*: 42 records. In this way, 64 records initially satisfied the search requirements.

Each query's result was stored in an Excel file organized by the origin database, and each item was individually inspected to verify that it was associated with a journal article. If a record was not associated with a research article from a journal, it was removed from the list.

Table 5 shows the reasons and quantities for each record's retention jointly with the removal. On the one hand, those records referred to RTRM based on EC were retained. On the other hand, documents associated with the following types were removed 1) *Duplicated*: The same article was available in the IEEE and Scopus results. Only one copy was retained; 2) *Not related content*: The article specifically did not refer to RTRM based on EC. Keywords were only mentioned in the narrative; 3) *Proceedings*: It was not an article but a conference proceeding; 4) *Review*: The article was not a research article. It contained a survey or review of the subject but not how to monitor the system reliability through EC in real-time.

Table 5. Results

	Scopus	IEEE
Retained	34	16
Duplicated	0	6
Not Related	8	0
Proceedings	0	0
Review	0	0
<i>Total</i>	<i>42</i>	<i>22</i>

Thus, 58 records were retained from the different databases, removing eight duplicates and obtaining 50 unique articles (See the first line under the caption in Table 5).

4.2. Dimensions and Characteristics. From the research questions, it was possible to define the dimensions associated with each one jointly with the categories to classify each record from the 50. Table 6 describes the dimensions and categories established by each research question to address the synthesis process indicated in Figure 1.

Table 6. Dimensions and Categories

Research Question	Dimension	Categories
RQ1	Research	Artificial Intelligence, Real-time, Reliability, Monitoring, Edge Devices
RQ2	Approach	Push-based, Pull-based, Distributed, Centralized
RQ3	Decision-Making	Centralized, Federated, Collaborative
RQ4	Methods	Direct Push, Direct Pull, Hierarchical Push, Hierarchical Pull, Query-based local
RQ5	Applications	Area, Subject

About RQ1, the approaches were 1) “*AI*” when models are applied to monitor, analyze, or forecast some properties associated with the system reliability, 2) “*Real-time*” when data are processed and analyzed when they arrive, 3) “*Reliability*” when the monitoring strategy monitors the system reliability itself in addition to the application aim, 4) “*Monitoring*” indicates a formalized strategy to observe (continuously and actively) the system's reliability, 5) “*Edge Computing*” when edge devices are articulated with the monitoring strategy focused on the system reliability. RQ2 refers to the approach used for monitoring the system's reliability. It could be “*Push-based*” when the edge devices push data to a reliability monitoring component; “*Pull-based*” when a reliability monitoring component scratches data from the edge devices to ensure the system's reliability; “*Distributed*” when the reliability monitoring is associated with some swarm intelligence where data are kept on the origin; “*Centralized*”: It indicates an approach where system reliability monitoring is associated with a central component, and edge devices play the role of obedience. RQ3 was associated with the decision-making strategy, indicating “*Centralized*” for those strategies where a significant piece makes decisions based on collected data, “*Federated*” when the decision-making strategy implies a set of autonomous and distributed features which decide based on some

consensus and feedback mechanism between them, and “*Collaborative*” is when two or more components exchange data and knowledge optionally before choosing between them. In this approach, collaboration is optional. A system component could act unilaterally or with a few other devices. RQ4 was focused on the data collection strategy and decision-making process. “*Direct Push*” indicates data are pushed to a central reliability monitoring component without intermediaries or data aggregation. “*Direct Pull*” is associated with a centralized reliability monitoring component scratching data directly from the data sources. “*Hierarchical Push*” is when data walks through a data processing hierarchy from the data source until reaching the wished reliability control level. “*Hierarchical Pull*” focuses on data traveling in a data processing hierarchy, but a certain reliability control level pulls data from the source in stages. “*Query-based local*” indicates that data resides locally (never moved), and the data source only answers queries using the local data. Finally, RQ5 was related to the applications. Thus, “*Area*” represents the field in which the system reliability monitoring is applied, while “*Subject*” refers to the specific application.

Using the ISSN (International Standard Serial Number) associated with the journal of each selected article, the quarter to which the journal belongs was obtained from Scimago Journal and Country Rank (SJCR) [41] and incorporated for the analysis. SJCR is a publicly available portal that includes the journals and country scientific indicators created using information from the Scopus database (Elsevier B.V.). The platform uses the Scimago Journal Rank indicator [42] to show the visibility of journals in the Scopus database from 1996. The indicator considers the journal prestige, the number of associated documents in the database, and the citation prestige given by “importance” and “closeness” according to the received citations. Thus, journals could fall into four categories a) Quarter 1 (Q1) indicates the Top-25% percent most visible journals for an area or discipline, 2) Quarter 2 (Q2) represents journals located in the region of (25; 50] %, 3) Quarter 3 (Q3) indicates those journals integrating the zone of (50; 75] %, and 4) Quarter 4 (Q4) composed of journals falling into the interval (75; 100] %. Eventually, it is possible that a journal is not registered or present in the Scopus database (e.g., because it is a new journal); in such a case, missing space is indicated.

Figure 2 describes the distribution by year of the retained articles from the queries, indicating the quarter to which the associated journal belongs. Since 2013 this field has taken sustained growth. As mentioned, the basic Scimago’s journal categories go from Q1 to Q4. However, some journals contain a mix of types (e.g., Q1/Q2), indicating that some journal subjects are assessed and fall under a specific category (e.g., Q1). Still,

other ones fall under a different category (e.g., Q2). Some journals do not have a Scimago category because they are new. The previous figure indicates this situation under the “Missing” category.

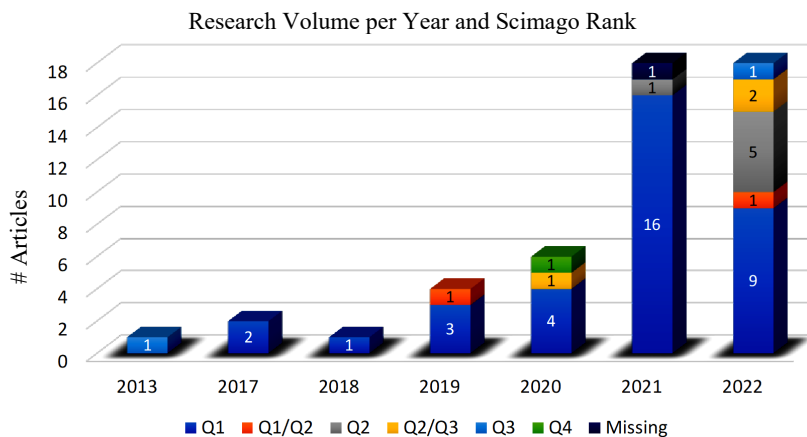


Fig. 2. Articles organized by year and Scimago's quarter to which the journal belongs

Through the individual reading of each article, they were classified according to the categories specified for each dimension. Regarding the research dimension, 74% of analyzed works incorporated some AI model, 94% focused on the real-time aspects, 90% analyzed specific reliability properties, 70% described a formalized monitoring strategy, and 98% explained the articulation between edge devices to IoT and Cloud computing.

The dimension related to the reliability monitoring approach shows that 75.86% of works follow a Push-based strategy. It is logical because there is no control over exogen factors affecting the deployment field (or decisions from third parties) in a distributed and heterogeneous environment. 12.07% of works incorporate a central component for system reliability monitoring. 10.34% of analyzed documents describe a Pull-based associated with the system reliability monitoring component. Lastly, 1.72% of works outline a distributed strategy for dealing with system reliability monitoring.

About the Decision-making dimension, 71.43% of works are associated with a centralized schema to make decisions on system reliability, 17.86% describe a distributed approach, and 10.71% introduced a federated strategy.

From the point of view of data collection methods, 44.23% address the data collection using direct-push strategies, and 42.31% describe a hierarchical-push approach (e.g., using gateways or variations of aggregation

schemas). It represents 86.54% of works using push strategies to provide data for system reliability monitoring. 7.69% of results deal with pull-based collection schemas, and 3.85% use a hierarchical-pull approach (i.e., 11.54% of works address the data collection using pull-based methods). 1.92% of results incorporate local data management articulated with the system reliability monitoring component through a Query-based schema.

Figure 3 describes the proportion of application areas based on the general area concept characterized by the fifth dimension.

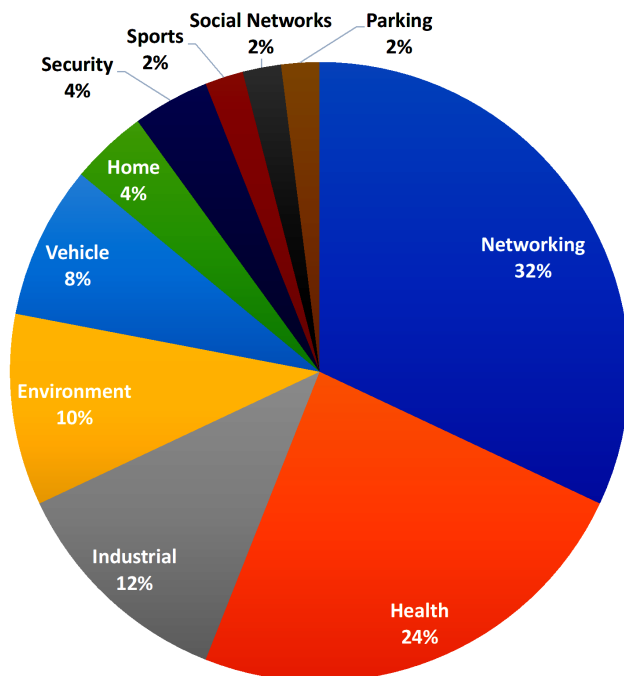


Fig. 3. Main Application Areas

5. A comparative perspective of the main approaches. Each of the obtained articles was synthetically described and analyzed from the dimensions introduced in Table 5. For better organizing, they will be presented according to the generated groups in the previous section and following the order criteria.

5.1. Articles between 2013 and 2019. Table 7 describes articles between 2013 and 2019, ordered by year, quarter, and the number of citations, indicating the decision-making approach.

Table 7. Prioritizing for Articles between 2013 and 2019

Article	Decision-making	Area	Scimago	Citation
IIOT system environmental monitoring using IPSO weight factor estimation [43]	2013	Centralized	System environmental monitoring	Q3
Self-regulating supply–demand systems [44]	2017	Collaborative	Energy supply–demand Matching	Q1
Real-Time Signal Quality-Aware ECG Telemetry System for IoT-Based Health Care Monitoring [45]	2017	Centralized	ECG Telemetry System	Q1
PsCPS: A distributed platform for cloud and fog integrated smart cyber-physical systems [46]	2018	Centralized	Cyber-physical Systems	Q1
Contextaide: End-to-End Architecture for Mobile Crowd-sensing Applications [47]	2019	Centralized & Collaborative	Perpetrator Tracking	Q1
Fault-Tolerant mHealth Framework in the Context of IoT-Based Real-Time Wearable Health Data Sensors [48]	2019	Federated & Collaborative	mHealth. Reliability	Q1
QoS-Adaptive Approximate Real-Time Computation for Mobility-Aware IoT Lifetime Optimization [49]	2019	Centralized	Task Schedule based on QoS requirements.	Q1/Q2
Deep Learning for Hybrid 5G Services in Mobile Edge Computing [50]	2019	Centralized	Network. Energy consumption optimization	Q1

Study [43] describes an improvement in particle swarm optimization (PSO). The authors claim that Improved PSO increased the measurement preciseness via weight factors estimated via experimental simulations. It allows the integration of weight factors, information source fusion reliability, information redundancy, and hierarchical structure integration in uncertain fusion cases.

In [44], the authors describe a generic decentralized self-regulatory framework shaped around standardized control system concepts and IoT. It involves a decentralized combinatorial optimization mechanism that matches supply–demand under different regulatory scenarios. Also, they describe an evaluation methodology jointly with the proposal that allows the systematic assessment of constraints.

In [45], it is described the signal quality-aware IoT-enabled electrocardiogram (ECG) telemetry system for continuous cardiac health monitoring applications. It comprises modules such as ECG signal sensing, automated Signal Quality Assessment (SQA), and signal-quality aware (SQA_w) ECG analysis and transmission. It aims to design and develop a lightweight ECG SQA method for Real-time classifying based on ECG signals.

It proposes a distributed platform for cloud and fog-integrated smart Cyber-Physical Systems (sCPS), named PsCPS [46]. It can provide services to address integration challenges among multiple clouds, multiple fog nodes, and sCPS subsystems. It comprises agents providing services for cloud and fog-integrated sCPS in the participating nodes. Agents can be developed, implemented, controlled, and managed in different ways (i.e., a set of single agents, multi-agent systems, or hierarchical multi-agent systems).

In [47], the authors introduce a design approach for tracking using Mobile-Crowd Sensing (MCS) for complex processing in real-time named ContextAiDe. It comprises 1) An API for detailing contexts jointly with the MCS applications, 2) A middleware for enabling context-optimized and fault-tolerant distributed executions, and 3) An optimization engine for providing the action paths. The authors claim that the strategy is context-optimized recruitment for the execution of computation- and communication-heavy MCS applications in an edge environment. It analyzes two contexts: a) Exact: which has a set of constraints that must be satisfied, and b) Preferred: which has constraints that may be satisfied to a certain level. Thus, ContextAiDe can optimize the operational overheads to enable real-time operation by adjusting the preferred contexts.

In [48], the authors describe a framework to deal with different concerns associated with mHealth applications. They explain a risk local triage algorithm (Risk-Level Localization Triage – RLLT). It can exclude the control process of patient triage and warn about failures related to wearable sensors. RLLT carries out the initial step towards detecting a patient's emergency case and then identifying the healthcare service package of the risk level. Thus, decision-makers are assisted with hospital selection based on the risk-level package, considering the time to arrive at every hospital related to the patient. Also, a mobile app can connect with the hospital's servers to verify the availability of healthcare services given a risk-level package to guide the feasible options.

In [49], the authors address the mobility-aware network lifetime maximization for battery-powered IoT applications. They focus on estimating real-time computation subject to quality-of-service constraints. They discuss a two-fold strategy. An offline step where a task schedule based on mixed-integer linear programming tries to optimize the network lifetime. An online step deals with a heuristic approach for adapting the task execution based on QoS requirements.

In [50], the authors introduce a strategy to minimize normalized energy consumption. In this context, energy consumption refers to the consumption per bit. The optimization is pursued through optimizing user association, resource allocation, and offloading probabilities subject to the

quality-of-service requirements. On the one hand, the mobility management entity manages the user association. On the other hand, every access point determines the resource allocation and offloading likelihoods. A deep neural network is trained offline in a central server based on a digital twin approach representing the current state of the network. Thus, the Mobility management entity uses a DNN model to obtain the user association scheme in real time. Next, an optimization algorithm analyzes the best resource allocation and offloading likelihoods at each access point.

5.2. Articles in 2020. Table 8 describes articles in 2020 ordered by quarter and the number of citations, indicating the decision-making approach.

Table 8. Prioritizing for Articles in 2020

Article	Decision-making	Area	Scimago	Citation
Joint Optimization of Offloading Utility and Privacy for Edge Computing Enabled IoT [51]	Centralized & Collaborative	Network. Resource Optimization	Q1	102
Towards collaborative intelligent IoT eHealth: From device to fog, and cloud [52]	Centralized & Collaborative	ECG-based arrhythmia detection	Q2/Q3	47
A Joint Deep Learning and Internet of Medical Things Driven Framework for Elderly Patients [53]	Centralized	Elderly Patients Monitoring	Q1	46
AI-Enabled Reliable Channel Modeling Architecture for Fog Computing Vehicular Networks [54]	Centralized	Vehicular Monitoring	Q1	26
Quantum computing-inspired network optimization for IoT applications [55]	Centralized	Sensor Networks > Improve the RT Data accuracy	Q1	23
Research on environmental monitoring trend analysis based on internet of things visualization technology [56]	Centralized	Environmental Monitoring	Q4	1

In [51], it gets described as a two-phase offloading optimization strategy. It aims for joint optimization of offloading utility and privacy in the Edge. The Resource utilization maximization is reached through a Utility-Aware Task Offloading (UTO) method. On the other hand, a balance between privacy preservation and execution performance is achieved using a mutual optimization method.

In [52], it describes a holistic AI-driven IoT eHealth architecture based on the concept of the Collaborative Machine Learning approach. Intelligence is distributed across different layers: Devices, Edge/Fog, and Cloud. The proposed approach enables healthcare professionals to continuously monitor the health-related data of subjects anywhere at any time, providing real-time

actionable insights for improving decision-making. The proposal is applied to the ECG-based arrhythmia detection case study.

In [53], the authors describe a sustainable, reliable, and optimized strategy for cardiac activity monitoring based on wearable sensors and image processing in elderly patients. A self-adaptive power control-based Enhanced Efficient-Aware (EEA) approach improves energy consumption and battery lifetime. In this context, it introduces a framework and architecture for cardiac image processing based on deep learning and the Internet of Medical Things. Sustainability and reliability are described around a model for optimizing battery usage and network optimization.

In [54], the authors propose an AI-based, Reliable Interference-free Mobility Management Algorithm (RIMMA) for fog computing intra-vehicular networks. Contributions are organized around four main axes. The first one they claim is that RIMMA, jointly with fog computing, improves computation, communication, cooperation, and storage space. The second one, it gets proposed a reliable and delay-tolerant wireless channel model with better QoS for passengers. The third one is a reliable and efficient multi-layer fog-driven inter-vehicular framework. The fourth one works on the QoS optimization based on mobility, reliability, and packet loss ratio.

In [55], a Quantum Computing-inspired Optimization (IoT-QCiO) technique is introduced. It focuses on maximizing Data Accuracy (DA) in a real-time environment of IoT applications. The proposed model adds quantum formalization of sensor-specific parameters to quantify IoT devices in terms of Sensors in the Vicinity (SIV) and Optimal Sensor Space (OSS). The algorithm optimization is based on data cost, accuracy, and temporal efficiency.

In [56], it introduces an environmental monitoring trend analysis algorithm based on the Internet of Things visualization technology. The proposed algorithm first divides the environmental monitoring area into several different clusters. Different types of sensors provide environmental data that are associated with a group. A mobile agent node is responsible for a monitoring area. It aims to build a two-dimensional positioning table based on the network energy consumption. Thus, it applies the Rosen gradient projection method for choosing the optimal path. With such information, it builds the mobile plan for data collection and forwarding to the control center. Finally, data are centrally analyzed by the control center.

5.3. Articles in 2021. Table 9 describes articles in 2021 ordered by quarter and the number of citations, indicating the decision-making approach and application area. Works with at least one mention are incorporated into the table, and the other ones are outlined in this section after those related to the table are presented.

Table 9. Prioritizing for Articles in 2021

Article	Decision-making	Area	Scimago	Citation
Digital electronics in fibers enable fabric-based machine-learning inference [57]	Centralized	Digital Electronics in fibers	Q1	27
A Smart, Efficient, and Reliable Parking Surveillance System with Edge Artificial Intelligence on IoT Devices [58]	Centralized	Parking Surveillance	Q1	26
Enabling Secure Authentication in Industrial IoT with Transfer Learning Empowered Blockchain [59]	Federated & Collaborative	Secure Authentication	Q1	20
Development of a speed invariant deep learning model with application to condition monitoring of rotating machinery [60]	Centralized	Machinery Clever Monitoring.	Q1	15
IHSF: An Intelligent Solution for Improved Performance of Reliable and Time-Sensitive Flows in Hybrid SDN-Based FC IoT Systems [61]	Federated	Congestion Control	Q1	11
An IoT-based deep learning approach to analyze indoor thermal comfort of disabled people [62]	Centralized	Indoor Monitoring	Q1	10
Fog-centric IoT based smart healthcare support service for monitoring and controlling an epidemic of Swine Flu virus [63]	Centralized	Monitoring of Swine Flu Virus	Q2	8
Intelligent system of game-theory-based decision making in smart sports industry [64]	Centralized	Athlete Performance Monitoring	Q1	5
Learning Spatiotemporal Latent Factors of Traffic via Regularized Tensor Factorization: Imputing Missing Values and Forecasting [65]	Centralized	Road Status Forecast	Q1	5
Hybrid Auto-Scaled Service-Cloud-Based Predictive Workload Modeling and Analysis for Smart Campus System [66]	Centralized	Scalability. Workload Management	Q1	5
Real-time energy consumption detection simulation of network node in internet of things based on artificial intelligence [67]	Centralized	Network. Resource Monitoring	Q1	4
Monitoring Cyber SentiHate Social Behavior During COVID-19 Pandemic in North America [68]	Centralized	Online Social Network Monitoring	Q1	3
Random Forest for Data Aggregation to Monitor and Predict COVID-19 Using Edge Networks [69]	Centralized	COVID-19 Monitor and Prediction	Missing	3
Scheduling Observers over a Shared Channel with Hard Delivery Deadlines [70]	Centralized	Sensor Network Monitoring & Optimizing	Q1	2

Continuation of Table 9

Article	Decision-making	Area	Scimago	Citation
Energy-Aware Distributed Edge ML for mHealth Applications with Strict Latency Requirements [71]	Federated & Collaborative	Epileptic seizures prediction	Q1	2
Information Security Monitoring and Management Method Based on Big Data in the Internet of Things Environment [72]	Centralized	Environmental Monitoring	Q1	2
Task Allocation Mechanism for Cable Real-Time Online Monitoring Business Based on Edge Computing [73]	Centralized	Cable Real-time Online Monitoring Business	Q1	1

In [57], the authors discuss the fabrication approach of digital fibers and its impact on Data Sensor-based decision-making using machine learning. Measures are collected from devices in the thread and stored locally in the same yarn. It allows reaching intra-fiber communications between devices via digital signals. A pre-trained CNN model gets embedded in the fiber and articulated to local data for decision-making.

In [58], the authors discuss the feasibility of using edge computing for intelligent parking occupancy detection using the real-time video feed. It incorporates AI at the edge by implementing an enhanced Single-Shot multi-box Detector (SSD).

The article introduces an authentication strategy supported by Transfer Learning and Blockchain (ATLB) [59]. On the one hand, blockchain technology is applied to privacy preservation and avoids central control in industrial applications. On the other hand, it builds trustworthy blockchains for privacy preservation in industrial applications. In terms of privacy preservation, it introduces different blockchains (e.g., inner and outer) for user authentication strategies to deal with collusion and Sybil attacks. A user-per-region credit approach allows for improving authentication accuracy. It uses transfer learning to minimize the training times between regions.

In [60], the authors describe dealing with model invariance to changing speed via a deep learning method. It can detect a mechanical imbalance (i.e., targeted fault) under varying speed settings. They study speed invariance by processing experimental data obtained from a motor test bed. Also, time-series and time-frequency data are applied to long short-term memory and convolutional neural network, respectively, to evaluate their performance.

In [61], the authors introduce an approach named an Intelligent Solution for Improved Performance of Reliable and Time-sensitive Flows (IPRTF) in Hybrid SDN-based Fog (IHSF) computing IoT systems. The

proposal is composed of three solutions. The first one is related to an algorithm to deploy Software-defined Network (SDN) switches between legacy switches to improve network observability. The second one is associated with a $\{K\}$ -nearest neighbor regression algorithm to Real-time prediction of the legacy links reliability at the SDN controller based on historical data. Thus, the SDN controller can make timely decisions, improving system performance. The third focuses on a Reliable and Time-Sensitive Deep Deterministic Policy Gradient (RT-DDPG) algorithm. It optimally computes forwarding paths in hybrid SDN-F for time-critical traffic flows generated by IoT applications.

In [62], a new learning model is proposed using a deep neural network. The outlined model can predict the indoor thermal comfort of people with different abilities in real time to facilitate remote monitoring. The data collection approach can focus on targeted data before moving them to cloud servers for further analysis.

In [63], the authors explain a fog-centric IoT-based intelligent healthcare support service to monitor and control the Swine Flu virus epidemic. They introduce a framework articulated with fog computing for delay-sensitive applications. Also, they describe the feasibility of using a hybrid classifier to classify the swine flu patient at an early stage and generate alerts to the health officials and patients' guardians.

In [64], an IoT-inspired framework gets proposed for real-time athlete performance analysis. Data from IoT devices represent the base for quantifying the athlete's performance in terms of probability parameters of Probabilistic Measure of Performance (PMP) and Level of Performance Measure (LoPM). The authors introduce a two-player game-theory-based mathematical framework for efficient decision modeling by monitoring officials.

In [65], the authors address the challenge of missing data or noisy information in the context of real-time monitoring. After addressing the previous concerns, the underlying idea is to forecast a city's road status (e.g., congestion). A directed graph models the road network; the nodes are intersections, while the edges represent road segments. The authors assume a set of sensors deployed in the city aligned with the road network. They propose a Temporal Regularized Tensor Factorization (TRTF) framework to account for the spatial structure and temporal dependencies. Also, they describe a data-driven graph-based autoregressive model where weights are learned from data to account for positive and negative correlations.

In [66], it gets introduced a bursts-aware auto-scaling strategy for detecting bursts on dynamic workloads based on resource estimation, decision-making scaling, and workload prediction. The authors pursue

keeping the Quality of Service (QoS) through a hybrid auto-scaled service cloud model dealing with horizontal and vertical scaling concerns. It is addressed through an ensemble algorithm for estimating defined workloads. It allows anticipating the resource management associated with workloads and load balancing for horizontal auto-scaling.

In [67], an artificial intelligence-based Internet of Things control method is proposed. It is aware of the information transmission topology structure to analyze the total energy consumption of the node during transmission. Thus, it estimates the load in each communication cycle to study the energy consumption control of multiple and single transmission path nodes. It fosters the improvement of the resource schedule and energy consumption in mobile network nodes.

In [68], the authors describe a framework for online social network real-time monitoring. It allows for data acquisition, processing, interpreting, and decision-making on the fly when data arrives. Bidirectional Encoder Representations from Transformers (BERT) are applied for natural language processing. BERT-based classifiers discriminate hate and sentiments considering iconic emotions (e.g., emojis). Supervised and unsupervised models are complementarily used for online analysis the social behavior.

In [69], the authors describe an e-healthcare framework based on edge computing. It aims to monitor online health data to predict the risk level of COVID-19 patients. A gateway is incorporated to synthesize data based on the original ones, transmitting those meaningful based on the statistical mean. It allows for minimizing data transmissions from every patient (saving energy), optimizing bandwidth use, and avoiding unnecessary redundancy. Geo-distributed edge servers are used to predict the risk level of each patient with Random Forest (RF) using the synthesized data.

In [70], the authors introduce a framework to formulate the Observer Selection Problem (OSP) through which the controller schedules a sequence of observations that maximize its knowledge about the system's current state. The transmission among observers, controllers, and actuators gets carried out through ultra-low-latency wireless communications. The proposed algorithm systematically prunes the search space to improve the knowledge based on the current situation. The authors claim that its work differs from others because of the extent of the controller's knowledge about the state of the system it controls. Alternative approaches work on real-time communications in that communication reliability is monitored by packet loss or error rate.

In [71], the authors describe an approach for gathering latency requirements of mHealth applications based on user equipment and edge server computing under energy constraints. On the one hand, the authors

address an optimization problem to locate the feature extraction and classification process cleverly between the user equipment and edge devices. It is a Power-aware optimization approach where the inferential mechanism is as close to the user as possible.

Paper [72] describes a framework for real-time environmental monitoring based on IoT. It addresses the basic principles of constructing the evaluation index system while introducing and outlining a reference architecture.

In [73], the authors describe a task allocation mechanism for cable real-time online monitoring business based on edge computing. They define a task allocation model based on the linear distribution characteristics of the cable, the edge node's statuses, the task processing overhead, and the scheduling strategy of delay-sensitive tasks. Next, they introduce a task allocation strategy based on improved discrete particle swarm optimization. The authors claim that the system focuses on the job queuing problem in edge nodes and the optimized task allocation problem among edge nodes.

In [74], it gets presented a strategy for generating a Device-Specific Identifier (i.e., IoT-ID). It captures the device's characteristics and can be used for device identification. The underlying assumption is that IoT-ID is based on Physically Unclonable Functions (PUFs). Thus, it is possible to exploit variations in the manufacturing process to derive a unique fingerprint for integrated circuits. The authors claim that the strategy is non-invasive and can be invoked using simple software APIs running on components of COTS (Commercially Off the Shelf). Among the mentioned properties, they highlight the following: constructability, real-time, uniqueness, and reproducibility.

5.4. Articles in 2022 and the beginning of 2023. Table 10 describes articles in 2022 and the beginning of 2023 that have received at least one citation. This section will initially describe those articles. Next, it is outlined the remaining works (without received citations).

In [75], the authors refer to the impact of the Point of Presence (PoP) in healthcare and how it gains importance as wearable devices and mobile apps are entrusted with RTRM and diagnosis of patients. Novelty relies on the utility value of sensors data improvement through the Laplacian mechanism of preserved Personally Identifiable Information (PII) response to each query from the Edge Open and Distance Learning (ODL).

In [76], the authors address a framework to optimize energy efficiency, battery lifetime, and reliability for intelligent and connected healthcare. It explains the Adaptive Transmission Data Rate (ATDR) mechanism that works on the average constant energy consumption by varying the active time of the sensor node. It allows for optimizing the

energy over the dynamic wireless channel. It explains a Self-Adaptive Routing Algorithm (SARA) to adopt a dynamic source routing mechanism with an energy-efficient and shortest path.

Table 10. Prioritizing for Articles in 2022 – 2023

Article	Decision-making	Area	Scimago	Citation
Emphasizing privacy and security of edge intelligence with machine learning for healthcare [75]	Federated	Patient Monitoring	Q2	13
AI-driven adaptive reliable and sustainable approach for internet of things enabled healthcare system [76]	Centralized	Body Sensor Network	Q2/Q3	8
Smart IoT and Machine Learning-based Framework for Water Quality Assessment and Device Component Monitoring [77]	Centralized	Water Quality Assessment	Q1	5
An innovative edge-based Internet of Energy solution for promoting energy saving in buildings [78]	Centralized	Domestic Energy Saving	Q1	5
Data fusion-based machine learning architecture for intrusion detection [79]	Centralized	Intrusion Detection	Q2	5
Real-Time Fault Detection and Condition Monitoring for Industrial Autonomous Transfer Vehicles Utilizing Edge Artificial Intelligence [80]	Federated	Industrial Autonomous Transfer Vehicles	Q1/Q2	3
Clinical Care of Hyperthyroidism Using Wearable Medical Devices in a Medical IoT Scenario [81]	Centralized	Clinical Care of Hyperthyroidism	Q2	2
MABASR - A Robust Wireless Interface Selection Policy for Heterogeneous Vehicular Networks [82]	Centralized	Selection Policy for Heterogeneous Vehicular Networks	Q1	1
Online Partial Offloading and Task Scheduling in SDN-Fog Networks with Deep Recurrent Reinforcement Learning [83]	Collaborative	Workload Orchestration	Q1	1
Hybrid multi-objective-optimization algorithm for energy efficient priority-based QoS routing in IoT networks [84]	Collaborative	Congestion Control	Q2	1
A Novel CNN-TLSTM Approach for Dengue Disease Identification and Prevention using IoT-Fog Cloud Architecture [85]	Centralized	Dengue Disease Monitoring	Q2/Q3	1

In [77], an IoT-based real-time framework is outlined to perform water quality management, monitoring, and alerts for taking actions based

on contamination and toxic parameter levels, and device and application performance as the first part of the proposed work. Machine learning models analyze water quality trends and device monitoring and management architecture. It is supported by a data push strategy associated with dashboards and rules working on a cloud platform.

In [78], it gets introduced the M2SP-EdgeIoT system focused on domestic energy saving and monitoring. It contemplates data collection, Data-based non-intrusive load monitoring, anomaly detection (using Autoencoders), and a data-guided recommender system for improving energy consumption.

In [79], the authors discuss the methodology of Real-Time Sequential Deep Extreme Learning Machine (RTS-DELM) implemented in wireless IoT enabled sensor networks for the detection of any intrusion activity. They claim to reach an approach that not only makes the casting of parallel data fusion networks but also renders their computations more effective.

In [80], it gets explained as a generic real-time fault diagnosis and condition monitoring system. It uses edge AI complementarity to the monitoring and proposes a data distributor open-source middleware platform called FIWARE. The application field is related to those autonomous transfer vehicle (ATV) types of equipment targeting a smart factory use case. Anomaly detection in ATVs is analyzed through a deep learning-based fault diagnosis method performed in the Edge AI unit. At the same time, results are sent to the storage through a data pipeline.

In [81], the authors describe a monitoring strategy monitor for patients with hyperthyroidism based on wearable medical devices. An architecture based on gateways articulates different levels of devices that communicate fused data streams (enriched by a metadata-based data model). It is supported by MQTT and RTMP protocols for messaging and video transmissions.

In [82], the authors describe a Multi-Armed Bandit Adaptive Similarity-based Regressor (MABASR) for interface selection. The interface selection policy is analyzed and learned through the relationship between channel quality parameters and uplink data rate.

Study [83] proposes a computational offloading and CPU resource scheduling method for minimizing energy consumption. It considers off-loadable and non-off-loadable tasks to characterize and optimize the allocation. On the one hand, a node will contribute to a common goal based on cooperation without any assumption about previous knowledge of statistics or global observations. On the other hand, it applies a deep recurrent Q-network (DRQN) to address partial observability, improving the computational complexity.

In [84], it gets explained the Energy-Efficient Priority-based Multi-Objective QoS routing (PMQoS) mechanism. It aims to ensure energy and

QoS in IoT networks. The proposed system regulates the routing performance based on the QoS parameters, using an optimization technique for three hybrid algorithms (WLFA-Whale Lion Fireworks optimization algorithm with Fitness Function Routing mechanisms). The WLFA focuses on preventing congestion and minimizing localization errors by using and selecting the shortest path through the network.

In [85], an IoT-based proposal is outlined for monitoring Dengue disease. Data-based diagnosis is based on a Hybrid CNN with Tanh Long and Short-Term Memory (TLSTM) using the Adaptive Teaching Learning based Optimization algorithm.

In [86], the authors propose a routing protocol for IoT networks with low power consumption. A secure RPL (Routing Protocol for Low power Lossy network) protocol is introduced to work on congestion control in IoT environments. Authors describe 5 phases a) All nodes register in the sink node, which computes a hash based on the provided data (ID, IP, MAC, Rank, and PUF); b) It gets presented a Multi-context aware parent selection using fuzzy logic; c) The congestion detection is based on agents; d) A lightweight CNN is applied for deep packets analysis; e) An adaptive trickle timer works based on different parameters such as traffic intensity, etc.

In [87], the authors introduce a routing schedule method based on Time-sensitive networking (TSN) for IIOT applications in underground mining. It discusses two alternatives to face the challenge associated with delays, walls, and sensor data collections a) a greedy algorithm based on the shortest distance, and b) a heuristic algorithm based on ant colony optimization.

In [88], the authors introduce a resource-efficient reliability model for MAS (Multi-Agent System) IoT systems for monitoring purposes. The proposed model starts with SLA requirements and characterizes it under the assumption of linear time complexity. It provides simplicity of computations and input metrics for aligning values in each range. Thus, it evaluates the time complexity, producing a measurement associated with the reliability of the model application.

In [89], the authors describe an identification method for rail vehicle running state using Tiny Machine Learning (TinyML). An IoT system is proposed focused on sizing and low energy consumption for data collection through Micro-Electro-Mechanical System (MEMS) sensors. The authors introduce a neural network (NN) model based on acceleration data for classifying the running states of rail vehicles. It represents a two-step approach where the model (i.e., NN) is trained after that is deployed on edge devices. An offset time window strategy addresses sensor results from edge devices and is uploaded to the cloud for visualization and analysis.

In [90], it presents a unique IoT-based sensor node framework (aka, iThing). It aims to predict the onboard battery State of Health (SOH) and Remaining Useful Life (RUL) with the least computational and memory load. The SOH and RUL prediction are implemented through a random learning-based method using the voltage and health metrics of the device.

In [91], and in [5], it gets described as an edge analytics-assisted monitoring solution to monitor several physical activities of the patient. It focuses on determining physical inactivity from their daily routine. It wears wearable sensors to monitor physical movements and GPU for efficient data processing. The study pursues to calculate the scale of the physical inactivity of the patient to make real-time health suggestions.

In [92], it gets proposed how well-established data management techniques from the fields of computer science can be applied to data management for hydrological modeling. It is not a real-time approach but takes advantage of metadata modeling to enrich primary data from sensors. Data are complemented with spatial and temporal information that supports different decision-making models (e.g., upstream drainage area determination).

In [93], shown how the Gulf of Mexico Avian Monitoring Network members used structured decision-making to identify bird monitoring priorities by using multiple tools and techniques to clearly define the problem and stakeholder objectives and to identify bird monitoring priorities at the scale of the entire northern Gulf of Mexico region to coordinate the process of broad-scale monitoring programs to address management, restoration, and scientific questions in other ecosystems and for other taxa.

Study [94] focused on the use of big data in the decision-making processes of healthcare management, the abovementioned three aspects significantly affect the identification of management models, the distribution of responsibilities among professionals (therefore the evaluation of the partial results of each area responsibility) and the processes of allocating resources within the health-care organization.

6. Results and discussions. Independently of the year, citations, and decision-making strategy, this section focuses on answering every research question introduced in Section III and breaks down accordingly.

6.1. RQ1: Research Volume. The research volume in RTRM on EC is associated with fields like Networking (32%), Health (24%), Industrial (12%), Environment (10%), and Vehicle (8%), among others, as shown in Figure 3. Only those five fields represent 86% of the research volume and catch the scientific community's interest.

Figure 4 indicates a surface diagram contrasting the publishing year jointly with the citation volume and Scimago's quarter. As it is possible to appreciate, on the one hand, the intense research on the specific subject

started to take an interest in 2013, with many citations focusing on Q1 journals. Even more, the volume of citations for recent years (e.g., 2021 and 2022) is close to the oldest works, which indicates an active development in the area. On the other hand, the citation over time shows an interesting growing perspective and opportunities in a context where the software and products reach (or propose) different As-a-Service strategies. Reliability is essential for achieving availability and scalability together [95].

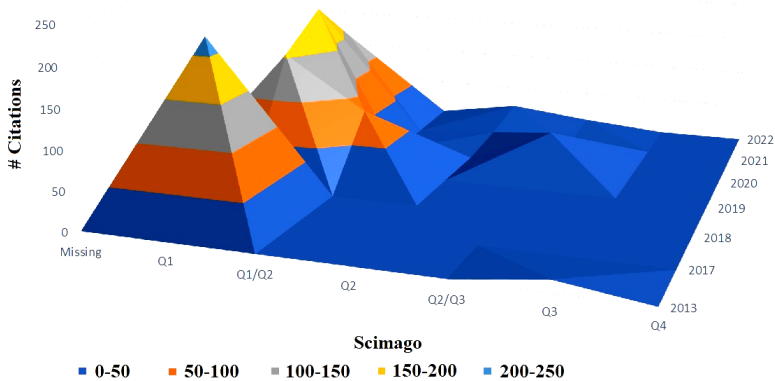


Fig. 4. Main Application Areas

Figure 5 shows that those results related to the research subject are published mainly in Q1 journals (70%), which are associated with a severe and strict publishing process jointly with a high interest in them from the scientific community and practitioners due to the quality of the content. It implies that the subject is worthy of addressing, while challenges and concerns are shared between stakeholders.

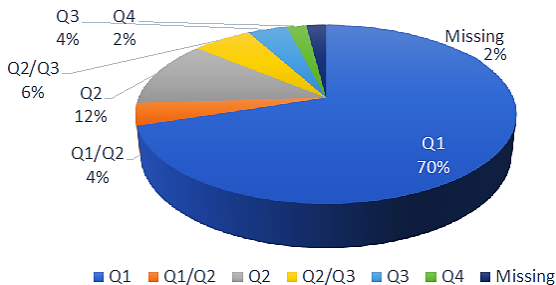


Fig. 5. Research Volume per Scimago Quarter

6.2. RQ2: System Reliability Monitoring Approaches. Figure 6 describes the proportion of data strategies associated with the AI models in reliability working on EC.

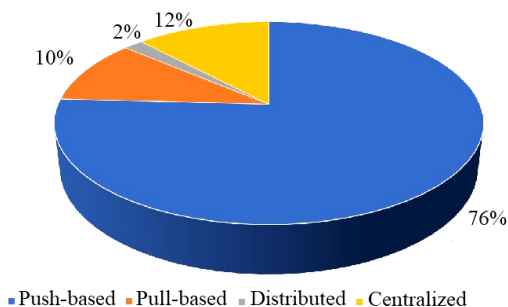


Fig. 6. Proportion of Data Strategy Related to AI Models in Reliability

The heterogeneous and distributed context in which EC is deployed implies that data interoperability is mandatory. In this context, how data is collected from heterogeneous data sources is essential to reach the aim and make proper decisions based on the last known data. According to analyzed works, 76% of the systems use push-based strategies for RTRM, while 12% use a centralized approach. Only 10% address this challenge through a Pull-based process because it requires specific entity data under monitoring and specific communications requirements (not always available depending on the use case or scenario in which a solution is deployed). 2% introduced a distributed schema where data are local to the data source, and the interaction with the reliability component was guided through queries. Thus, the reliability component sends a query. At the same time, the answer is replied to using local data (with its limitations of it because the data are always partial to the data source).

Figure 7 describes the evolution of RTRM strategies over time. As it is possible to appreciate, it was focused on Push-based strategies between 2013 and 2019. In 2020, a centralized approach indicated a possible necessity of consolidation before making decisions around the system's reliability.

However, it additionally emerges Pull-based and distributed techniques between 2021 and 2022. In the end, there is no good or bad alternative because it would depend on the use case and kind of application that defines the specific monitoring requirements. It implies that a Pull-based approach could be interesting to watch distributed data centers because they could support Virtual Private Network (VPN) connections. Still, it would not be the best approach in Low-cost, tiny, distributed devices focusing on body sensor networks.

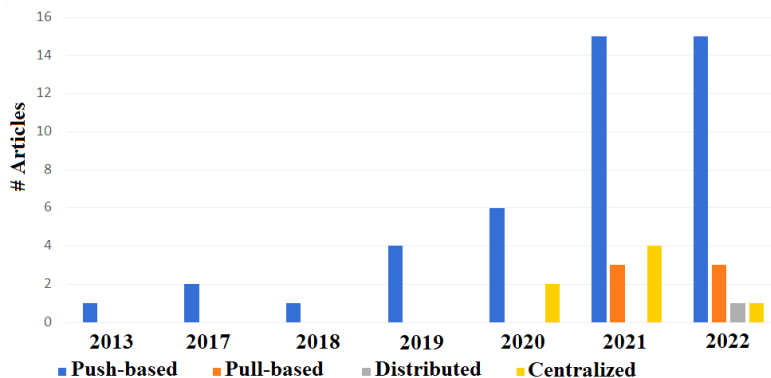


Fig. 7. Evolution of the Reliability Monitoring Strategies over Time

6.3. RQ3: Decision-making strategies. Even when edge computing is associated with a heterogeneous and distributed environment where data interoperability and data quality play an essential role, the main proportion of the system reliability decision-making approaches has been focused on centralized models from 2013 until now (Figure 8). It is a complex scenario because the ability to data analysis, make decisions, and act accordingly depends on a central component or element. Even when the central element could be redundant among other aspects, the figure of a central dependency is contradictory with the reliability perspective and a potential limitation to address the system scalability and high availability.

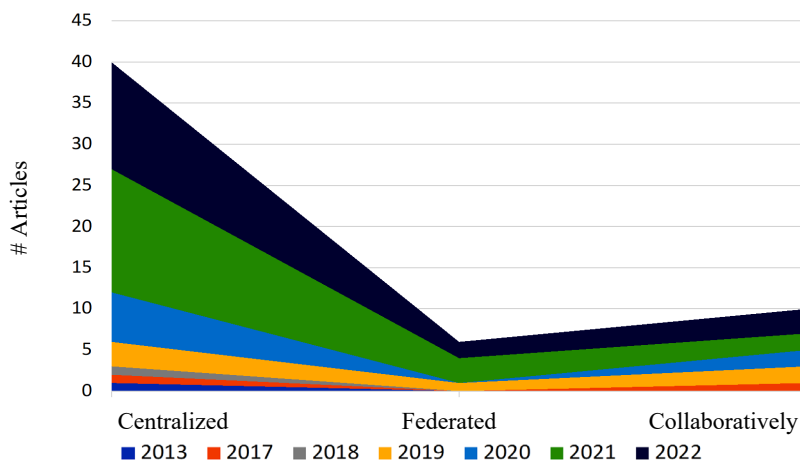


Fig. 8. The proportion of AI and Decision-making approaches per year

However, Federated and Collaborative decision-making approaches are taking interest from 2021 onwards. They propose a system in which the decision-making articulates alternative heterogeneous components monitoring different points of view in a distributed context and the possibility to collaborate, keeping a certain level of local autonomy.

The necessity of collaboration and avoiding central control has influenced Blockchain-based technologies significantly [96]. The idea is that the database content does not depend on a unique stakeholder; on the contrary, any addition or modification (or deletion) requires consensus among the participants. In the reliability context, it could be an exciting approach that the relative importance of each element in the complete system reliability could weigh.

6.4. RQ4: Data Collection Methods. Figure 9 describes the evolution of data collection methods associated with real-time monitoring systems. It was dominated by Direct and Hierarchical Push-based strategies between 2013 and 2020. Nevertheless, Direct-pull and Query-based approaches emerged as an alternative in 2021. 2022 is a context dominated by the Push approaches (be they hierarchical or not). It is consistent with the kind of areas in which the research volume is focusing, i.e., Networking (32%), Health (24%), Industrial (12%), Environment (10%), and Vehicles (8%).

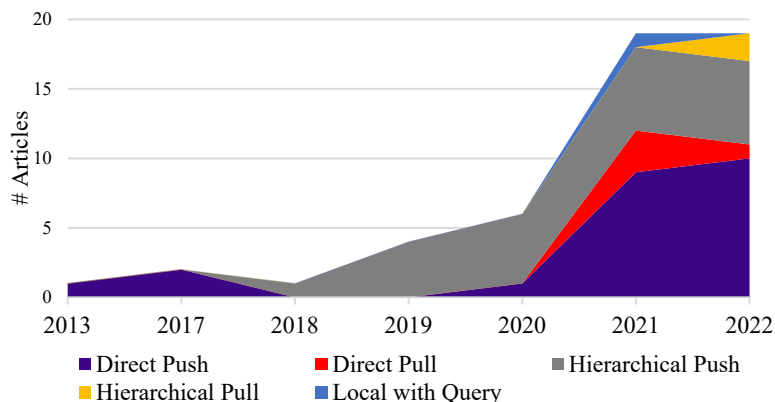


Fig. 9. Evolution of Data Collection Methods for Monitoring Reliability

An aspect worthy of mentioning is that independently of the increment of the research volume in this area, the push-based data collection approaches have become a central aspect to consider. On the other hand, those aspects related to data confidentiality, data sovereignty, and data ownership remain the main challenges and concerns.

6.5. RQ5: Applications. Figure 10 shows a tree map focusing on the main application areas dimensioned based on the citation volume, while inside each room, the most cited subjects are drawn per area.

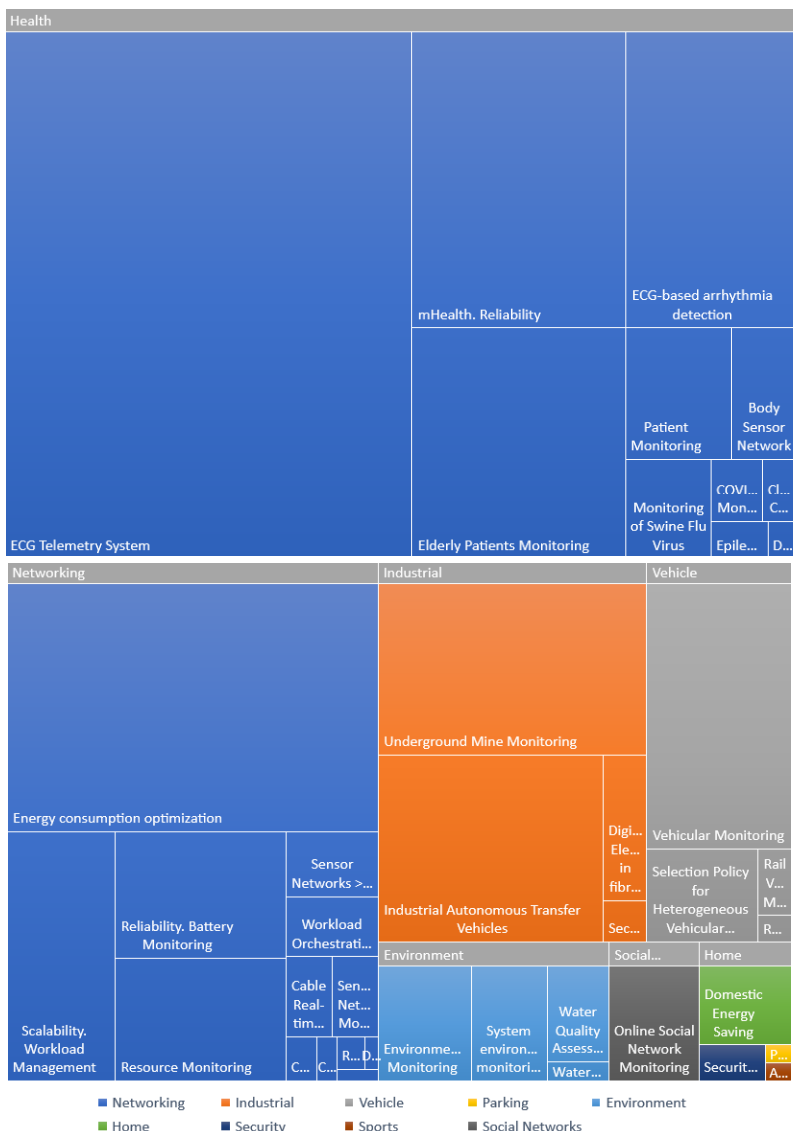


Fig. 10. Main Subjects per Area based on the Citation Volume

In the networking field, the highest cited works are related to resource optimization, task scheduling, energy saving, and data accuracy. On the other hand, the health field focused on ECG telemetry systems, mHealth Reliability, arrhythmia detection, and patient monitoring.

An exciting work (worthy of reading and capitalizing) is introduced in the industrial field related to digital electronics in fibers. It gathers reliability, AI, and local data processing embedded in clothes fibers. Security authentication, machinery monitoring, and energy supply-demand matching are also addressed.

As it is logical, the vehicle field focused on vehicular monitoring, road status monitoring, and parking surveillance. Analogously, environmental areas focus on environmental and water monitoring due to its importance to people's health and relative impact.

Figure 11 outlines the evolution of application areas per year. It is not strange that the beginning is associated with environmental monitoring because it comes naturally from the IoT environment and the possibility to implement real-time monitoring systems at a relatively low cost [18].

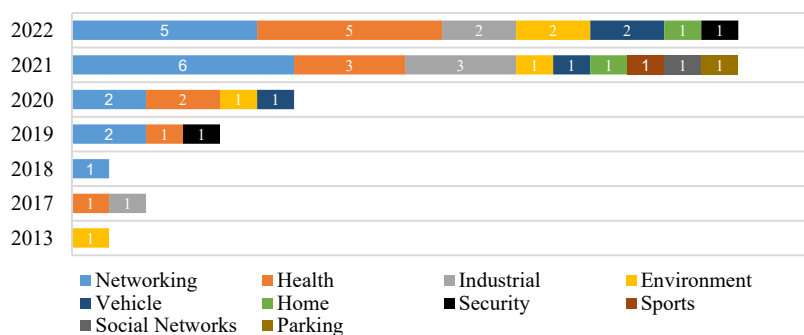


Fig. 11. Application Areas per Year

However, the interest in the application on networking and health has been worthy of highlighting since 2018 onwards. The last three years have provided the context to extend the applications and increase the aspects to be analyzed under the reliability light.

6.6. General Considerations. After the analysis of the different research questions and the involved articles, it emerges essential aspects that are worth mentioning:

1. A few works mentioned using metadata to enrich sensor data to improve the data and automatic interpretation [81], [92]. It is essential to ensure semantic data interoperability because it directly knocks the heart of

system reliability due to the Data-driven decision-making approaches. Definitely, it is not a general approach; it constitutes an open subject, a challenge due to its complexity, and an opportunity to improve decision-making in qualitative terms.

2. The analyzed works need more measurement frameworks, or some formalization focused on guiding the experimental design jointly with the measurement process. It is critical because the strong assumption is that data used for system reliability monitoring are trustworthy, consistent (cohesive and coherent), accurate, precise, comparable, calibrated, and opportune. It is a current challenge and an opportunity to improve the current real-time system reliability monitoring systems.

3. Only [52] mentions the idea of calibration, but it is not associated with the data sources but with input features. The calibration process is essential in any device trying to quantify some property under monitoring because it ensures that the device is aligned with a given reference pattern. The system's entropy could affect (directly or not) the system components producing deviations due to the wear. Data-driven decision-making assumes that the consumed information is adequate for supporting the decision process. Calibration strategies in sensible areas like health and networking are an opportunity.

4. The process measurement assumptions are not reviewed. When data is received from different data sources, the decision-maker assumes that the data come from a repeatable measurement process (It is possible to obtain new data with a compatible method), results are comparable (It is possible to contrast numeric values over time), and extensible (It is possible to add a new characteristic to quantify). It represents an opportunity to integrate distributed and heterogeneous environments where reliability needs to be monitored in real-time.

5. Analyzed works discussed collecting and analyzing data for making decisions based on the system's aim. However, they do not add details about how to ensure the system's reliability and how to monitor it. That is to say; they do not say anything about how to ensure that the system requirements work as expected over time. It means that observability focuses on the typical telemetry and derived actions rather than how to ensure that the telemetry strategy and the system requirements are reached consistently over time. The meta-system monitoring needs to be addressed. For instance, observability is crucial for monitoring a certain level of availability for a given service. Still, also it is essential to monitor the process of producing the Service Level Agreement (SLA) that originated the instantiation of the offered service. The first refers to a particular SLA

with a customer, but the second is associated with the matrix producing the offered SLA. Both must be monitored in terms of reliability.

6. In [69], the authors exposed the challenge of similar or redundant readings and the impact on data transmission latency and energy consumption. Semantic data interoperability is only sometimes considered, affecting how to optimize data retention and transmission policies from data sources (independent of the data collection methods). It is aggravated in data fusion scenarios because the potential integrations become relative and vague.

7. In [65], the authors described the data sparsity, incompleteness, and noise challenges due to the quick deployment of cheap sensors on the field. It is a good point because the IoT environment increases device density and coverage. Still, it needs to address the essential aspects of the data meaning and how to use them to support a decision-making process. Data quality aspects [17] could be analyzed more closely with system reliability.

As it is possible to appreciate, there are a lot of ongoing works on RTRM jointly with current concerns and opportunities. There is no standard approach around it, but there are multiple classes of proposals according to use cases or application areas. The subject has grown in importance in the last two years while leading international journals reported results that substantiate such growth.

7. Conclusion. This review showed the broad proposals and approaches of RTRM on EC. Most publications confirm the importance of the subject in areas such as Networking, Health, Industrial, Environment, and Vehicle. Nevertheless, there needs to be more consensus around monitoring reliability, ensuring syntactic and semantic data interoperability, formalizing the decision-making strategies in edge computing, and capitalizing on previous experiences and knowledge.

In terms of research volume, on the one hand, the intense research on the specific subject started to take an interest in 2013, where many citations focus on Q1 journals. On the other hand, the volume of citations for recent years (e.g., 2021, 2022 and 2023) is close to the oldest works, which indicates an active development in the area.

From the perspective of System Reliability Monitoring Approaches (SRMA): a) 76% of the systems use push-based strategies for reliability monitoring; b) 12% use a centralized approach; c) 10% address this challenge through a Pull-based approach; d) 2% introduced a distributed schema where data are local to the data source, and the interaction with the reliability component was guided through queries. The current dominant monitoring approach has been based on the Push strategy since 2013.

A centralized approach has dominated the decision-making strategies for reliability monitoring. However, the federated and collaborative

decision-making approaches are taking interest from 2021 onwards. They propose a system in which the decision-making articulates different heterogeneous components monitoring alternative points of view in a distributed context and the possibility to collaborate, keeping a certain level of local autonomy. The necessity of collaboration and avoiding central control has influenced Blockchain-based technologies in this area.

In terms of data collection methods, it is a clear leadership of Push-based data collection methods (hierarchical or not). The non-hierarchical push-based methods represent 52.63%, while the hierarchical approach is associated with 31.58. Thus, in general, the participation of Push-based data collection methods is 84.21%. The methods to use depend on the use cases and fields, among others. The remaining 15.79% is related to Pull-based data collection methods, composed of 5.26% for Direct Pull-based methods and 10.53% for Hierarchical Pull-based practices.

From the perspective of the application, Networking and Health areas represent 56% of published works. The percentage reaches 78% when the Industrial and Environment sectors are incorporated.

Beyond the research questions, there are essential aspects that need to be mentioned transversally as follows:

- A few works mentioned the use of metadata to enrich data coming from data sources to improve the data and automatic interpretation. Semantic data interoperability is an open opportunity in RTRM systems considering the distributed and heterogeneous environment associated with edge computing.

- Lack of measurement frameworks for formalizing a consistent, extensible, and repeatable measurement process. It represents a concern because it impacts the heart of Data-driven decision-making, but also an opportunity to improve reliability monitoring.

- Calibration is a subject analyzed on the surface in the works. However, it is essential to consider and maintain over time through different recalibration strategies. It is necessary to establish a reference pattern before obtaining measures. Also, this is a central resource when a metric is contrasted to decision criteria through indicators. Any data source (independently of the type) could fail or lose the calibration reference over time due to entropy.

- The underlying assumptions related to the measurement process are not analyzed. Neither is their impact on Data-driven decision-making. Thus, it represents an opportunity to improve reliability monitoring strategies through the online analysis of interest, comparability, accuracy, precision, trust, opportunity, coherence, cohesion, etc. It is a stage in which the strategy could discriminate among data, information, and knowledge.

– The analyzed works focused on the aimed systems and how to collect and analyze data (e.g., patient monitoring). However, they rarely contemplated the meta-system monitoring to say whether (or not) system reliability is working as expected. Thus, meta-system analysis for monitoring different system properties (in addition to reliability) is open in this environment.

– Data ownership, Data sovereignty, Data Confidentiality, and Data Interoperability (Syntactic and Semantic) are rarely analyzed in the context of online reliability monitoring in EC. It represents a huge opportunity to make a difference. The security around the RTRM strategy is an open topic.

Thus, RTRM in EC is an ongoing work. It lacks standards but has taken importance and interest in the last two years. Articles focused on Push-based data collection methods for supporting centralized decision-making strategies. It was concentrated and deployed mainly on networking, health, industrial, and environmental monitoring. However, there are multiple opportunities and paths to walk to improve it. E.g., data interoperability, federated and collaborative decision-making models, formalization of the experimental design for measurement process, data sovereignty, organizational memory to capitalize previous knowledge (and experiences), calibration and recalibration strategies for data sources.

References

1. Sun Y., Kadota I., Talak R., Modiano E. Age of Information: A New Metric for Information Freshness. Springer Cham, 2020. DOI: 10.2200/S00954ED2V01Y201909CNT023.
2. Li H., Li X., Cheng Q. A fine-grained privacy protection data aggregation scheme for outsourcing smart grid. *Frontiers of Computer Science*. 2023. vol. 17. no. 3. DOI: 10.1007/s11704-022-2003-y.
3. Murtadha M.K., Mushgil B.M. Flexible handover solution for vehicular ad-hoc networks based on software-defined networking and fog computing. *International Journal of Electrical and Computer Engineering*. 2023. vol. 13. no. 2. pp. 1570–1579. DOI: 10.11591/ijece.v13i2.
4. Zhang H., Qi Q., Ji W., Tao F. An update method for digital twin multi-dimension models. *Robotics and Computer-Integrated Manufacturing*. 2023. vol. 80. DOI: 10.1016/j.rcim.2022.102481.
5. Algiriyage N., Prasanna R., Stock K., Doyle E.E., Johnston D. DEES: a real-time system for event extraction from disaster-related web text. *Social Network Analysis and Mining*. 2023. vol. 13. no. 1. DOI: 10.1007/s13278-022-01007-2.
6. Mokhtar M.N.A.B.D., Ismail I., Hamzah W.M.A.F.W., Shamsuddin S.N.W., Arsad M.A.M. Real-Time Dream House Decorator in the Virtual Reality Environment. *International Conference on Business and Technology*. Cham: Springer International Publishing, 2021. vol. 487. pp. 525–537. DOI: 10.1007/978-3-031-08084-5_38.
7. Grover J. Industrial IoT and Its Applications. *IoT for Sustainable Smart Cities and Society*. 2022. pp. 107–124. DOI: 10.1007/978-3-030-89554-9_5.

8. Singh M., Srivastava R., Fuenmayor E., Kuts V., Qiao Y., Murray N., Devine D. Applications of Digital Twin across Industries: A Review. *Applied Sciences*. 2022. vol. 12. no. 11. DOI: 10.3390/app12115727.
9. Wei C., Xu J., Li Q., Jiang S. An Intelligent Wildfire Detection Approach through Cameras Based on Deep Learning. *Sustainability*. 2022. vol. 14. no. 23. DOI: 10.3390/su142315690.
10. Prabhu B.V.B., Lakshmi R., Ankitha R., Prateeksha M.S., Priya N.C. RescueNet: YOLO-based object detection model for detection and counting of flood survivors. *Modeling Earth Systems and Environment*. 2022. vol. 8. no. 4. pp. 4509–4516. DOI: 10.1007/s40808-022-01414-6.
11. Kumar N., Ramesh M.V. Accurate IoT Based Slope Instability Sensing System for Landslide Detection. *IEEE Sensors Journal*. 2022. vol. 22. no. 17. pp. 17151–17161. DOI: 10.1109/JSEN.2022.3189903.
12. Wei Z., Zhu M., Zhang N., Wang L., Zou Y., Meng Z., Wu H., Feng Z. UAV-Assisted Data Collection for Internet of Things: A Survey. *IEEE Internet of Things Journal*. 2022. vol. 9. no. 17. pp. 15460–15483. DOI: 10.1109/JIOT.2022.3176903.
13. Salau B.A., Rawal A., Rawat D.B. Recent Advances in Artificial Intelligence for Wireless Internet of Things and Cyber-Physical Systems: A Comprehensive Survey. *IEEE Internet of Things Journal*. 2022. vol. 9. no. 15. pp. 12916–12930. DOI: 10.1109/JIOT.2022.3170449.
14. Costa B., Bachiega J., De Carvalho L.R., Araujo A.P. Orchestration in Fog Computing: A Comprehensive Survey. *ACM Computing Surveys*. 2023. vol. 55. no. 2. DOI: 10.1145/3486221.
15. Niveditha P.S., John S.P., Simpson S.V. Review on Edge Computing-assisted d2d Networks. *International Conference on Innovative Computing and Communications: Proceedings of ICICC*. 2022. pp. 41–58. DOI: 10.1007/978-981-19-2821-5_4.
16. Nguyen T.A., Kaliappan V.K., Jeon S., Jeon K.-S., Lee J.-W., Min D. Blockchain Empowered Federated Learning with Edge Computing for Digital Twin Systems in Urban Air Mobility. In *Asia-Pacific International Symposium on Aerospace Technology*. 2023. pp. 935–950. DOI: 10.1007/978-981-19-2635-8_69.
17. ISO/IEC 25010. Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. 2011. Available at: <https://www.iso.org/standard/35733.html> (accessed 20.09.2023).
18. Divan M.J., Sanchez-Reynoso M.L., Panebianco J.E., Mendez M.J. IoT-Based Approaches for Monitoring the Particulate Matter and Its Impact on Health. *IEEE Internet of Things Journal*. 2021. vol. 8. no. 15. pp. 11983–12003. DOI: 10.1109/JIOT.2021.3068898.
19. Divan M.J. Data-driven decision making. *Proceeding of International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)*. 2017. pp. 50–56. DOI: 10.1109/ICTUS.2017.8285973.
20. Runeson P., Host M. Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering*. 2009. vol. 14. pp. 131–164. DOI: 10.1007/s10664-008-9102-8.
21. Verner J.M., Sampson J., Tosic V., Bakar N.A., Kitchenham B.A. Guidelines for industrially-based multiple case studies in software engineering. *Proceeding of Third International Conference on Research Challenges in Information Science*. 2009. pp. 313–324. DOI: 10.1109/RCIS.2009.5089295.
22. Petersen K., Vakkalanka S., Kuzniarz L. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and software technology*. 2015. vol. 64. pp. 1–18. DOI: 10.1016/j.infsof.2015.03.007.
23. Prajeesha, Anuradha M. EDGE Computing Application in SMART GRID-A Review. *Proceeding of the 2nd International Conference on Electronics and Sustainable*

- Communication Systems (ICESC'2021). 2021. pp. 397–402. DOI: 10.1109/ICESC51422.2021.9532792.
24. Anikwe C.V., Nweke H.F., Ikegwu A.C., Egwuonwu C.A., Onu F.U., Alo U.R., Teh Y.W. Mobile and wearable sensors for data-driven health monitoring system: State-of-the-art and prospect. *Expert Systems with Applications*. 2022. vol. 202. DOI: 10.1016/j.eswa.2022.117362.
25. Hu Z., Xu X., Zhang Y., Tang H., Cheng Y., Qian C., Khosravi M.R.. Cloud-edge cooperation for meteorological radar big data: a review of data quality control. *Complex and Intelligent Systems*. 2021. DOI: 10.1007/s40747-021-00581-w.
26. Groshev M., Guimaraes C., De La Oliva A., Gazda R. Dissecting the Impact of Information and Communication Technologies on Digital Twins as a Service. *IEEE Access*. 2021. vol. 9. pp. 102862–102876. DOI: 10.1109/ACCESS.2021.3098109.
27. Begum B.A. Nandury S.V. A Survey of Data Aggregation Protocols for Energy Conservation in WSN and IoT. *Wireless Communications and Mobile Computing*. 2022. vol. 2022. DOI: 10.1155/2022/8765335.
28. Gurewitz O., Shifrin M., Dvir E. Data Gathering Techniques in WSN: A Cross-Layer View. *Sensors*. 2022. vol. 22. no. 7. DOI: 10.3390/s22072650.
29. Bayih A.Z., Morales J., Assabie Y., de By R.A. Utilization of Internet of Things and Wireless Sensor Networks for Sustainable Smallholder Agriculture. *Sensors*. 2022. vol. 22. no. 9. DOI: 10.3390/s22093273.
30. Mirani A.A., Velasco-Hernandez G., Awasthi A., Walsh J. Key Challenges and Emerging Technologies in Industrial IoT Architectures: A Review. *Sensors*. 2022. vol. 22, no. 15. DOI: 10.3390/s22155836.
31. Bagwari S., Gehlot A., Singh R., Priyadarshi N., Khan B. Low-Cost Sensor-Based and LoRaWAN Opportunities for Landslide Monitoring Systems on IoT Platform: A Review. *IEEE Access*. 2022. vol. 10. pp. 7107–7127. DOI: 10.1109/ACCESS.2021.3137841.
32. Shahbazi Z., Byun Y.-C. Analysis of the Security and Reliability of Cryptocurrency Systems Using Knowledge Discovery and Machine Learning Methods. *Sensors*. 2022. vol. 22. no. 23. DOI: 10.3390/s22239083.
33. Amador-Dominguez E., Serrano E., Manrique D. GENI: A framework for the generation of explanations and insights of knowledge graph embedding predictions. *Neurocomputing*. 2023. vol. 521. pp. 199–212. DOI: 10.1016/j.neucom.2022.12.010.
34. Valente F., Paredes S., Henriques J., Rocha T., de Carvalho P., Morais J. Interpretability, personalization and reliability of a machine learning based clinical decision support system. *Data Min. Knowl. Discov*. 2022. vol. 36. no. 3. pp. 1140–1173. DOI: 10.1007/s10618-022-00821-8.
35. Yang B., Bai X., Zhang C. Data Collection Method of Energy Adaptive Distributed Wireless Sensor Networks Based on UAV. *Wirel. Commun. Mob. Comput*. 2022. vol. 2022. DOI: 10.1155/2022/3469221.
36. Bai Y., Cao L., Wang S., Ding H., Yue Y. Data Collection Strategy Based on OSELM and Gray Wolf Optimization Algorithm for Wireless Sensor Networks. *Comput. Intell. Neurosci*. 2022. vol. 2022. DOI: 10.1155/2022/4489436.
37. Wei D. et al. Power-Efficient Data Collection Scheme for AUV-Assisted Magnetic Induction and Acoustic Hybrid Internet of Underwater Things. *IEEE Internet Things J*. 2022. vol. 9. no. 14. pp. 11675–11684. DOI: 10.1109/JIOT.2021.3131679.
38. Benmansour F.L., Labraoui N.A. Comprehensive Review on Swarm Intelligence-Based Routing Protocols in Wireless Multimedia Sensor Networks. *Int. J. Wirel. Inf. Networks*. 2021. vol. 28. no. 2. pp. 175–198. DOI: 10.1007/s10776-021-00508-9.
39. Hannan M.A., Hassan K., Jern K.P. A review on sensors and systems in structural health monitoring: Current issues and challenges. *Smart Struct. Syst*. 2018. vol. 22. no. 5. pp. 509–525. DOI: 10.12989/sss.2018.22.5.509.

40. Ajakwe S.O., Nwakanma C.I., Kim D.-S., Lee J.-M. Key Wearable Device Technologies Parameters for Innovative Healthcare Delivery in B5G Network: A Review. *IEEE Access*. 2022. vol. 10. pp. 49956–49974. DOI: 10.1109/ACCESS.2022.3173643.
41. SJR – SCImago Journal and Country Rank [Official web site of SCImago]. 2020. Available at: www.scimagojr.com (accessed 23.09.2020).
42. Guerrero-Bote V.P., Moya-Aneón F. A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *J. Informetr.* 2012. vol. 6. no. 4. pp. 674–688. DOI: 10.1016/j.joi.2012.07.001.
43. Sung W., Hsu C. IOT system environmental monitoring using IPSO weight factor estimation. *Sens. Rev.* 2013. vol. 33. no. 3, pp. 246–256. DOI: 10.1108/02602281311324708.
44. Pournaras E., Yao M., Helbing D. Self-regulating supply–demand systems. *Futur. Gener. Comput. Syst.* 2017. vol. 76. pp. 73–91. DOI: 10.1016/j.future.2017.05.018.
45. Satija U., Ramkumar B., Manikandan M.S. Real-Time Signal Quality-Aware ECG Telemetry System for IoT-Based Health Care Monitoring. *IEEE Internet Things J.* 2017. vol. 4. no. 3. pp. 815–823. DOI: 10.1109/JIOT.2017.2670022.
46. Al-Jaroodi J., Mohamed N. PsCPS: A Distributed Platform for Cloud and Fog Integrated Smart Cyber-Physical Systems. *IEEE Access*. 2018. vol. 6. pp. 41432–41449. DOI: 10.1109/ACCESS.2018.2856509.
47. Pore M., Chakati V., Banerjee A., Gupta S.K.S. ContextAiDe. *ACM Trans. Internet Technol.* 2019. vol. 19. no. 2. pp. 1–23. DOI: 10.1145/3301444.
48. Albahri O.S. et al. Fault-Tolerant mHealth Framework in the Context of IoT-Based Real-Time Wearable Health Data Sensors. *IEEE Access*. 2019. vol. 7. pp. 50052–50080. DOI: 10.1109/ACCESS.2019.2910411.
49. Cao K., Xu G., Zhou J., Wei T., Chen M., Hu S. QoS-Adaptive Approximate Real-Time Computation for Mobility-Aware IoT Lifetime Optimization. *IEEE Trans. Comput. Des. Integr. Circuits Syst.* 2019. vol. 38. no. 10. pp. 1799–1810. DOI: 10.1109/TCAD.2018.2873239.
50. Dong R., She C., Hardjawana W., Li Y., Vucetic B. Deep Learning for Hybrid 5G Services in Mobile Edge Computing Systems: Learn From a Digital Twin. *IEEE Trans. Wirel. Commun.* 2019. vol. 18. no. 10. pp. 4692–4707. DOI: 10.1109/TWC.2019.2927312.
51. Xu X., He C., Xu Z., Qi L., Wan S., Bhuiyan M.Z.A. Joint Optimization of Offloading Utility and Privacy for Edge Computing Enabled IoT. *IEEE Internet Things J.* 2020. vol. 7. no. 4. pp. 2622–2629. DOI: 10.1109/JIOT.2019.2944007.
52. Farahani B., Barzegari M., Aliee F.S., Shaik K.A. Towards collaborative intelligent IoT eHealth: From device to fog, and cloud. *Microprocess. Microsyst.* 2020. vol. 72. p. 102938. DOI: 10.1016/j.micpro.2019.102938.
53. Zhang T. et al. A Joint Deep Learning and Internet of Medical Things Driven Framework for Elderly Patients. *IEEE Access*. 2020. vol. 8. pp. 75822–75832. DOI: 10.1109/ACCESS.2020.2989143.
54. Sodhro A.H., Sodhro G.H., Guizani M., Pirbhulal S., Boukerche A. AI-Enabled Reliable Channel Modeling Architecture for Fog Computing Vehicular Networks. *IEEE Wirel. Commun.* 2020. vol. 27. no. 2. pp. 14–21. DOI: 10.1109/MWC.001.1900311.
55. Bhatia M., Sood S.K. Quantum Computing-Inspired Network Optimization for IoT Applications. *IEEE Internet Things J.* 2020. vol. 7. no. 6. pp. 5590–5598. DOI: 10.1109/JIOT.2020.2979887.
56. Zhang J. Research on environmental monitoring trend analysis based on internet of things visualization technology. *Fresenius Environ. Bull.* 2020. vol. 29. no. 2. pp. 1054 – 1062. [Online]. Available at: www.scopus.com/inward/record.uri?eid=2-s2.0-

- 85090455681&partnerID=40&md5=7c5626045c3f39320bb91185bf694295 (accessed 11.05.2023).
57. Loke G. et al. Digital electronics in fibers enable fabric-based machine-learning inference. *Nat. Commun.* 2021. vol. 12. no. 1. p. 3317. DOI: 10.1038/s41467-021-23628-5.
 58. Ke R., Zhuang Y., Pu Z., Wang Y. A Smart, Efficient, and Reliable Parking Surveillance System with Edge Artificial Intelligence on IoT Devices. *IEEE Trans. Intell. Transp. Syst.* 2021. vol. 22. no. 8. pp. 4962–4974. DOI: 10.1109/TITS.2020.2984197.
 59. Wang X., Garg S., Lin H., Piran M.J., Hu J., Hossain M.S. Enabling Secure Authentication in Industrial IoT with Transfer Learning Empowered Blockchain. *IEEE Trans. Ind. Informatics.* 2021. vol. 17. no. 11. pp. 7725–7733. DOI: 10.1109/TII.2021.3049405.
 60. Lee W.J., Xia K., Denton N.L., Ribeiro B., Sutherland J.W. Development of a speed invariant deep learning model with application to condition monitoring of rotating machinery. *J. Intell. Manuf.* 2021. vol. 32. no. 2. pp. 393–406. DOI: 10.1007/s10845-020-01578-x.
 61. Ibrar M., Wang L., Muntean G.-M., Chen J., Shah N., Akbar A. IHSF: An Intelligent Solution for Improved Performance of Reliable and Time-Sensitive Flows in Hybrid SDN-Based FC IoT Systems. *IEEE Internet Things J.* 2021. vol. 8. no. 5. pp. 3130–3142. DOI: 10.1109/JIOT.2020.3024560.
 62. Brik B., Esseghir M., Merghem-Boulahia L., Snoussi H. An IoT-based deep learning approach to analyse indoor thermal comfort of disabled people. *Build. Environ.* 2021. vol. 203. p. 108056. DOI: 10.1016/j.buildenv.2021.108056.
 63. Singh P.D., Kaur R., Singh K.D., Dhiman G., Soni M. Fog-centric IoT based smart healthcare support service for monitoring and controlling an epidemic of Swine Flu virus. *Informatics Med. Unlocked.* 2021. vol. 26. p. 100636. DOI: 10.1016/j.imu.2021.100636.
 64. Bhatia M. Intelligent System of Game-Theory-Based Decision Making in Smart Sports Industry. *ACM Trans. Intell. Syst. Technol.* 2021. vol. 12. no. 3. pp. 1–23. DOI: 10.1145/3447986.
 65. Baggag A. et al. Learning Spatiotemporal Latent Factors of Traffic via Regularized Tensor Factorization: Imputing Missing Values and Forecasting. *IEEE Trans. Knowl. Data Eng.* 2021. vol. 33. no. 6. pp. 2573–2587. DOI: 10.1109/TKDE.2019.2954868.
 66. Razzaq M.A., Mahar J.A., Ahmad M., Saher N., Mehmood A., Choi G.S. Hybrid Auto-Scaled Service-Cloud-Based Predictive Workload Modeling and Analysis for Smart Campus System. *IEEE Access.* 2021. vol. 9. pp. 42081–42089. DOI: 10.1109/ACCESS.2021.3065597.
 67. Jin H., Zhao J. Real-time energy consumption detection simulation of network node in internet of things based on artificial intelligence. *Sustainable Energy Technologies and Assessments.* 2021. vol. 44. no. 101004. DOI: 10.1016/j.seta.2021.101004.
 68. Alzamzami F., El Saddik A. Monitoring Cyber SentiHate Social Behavior During COVID-19 Pandemic in North America. *IEEE Access.* 2021. vol. 9. pp. 91184–91208. DOI: 10.1109/ACCESS.2021.3088410.
 69. Adhikari M., Ambigavathi M., Menon V.G., Hammoudeh M. Random Forest for Data Aggregation to Monitor and Predict COVID-19 Using Edge Networks. *IEEE Internet Things Mag.* 2021. vol. 4. no. 2. pp. 40–44. DOI: 10.1109/IOTM.0001.2100052.
 70. Jurdi R., Andrews J.G., Heath R.W. Scheduling Observers Over a Shared Channel With Hard Delivery Deadlines. *IEEE Trans. Commun.* 2021. vol. 69. no. 1. pp. 133–148. DOI: 10.1109/TCOMM.2020.3032172.
 71. Hashash O., Sharafeddine S., Dawy Z., Mohamed A., Yaacoub E. Energy-Aware Distributed Edge ML for mHealth Applications with Strict Latency Requirements.

- IEEE Wirel. Commun. Lett. 2021. vol. 10. no. 12. pp. 2791–2794. DOI: 10.1109/LWC.2021.3117876.
72. Liang W., Li W., Feng L. Information Security Monitoring and Management Method Based on Big Data in the Internet of Things Environment. IEEE Access. 2021. vol. 9. pp. 39798–39812. DOI: 10.1109/ACCESS.2021.3064350.
73. Shao S., Zhang Q., Guo S., Qi F. Task Allocation Mechanism for Cable Real-Time Online Monitoring Business Based on Edge Computing. IEEE Syst. J. 2021. vol. 15. no. 1. pp. 1344–1355. DOI: 10.1109/JSYST.2020.2988266.
74. Vaidya G., Nambi A., Prabhakar T.V., Kumar V.T., Sudhakara S. Towards generating a reliable device-specific identifier for IoT devices. Pervasive Mob. Comput. 2021. vol. 76. DOI: 10.1016/j.pmcj.2021.101445.
75. Rajendran S. et al. Emphasizing privacy and security of edge intelligence with machine learning for healthcare. Int. J. Intell. Comput. Cybern. 2022. vol. 15. no. 1. pp. 92–109. DOI: 10.1108/IJICC-05-2021-0099.
76. Zahid N., Sodhro A.H., Kambh U.R., Alkhayyat A., Wang L. AI-driven adaptive reliable and sustainable approach for internet of things enabled healthcare system. Math. Biosci. Eng. 2022. vol. 19. no. 4. pp. 3953–3971. DOI: 10.3934/mbe.2022182.
77. Bhardwaj A. et al. Smart IoT and Machine Learning-based Framework for Water Quality Assessment and Device Component Monitoring. Environ. Sci. Pollut. Res. 2022. vol. 29. no. 30. pp. 46018–46036. DOI: 10.1007/s11356-022-19014-3.
78. Alsalemi A., Himeur Y., Bensaali F., Amira A. An innovative edge-based Internet of Energy solution for promoting energy saving in buildings. Sustainable Cities and Society. 2022. vol. 78. no. 103571. DOI: 10.1016/j.scs.2021.103571.
79. Khan M.A., Ghazal T.M., Lee S.-W., Rehman A. Data Fusion-Based Machine Learning Architecture for Intrusion Detection. Comput. Mater. Contin. 2022. vol. 70. no. 2. pp. 3399–3413. DOI: 10.32604/cmc.2022.020173.
80. Gültekin Ö., Cinar E., Özkan K., Yazıcı A. Real-Time Fault Detection and Condition Monitoring for Industrial Autonomous Transfer Vehicles Utilizing Edge Artificial Intelligence. Sensors. 2022. vol. 22(9). no. 3208. DOI: 10.3390/s22093208.
81. Wei L., Hou S., Liu Q. Clinical Care of Hyperthyroidism Using Wearable Medical Devices in a Medical IoT Scenario. Journal of Healthcare Engineering. 2022. vol. 2022. DOI: 10.1155/2022/5951326.
82. Nikolov G., Kuhn M., Mccgibney A., Wenning B.-L. MABASR – A Robust Wireless Interface Selection Policy for Heterogeneous Vehicular Networks. IEEE Access. 2022. vol. 10. pp. 26068–26077. DOI: 10.1109/ACCESS.2022.3156597.
83. Baek J., Kaddoum G. Online Partial Offloading and Task Scheduling in SDN-Fog Networks With Deep Recurrent Reinforcement Learning. IEEE Internet Things J. 2022. vol. 9. no. 13. pp. 11578–11589. DOI: 10.1109/JIOT.2021.3130474.
84. Thenmozhi R., Sakthivel P., Kulothungan K. Hybrid multi-objective-optimization algorithm for energy efficient priority-based QoS routing in IoT networks. Wireless Networks. 2022. DOI: 10.1007/s11276-021-02848-z.
85. Manoharan S.N., Kumar K.M.V.M., Vadivelan N.A. Novel CNN-TLSTM Approach for Dengue Disease Identification and Prevention using IoT-Fog Cloud Architecture. Neural Processing Letters. 2022. vol. 55. no. 2. pp. 1951–1973. DOI: 10.1007/s11063-022-10971-x.
86. Sithik M.M., Kumar B.M. Intelligent agent based virtual clustering and multi-context aware routing for congestion mitigation in secure RPL-IoT environment. Ad Hoc Networks. 2022. vol. 137. DOI: 10.1016/j.adhoc.2022.102972.
87. Zhang Y., Wu J., Liu M., Tan A. TSN-based routing and scheduling scheme for Industrial Internet of Things in underground mining. Engineering Applications of Artificial Intelligence. 2022. vol. 115. DOI: 10.1016/j.engappai.2022.105314.

88. Eroshkin I., Vojtech L., Neruda M. Resource Efficient Real-Time Reliability Model for Multi-Agent IoT Systems. IEEE Access. 2022. vol. 10. pp. 2578–2590. DOI: 10.1109/ACCESS.2021.3138931.
89. Zhou S., Du Y., Chen B., Li Y., Luan X. An Intelligent IoT Sensing System for Rail Vehicle Running States Based on TinyML. IEEE Access. 2022. vol. 10. pp. 98860–98871. DOI: 10.1109/ACCESS.2022.3206954.
90. Sinha A., Das D., Ududalpalay V., Mohanty S.P. iThing: Designing Next-Generation Things with Battery Health Self-Monitoring Capabilities for Sustainable IIoT. IEEE Trans. Instrum. Meas. 2022. vol. 71. pp. 1–9. DOI: 10.1109/TIM.2022.3216594.
91. Manocha A., Singh R. A Novel Edge Analytics Assisted Motor Movement Recognition Framework Using Multi-Stage Convo-GRU Model. Mob. Networks Appl. 2022. vol. 27. no. 2. pp. 657–676. DOI: 10.1007/s11036-019-01321-8.
92. Bollen E. et al. A database system for querying of river networks: facilitating monitoring and prediction applications. Water Supply. 2022. vol. 22. no. 3. pp. 2832–2846. DOI: 10.2166/ws.2021.433.
93. Fournier A.M.V. et al. Structured Decision Making to Prioritize Regional Bird Monitoring Needs. INFORMS Journal on Applied Analytics. 2023. vol. 53(3). pp. 207–217. DOI: 10.1287/inte.2022.1154.
94. Fanelli S., Pratici L., Salvatore F.P., Donelli C.C., Zangrandi A. Big data analysis for decision-making processes: challenges and opportunities for the management of health-care organizations. Management Research Review. 2023. vol. 46. no. 3. pp. 369–389. DOI: 10.1108/MRR-09-2021-0648.
95. Overdal M., Haddara M., Langseth M. Exploring Public Cloud-ERP Systems' Impact on Organizational Performance. Lect. Notes Networks Syst. 2023. vol. 561. pp. 121–137. DOI: 10.1007/978-3-031-18344-7_8.
96. Reyna A., Martín C., Chen J., Soler E., Díaz M. On blockchain and its integration with IoT. Challenges and opportunities. Futur. Gener. Comput. Syst. 2018. vol. 88. pp. 173–190. DOI: 10.1016/j.future.2018.05.046.

Diván Mario José — Ph.D., Honored professor, Sr. solutions architect, Department of edge reliability engineering, Intel Corporation. Research interests: cyber security, management, scalability and monetization of IT service platforms. The number of publications — 91. mario.jose.divan.koller@intel.com; 3, 25th Ave., Jones Farm Campus, 97124, Hillsboro, Oregon, United States of America; office phone: +1(503)264-0180.

Shchemelinin Dmitry — Ph.D., Dr.Sci., Head (lead) of the department, Department of cloud site reliability engineering, Intel Corporation. Research interests: monitoring, state management and scalability of globally distributed computing systems by using methods of artificial intelligence, big data collection and analysis. The number of publications — 78. dshchemel@gmail.com; 3, 25th Ave., Jones Farm Campus, 97124, Hillsboro, Oregon, United States of America; office phone: +1(408)981-5963.

Carranza Marcos E. — Principal engineer, Department of edge reliability engineering, Intel Corporation. Research interests: cyber security, management, scalability and monetization of IT service platforms. The number of publications — 32. marcos.e.carranza@intel.com; 3, 25th Ave., Jones Farm Campus, 97124, Hillsboro, Oregon, United States of America; office phone: +1(503)764-7952.

Martinez-Spessot Cesar Ignacio — Head (lead) of the department, Department of edge reliability engineering, Intel Corporation. Research interests: application of digital technologies by using artificial intelligence methods, development of the internet of things ecosystems and applications. The number of publications — 31. cesar.martinez@intel.com; 3, 25th Ave., Jones

Farm Campus, 97124, Hillsboro, Oregon, United States of America; office phone: +1(503)7120559.

Buinevich Mikhail — Ph.D., Dr.Sci., Professor of the department, Department of applied mathematics and information technologies, Saint-Petersburg University of State Fire Service of EMERCOM of Russia. Research interests: cyber antifragility of info telecommunications systems, artificial intelligence methods in processing data state of complex systems, training of specialists in information security, cyber polygon. The number of publications — 170. bmv1958@yandex.ru; 149, Moskovskij Ave., 196105, St. Petersburg, Russia; office phone: +7(812)388-8639.

Acknowledgements. This work was fully supported by Intel Corporation.

М.Х. ДИВАН, Д.А. ЩЕМЕЛИНИН, М. КАРРАНСА,
Ц.И. МАРТИНЕС-СПЕССОТ, М.В. БУЙНЕВИЧ

**МОНИТОРИНГ НАДЕЖНОСТИ ПОЛЬЗОВАТЕЛЬСКИХ
ВЫЧИСЛИТЕЛЬНЫХ УСТРОЙСТВ В РЕЖИМЕ РЕАЛЬНОГО
ВРЕМЕНИ: СИСТЕМАТИЧЕСКОЕ ОТОБРАЖЕНИЕ**

Диван М.Х., Щемелинин Д.А., Карранса М., Мартинес-Спессот Ц.И., Буйневич М.В.
Мониторинг надежности пользовательских вычислительных устройств в режиме реального времени: систематическое отображение.

Аннотация. Данный исследовательский обзор сосредоточен на мониторинге надежности вычислительных систем в режиме реального времени на стороне пользователя. В условиях гетерогенной и распределенной вычислительной среды, где отсутствует централизованный контроль, исследуется использование моделей искусственного интеллекта для поддержки процессов принятия решений в мониторинге надежности системы. Методология исследования основана на систематическом отображении предыдущих исследований, опубликованных в научных базах данных IEEE и Scopus. Анализ проведен на основе 50 научных статей, опубликованных с 2013 по 2022 годы, показал растущий научный интерес к данной области. Основное применение исследуемого метода связано с сетевыми технологиями и здравоохранением. Данный метод нацелен на интеграцию сети медицинских сенсоров и управляющих данных с пользовательскими вычислительными устройствами. Однако этот метод также применяется в промышленном и экологическом мониторинге. Выводы исследования показывают, что мониторинг надежности пользовательских вычислительных устройств в режиме реального времени находится на начальной стадии развития. Он не имеет стандартов, но за последние два года приобрел значительное значение и интерес. Большинство исследуемых статей сосредоточены на методах сбора данных с использованием уведомлений для поддержки централизованных стратегий принятия решений. Однако, существует множество возможностей для дальнейшего развития данного метода, таких как совместимость данных, федеративные и совместные модели принятия решений, формализация экспериментального дизайна, суверенитет данных, систематизация базы данных для использования предыдущих знаний и опыта, стратегии калибровки и повторной корректировки для источников данных.

Ключевые слова: реальное время, надежность системы, мониторинг, вычисления на стороне пользователя, систематическое исследование.

Литература

1. Sun Y., Kadota I., Talak R., Modiano E. Age of Information: A New Metric for Information Freshness // Springer Cham, 2020. DOI: 10.2200/S00954ED2V01Y201909CNT023.
2. Li H., Li X., Cheng Q. A fine-grained privacy protection data aggregation scheme for outsourcing smart grid // Frontiers of Computer Science. 2023. vol. 17. no. 3. DOI: 10.1007/s11704-022-2003-y.
3. Murtadha M.K., Mushgil B.M. Flexible handover solution for vehicular ad-hoc networks based on software-defined networking and fog computing // International Journal of Electrical and Computer Engineering. 2023. vol. 13. no. 2. pp. 1570–1579. DOI: 10.11591/ijece.v13i2.

4. Zhang H., Qi Q., Ji W., Tao F. An update method for digital twin multi-dimension models // *Robotics and Computer-Integrated Manufacturing*. 2023. vol. 80. DOI: 10.1016/j.rcim.2022.102481.
5. Algiriyage N., Prasanna R., Stock K., Doyle E.E., Johnston D. DEES: a real-time system for event extraction from disaster-related web text // *Social Network Analysis and Mining*. 2023. vol. 13. no. 1. DOI: 10.1007/s13278-022-01007-2.
6. Mokhtar M.N.A.B.D., Ismail I., Hamzah W.M.A.F.W., Shamsuddin S.N.W., Arsad M.A.M. Real-Time Dream House Decorator in the Virtual Reality Environment // *International Conference on Business and Technology*. Cham: Springer International Publishing, 2021. vol. 487. pp. 525–537. DOI: 10.1007/978-3-031-08084-5_38.
7. Grover J. Industrial IoT and Its Applications // *IoT for Sustainable Smart Cities and Society*. 2022. pp. 107–124. DOI: 10.1007/978-3-030-89554-9_5.
8. Singh M., Srivastava R., Fuenmayor E., Kuts V., Qiao Y., Murray N., Devine D. Applications of Digital Twin across Industries: A Review // *Applied Sciences*. 2022. vol. 12. no. 11. DOI: 10.3390/app12115727.
9. Wei C., Xu J., Li Q., Jiang S. An Intelligent Wildfire Detection Approach through Cameras Based on Deep Learning // *Sustainability*. 2022. vol. 14. no. 23. DOI: 10.3390/su142315690.
10. Prabhu B.VB., Lakshmi R., Ankitha R., Prateeksha M.S., Priya N.C. RescueNet: YOLO-based object detection model for detection and counting of flood survivors // *Modeling Earth Systems and Environment*. 2022. vol. 8. no. 4. pp. 4509–4516. DOI: 10.1007/s40808-022-01414-6.
11. Kumar N., Ramesh M.V. Accurate IoT Based Slope Instability Sensing System for Landslide Detection // *IEEE Sensors Journal*. 2022. vol. 22. no. 17. pp. 17151–17161. DOI: 10.1109/JSEN.2022.3189903.
12. Wei Z., Zhu M., Zhang N., Wang L., Zou Y., Meng Z., Wu H., Feng Z. UAV-Assisted Data Collection for Internet of Things: A Survey // *IEEE Internet of Things Journal*. 2022. vol. 9. no. 17. pp. 15460–15483. DOI: 10.1109/JIOT.2022.3176903.
13. Salau B.A., Rawal A., Rawat D.B. Recent Advances in Artificial Intelligence for Wireless Internet of Things and Cyber-Physical Systems: A Comprehensive Survey // *IEEE Internet of Things Journal*. 2022. vol. 9. no. 15. pp. 12916–12930. DOI: 10.1109/JIOT.2022.3170449.
14. Costa B., Bachiega J., De Carvalho L.R., Araujo A.P. Orchestration in Fog Computing: A Comprehensive Survey // *ACM Computing Surveys*. 2023. vol. 55. no. 2. DOI: 10.1145/3486221.
15. Niveditha P.S., John S.P., Simpson S.V. Review on Edge Computing-assisted d2d Networks // *International Conference on Innovative Computing and Communications: Proceedings of ICICC*. 2022. pp. 41–58. DOI: 10.1007/978-981-19-2821-5_4.
16. Nguyen T.A., Kaliappan V.K., Jeon S., Jeon K.-S., Lee J.-W., Min D. Blockchain Empowered Federated Learning with Edge Computing for Digital Twin Systems in Urban Air Mobility // *In Asia-Pacific International Symposium on Aerospace Technology*. 2023. pp. 935–950. DOI: 10.1007/978-981-19-2635-8_69.
17. ISO/IEC 25010. Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. 2011. Available at: <https://www.iso.org/standard/35733.html> (accessed 20.09.2023).
18. Divan M.J., Sanchez-Reynoso M.L., Panebianco J.E., Mendez M.J. IoT-Based Approaches for Monitoring the Particulate Matter and Its Impact on Health // *IEEE Internet of Things Journal*. 2021. vol. 8. no. 15. pp. 11983–12003. DOI: 10.1109/JIOT.2021.3068898.
19. Divan M.J. Data-driven decision making // *Proceeding of International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)*. 2017. pp. 50–56. DOI: 10.1109/ICTUS.2017.8285973.

20. Runeson P., Host M. Guidelines for conducting and reporting case study research in software engineering // *Empirical software engineering*. 2009. vol. 14. pp. 131–164. DOI: 10.1007/s10664-008-9102-8.
21. Verner J.M., Sampson J., Tosic V., Bakar N.A., Kitchenham B.A. Guidelines for industrially-based multiple case studies in software engineering // *Proceeding of Third International Conference on Research Challenges in Information Science*. 2009. pp. 313–324. DOI: 10.1109/RCIS.2009.5089295.
22. Petersen K., Vakkalanka S., Kuzniarz L. Guidelines for conducting systematic mapping studies in software engineering: An update // *Information and software technology*. 2015. vol. 64. pp. 1–18. DOI: 10.1016/j.infsof.2015.03.007.
23. Prajeesha, Anuradha M. EDGE Computing Application in SMART GRID-A Review // *Proceeding of the 2nd International Conference on Electronics and Sustainable Communication Systems (ICESC'2021)*. 2021. pp. 397–402. DOI: 10.1109/ICESC51422.2021.9532792.
24. Anikwe C.V., Nweke H.F., Ikegwu A.C., Egwuonwu C.A., Onu F.U., Alo U.R., Teh Y.W. Mobile and wearable sensors for data-driven health monitoring system: State-of-the-art and prospect // *Expert Systems with Applications*. 2022. vol. 202. DOI: 10.1016/j.eswa.2022.117362.
25. Hu Z., Xu X., Zhang Y., Tang H., Cheng Y., Qian C., Khosravi M.R.. Cloud-edge cooperation for meteorological radar big data: a review of data quality control // *Complex and Intelligent Systems*. 2021. DOI: 10.1007/s40747-021-00581-w.
26. Groshev M., Guimaraes C., De La Oliva A., Gazda R. Dissecting the Impact of Information and Communication Technologies on Digital Twins as a Service // *IEEE Access*. 2021. vol. 9. pp. 102862–102876. DOI: 10.1109/ACCESS.2021.3098109.
27. Begum B.A. Nandury S.V. A Survey of Data Aggregation Protocols for Energy Conservation in WSN and IoT // *Wireless Communications and Mobile Computing*. 2022. vol. 2022. DOI: 10.1155/2022/8765335.
28. Gurewitz O., Shifrin M., Dvir E. Data Gathering Techniques in WSN: A Cross-Layer View // *Sensors*. 2022. vol. 22. no. 7. DOI: 10.3390/s22072650.
29. Bayih A.Z., Morales J., Assabie Y., de By R.A. Utilization of Internet of Things and Wirelless Sensor Networks for Sustainable Smallholder Agriculture // *Sensors*. 2022. vol. 22. no. 9. DOI: 10.3390/s22093273.
30. Mirani A.A., Velasco-Hernandez G., Awasthi A., Walsh J. Key Challenges and Emerging Technologies in Industrial IoT Architectures: A Review // *Sensors*. 2022. vol. 22, no. 15. DOI: 10.3390/s22155836.
31. Bagwari S., Gehlot A., Singh R., Priyadarshi N., Khan B. Low-Cost Sensor-Based and LoRaWAN Opportunities for Landslide Monitoring Systems on IoT Platform: A Review // *IEEE Access*. 2022. vol. 10. pp. 7107–7127. DOI: 10.1109/ACCESS.2021.3137841.
32. Shahbazi Z., Byun Y.-C. Analysis of the Security and Reliability of Cryptocurrency Systems Using Knowledge Discovery and Machine Learning Methods // *Sensors*. 2022. vol. 22. no. 23. DOI: 10.3390/s22239083.
33. Amador-Domínguez E., Serrano E., Manrique D. GENI: A framework for the generation of explanations and insights of knowledge graph embedding predictions // *Neurocomputing*. 2023. vol. 521. pp. 199–212. DOI: 10.1016/j.neucom.2022.12.010.
34. Valente F., Paredes S., Henriques J., Rocha T., de Carvalho P., Morais J. Interpretability, personalization and reliability of a machine learning based clinical decision support system // *Data Min. Knowl. Discov*. 2022. vol. 36. no. 3. pp. 1140–1173. DOI: 10.1007/s10618-022-00821-8.
35. Yang B., Bai X., Zhang C. Data Collection Method of Energy Adaptive Distributed Wireless Sensor Networks Based on UAV // *Wirel. Commun. Mob. Comput*. 2022. vol. 2022. DOI: 10.1155/2022/3469221.

36. Bai Y., Cao L., Wang S., Ding H., Yue Y. Data Collection Strategy Based on OSELM and Gray Wolf Optimization Algorithm for Wireless Sensor Networks // *Comput. Intell. Neurosci.* 2022. vol. 2022. DOI: 10.1155/2022/4489436.
37. Wei D. et al. Power-Efficient Data Collection Scheme for AUV-Assisted Magnetic Induction and Acoustic Hybrid Internet of Underwater Things // *IEEE Internet Things J.* 2022. vol. 9. no. 14. pp. 11675–11684. DOI: 10.1109/JIOT.2021.3131679.
38. Benmansour F.L., Labraoui N.A. Comprehensive Review on Swarm Intelligence-Based Routing Protocols in Wireless Multimedia Sensor Networks // *Int. J. Wirel. Inf. Networks.* 2021. vol. 28. no. 2. pp. 175–198. DOI: 10.1007/s10776-021-00508-9.
39. Hannan M.A., Hassan K., Jern K.P. A review on sensors and systems in structural health monitoring: Current issues and challenges // *Smart Struct. Syst.* 2018. vol. 22. no. 5. pp. 509–525. DOI: 10.12989/sss.2018.22.5.509.
40. Ajakwe S.O., Nwakanma C.I., Kim D.-S., Lee J.-M. Key Wearable Device Technologies Parameters for Innovative Healthcare Delivery in B5G Network: A Review // *IEEE Access.* 2022. vol. 10. pp. 49956–49974. DOI: 10.1109/ACCESS.2022.3173643.
41. SJR – SCImago Journal and Country Rank [Official web site of SCImago]. 2020. Available at: www.scimagojr.com (accessed 23.09.2020).
42. Guerrero-Bote V.P., Moya-Anegón F. A further step forward in measuring journals' scientific prestige: The SJR2 indicator // *J. Informetr.* 2012. vol. 6. no. 4. pp. 674–688. DOI: 10.1016/j.joi.2012.07.001.
43. Sung W., Hsu C. IOT system environmental monitoring using IPSO weight factor estimation // *Sens. Rev.* 2013. vol. 33. no. 3. pp. 246–256. DOI: 10.1108/02602281311324708.
44. Pournaras E., Yao M., Helbing D. Self-regulating supply–demand systems // *Futur. Gener. Comput. Syst.* 2017. vol. 76. pp. 73–91. DOI: 10.1016/j.future.2017.05.018.
45. Satija U., Ramkumar B., Manikandan M.S. Real-Time Signal Quality-Aware ECG Telemetry System for IoT-Based Health Care Monitoring // *IEEE Internet Things J.* 2017. vol. 4. no. 3. pp. 815–823. DOI: 10.1109/JIOT.2017.2670022.
46. Al-Jaroodi J., Mohamed N. PsCPS: A Distributed Platform for Cloud and Fog Integrated Smart Cyber-Physical Systems // *IEEE Access.* 2018. vol. 6. pp. 41432–41449. DOI: 10.1109/ACCESS.2018.2856509.
47. Pore M., Chakati V., Banerjee A., Gupta S.K.S. ContextAiDe // *ACM Trans. Internet Technol.* 2019. vol. 19. no. 2. pp. 1–23. DOI: 10.1145/3301444.
48. Albahri O.S. et al. Fault-Tolerant mHealth Framework in the Context of IoT-Based Real-Time Wearable Health Data Sensors // *IEEE Access.* 2019. vol. 7. pp. 50052–50080. DOI: 10.1109/ACCESS.2019.2910411.
49. Cao K., Xu G., Zhou J., Wei T., Chen M., Hu S. QoS-Adaptive Approximate Real-Time Computation for Mobility-Aware IoT Lifetime Optimization // *IEEE Trans. Comput. Des. Integr. Circuits Syst.* 2019. vol. 38. no. 10. pp. 1799–1810. DOI: 10.1109/TCAD.2018.2873239.
50. Dong R., She C., Hardjawana W., Li Y., Vucetic B. Deep Learning for Hybrid 5G Services in Mobile Edge Computing Systems: Learn From a Digital Twin // *IEEE Trans. Wirel. Commun.* 2019. vol. 18. no. 10. pp. 4692–4707. DOI: 10.1109/TWC.2019.2927312.
51. Xu X., He C., Xu Z., Qi L., Wan S., Bhuiyan M.Z.A. Joint Optimization of Offloading Utility and Privacy for Edge Computing Enabled IoT // *IEEE Internet Things J.* 2020. vol. 7. no. 4. pp. 2622–2629. DOI: 10.1109/JIOT.2019.2944007.
52. Farahani B., Barzegari M., Aliee F.S., Shaik K.A. Towards collaborative intelligent IoT eHealth: From device to fog, and cloud // *Microprocess. Microsyst.* 2020. vol. 72. p. 102938. DOI: 10.1016/j.micpro.2019.102938.

53. Zhang T. et al. A Joint Deep Learning and Internet of Medical Things Driven Framework for Elderly Patients // IEEE Access. 2020. vol. 8. pp. 75822–75832. DOI: 10.1109/ACCESS.2020.2989143.
54. Sodhro A.H., Sodhro G.H., Guizani M., Pirbhulal S., Boukerche A. AI-Enabled Reliable Channel Modeling Architecture for Fog Computing Vehicular Networks // IEEE Wirel. Commun. 2020. vol. 27. no. 2. pp. 14–21. DOI: 10.1109/MWC.001.1900311.
55. Bhatia M., Sood S.K. Quantum Computing-Inspired Network Optimization for IoT Applications // IEEE Internet Things J. 2020. vol. 7. no. 6. pp. 5590–5598. DOI: 10.1109/JIOT.2020.2979887.
56. Zhang J. Research on environmental monitoring trend analysis based on internet of things visualization technology // Fresenius Environ. Bull. 2020. vol. 29. no. 2. pp. 1054 – 1062. [Online]. Available at: www.scopus.com/inward/record.uri?eid=2-s2.0-85090455681&partnerID=40&md5=7c5626045c3f39320bb91185bf694295 (accessed 11.05.2023).
57. Loke G. et al. Digital electronics in fibers enable fabric-based machine-learning inference // Nat. Commun. 2021. vol. 12. no. 1. p. 3317. DOI: 10.1038/s41467-021-23628-5.
58. Ke R., Zhuang Y., Pu Z., Wang Y. A Smart, Efficient, and Reliable Parking Surveillance System with Edge Artificial Intelligence on IoT Devices // IEEE Trans. Intell. Transp. Syst. 2021. vol. 22. no. 8. pp. 4962–4974. DOI: 10.1109/TITS.2020.2984197.
59. Wang X., Garg S., Lin H., Piran M.J., Hu J., Hossain M.S. Enabling Secure Authentication in Industrial IoT with Transfer Learning Empowered Blockchain // IEEE Trans. Ind. Informatics. 2021. vol. 17. no. 11. pp. 7725–7733. DOI: 10.1109/TII.2021.3049405.
60. Lee W.J., Xia K., Denton N.L., Ribeiro B., Sutherland J.W. Development of a speed invariant deep learning model with application to condition monitoring of rotating machinery // J. Intell. Manuf. 2021. vol. 32. no. 2. pp. 393–406. DOI: 10.1007/s10845-020-01578-x.
61. Ibrar M., Wang L., Muntean G.-M., Chen J., Shah N., Akbar A. IHSF: An Intelligent Solution for Improved Performance of Reliable and Time-Sensitive Flows in Hybrid SDN-Based FC IoT Systems // IEEE Internet Things J. 2021. vol. 8. no. 5. pp. 3130–3142. DOI: 10.1109/JIOT.2020.3024560.
62. Brik B., Esseghir M., Merghem-Boulahia L., Snoussi H. An IoT-based deep learning approach to analyse indoor thermal comfort of disabled people // Build. Environ. 2021. vol. 203. p. 108056. DOI: 10.1016/j.buildenv.2021.108056.
63. Singh P.D., Kaur R., Singh K.D., Dhiman G., Soni M. Fog-centric IoT based smart healthcare support service for monitoring and controlling an epidemic of Swine Flu virus // Informatics Med. Unlocked. 2021. vol. 26. p. 100636. DOI: 10.1016/j.imu.2021.100636.
64. Bhatia M. Intelligent System of Game-Theory-Based Decision Making in Smart Sports Industry // ACM Trans. Intell. Syst. Technol. 2021. vol. 12. no. 3. pp. 1–23. DOI: 10.1145/3447986.
65. Baggag A. et al. Learning Spatiotemporal Latent Factors of Traffic via Regularized Tensor Factorization: Imputing Missing Values and Forecasting // IEEE Trans. Knowl. Data Eng. 2021. vol. 33. no. 6. pp. 2573–2587. DOI: 10.1109/TKDE.2019.2954868.
66. Razzaq M.A., Mahar J.A., Ahmad M., Saher N., Mehmood A., Choi G.S. Hybrid Auto-Scaled Service-Cloud-Based Predictive Workload Modeling and Analysis for Smart Campus System // IEEE Access. 2021. vol. 9. pp. 42081–42089. DOI: 10.1109/ACCESS.2021.3065597.

67. Jin H., Zhao J. Real-time energy consumption detection simulation of network node in internet of things based on artificial intelligence // *Sustainable Energy Technologies and Assessments*. 2021. vol. 44. no. 101004. DOI: 10.1016/j.seta.2021.101004.
68. Alzamzami F., El Saddik A. Monitoring Cyber SentiHate Social Behavior During COVID-19 Pandemic in North America // *IEEE Access*. 2021. vol. 9. pp. 91184–91208. DOI: 10.1109/ACCESS.2021.3088410.
69. Adhikari M., Ambigavathi M., Menon V.G., Hammoudeh M. Random Forest for Data Aggregation to Monitor and Predict COVID-19 Using Edge Networks // *IEEE Internet Things Mag.* 2021. vol. 4. no. 2. pp. 40–44. DOI: 10.1109/IOTM.0001.2100052.
70. Jurdi R., Andrews J.G., Heath R.W. Scheduling Observers Over a Shared Channel With Hard Delivery Deadlines // *IEEE Trans. Commun.* 2021. vol. 69. no. 1. pp. 133–148. DOI: 10.1109/TCOMM.2020.3032172.
71. Hashash O., Sharafeddine S., Dawy Z., Mohamed A., Yaacoub E. Energy-Aware Distributed Edge ML for mHealth Applications with Strict Latency Requirements // *IEEE Wirel. Commun. Lett.* 2021. vol. 10. no. 12. pp. 2791–2794. DOI: 10.1109/LWC.2021.3117876.
72. Liang W., Li W., Feng L. Information Security Monitoring and Management Method Based on Big Data in the Internet of Things Environment // *IEEE Access*. 2021. vol. 9. pp. 39798–39812. DOI: 10.1109/ACCESS.2021.3064350.
73. Shao S., Zhang Q., Guo S., Qi F. Task Allocation Mechanism for Cable Real-Time Online Monitoring Business Based on Edge Computing // *IEEE Syst. J.* 2021. vol. 15. no. 1. pp. 1344–1355. DOI: 10.1109/JSYST.2020.2988266.
74. Vaidya G., Nambi A., Prabhakar T.V., Kumar V.T., Sudhakara S. Towards generating a reliable device-specific identifier for IoT devices // *Pervasive Mob. Comput.* 2021. vol. 76. DOI: 10.1016/j.pmcj.2021.101445.
75. Rajendran S. et al. Emphasizing privacy and security of edge intelligence with machine learning for healthcare // *Int. J. Intell. Comput. Cybern.* 2022. vol. 15. no. 1. pp. 92–109. DOI: 10.1108/IJICC-05-2021-0099.
76. Zahid N., Sodhro A.H., Kamboh U.R., Alkhayyat A., Wang L. AI-driven adaptive reliable and sustainable approach for internet of things enabled healthcare system // *Math. Biosci. Eng.* 2022. vol. 19. no. 4. pp. 3953–3971. DOI: 10.3934/mbe.2022182.
77. Bhardwaj A. et al. Smart IoT and Machine Learning-based Framework for Water Quality Assessment and Device Component Monitoring // *Environ. Sci. Pollut. Res.* 2022. vol. 29. no. 30. pp. 46018–46036. DOI: 10.1007/s11356-022-19014-3.
78. Alsalemi A., Himeur Y., Bensaali F., Amira A. An innovative edge-based Internet of Energy solution for promoting energy saving in buildings // *Sustainable Cities and Society*. 2022. vol. 78. no. 103571. DOI: 10.1016/j.scs.2021.103571.
79. Khan M.A., Ghazal T.M., Lee S.-W., Rehman A. Data Fusion-Based Machine Learning Architecture for Intrusion Detection // *Comput. Mater. Contin.* 2022. vol. 70. no. 2. pp. 3399–3413. DOI: 10.32604/cmc.2022.020173.
80. Gültekin Ö., Cinar E., Özkan K., Yazıcı A. Real-Time Fault Detection and Condition Monitoring for Industrial Autonomous Transfer Vehicles Utilizing Edge Artificial Intelligence // *Sensors*. 2022. vol. 22(9). no. 3208. DOI: 10.3390/s22093208.
81. Wei L., Hou S., Liu Q. Clinical Care of Hyperthyroidism Using Wearable Medical Devices in a Medical IoT Scenario // *Journal of Healthcare Engineering*. 2022. vol. 2022. DOI: 10.1155/2022/5951326.
82. Nikolov G., Kuhn M., McGibney A., Wenning B.-L. MABASR – A Robust Wireless Interface Selection Policy for Heterogeneous Vehicular Networks // *IEEE Access*. 2022. vol. 10. pp. 26068–26077. DOI: 10.1109/ACCESS.2022.3156597.

83. Baek J., Kaddoum G. Online Partial Offloading and Task Scheduling in SDN-Fog Networks With Deep Recurrent Reinforcement Learning // IEEE Internet Things J. 2022. vol. 9. no. 13. pp. 11578–11589. DOI: 10.1109/JIOT.2021.3130474.
84. Thenmozhi R., Sakthivel P., Kulothungan K. Hybrid multi-objective-optimization algorithm for energy efficient priority-based QoS routing in IoT networks // Wireless Networks. 2022. DOI: 10.1007/s11276-021-02848-z.
85. Manoharan S.N., Kumar K.M.V.M., Vadivelan N.A. Novel CNN-TLSTM Approach for Dengue Disease Identification and Prevention using IoT-Fog Cloud Architecture // Neural Processing Letters. 2022. vol. 55. no. 2. pp. 1951–1973. DOI: 10.1007/s11063-022-10971-x.
86. Sithik M.M., Kumar B.M. Intelligent agent based virtual clustering and multi-context aware routing for congestion mitigation in secure RPL-IoT environment // Ad Hoc Networks. 2022. vol. 137. DOI: 10.1016/j.adhoc.2022.102972.
87. Zhang Y., Wu J., Liu M., Tan A. TSN-based routing and scheduling scheme for Industrial Internet of Things in underground mining // Engineering Applications of Artificial Intelligence. 2022. vol. 115. DOI: 10.1016/j.engappai.2022.105314.
88. Eroshkin I., Vojtech L., Neruda M. Resource Efficient Real-Time Reliability Model for Multi-Agent IoT Systems // IEEE Access. 2022. vol. 10. pp. 2578–2590. DOI: 10.1109/ACCESS.2021.3138931.
89. Zhou S., Du Y., Chen B., Li Y., Luan X. An Intelligent IoT Sensing System for Rail Vehicle Running States Based on TinyML // IEEE Access. 2022. vol. 10. pp. 98860–98871. DOI: 10.1109/ACCESS.2022.3206954.
90. Sinha A., Das D., Udutalapally V., Mohanty S.P. iThing: Designing Next-Generation Things with Battery Health Self-Monitoring Capabilities for Sustainable IIoT // IEEE Trans. Instrum. Meas. 2022. vol. 71. pp. 1–9. DOI: 10.1109/TIM.2022.3216594.
91. Manocha A., Singh R. A Novel Edge Analytics Assisted Motor Movement Recognition Framework Using Multi-Stage Convo-GRU Model // Mob. Networks Appl. 2022. vol. 27. no. 2. pp. 657–676. DOI: 10.1007/s11036-019-01321-8.
92. Bollen E. et al. A database system for querying of river networks: facilitating monitoring and prediction applications // Water Supply. 2022. vol. 22. no. 3. pp. 2832–2846. DOI: 10.2166/ws.2021.433.
93. Fournier A.M.V. et al. Structured Decision Making to Prioritize Regional Bird Monitoring Needs // INFORMS Journal on Applied Analytics. 2023. vol. 53(3). pp. 207–217. DOI: 10.1287/inte.2022.1154.
94. Fanelli S., Pratici L., Salvatore F.P., Donelli C.C., Zangrandi A. Big data analysis for decision-making processes: challenges and opportunities for the management of health-care organizations // Management Research Review. 2023. vol. 46. no. 3. pp. 369–389. DOI: 10.1108/MRR-09-2021-0648.
95. Overdal M., Haddara M., Langseth M. Exploring Public Cloud-ERP Systems' Impact on Organizational Performance // Lect. Notes Networks Syst. 2023. vol. 561. pp. 121–137. DOI: 10.1007/978-3-031-18344-7_8.
96. Reyna A., Martin C., Chen J., Soler E., Díaz M. On blockchain and its integration with IoT. Challenges and opportunities // Futur. Gener. Comput. Syst. 2018. vol. 88. pp. 173–190. DOI: 10.1016/j.future.2018.05.046.

Диван Марио Хосе — Ph.D., почетный профессор, старший архитектор решений, отделение разработки надежности вычислительных сред, Корпорация Intel. Область научных интересов: кибербезопасность, управление, масштабируемость и монетизация ИТ-сервисных платформ. Число научных публикаций — 91. mario.jose.divan.koller@intel.com; 25-я авеню, Кампус Джонс Фарм, 3, 97124, Хиллсборо, штат Орегон, Соединённые штаты Америки; р.т.: +1(503)264-0180.

Щемелинин Дмитрий Александрович — д-р техн. наук, руководитель направления, отделение разработки надежности облачных платформ, Корпорация Intel. Область научных интересов: мониторинг, управление состоянием и масштабируемость глобально распределенных вычислительных комплексов с применением методов искусственного интеллекта, сбор и анализ больших данных. Число научных публикаций — 78. dshchmel@gmail.com; 25-я авеню, Кампус Джонс Фарм, 3, 97124, Хиллсборо, штат Орегон, Соединённые штаты Америки; р.т.: +1(408)981-5963.

Карранса Маркос — главный инженер, отделение разработки надежности вычислительных сред, корпорация Intel. Область научных интересов: кибербезопасность, управление, масштабируемость и монетизация ИТ-сервисных платформ. Число научных публикаций — 32. marcos.e.carranza@intel.com; 25-я авеню, Кампус Джонс Фарм, 3, 97124, Хиллсборо, штат Орегон, Соединённые штаты Америки; р.т.: +1(503)764-7952.

Мартинес-Спессот Цезарь Игнасио — руководитель направления, отделение разработки надежности вычислительных сред, Корпорация Intel. Область научных интересов: применение цифровых технологий для улучшения бизнес-процессов с применением методов искусственного интеллекта, развитие экосистем и приложений Интернета вещей. Число научных публикаций — 31. cesar.martinez@intel.com; 25-я авеню, Кампус Джонс Фарм, 3, 97124, Хиллсборо, штат Орегон, Соединённые штаты Америки; р.т.: +1(503)7120559.

Буйневич Михаил Викторович — д-р техн. наук, профессор кафедры, кафедра прикладной математики и информационных технологий, Санкт-Петербургский университет государственной противопожарной службы МЧС России. Область научных интересов: киберустойчивость информационно-телекоммуникационных систем, методы искусственного интеллекта в обработке данных о состоянии сложных систем, подготовка специалистов в области информационной безопасности, киберполигон. Число научных публикаций — 170. bmv1958@yandex.ru; Московский проспект, 149, 196105, Санкт-Петербург, Россия; р.т.: +7(812)388-8639.

Поддержка исследований. Эта работа была полностью поддержана корпорацией Intel.

K. KRISHNA, R. THIRUMURU
**A BALANCED INTRUSION DETECTION SYSTEM FOR
WIRELESS SENSOR NETWORKS IN A BIG DATA
ENVIRONMENT USING CNN-SVM MODEL**

Krishna K., Thirumuru R. A Balanced Intrusion Detection System for Wireless Sensor Networks in a Big Data Environment Using CNN-SVM Model.

Abstract. Wireless Sensor Networks (WSNs) were exposed to several distinct safety issues and attacks regarding gathering and sending data. In this scenario, one of the most prevalent WSN assaults that may target any tier of the protocol stack is the Denial of Service (DoS) attack. The current research suggested various strategies to find the attack in the network. However, it has classification challenges. An effective ensemble deep learning-based intrusion detection system to identify the assault in the WSN network was, therefore, suggested in this research to address this issue. The data pre-processing involves converting qualitative data into numeric data using the One-Hot Encoding technique. Following that, Normalization Process was carried out. Then Manta-Ray Foraging Optimization is suggested to choose the best subset of features. Then Synthetic Minority Oversampling Technique (SMOTE) oversampling creates a new minority sample to balance the processed dataset. Finally, CNN-SVM classifier is proposed to classify the attack kinds. The Accuracy, F-Measure, Precision, and Recall metrics were used to assess the outcomes of 99.75%, 99.21%, 100%, and 99.6%, respectively. Compared to existing approaches, the proposed method has shown to be extremely effective in detecting DoS attacks in WSNs.

Keywords: WSN, DoS attacks, artificial intelligence, deep learning, Convolutional Neural Networks (CNN), Support Vector Machine (SVM).

1. Introduction. Networking systems are used by people worldwide to solve problems, find fresh perspectives, and fulfil fundamental needs. One of the most recent and well-liked technical developments is the creation of sensors, which permit users to accept data from a long distance. Devices of the Internet of Things (IoT) are using sensors more frequently these days [1]. Wireless Sensor Networks are regarded as an important research topic. Many uses, including telecommunications, the military, healthcare, research, and agriculture, can benefit from the technology [2] and see natural catastrophes like earthquakes, floods, and volcanoes [3]. All of the nodes involved in the design of these applications are presumed to be reliable and cooperative. However, in actual deployments, nodes are subject to various intrusions and assaults that can substantially impact the system's performance and the network's capacity to function. Sadly, protecting this form of network from numerous harmful assaults is a challenging issue [4]. A DoS assault is the most frequent danger that WSNs encounter [5]. DoS attacks may harm a network service by blocking real traffic from connecting to the network by saturating it with many fraudulent requests. DoS assaults may occur at any

tier of the TCP/IP protocol stack [6, 7]. The WSN's security must be verified using the Intrusion Detection System (IDS).

IDSs keep an eye on system activity to spot and stop malicious traffic. Attacks may be immediately detected by choosing the usual network traffic design and size. IDS monitors and examines network-generated events to spot anything unusual and notify sensor nodes of an intrusion [8]. Data received from sensors may be used in real time by intrusion detection/prevention systems and cyber threat analysts to identify useful information. With the aid of this data, security solutions can be created to identify weaknesses and assaults.

One of the most popular ways to help IDSs better identify and distinguish intruders is to combine them with artificial intelligence (AI) approaches [9]. To detect DoS in WSN, the AI-based classification algorithm has been applied. This study used deep learning techniques to assess the defences against DoS attacks in WSN. To verify the efficiency of these defence methods in identifying and categorizing the many types of DoS assaults, they were assessed on a unique Wireless Sensor Network DataSet called WSN-DS, which included several normal and attack circumstances. The primary goal of attack detection, classified as a classification challenge, is determining if an assault is Flooding, Black Hole, Normal, Time Division Multiple Access (TDMA), or Gray Hole. Thus, this study suggested a unique deep-learning architecture to address the issues of attack detection. This study's primary contribution is as follows.

Pre-processing the data represents the initial stage of the suggested method. Here, the One-Hot Encoding approach changed the categorical data into numeral information. Following that, the normalization process is carried out to determine the normalized value using average and standard deviation values.

- Then, to select the best subset of aspects, this research proposed a Manta Ray Foraging Optimization, which makes every possible feature combination and returns the set of attributes that performs the best. Moreover, to balance the dataset, SMOTE oversampling approach is employed, which increases the classification performance.

- Finally, this research proposed a CNN-SVM-based deep learning approach to classify the attacks. Therefore, the suggested multi-class intrusion detection technique can properly categorize particular class assaults.

- This article is organized as follows: Section 2 reviews the existing artificial intelligence-based malicious threat detection approaches in WSN. Section 3 describes the proposed deep learning-based attack detection framework. Next, model implementation and their consecutive outcomes were discussed in Section 4. Finally, Section 5 concludes this research paper.

2. Literature survey. Articles on wireless sensor networks has been published more often in recent years WSN. Some efficient anomaly detection techniques, such as those based on the composition and properties of WSNs, classification methods, clustering algorithms, machine learning algorithms, and statistical learning models, are suggested in the literature. Despite the benefits of WSN, several security flaws make it vulnerable to DoS assaults. Many technological developments have made identifying and defending against DoS attacks simpler. However, the most efficient strategies for averting such safety problems and DoS attacks were supplied via deep learning.

Using edge and cloud data analysis, the authors in [10] presented a new method for automatic anomaly detection for heterogeneous sensor networks. The latter employs an artificial neural network method that is completely unsupervised, whilst the former uses a multi-parameterized edit distance algorithm created inside a residential structure and subsequently altered with various fictitious impairments. The findings demonstrate that the suggested technique can appropriately recognize abnormalities and self-adapt to environmental fluctuations.

In [11] the authors developed a highly scalable architecture for a hybrid intrusion detection warning system to create an adaptable and efficient IDS that can categorize unexpected and surprising cyberattacks. This system learns how to represent abstract and high-dimensional features by passing the IDS data through several hidden layers. However, the suggested system does not provide comprehensive details on the makeup and traits of the malware.

A new model, the genetic algorithm and an extreme gradient boosting (GXGBoost) model, was suggested by the authors in [12], using a XGBoost classifier to identify intrusion assaults. In the extremely inconsistent data flow of WSNs, this model seeks to develop the capability of conventional approaches to identifying certain threats. To lessen the number of aspects and increase the efficacy of incursion detection in WSN, this model does not concentrate on feature selection approaches.

To maintain the dataset's balance and develop the intrusion detection classifier, in study [13] developed a way using the SMOTE. However, this technique must further enhance its ability to recognize wireless sensor network intrusion data.

In paper [14] presented the backward sequence selection with Light Gradient Boosting Machine (SLGBM) approach to detect intrusions in wireless sensor networks. The Sequence Backward Selection (SBS) technique was initially used to minimize the feature space's data dimension to lower the real traffic data's computing cost. To identify various network

assaults, a LightGBM algorithm is then used. But this approach does not employ a distributed mechanism to further cut down on time.

For intrusion detection, in [15] SVM had the maximum accuracy compared to other algorithms, whereas Gradient Boosting Classifier had the minimum accuracy rating. It was discovered after examining the efficacy of several machine-learning techniques. However, this study is still not focused on multi-class classification and considering features to choose the most crucial aspects and improve accuracy.

In paper [16] the LightGBM method and autoencoders were suggested for constructing a network intrusion detection system, with the latter being used for feature selection and the former for training and detection. A foundation for intrusion detection is provided when data collection, including network intrusion behaviours, is fed into an autoencoder because there is a significant difference in reconstructural error between the input data and the output data produced by the autoencoder. An adequate threshold is determined based on the reconstruction error to discriminate between defensive and offensive actions symmetrically. However, this strategy was limited to binary class classification.

A unique feature selection technique called dynamic recursive feature selection algorithm (DRFSA), which determines how many characteristics are appropriate for categorization, was suggested by the authors in their paper published in Science [17]. Employing fuzzy temporal constraints as a smart addition to the decision tree algorithm is also suggested to categorize network users and traffic better accurately. Convolution neural networks are also used for data classification in huge volumes. No distributed environment was used to implement this work.

In paper [18] the authors suggested a network intrusion detection system based on LightGBM and adaptive synthetic (ADASYN) oversampling technology. The authors initially used information pre-processing to normalize and one-hot encode the original data to lessen the impact of the top or lowest ranking for all requirements. Next, the ADASYN oversampling method was used to deal with the poor rate of minority assault detection brought on by an imbalance in the training data. Finally, the system's temporal complexity is reduced while preserving detection accuracy by using the LightGBM ensemble learning model. A big data environment is not considered, though.

In study [19] the authors suggested a classification-based network attack detection method for massive amounts of data. The SMOTE + Tomek-Link – Hybrid Deep Learning (STL-HDL) method combines a hybrid deep learning network here with data balancing. The STL ensemble method was employed in the research for data balancing. The Convolutional

Neural Network-Long Short Term Memory (CNN-LSTM) is utilized for classification. However, a WSN-DS real-time dataset has not been utilized to evaluate how well the proposed strategy worked.

In study [20] the authors suggested a passive-aggressive online classifier and an information gain ratio-based model. First, the important parts of the sensor data are chosen using the information gain ratio. Second, several forms of Denial of Service assaults have been identified and categorized using the online Passive Aggressive algorithm. However, the data balance has not been taken into account.

Study [21] suggested a light-weight Intelligent Intrusion Detection Model for WSN. Utilizing the k-nearest neighbour algorithm (kNN) and the sine-cosine algorithm (SCA) allows for the intelligent identification of various threats, including unknown assaults, while significantly improving classification accuracy and reducing false alarm rates. The polymorphic mutation (PM) technique is employed to compensate for the decrease in optimization precision, and compact machine SCA (CSCA) is employed to reduce computation time and space. On the other hand, the big data environment is not considered.

In [22] the authors developed a memory-efficient, light-weight anomaly detection system that maintains high accuracy while reducing computing complexity. Concepts of dimension reduction and one-class learning were used in the study. The One-Class Support Vector Machine (OCSVM) was employed for one-class learning and size reduction, and the Candid Covariance-Free Incremental Principal Component Analysis (CCIPCA) approach was employed. The normal reference model must be adaptively learned over time to guarantee that sensor data is accurate because it gets rigid.

In a WSN, in study [23] the authors suggested a technique used for monitoring and thwarting DoS and distributed DoS assaults. The suggested approach uses an RNN as a classifier. However, the dispersed environment is not taken into account in this work.

Paper [24] provided a unique perspective on the malicious security threats in WSNs using an ensemble learning-based quick intrusion detection system that can handle continuous and dynamic data streaming. Although it generates superior results, the pre-processing with data reduction and parameter adjustment is not carried out, which might increase the classifier's effectiveness.

In [25] the authors developed a CNN-LSTM network to identify and categorize DoS intrusion threats. To enhance the accuracy of the classification findings, this strategy does not balance the dataset and yields worse performance outcomes.

To identify DoS assaults targeted specifically at WSNs, the authors in [26] suggested a special DoS Intrusion Detection System (DDS). STLGBM-DDS, the proposed system, integrates the LightGBM machine learning algorithm, data balance and feature selection techniques to create an ensemble intrusion detection system. It was created using the big data platform Apache Spark. The data imbalance's impact on system performance was reduced by treating the data imbalance utilizing the SMOTE and the STL sampling methods. Information Gain Ratio is also employed as a feature selection strategy during the data preparation. However, using incorrect neural network parameters has impacted classification accuracy.

Considering the literature review cited earlier, almost all studies that advocate for deploying intrusion detection systems acknowledge that the data imbalance problem is still unresolved. Furthermore, most research has overlooked feature selection. To illustrate the need for large data environments as the amount of WSNs data increases daily, not all required work is completed in one.

3. Proposed Method. WSN deployments have increased dramatically during the past several years. WSNs are being used in various applications across numerous sectors because of their tiny size and low cost. WSNs are exposed to several distinct security concerns and attacks regarding data collection and transmission. The DoS attack is among the most widespread WSN assaults that can target any protocol stack layer. Attack detection is a crucial responsibility in this case for protecting the network and the data. To provide security, this research proposed a deep learning framework. This framework consists of the following steps: data pre-processing, data normalization, feature selection, data balancing, and attack classification shown in Figure 1.

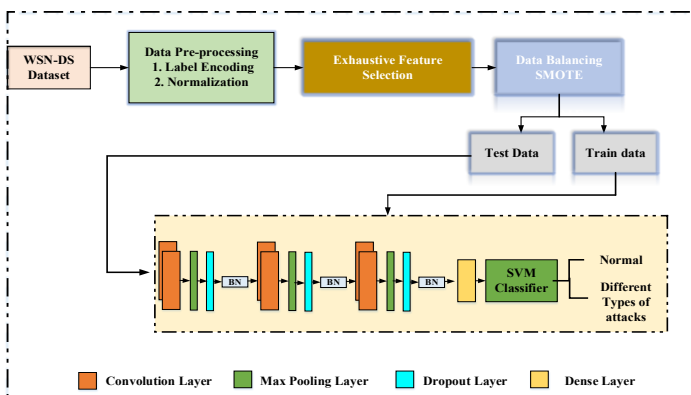


Fig. 1. Architecture of the proposed approach

The initial data pre-processing stage involves preparing the raw dataset for categorization algorithms – the One-Hot Encoding technique, which transforms the categorical data into numeric data. Then the Normalization Process was carried out to determine the normalized value using mean and standard deviation values. After normalization, removing unnecessary and redundant features and choosing the best subset of features is crucial using a feature selection approach that uses a process called Manta-Ray Foraging based Feature Selection that assesses each feature set using a fitness function. It implies that this method attempts to make every possible feature combination and returns the set of attributes that performs the best. After that, we must balance the imbalanced dataset to increase classification performance. The assault detection dataset contains imbalanced classes, which hurts classification performance. Oversampling with under-sampling approaches was used in our research. Noise and unwanted data size growth are caused when oversampling techniques are used alone. The imbalanced class problem is addressed in this research using the SMOTE approach. Following that, the dataset must be divided for testing and training purposes.

Attacks are finally identified using the CNN-SVM-based Deep Learning approach, in which the Convolution layer, Maxpooling layer, Batch Normalization, Dropout layer, and Dense layer are employed. Finally, the SVM classifier is utilized to classify the attack kinds. The ADAM optimizer will tune the deep learning model's hyper-parameter to detect intrusions with a high detection accuracy by minimizing the error rate. As a result, the proposed multi-class intrusion detection approach can accurately classify the specific classes such as Flooding, Blackhole, Normal, scheduling, or Grayhole. The following sections are arranged to explain the proposed method process in detail.

3.1. Dataset Description. A WSN-DS dataset was constructed in [27] to detect and classify the attacks on WSNs specifically. The Low-energy adaptive clustering hierarchy (LEACH) protocol was selected to create the dataset since it is one of the most popular and regularly utilized routing protocols in WSNs. The data were gathered in the NS-2 simulation environment. The dataset includes 23 characteristics that were obtained using the LEACH routing technique. The LEACH protocol is a routing protocol that employs 23 characteristics to indicate the status of each sensor node in the wireless network. There are 23 features, including Id, Time, Is_CH, Who_CH, RSSI, Dist_To_CH, M_D_CH, and A_D_CH. Energy that is now being used, consumed, ADV_S, ADV_R, JOIN_S, JOIN_R, ADV_SCH_S, ADV_SCH_R, rank, DATA_S, DATA_R, Data_Sent_BS, Dist_CH_BS, Send_code, and attack_type. However, as indicated in Table 1, there are only 19 features total in the dataset file, in addition to the class label.

Table 1. A detailed description of the attributes of the WSN-DS dataset

No	Attribute Name	Attribute Description
1	Id (Identification)	To recognize the sensor node in any round and at any stage, it is a special ID
2	Time	The sensor node's current simulation time is displayed
3	Is_CH	It acts as a flag to show whether the node is a CH (cluster head). Value 0 denotes a normal node, whereas value 1 denotes a CH
4	who_CH	It is a CH ID for the current round
5	Dist_To_CH	In the current round, it is the separation between the node and its CH
6	ADV_S (Advertise_Send)	The number was used to promote CH's broadcast messages delivered to nodes
7	ADV_R (Advertise_Receive)	The number was used to promote CH communications obtained from CHs
8	JOIN_S (Join_Send)	The amount of join request messages delivered to the CH by the nodes
9	JOIN_R (Join_Receive)	It represents the number of join request messages the CH has received from the nodes
10	SCH_S (Schedule_Send)	It depends on how many advertised TDMA (Time Division Multiple Access) scheduled broadcast messages are sent to the nodes
11	SCH_R (Schedule_Receive)	It is the quantity of advertised TDMA scheduling messages that the CH has received
12	Rank	It is this node's position in the TDMA schedule
13	DATA_S (Data packets_Send)	It refers to how many data packets a sensor delivers to its CH
14	DATA_R (Data packets_Receive)	It measures how many data packets were received from CH
15	Data_Sent_BS (Data packets_Send_Base station)	It refers to how many data packets were transferred to the BS
16	Dist_CH_BS (Distance_Cluster Head_Base station)	It measures the separation between the CH and the BS (base station).
17	Send_Code	It is the cluster-sending node
18	Consumed_Energy	It represents the energy that the sensor node used during the previous round
19	Attack_Type	Type of Attack (A black hole, Grayhole, Flooding, TDMA/Scheduling, Normal)

The WSN-DS dataset has 374.661 samples. The samples in the dataset are separated into five groups, four of which correspond to different forms of DoS attacks: Black hole, grey hole, flooding, TDMA/Scheduling, and normal. Table 2 displays the precise data distribution of each class.

Table 2. The number of samples in each class of WSN-DS Attack

Class	Number of Samples	Proportion (%)
Normal	340066	90.77
Gray hole	10049	2.68
Black hole	14596	3.90
TDMA/Scheduling	3312	0.88
Flooding	6638	1.77
Total	374661	100

This WSN-DS dataset served as the basis for all evaluations in the fundamental research. The dataset was selected as that includes DoS assaults specific to WSNs.

3.2. Data Normalization and Encoding. Constructing the raw dataset suitable for classification algorithms is the first phase in data pre-processing. Due to this, the One-Hot Encoding technique was employed in this work to first convert the dataset's category values to quantitative data, after which the normalization step was carried out by Equation (1).

$$x'_z = |x_z - \mu| / \sigma, \quad (1)$$

where, x_z is the original value, x'_z are the normalized value, μ and σ are the mean and standard deviation values, respectively. As a result, each numerical value in the data set was changed to a number between 0 and 1.

For instance, the Attack Type attribute in the WSN-DS dataset contains textual descriptors like Normal, Grayhole, Blackhole, TDMA, and Flooding. AI-systems are prevented from using these textual descriptions to continue their calculations. Consequently, employing our suggested One-Hot Encoding in conjunction with the normalizing method is required to transform these descriptors into numeric values.

The process of One-Hot encoding can be applied in the WSN-DS dataset:

1. Identify the categorical variables: The attack type is described in textual format in this data set.
2. Create binary columns: Create a new binary column for each categorical variable for each unique category. In the case of "Attack Type,"

five binary columns will be created: "Normal," "Gray hole," "Black hole," "TDMA" and "Flooding".

3. Assign binary values: Assign a value of 1 to the appropriate binary column for each data instance corresponding to the respective category.

4. Remove the original categorical variable: Once the One-Hot encoding is applied, the original categorical variable.

The classification labels were changed into numbers via the One-Hot Encoding process, as shown in Table 3.

Table 3. Findings from Data Normalization and Encoding

Attribute Class	One-Hot Encoding Output
Normal	0
Gray hole	1
Black hole	2
TDMA	3
Flooding	4

Certain characteristics with large numerical values are prevented after normalization and encoding from impairing the algorithm's performance and having a negative impact. These data are then used in the feature selection method; this is covered in greater detail in the section below.

3.3. Feature Selection. Feature selection is an approach that determines the most effective feature subset by reducing noise and eliminating unnecessary and redundant features in the WSN-DS dataset. It can enhance computation performance by reducing the computational load. This research proposed a Manta Ray Foraging (MRF) optimization to choose the features, which examines each feature set that simulates manta ray foraging behaviours. Three foraging operators are used in this strategy: chain foraging, cyclone foraging, and somersault foraging.

3.3.1. Chain Foraging. The manta ray chain prefers to devour plankton as its main diet. Hence the MRF optimization algorithm indicates that a place with much plankton is the finest. According to the optimum location and the manta ray in front of it, the MRF algorithm modifies the manta ray's position by omitting the first one. The chain foraging update process is shown in Equation (2):

$$s_m^{iter+1} = \begin{cases} s_m^{iter} + rand_1(s_b^{iter} - s_m^{iter}) + \gamma(s_b^{iter} - s_m^{iter}), m = 1 \\ s_m^{iter} + rand_2(s_{m-1}^{iter} - s_m^{iter}) + \gamma(s_b^{iter} - s_m^{iter}), m = 2, \dots, M \end{cases}, \quad (2)$$

$$\gamma = 2 \times rand_3 \times \sqrt{|\log 2(rand_4)|}, \quad (3)$$

where, s_m^{iter} denotes where the m^{th} the manta ray is at iteration; $rand_{1,2,3,4}$ are randomly produced integers in the range $[0, 1]$ that are distinct from one another; γ is the weight coefficient; $s_b^{iteration}$ is the maximum level of plankton concentration. The optimum plankton site, therefore, dictates the position update.

3.3.2. Cyclone Foraging. In addition to moving in the path of the plankton, each manta ray follows the manta ray in front of it as it swims. Equation (4) provides a mathematical representation of the cyclone foraging updating strategy.

$$s_m^{iter+1} = \begin{cases} s_b^{iter} + rand_5(s_b^{iter} - s_m^{iter}) + \alpha(s_b^{iter} - s_m^{iter}), m = 1 \\ s_b^{iter} + rand_6(s_{m-1}^{iter} - s_m^{iter}) + \alpha(s_b^{iter} - s_m^{iter}), m = 2, \dots, M \end{cases}, \quad (4)$$

$$\alpha = 2 \times e^{rand_7} \times \left(\frac{Max.iter - iter + 1}{iter} \right) \sin(2 \times \pi \times rand_8), \quad (5)$$

where, α represent the weight coefficient; $Max.iter$ determines the most iterations possible; and $rand_{5,6}$ represent the different randomly produced numbers between $[0, 1]$.

The MRF algorithm's exploration and exploitation are mostly driven by this step, which uses the best plankton as a benchmark (forcing the manta rays to move to a random place in the explore regions that is both distant from their current location and the best site for prey). Equations (6) and (7) show this hypothesized process.

$$s_{rp}^{iter} = Lw + rand_9(UP - Lw), \quad (6)$$

$$s_m^{iter+1} = \begin{cases} s_{rp}^{iter} + rand_{10}(s_{rp}^{iter} - s_m^{iter}) + \alpha(s_{rp}^{iter} - s_m^{iter}), m = 1 \\ s_{rp}^{iter} + rand_{11}(s_{m-1}^{iter} - s_m^{iter}) + \alpha(s_{rp}^{iter} - s_m^{iter}), m = 2, \dots, M \end{cases}, \quad (7)$$

where, s_{rp}^{iter} is the fabricated random place within the permitted range; UP and Lw are the maximum and minimum values that can be found at a specific place; and $rand_{5,6,7,8,9,10,11}$ is the various randomly generated integers between 0 and 1.

3.3.3. Somersault Foraging. Equation 8 provides a mathematical description of this scenario where each manta ray circles this spot and drifts

to a different position. The reference point is the location with the greatest plankton concentration thus far.

$$s_m^{iter+1} = s_m^{iter} + smsf \times \left(rand_{12} \times s_b^{iter} - \left(rand_{13} \times s_m^{iter} \right) \right), \quad (8)$$

where $smsf$ is the somersault factor; and $rand_{12,13}$ are various randomly generated integers between 0 and 1.

Algorithm 1. MRF optimization algorithm

Input: Normalized data
Output: Extracted features
While criteria! Satisfied do
for m=1,..., M do
if $rand_1 < 0.5$ then
if $(iter/Max.iter) < rand_2$ then // Cyclone Foraging
Evaluate using equations (5 and 6)
else
Determine Equation (3)
end if
else // Chain Foraging
Calculate using Equation (1)
end if
// Calculate fitness for n^{th} manta ray $f(s_m^{iter+1})$
if $f(s_m^{iter+1}) < \tilde{f}(s_b^{iter})$ then
$s_b^{iter} = s_m^{iter+1}$
end if
end for
// Somersault Foraging
for m=1,..., M do
Evaluate using Equation (7)
// Calculate fitness for m^{th} manta ray $f(s_m^{iter+1})$
if $f(s_m^{iter+1}) < \tilde{f}(s_b^{iter})$ then
$s_b^{iter} = s_m^{iter+1}$
end if
end for
end while

The Id and ADV_R features in the WSN-DS dataset show substantial correlations among themselves, even though many other characteristics in the dataset do not.

The correlation between these two features can be calculated using Pearson's correlation coefficient to measure the linear relationship between 2 variables. The `corr()` function computes the pairwise correlations between the dataset's selected features (Id and ADV_R). The Id property was thus dropped from the dataset due to this action. Id, ADV_R, and SCH_R, respectively, have the attributes that have the least effect on the class, as shown in Table 4. As a result, the dataset did not include these attributes.

Table 4. Feature selection results of the proposed approach

Feature No	Attribute Name
2	Time
3	Is_CH
4	who_CH
5	Dist_To_CH
6	ADV_S
8	JOIN_S
9	JOIN_R
10	SCH_S
12	Rank
13	DATA_S
14	DATA_R
15	Data_Sent_BS
16	dist_CH_To_BS
17	Send_Code
18	Consumed_Energy

The number of features has decreased due to the feature selection procedure, leaving the WSN-DS dataset with only 16 features.

3.4. Data Balancing. The WSN-DS contains unevenly distributed classes, which has a detrimental effect on the precise classification. In study [28] the authors suggested the heuristic oversampling approach known as SMOTE to address the class imbalance in datasets. By oversampling the data from the minority class, this strategy creates fake data. By creating synthetic data, SMOTE also solves the overfitting issue brought on by random oversampling techniques. As a result, this study suggested using a SMOTE to modify the sparse spreading of the minority class samples and usually avoid overfitting.

SMOTE selects samples that are near the feature space. From the minority class, a random sample is taken, and the sample's nearest k neighbours are then identified. Following the random selection among the closest neighbours, the distinction between the two sample features is multiplied by a value between 0 and 1 and added to the selected sample value. A line then connects the two sample features, and synthetically samples are produced this way. The kNN algorithm are randomly picked the neighbours depending on the oversampling needed.

The following definition applies to SMOTE samples, which are linear combinations of the minority class's two comparable samples $(r^S - r)$:

$$n = r + v \times (r^S - r), \quad 0 \leq d \leq 1, \quad (9)$$

where v is the variation between the two samples; and r^S is the randomly chosen model of r based on the closest neighbour digit. In order to address the class imbalance issue, SMOTE introduces new samples at random between neighbourhood samples from minority classes, greatly boosting the percentage of samples from minority classes. Figure 2 shows the raw dataset, and Figure 3 illustrates the after-data balancing.

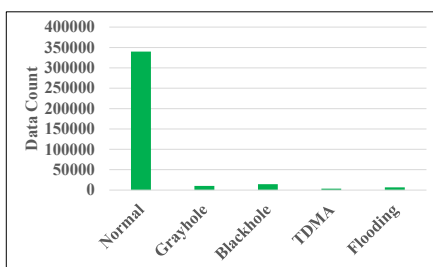


Fig. 2. Dataset before data balancing

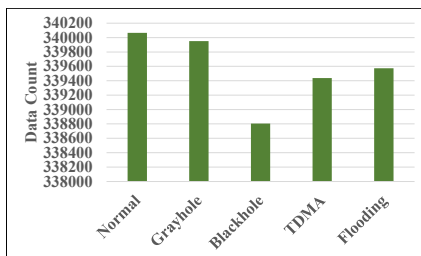


Fig. 3. Dataset after data balancing

Following that, the dataset is split into sections for testing and training. Finally, to classify the attacks, this research proposed a deep learning-based framework described in the following section.

3.5 CNN-SVM-based classification. Attacks are finally identified using the CNN-SVM based Deep Learning approach, in which the Convolution layer, max-pooling layer, batch normalization, Dropout layer, and dense layer are employed. Finally, the SVM classifier is utilized to provide the classification of the attack kinds.

CNN: The most analytical information can be gathered from raw data by CNN. In the proposed framework, the most recognizable features are taken from the raw data using a 5×5 kernel/filter. The input layer's $n \times n$ input neurons are convolved with an $m \times m$ filter in the convolutional layer, producing an output with the dimensions $(n-m+1) \times (n-m+1)$. Each layer's output serves as the following layer's input.

SVM: SVM attempts to denote a multi-dimensional dataset in a space where a hyperplane separates the data into different classes. The SVM classifier can minimize the generalization error.

CNN-SVM: The hybrid CNN-SVM model, SVM functions as a binary classifier and takes the place of CNN's softmax layer in this scenario, is introduced in the current work. Although CNN is a feature extractor, SVM is a binary classifier. Figure 4 describes a suggested hybrid CNN-SVM model's design.

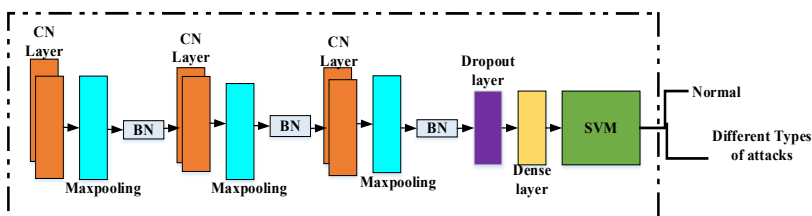


Fig. 4. Hybrid CNN-SVM approach

A 28×28 normalized data matrix is provided to the CNN layer as an input. The convolutional layers have a stride of size two and a 5×5 convolutional filtering. The feature map layer's convolution layers 1, 2, and 3 extract values that are also considered distinguishing properties of the input data. Here, dropout is a powerful regularization technique that can significantly reduce overfitting in Deep Neural Network (DNNs). By randomly dropping out neurons during training, the network learns to generalize better to new data. The CNN is trained after going through multiple epochs and until the training process converges. Here, the SVM classifier takes the place of CNN's last layer. The SVM classifier considers

the input features from the convolution layer 3 data as input. These new automatically generated training data features are initially used to train the SVM classifier. To identify the assaults utilized for testing, the trained SVM classifier is employed.

Table 5. Model Hyperparameter

Hyperparameter	Value
Iterations	100
Activation Function	ReLU
Loss Function	Categorical cross-entropy
Optimization algorithm	Adam
Learning rate	0.01

By reducing the error rate, as shown in Table 5, the hyper-parameter of the deep learning model will be tuned by the ADAM optimizer to identify intrusions with high detection accuracy. Consequently, the proposed CNN-SVM method can accurately categorize certain categories like Flooding, Blackhole, Normal, TDMA, or Grayhole.

4. Experiments and Evaluations. This segment details the performance and comparative findings of the suggested strategy and the implementation outcomes. The proposed method is implemented on the balanced WSN-DS dataset. For the evaluation purpose, we have split the dataset into testing and training in the ratio of 50:50. The simulation was carried out using the MATLAB 2021a tool operated in Windows 7 operating systems with 8GB RAM size of Intel premium processor.

4.1. Evaluation Parameters. This section describes the various performance indicators of the recommended deep learning system for classifying attacks in WSN.

Figure 5 demonstrates that after 32 iterations, the MRF optimization technique decreases the fitness function and converges on the value. To achieve the best characteristics, the optimization was done 100 times. As a result, the MRF technique has a good performance in terms of feature selection.

Figure 6 shows the multi-intrusion detection system outcomes of the suggested CNN-SVM-based method. Figure 6 classifies the assaults in the WSN-DS dataset. A confusion matrix displays the assaults properly detected on the main diagonal (top left to bottom right). The erroneous labels are visible in the remaining cells, referred to as true negatives or false negatives. The suggested hybrid model, therefore, produces superior outcomes. From that confusion matrix, we obtain the following performance values.

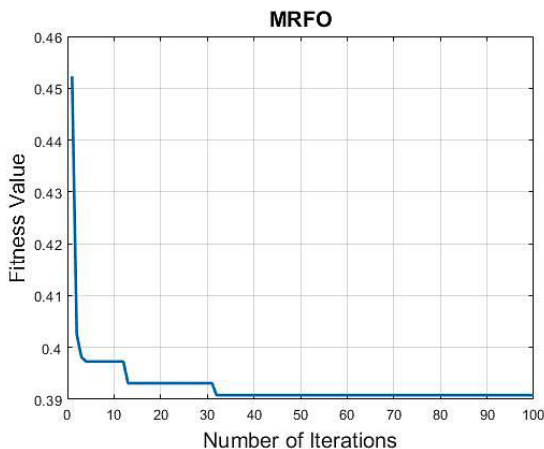


Fig. 5. Results of Manta Ray Foraging Optimization

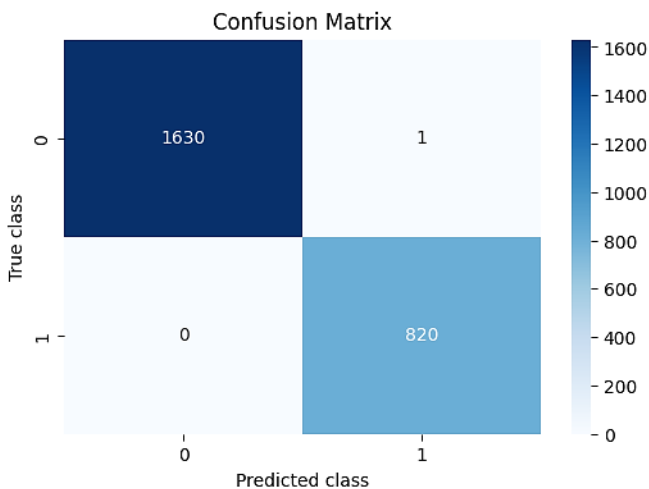


Fig. 6. Confusion Matrix

The efficiency of the unique strategy for attack categorization is assessed using the performance of our suggested method and several measures, including accuracy, F1 Score, precision, and recall. The following formulas were used to assess the Accuracy, Precision, Recall, and F1 Score performance metrics.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}, \quad (10)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}, \quad (11)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}, \quad (12)$$

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall}. \quad (13)$$

The suggested method's performance evaluation measures are shown in Figure 7.

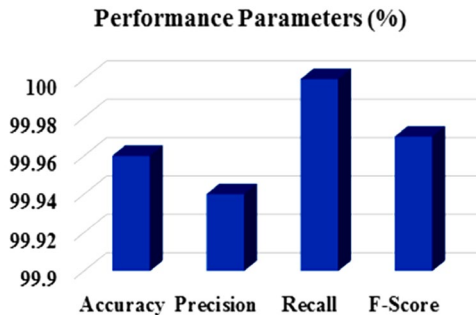


Fig. 7. Performance parameters of the proposed approach

The resulting values of accuracy, F1 Score, recall and F-score are 99.75%, 99.60%, 99.21% and 100.0%, tabulated in Table 6.

Table 6. Performance Evaluation

Performance Parameters	Value (%)
Accuracy	99.96
Precision	99.94
Recall	100
F-Score	99.97

By applying data normalization and encoding, the performance of our suggested method achieves greater accuracy, F1 score, precision, recall,

feature selection using the MRF approach, and data balancing and classification of attacks using the CNN-SVM approach.

4.2. Comparison Analysis. This segment contrasts the suggested method with accepted practices to evaluate it against current practices. The following algorithms have been reported: Decision Tree (DT) [26], Random Forest (RF) [26], Naive Bayes (NB) [26], Logistic Regression (LR) [26], Multi-Layer Perceptron (MLP) [26], CNN [26], LSTM [26], and CNN-LSTM [26].

Figure 8 displays a comparison of all accuracy levels. Using CNN-SVM enhances the accuracy of the suggested method. As related to the baseline, our suggested technique achieves greater accuracy than DT, RF, NB, LR, MLP, CNN, LSTM, and CNN-LSTM such as 97.53%, 97.67%, 92.53%, 97.16%, 96.24%, 98.03%, 99.31%, and 98.04%.

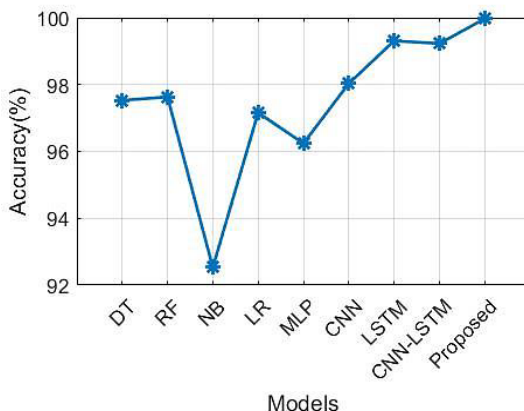


Fig. 8. Comparison of Accuracy

Consequently, our original distinguishing technique has a 99.96% accuracy rate, outperforming conventional techniques.

Figure 9 shows the comparison of complete precision. Using CNN-SVM enhances the suggested method's accuracy. As compared to the baseline, our suggested strategy achieves greater precision than DT, RF, NB, LR, MLP, CNN, LSTM, and CNN-LSTM such as 0.97%, 0.98%, 0.88%, 0.97%, 0.96%, 0.98%, 0.99%, and 0.98%. Our innovative distinguishing method has a 0.9994% precision, outperforming conventional techniques.

Figure 10 shows the entire recall comparison. Using CNN-SVM increases the recall of the suggested method. We are comparing the recall of

our suggested method to the baseline DT, RF, NB, LR, MLP, CNN, LSTM, and CNN-LSTM such as 0.98%, 0.98%, 0.93%, 0.97%, 0.96%, 0.98%, 0.99%, and 0.98%. Consequently, our original, distinctive technique has a 100.0% recall compared to standard methods.

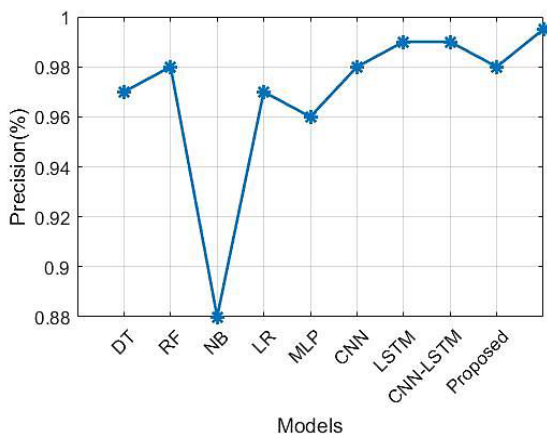


Fig. 9. Comparison of Precision

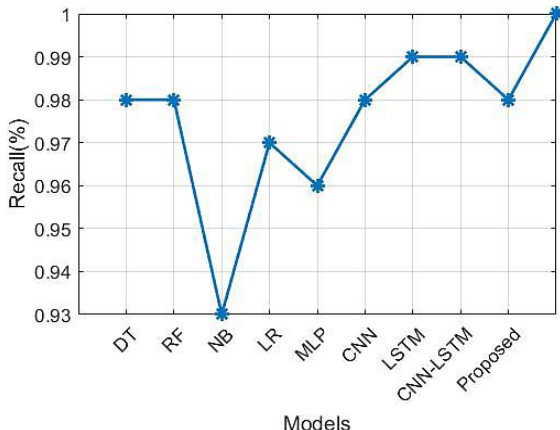


Fig. 10. Comparison of Recall

Figure 11 displays the comparison of the total F-Score. Using CNN-SVM raises the F-Score of the suggested method. As compared to the baseline, our suggested method outperforms DT, RF, NB, LR, MLP, CNN, LSTM, and CNN-LSTM such as 0.97%, 0.97%, 0.90%, 0.97%, 0.96%,

0.98%, 0.98%, and 0.98%. Because of this, our innovative, distinctive method has a 0.9997% F-Score, which is better than baseline methods. The overall comparison of the performance parameters is tabulated in Table 7.

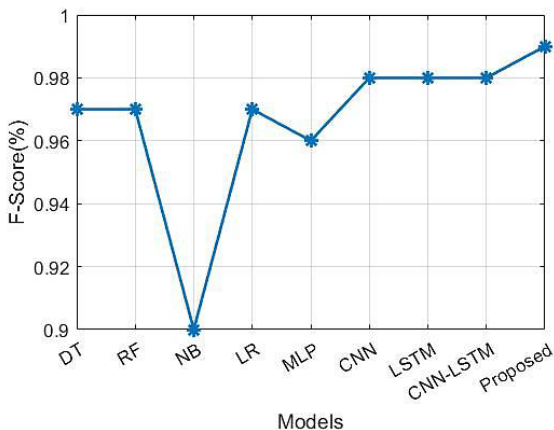


Fig. 11. Comparison of F-Score

Table 7. Comparison of Performance Parameters

Methods	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
DT	97.53	97	98	97
Random Forest	97.67	98	98	97
Naïve Bayes	92.53	88	93	90
Logistic Regression	97.16	97	97	97
Multi-layer Perceptron	96.24	96	96	96
CNN	98.03	98	98	98
LSTM	99.31	99	99	98
CNN-LSTM	98.04	98	98	98
Proposed	99.75	99.21	100	99.6

4.3. Discussion. The striking aspect of the results is that, on the WSN-DS dataset, the CNN and LSTM deep learning algorithms outperform the CNN-SVM hybrid strategy separately. Data balancing enhances the performance of the DoS intrusion detection system greatly. In research published in the literature, hybrid approaches frequently outperformed individual deep learning techniques, whereas individual methods outperformed hybrid methods on the WSN-DS dataset with the feature

selection algorithm. The peculiar feature structure of the WSN-DS dataset is thought to be the root cause of this circumstance. The NB algorithm had the worst results. Since it could not recognize data from the TDMA and Flooding classes with a high rate, the Naive Bayes method performed the worst. Figures 8–11 and Table 7 compare classification algorithms based on accuracy, precision, recall, and F1-Score characteristics. When the outcomes of these parameters are analyzed, the suggested technique yields the best results for each parameter.

5. Conclusion. This work proposes a novel classification-based DoS intrusion detection method to identify DoS assaults targeted at WSNs. The suggested CNN-SVM technique combines data balance and feature selection operations in the deep learning-based approach. The SMOTE ensemble technique was utilized in this work to balance the data. For the attack categorization procedure, CNN-SVM was used. Also, the impact of each feature was evaluated, and using the Manta Ray Optimization feature selection approach, the number of features was lessened from 19 to 16. It was observed that the feature selection process led to an improvement in the algorithm's performance. The WSN-DS dataset was used for experimental experiments. The metrics Accuracy, Precision, Recall, and F-Score were used to assess the performance of the suggested technique. According to the test findings, the proposed approach performed better than the alternatives, with an accuracy, F-Measure, Precision, and Recall of 99.75%, 99.21%, 100%, and 99.6%. In the future, it is intended to integrate the machine learning algorithm with various deep learning techniques, such as Gated Recurrent Unit (GRU) and Auto Encoder (AE), to increase performance and analyze the outcomes. Furthermore, several oversampling and undersampling approaches will be explored for data balance. Furthermore, the suggested method's performance on various datasets will be evaluated. Finally, research is being conducted to improve the dependability and transparency of intrusion detection systems using Explainable Artificial Intelligence approaches.

References

1. Kopetz H. Internet of things. Real-time systems. 2011. pp. 307–323. DOI: 10.1007/978-1-4419-8237-7_13.
2. Alsulaiman L., Al-Ahmadi S. Performance evaluation of machine learning techniques for DOS detection in wireless sensor network. arXiv preprint arXiv:2104.01963. 2021.
3. Aziz N.A.A., Aziz K.A. Managing disaster with wireless sensor networks. 13th international conference on advanced communication technology (ICACT2011). 2011. pp. 202–207.
4. Butun I., Morgera S.D., Sankar R. A survey of intrusion detection systems in wireless sensor networks. IEEE Communications surveys and tutorials. 2013. vol. 16. no. 1. pp. 266–282.

5. Pelechrinis K., Iliofotou M., Krishnamurthy S.V. Denial of service attacks in wireless networks: The case of jammers. *IEEE Communications surveys and tutorials*. 2010. vol. 13. no. 2. pp. 245–257.
6. López J., Zhou J. (Eds.). *Wireless sensor network security*. Ios Press, 2008. 320 p.
7. Das S.K., Kant K., Zhang N. *Handbook on securing cyber-physical critical infrastructure*. Elsevier Inc., 2012. 848 p.
8. Rassam M.A., Maarof M.A., Zainal A. A survey of intrusion detection schemes in wireless sensor networks. *American Journal of Applied Sciences*. 2012. vol. 9. no. 10. pp. 1636–1652.
9. Mahbooba B., Sahal R., Alosaimi W., Serrano M., Alosaimi W. Trust in intrusion detection systems: an investigation of performance analysis for machine learning and deep learning models. *Complexity*. 2021. vol. 2021. 23 p. DOI: 10.1155/2021/5538896.
10. Cauteruccio F., Fortino G., Guerrieri A., Liotta A., Mocanu D.C., Perra C., Terracina G., Vega M.T. Short-long term anomaly detection in wireless sensor networks based on machine learning and multi-parameterized edit distance. *Information Fusion*. 2019. vol. 52. pp. 13–30.
11. Vinayakumar R., Alazab M., Soman K.P., Poornachandran P., Al-Nemrat A., Venkatraman S. Deep learning approach for intelligent intrusion 936 detection system. *IEEE Access*. 2019. vol. 7. pp. 41525–41550.
12. Alqahtani M., Gumaei A., Mathkour H., Maher Ben Ismail M. A genetic-based extreme gradient boosting model for detecting intrusions in wireless sensor networks. *Sensors*. 2019. vol. 19(20). no. 4383. DOI: 10.3390/s19204383.
13. Tan X., Su S., Huang Z., Guo X., Zuo Z., Sun X., Li L. Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm. *Sensors*. 2019. vol. 19(1). DOI: 10.3390/s19010203.
14. Jiang S., Zhao J., Xu X. SLGBM: An intrusion detection mechanism for wireless sensor networks in smart environments. *IEEE Access*. 2020. vol. 8. pp. 169548–169558.
15. Abhale A.B., Manivannan S.S. Supervised machine learning classification algorithmic approach for finding anomaly type of intrusion detection in wireless sensor network. *Optical Memory and Neural Networks*. 2020. vol. 29. pp. 244–256.
16. Tang C., Luktarhan N., Zhao Y. An efficient intrusion detection method based on LightGBM and autoencoder. *Symmetry*. 2020. vol. 12(9). no. 1458. DOI: 10.3390/sym12091458.
17. Nancy P., Muthurajkumar S., Ganapathy S., Santhosh Kumar S.V.N., Selvi M., Arputharaj K. Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks. *IET Communications*. 2020. vol. 14. no. 5. pp. 888–895.
18. Liu J., Gao Y., Hu F. A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. *Computers and Security*. 2021. vol. 106. no. 102289.
19. Al S., Dener M. STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment. *Computers and Security*. 2021. vol. 110. no. 102435.
20. Ifzarne S., Tabbaa H., Hafidi I., Lamghari N. Anomaly detection using machine learning techniques in wireless sensor networks. *Journal of Physics: Conference Series*. 2021. vol. 1743(1). no. 012021. DOI: 10.1088/1742-6596/1743/1/012021.
21. Pan J.S., Fan F., Chu S.C., Zhao H.Q., Liu G.Y. A Light-weight Intelligent Intrusion Detection Model for Wireless Sensor Networks. *Security and Communication Networks*. 2021. vol. 2021(2). 15 p. DOI: 10.1155/2021/5540895.

22. Zamry N.M., Zainal A., Rassam M.A., Alkhamash E.H., Ghaleb F.A., Saeed F. Light-weight Anomaly Detection Scheme Using Incremental Principal Component Analysis and Support Vector Machine. *Sensors*. 2021. vol. 21(23). no. 8017. DOI: 10.3390/s21238017.
23. Yadav A., Kumar A. Intrusion Detection and Prevention Using RNN in WSN. *Proceedings of Inventive Computation and Information Technologies (ICICIT)*. 2022. pp. 531–539.
24. Tabbaa H., Ifzarne S., Hafidi I. An Online Ensemble Learning Model for Detecting Attacks in Wireless Sensor Networks. *arXiv preprint arXiv:2204.13814*. 2022. 15 p.
25. Salmi S., Oughdir L. CNN-LSTM Based Approach for Dos Attacks Detection in Wireless Sensor Networks. *International Journal of Advanced Computer Science and Applications*. 2022. vol. 13. no. 4. pp. 835–842.
26. Dener M., Al S., Orman A. STLGBM-DDS: An Efficient Data Balanced DoS Detection System for Wireless Sensor Networks on Big Data Environment. *IEEE Access*. 2022. vol. 10. pp. 92931–92945.
27. Almomani I., Al-Kasasbeh B., Al-Akhras M. WSN-DS: A dataset for intrusion detection systems in wireless sensor networks. *Journal of Sensors*. 2016. vol. 2016. 16 p. DOI: 10.1155/2016/4731953.
28. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*. 2002. vol. 16. pp. 321–357.

Krishna Kuraganty Phani Rama — Assistant professor, Electronics and communication department, Prasad V. Potluri Siddhartha Institute of Technology; Researcher, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeshwaram. Research interests: wireless sensor networks, wireless communication. The number of publications — 17. kprkrishna007@gmail.com; Teacher's Colony, Bhavanipuram, 520012, Vijayawada, India; office phone: +91(866)258-1699.

Thirumuru Ramakrishna — Ph.D., Professor, Electronics and communication department, Koneru Lakshmaiah Education Foundation (Deemed to be University), Vaddeshwaram. Research interests: speech signal processing, wireless sensor networks, wireless communication. The number of publications — 25. ramakrishnaece@kluniversity.in; Green Fields, Vaddeshwaram, 522302, Guntur, Russia; office phone: +91(8645)350-0200.

К. КРИШНА, Р. ТИРУМУРУ

СБАЛАНСИРОВАННАЯ СИСТЕМА ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ ДЛЯ БЕСПРОВОДНЫХ СЕНСОРНЫХ СЕТЕЙ В СРЕДЕ БОЛЬШИХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ CNN-SVM

Кришна К., Тирумuru Р. Сбалансированная система обнаружения вторжений для беспроводных сенсорных сетей в среде больших данных с использованием модели CNN-SVM.

Аннотация. Беспроводные сенсорные сети (WSN) подвергались нескольким различным проблемам безопасности и атакам, связанным со сбором и отправкой данных. В этом сценарии одной из наиболее распространенных атак WSN, которая может быть нацелена на любой уровень стека протоколов, является атака типа «отказ в обслуживании» (DoS). Текущее исследование предлагает различные стратегии обнаружения атаки в сети. Однако у него есть проблемы с классификацией. Поэтому в этом исследовании для решения этой проблемы была предложена эффективная система обнаружения вторжений на основе ансамблевого глубокого обучения для выявления атак в сети WSN. Предварительная обработка данных включает преобразование качественных данных в числовые с использованием метода One-Hot Encoding. После этого был проведен процесс нормализации. Затем предлагается выбрать лучшее подмножество функций с помощью Manta-Ray Foraging Optimization. Затем метод передискретизации синтетического меньшинства (SMOTE) создает новую выборку меньшинства для балансировки обработанного набора данных. Наконец, предлагается классификатор CNN-SVM для классификации видов атак. Метрики Точность, F-мера, Прецизионность и Отзыв использовались для оценки результатов 99,75%, 99,21%, 100% и 99,6% соответственно. По сравнению с существующими подходами предложенный метод оказался чрезвычайно эффективным при обнаружении DoS-атак в WSN.

Ключевые слова: беспроводная сенсорная сеть, DoS-атаки, искусственный интеллект, глубокое обучение, сверточные нейронные сети, метод опорных векторов.

Литература

1. Kopetz H. Internet of things. Real-time systems. 2011. pp. 307–323. DOI: 10.1007/978-1-4419-8237-7_13.
2. Alsulaiman L., Al-Ahmadi S. Performance evaluation of machine learning techniques for DOS detection in wireless sensor network. arXiv preprint arXiv:2104.01963. 2021.
3. Aziz N.A.A., Aziz K.A. Managing disaster with wireless sensor networks. 13th international conference on advanced communication technology (ICACT2011). 2011. pp. 202–207.
4. Butun I., Morgera S.D., Sankar R. A survey of intrusion detection systems in wireless sensor networks. IEEE Communications surveys and tutorials. 2013. vol. 16. no. 1. pp. 266–282.
5. Pelechrinis K., Iliofotou M., Krishnamurthy S.V. Denial of service attacks in wireless networks: The case of jammers. IEEE Communications surveys and tutorials. 2010. vol. 13. no. 2. pp. 245–257.
6. López J., Zhou J. (Eds.). Wireless sensor network security. Ios Press, 2008. 320 p.
7. Das S.K., Kant K., Zhang N. Handbook on securing cyber-physical critical infrastructure. Elsevier Inc., 2012. 848 p.

8. Rassam M.A., Maarof M.A., Zainal A. A survey of intrusion detection schemes in wireless sensor networks. *American Journal of Applied Sciences*. 2012. vol. 9. no. 10. pp. 1636–1652.
9. Mahbooba B., Sahal R., Alosaimi W., Serrano M., Alosaimi W. Trust in intrusion detection systems: an investigation of performance analysis for machine learning and deep learning models. *Complexity*. 2021. vol. 2021. 23 p. DOI: 10.1155/2021/5538896.
10. Cauteruccio F., Fortino G., Guerrieri A., Liotta A., Mocanu D.C., Perra C., Terracina G., Vega M.T. Short-long term anomaly detection in wireless sensor networks based on machine learning and multi-parameterized edit distance. *Information Fusion*. 2019. vol. 52. pp. 13–30.
11. Vinayakumar R., Alazab M., Soman K.P., Poornachandran P., Al-Nemrat A., Venkatraman S. Deep learning approach for intelligent intrusion 936 detection system. *IEEE Access*. 2019. vol. 7. pp. 41525–41550.
12. Alqahtani M., Gumaï A., Mathkour H., Maher Ben Ismail M. A genetic-based extreme gradient boosting model for detecting intrusions in wireless sensor networks. *Sensors*. 2019. vol. 19(20). no. 4383. DOI: 10.3390/s19204383.
13. Tan X., Su S., Huang Z., Guo X., Zuo Z., Sun X., Li L. Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm. *Sensors*. 2019. vol. 19(1). DOI: 10.3390/s19010203.
14. Jiang S., Zhao J., Xu X. SLGBM: An intrusion detection mechanism for wireless sensor networks in smart environments. *IEEE Access*. 2020. vol. 8. pp. 169548–169558.
15. Abhale A.B., Manivannan S.S. Supervised machine learning classification algorithmic approach for finding anomaly type of intrusion detection in wireless sensor network. *Optical Memory and Neural Networks*. 2020. vol. 29. pp. 244–256.
16. Tang C., Luktarhan N., Zhao Y. An efficient intrusion detection method based on LightGBM and autoencoder. *Symmetry*. 2020. vol. 12(9). no. 1458. DOI: 10.3390/sym12091458.
17. Nancy P., Muthurajkumar S., Ganapathy S., Santhosh Kumar S.V.N., Selvi M., Arputharaj K. Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks. *IET Communications*. 2020. vol. 14. no. 5. pp. 888–895.
18. Liu J., Gao Y., Hu F. A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. *Computers and Security*. 2021. vol. 106. no. 102289.
19. Al S., Dener M. STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment. *Computers and Security*. 2021. vol. 110. no. 102435.
20. Ifzarne S., Tabbaa H., Hafidi I., Lamghari N. Anomaly detection using machine learning techniques in wireless sensor networks. *Journal of Physics: Conference Series*. 2021. vol. 1743(1). no. 012021. DOI: 10.1088/1742-6596/1743/1/012021.
21. Pan J.S., Fan F., Chu S.C., Zhao H.Q., Liu G.Y. A Light-weight Intelligent Intrusion Detection Model for Wireless Sensor Networks. *Security and Communication Networks*. 2021. vol. 2021(2). 15 p. DOI: 10.1155/2021/5540895.
22. Zamry N.M., Zainal A., Rassam M.A., Alkhamash E.H., Ghaleb F.A., Saeed F. Light-weight Anomaly Detection Scheme Using Incremental Principal Component Analysis and Support Vector Machine. *Sensors*. 2021. vol. 21(23). no. 8017. DOI: 10.3390/s21238017.
23. Yadav A., Kumar A. Intrusion Detection and Prevention Using RNN in WSN. *Proceedings of Inventive Computation and Information Technologies (ICICIT)*. 2022. pp. 531–539.

24. Tabbaa H., Ifzarne S., Hafidi I. An Online Ensemble Learning Model for Detecting Attacks in Wireless Sensor Networks. arXiv preprint arXiv:2204.13814. 2022. 15 p.
25. Salmi S., Oughdir L. CNN-LSTM Based Approach for Dos Attacks Detection in Wireless Sensor Networks. International Journal of Advanced Computer Science and Applications. 2022. vol. 13. no. 4. pp. 835–842.
26. Dener M., Al S., Orman A. STLGBM-DDS: An Efficient Data Balanced DoS Detection System for Wireless Sensor Networks on Big Data Environment. IEEE Access. 2022. vol. 10. pp. 92931–92945.
27. Almomani I., Al-Kasasbeh B., Al-Akhras M. WSN-DS: A dataset for intrusion detection systems in wireless sensor networks. Journal of Sensors. 2016. vol. 2016. 16 p. DOI: 10.1155/2016/4731953.
28. Chawla N.V., Bowyer K.W., Hall L.O., Kegelmeyer W.P. SMOTE: Synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002. vol. 16. pp. 321–357.

Кришна Кураганти Пхани Рама — доцент, отдел электроники и связи, Прасад В. Потлури Технологический институт Сиддхартхи; научный сотрудник, Образовательный фонд Конеру Лакшмайи (Считается университетом). Область научных интересов: беспроводные сенсорные сети, беспроводная связь. Число научных публикаций — 17. kprkrishna007@gmail.com; Учительская колония, Бхаванипурам, 520012, Виджаявада, Индия; р.т.: +91(866)258-1699.

Тирумур Рамакришна — Ph.D., профессор, отдел электроники и связи, Образовательный фонд Конеру Лакшмайи (Считается университетом). Область научных интересов: обработка речевых сигналов, беспроводные сенсорные сети, беспроводная связь. Число научных публикаций — 25. ramakrishnaece@kluniversity.in; Зеленые поля, Ваддесварам, 522302, Гунтур, Россия; р.т.: +91(8645)350-0200.

А.Ю. ПОЛЕТАЕВ, И.В. ПАРАМОНОВ, Е.И. БОЙЧУК
**АЛГОРИТМ ПОСТРОЕНИЯ ДЕРЕВА СИНТАКСИЧЕСКИХ
ЕДИНИЦ РУССКОЯЗЫЧНОГО ПРЕДЛОЖЕНИЯ ПО ДЕРЕВУ
СИНТАКСИЧЕСКИХ СВЯЗЕЙ**

Полетаев А.Ю., Парамонов И.В., Бойчук Е.И. Алгоритм построения дерева синтаксических единиц русскоязычного предложения по дереву синтаксических связей.

Аннотация. Автоматический синтаксический анализ предложения – одна из важных задач компьютерной лингвистики. В настоящее время для русского языка отсутствуют общедоступные и пригодные для практического применения анализаторы синтаксической структуры. Создание таких анализаторов «с нуля» требует составления корпуса деревьев, размеченного в соответствии с заданной формальной грамматикой, что представляет собой крайне трудоёмкую задачу. Однако, поскольку для русского языка существует несколько анализаторов синтаксических связей, представляется полезным использовать результаты их работы для анализа синтаксической структуры предложений. В настоящей работе предлагается алгоритм, позволяющий построить дерево синтаксических единиц русскоязычного предложения по данному дереву синтаксических связей. Алгоритм использует грамматику, сформулированную в соответствии с классическим справочником Д.Э. Розенталя. Приведены результаты экспериментов по оценке качества работы предложенного алгоритма на корпусе из 300 предложений на русском языке. 200 предложений были выбраны из вышеупомянутого справочника и 100 из открытого корпуса публицистических текстов OpenCorpora. В ходе экспериментов предложения подавались на вход анализаторов из состава библиотек Stanza, SpaCy и Natasha, после чего полученные деревья синтаксических связей обрабатывались предложенным алгоритмом. Полученные в результате обработки дерева синтаксических единиц сравнивались с размеченными вручную экспертами-филологами. Наилучшее качество было получено при использовании анализатора синтаксических связей из библиотеки Stanza: F_1 -мера построения синтаксических единиц составила 0.85, а точность определения членов предложения – 0.93, чего должно быть достаточно для решения многих практических задач в таких областях, как извлечение событий, информационный поиск, анализ тональности.

Ключевые слова: компьютерная лингвистика, обработка естественного языка, синтаксический анализ, дерево синтаксических единиц, дерево синтаксических связей, формальная грамматика.

1. Введение. Для решения некоторых задач обработки естественного языка, например, извлечения событий, информационного

поиска, анализа тональности, требуются методы автоматического анализа предложений. Одна из важнейших задач анализа предложения – синтаксический анализ, включающий в себя как анализ синтаксических связей (dependency parsing), так и анализ синтаксической структуры (constituency parsing) [1].

Синтаксические связи между словами предложения могут быть представлены в виде дерева синтаксических связей (дерева зависимостей, дерева подчинения, dependency tree) [2]. Пример дерева синтаксических связей для предложения на русском языке приведён на рисунке 1.

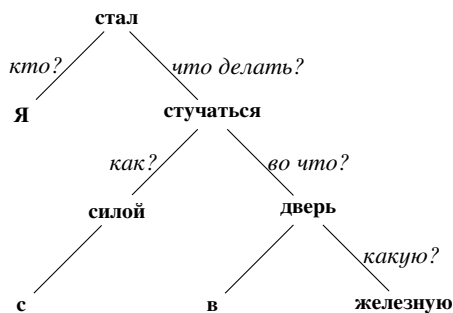


Рис. 1. Дерево синтаксических связей предложения
Я стал с силой стучаться в железную дверь

Синтаксическая структура предложения может быть представлена в виде дерева синтаксических единиц (системы синтаксических групп, дерева непосредственно составляющих, constituency tree), каждый узел которого – синтаксическая единица [3, 4], а потомки этого узла – формирующие её меньшие синтаксические единицы. Корнем дерева синтаксических единиц является всё предложение в целом как наиболее крупная синтаксическая единица, а листьями – отдельные слова. Пример дерева синтаксических единиц для предложения на русском языке приведён на рисунке 2.

Поскольку дерево синтаксических единиц описывает не только отдельные слова, но и то, как они объединяются в более крупные единицы синтаксиса, а это объединение в естественном языке происходит в соответствии с его грамматикой, для построения дерева синтаксических единиц необходима формализация этой грамматики [1]. Нужно отметить, что для одного и того же предложения по разным формальным грамматикам будут построены разные деревья синтаксических единиц [2].

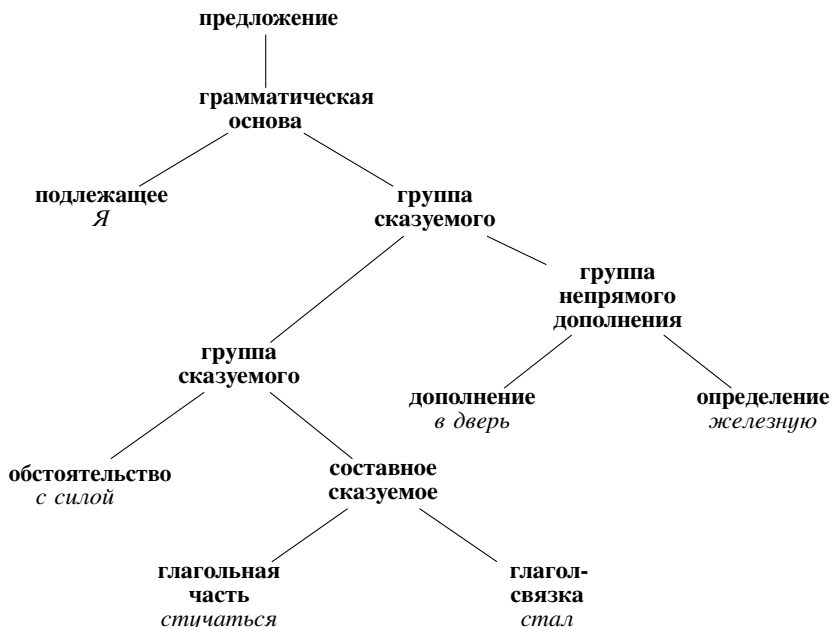


Рис. 2. Дерево синтаксических единиц предложения
Я стал с силой стучаться в железную дверь

Лингвисты изучают естественный язык с помощью анализа синтаксических единиц (immediate constituent analysis) как минимум с середины XX века [5]. Данный метод используется как в исследованиях, объектом которых является естественный язык сам по себе, так и в исследованиях, использующих тексты на естественном языке как источник информации о других процессах и явлениях [6 – 10]. Также информация о синтаксической структуре предложения может не анализироваться исследователями напрямую, а использоваться как источник данных для алгоритмов машинного обучения [11].

Важно отметить, что применимость дерева синтаксических единиц для решения прикладных задач связана с формальной грамматикой, в соответствии с которой оно построено: например, если грамматика не разделяет сложносочинённые и сложноподчинённые предложения, то исследование, для которого это разделение имеет значение, провести просто не получится. С другой стороны, чем сложнее грамматика, тем сложнее на её основе проводить автоматический анализ и построение деревьев синтаксических единиц.

В настоящий момент для русского языка отсутствуют общедоступные и пригодные для практического применения анализаторы синтаксической структуры. Создание алгоритмов для такого анализа требует составления корпуса деревьев (treebank), размеченного в соответствии с формальной грамматикой, что представляет собой крайне трудоёмкую задачу – например, создание Penn Treebank потребовало 8 лет работы [12]. Однако, поскольку для русского языка доступно несколько парсеров синтаксических связей, представляется полезным использовать результаты анализа синтаксических связей слов в предложении для анализа его синтаксической структуры.

Цель данной работы – предложить алгоритм, позволяющий построить дерево синтаксических единиц русскоязычного предложения, соответствующее заданной грамматике, по данному дереву синтаксических связей, а также оценить эффективность предложенного алгоритма. Формальная грамматика для работы была сформулирована в соответствии с классической работой Д.Э. Розенталя, И.Б. Голуб и М.А. Теленковой [13] с некоторыми упрощениями. Эта модель языка с одной стороны достаточно точна, с другой – проста, но может при необходимости быть уточнена и дополнена для решения различных прикладных задач.

Оставшаяся часть работы организована следующим образом. В разделе 2 приведён обзор связанных работ. Раздел 3 посвящён используемой в работе формальной грамматике. В разделе 4 описан предлагаемый алгоритм построения деревьев синтаксических единиц. Раздел 5 содержит результаты экспериментов по оценке качества построения деревьев синтаксических единиц на размеченных наборах предложений при использовании различных анализаторов синтаксических связей. В заключении подведены итоги работы.

2. Связанные работы. Использование деревьев синтаксических единиц для представления синтаксической структуры предложения было предложено Н. Хомским [14, 15]. Им также была обоснована значимость используемой для синтаксического анализа грамматики естественного языка и сформулированы основные принципы построения синтаксических деревьев.

Большая часть исследований деревьев синтаксических единиц выполнены для английского языка, в частности существуют достаточно подробно проработанные грамматики, например, HPSG (Head-Driven Phrase Structure Grammar), которая учитывает, кроме синтаксиса, некоторые элементы семантики предложений [16]. Однако из-за своей сложности такие грамматики достаточно сложны для использования,

поэтому, например, крупнейший набор деревьев синтаксических единиц Penn Treebank размечен в соответствии с достаточно сильно упрощённой грамматикой [17, 18].

В работах [19 – 21] на примере английского языка рассматриваются основные современные подходы к автоматическому построению деревьев синтаксических единиц. При использовании различных нейронных сетей на некоторых корпусах, например, Penn Treebank, возможно достижение крайне высокого качества построения деревьев синтаксических единиц – F_1 -мера выше 0.95. Однако в данных работах также показано, что качество синтаксического анализа достаточно сильно зависит от используемого корпуса, и, например, модель BERT, с помощью которой на Penn Treebank была достигнута F_1 -мера 0.96, при построении деревьев синтаксических единиц на корпусе предложений из художественных текстов показала гораздо более низкое качество, с F_1 -мерой всего 0.86 [21]. В работе [20] также отмечается, что качество построения деревьев синтаксических единиц может достаточно сильно отличаться в зависимости от языка: для китайского языка F_1 -мера в схожих условиях оказывается на 5–7 % ниже, чем для английского.

Первые попытки применить наработки Хомского для исследования русского языка относятся ещё к 60-м годам XX века [22], наиболее развитая грамматика синтаксических единиц, опирающаяся на накопленные лингвистами знания, была разработана и описана в работах А.В. Гладкого [23, 24]. Однако, из-за несовершенства вычислительной техники того периода и отсутствия достаточных лингвистических ресурсов, а также сложности самой грамматики, не было разработано автоматических синтаксических анализаторов, строящих деревья синтаксических единиц в соответствии с этой грамматикой.

В работах [25, 26] описывается созданный И.А. Кагировым и А.Б. Леонтьевой анализатор, строящий деревья синтаксических единиц на основе частеречной разметки, морфологической информации и анализа синтаксических связей. К достоинствам этих работ следует отнести предложенную формальную грамматику составляющих, достаточно хорошо соответствующую теории синтаксиса русского языка, но, вместе с тем, достаточно простую для автоматического анализа. Главным недостатком является то, что эксперименты проводились только для простых предложений.

Также необходимо отметить проект «Диалинг» и связанные с ним исследования [27]. Несмотря на то, что предлагаемая его авторами система клауз относится скорее к анализу семантики, опыт построения систем автоматического анализа на её основе показывает, что в настоящий момент

качество доступных инструментов автоматического анализа отдельных слов и связей между ними уже достаточно высоко, чтобы результаты их работы можно было использовать в качестве входных данных для автоматического анализа сложных предложений [28, 29].

3. Используемая грамматика. Крайне важен выбор формальной грамматики, описывающей синтаксическую структуру предложений. С одной стороны, она должна быть достаточно подробной, чтобы дерево синтаксических единиц предоставляло как можно больше информации. С другой стороны, чем сложнее грамматика, тем сложнее точно построить дерево на её основе, возрастает число ошибок и снижается надёжность предоставляемой деревом информации.

Для целей данного исследования была формализована грамматика русскоязычных предложений из классической работы [13]. Далее приводится описание полученного результата, а также упрощений, допущенных для лучшей формализации.

Для того, чтобы привести в рамках данной статьи полное описание полученной формальной грамматики, используется нескольких принципов. Во-первых, поскольку в русском языке нет строгого порядка слов, и, например, определение может находиться как перед определяемым словом, так и после, большая часть правил должна содержать последовательности как для прямого, так и для обратного порядка слов. Чтобы сократить запись, условимся, что порядок символов в последовательностях может быть произвольным, если в описании соответствующего правила не указано иное. Во-вторых, чтобы не вводить дополнительные фиктивные типы символов для обеспечения порядка применения продукций, необходимого для обеспечения однозначности разбора, условимся, что продукции приводятся в порядке приоритета их применения при нисходящем синтаксическом анализе, от наиболее приоритетного к наименее приоритетному.

Чтобы обеспечить однозначность разбора синтаксической структуры вне зависимости от порядка слов, было принято, что синтаксические единицы, обозначающие свойства явлений, в дереве должны находиться ближе к единицам, называющим эти явления, чем единицы, называющие другие явления. Например, вначале в одну синтаксическую единицу должны объединяться подлежащее и связанное с ним определение, и только затем будет выделена синтаксическая единица дополнения, связанного с подлежащим:

((((Лучший студент) группы) (сдал экзамен))).

3.1. Общая структура предложения. Предложение в целом состоит из грамматической основы и знака препинания, стоящего после неё. Знак препинания не обязателен, поскольку он часто опускается в сетевых текстах [30]:

предложение ::= грамматическая-основа [знак-препинания]

Грамматическая основа предложения состоит из группы подлежащего и группы сказуемого, если предложение двусоставное; только из группы подлежащего или группы сказуемого, если предложение односоставное. Грамматическая основа сложносочинённого предложения выражается специальной синтаксической единицей, объединяющей две и более грамматических основы с помощью сочинительного союза. В расчленённом сложноподчинённом предложении придаточное предложение относится ко всему главному предложению в целом, поэтому грамматическая основа такого предложения также будет состоять из двух отдельных грамматических основ, связанных подчинительным союзом. Любая грамматическая основа может предваряться вводной конструкцией, выраженной отдельной синтаксической единицей.

Для упрощения грамматики не выделяется отдельный тип бессоюзных сложных предложений, так как, согласно [13], все бессоюзные сложные предложения могут быть соотнесены либо со сложносочинёнными, либо со сложноподчинёнными предложениями, а знаки препинания в них, соответственно, с сочинительными либо подчинительными союзами.

Описанное выше выражаются следующими формальными правилами:

грамматическая-основа ::= вводная-конструкция грамматическая-основа
 | сложносочинённое-предложение
 | расчленённое-сложноподчинённое-предложение
 | группа-подлежащего группа-сказуемого
 | группа-сказуемого
 | группа-подлежащего
 сложносочинённое-предложение ::= грамматическая-основа
 грамматическая-основа { грамматическая-основа } сочинительный-союз
 расчленённое-сложноподчинённое-предложение ::= грамматическая-основа
 подчинительный-союз грамматическая-основа

3.2. Группа подлежащего. Группа подлежащего может включать в себя, кроме самого подлежащего, определение, поясняющее называемый подлежащим предмет, например, хороший ученик; не прямое дополнение,

называющее предмет, связанный с подлежащим, например, ученик школы; а также прямое дополнение и обстоятельство, если подлежащее называет действие, например, быстро сдать экзамен оказалось сложно или покраска забора затянулась. Смысл подлежащего может уточняться присоединённым к нему придаточным предложением. Как и синтаксические группы любых других членов предложения, группа подлежащего может включать в себя частицу, предваряться вводной конструкцией, а также быть заключённой в кавычки. Структура группы подлежащего формально описывается следующим правилом:

группа-подлежащего ::= вводная-конструкция группа-подлежащего
 | однородные-подлежащие
 | группа-подлежащего подчинительный-союз грамматическая-основа
 | группа-подлежащего группа-непрямого-дополнения
 | группа-подлежащего группа-обстоятельства
 | группа-подлежащего группа-определения
 | группа-подлежащего группа-прямого-дополнения
 | группа-подлежащего частица
 | кавычка группа-подлежащего кавычка
 | подлежащее

Однородные подлежащие, как и другие однородные члены предложения, формируют синтаксическую группу, включающую в себя несколько групп подлежащих и соединяющий их сочинительный союз. Как и в случае со сложносочинёнными предложениями, для упрощения грамматики знаки препинания, замещающие сочинительные союзы, приравниваются к сочинительным союзам: *стол, стул и кровать = стол, и стул, и кровать*. Далее в тексте опускаются правила для синтаксических групп однородных членов, строящиеся аналогично группе однородных подлежащих. Структура синтаксической группы однородных подлежащих формально описывается следующим правилом:

однородные-подлежащие ::= группа-подлежащего группа-подлежащего
 {группа-подлежащего} сочинительный-союз

3.3. Группа сказуемого. Сказуемое может быть как простым, так и составным, состоящим из главной части – глагольной или именной – и глагола-связки. В случае глагольной главной части говорят о составном глагольном сказуемом: *начать рисовать*; в случае именной – о составном именном сказуемом: *быть студентом*. Как глагольная, так и именная части могут состоять из нескольких однородных членов: *начали петь и танцевать, был важен и уместен*. Глагольная главная часть может быть составной: могу согласиться уйти.

Для более формального отделения составных сказуемых от других синтаксических единиц был ограничен список глаголов, которые могут быть связками.

– Модальные глаголы: *быть, стать, оказаться, начать, продолжить, мочь, позволить, сметь, устать, перестать, являться, пытаться.*

– Глаголы чувственного восприятия и близкие к ним по семантике: *любить, видеть, смотреть, слышать, чувствовать, нюхать, трогать, считать, представлять.*

Фразы с глаголами, не входящими в данный список, в рамках данной грамматики не являются составными сказуемыми. Например, схожая по строению с составным сказуемым фраза *пожимать плечами* считается сказуемым *пожимать* и дополнением *плечами*. Связка может быть нулевой, когда на её месте подразумевается соответствующий модальный глагол: *Я (есть) студент.* Глагол-связка также может быть составной: *хочу быть генералом.*

Группа сказуемого может, кроме сказуемого, включать прямые дополнения, как правило, называющие объект действия, такие как *сдаёт экзамен*; обстоятельства, характеризующие свойства действия, например, *сдаёт легко*; не прямые дополнения, называющие объекты, связанные с действием, например, *вышел из дома*. Также в состав группы сказуемого может входить определение, сообщающее о свойстве именной части составного сказуемого, такое как *был лучшим учеником.*

Структура группы сказуемого описывается следующим набором формальных правил:

```

группа-сказуемого ::= вводная-конструкция группа-сказуемого
| однородные-сказуемые
| группа-сказуемого подчинительный-союз грамматическая-основа
| группа-сказуемого группа-непрямого-дополнения
| группа-сказуемого группа-определения
| группа-сказуемого группа-обстоятельства
| группа-сказуемого группа-прямого-дополнения
| группа-сказуемого частица
| составное-сказуемое
| сказуемое
составное-сказуемое ::= ( глагольная-часть | именная-часть ) [ глагол-связка ]
глагольная-часть ::= однородные-глагольные-части
| составное-сказуемое
| глагол-начальной-формы
именная-часть ::= однородные-именные-части
| предлог именная-часть

```

| именная-часть частица
 | кавычка именная-часть кавычка
 | именная-часть-речи
 глагол-связка ::= составное-сказуемое
 | модальный-глагол

Важное отличие синтаксической группы однородных сказуемых от синтаксических групп других однородных членов – к однородным сказуемым может относиться общее непрямое дополнение, как в предложении *Я живу и работаю в Ярославле*. Формально это описывается правилом:

однородные-сказуемые ::= группа-непрямого-дополнения однородные-сказуемые
 | группа-сказуемого группа-сказуемого
 {группа-сказуемого} сочинительный-союз

3.4. Группа дополнения. Синтаксические группы прямых и косвенных дополнений грамматикой описываются одинаково. Поскольку дополнение, как и подлежащее, называет явление, то и синтаксическая группа дополнения описывается почти так же, как и группа подлежащего. Отличий всего два: во-первых, в составе группы дополнения может быть предлог, во-вторых, дополнение может быть выражено синтаксической группой прямой речи. Формально это выражается следующими правилами:

группа-прямого-дополнения ::= группа-дополнения
 группа-косвенного-дополнения ::= группа-дополнения
 группа-дополнения ::= вводная-конструкция группа-дополнения
 | прямая-речь
 | однородные-дополнения
 | группа-дополнения подчинительный-союз грамматическая-основа
 | группа-дополнения группа-косвенного-дополнения
 | предлог группа-дополнения
 | группа-дополнения группа-обстоятельства
 | группа-дополнения группа-определения
 | группа-дополнения группа-прямого-дополнения
 | группа-дополнения частица
 | кавычка группа-дополнения кавычка
 | дополнение

Важная особенность синтаксической группы однородных дополнений: к однородным дополнениям может относиться общий предлог, например, *рыбу ловят в озёрах и прудах*. Формально это можно

описать правилом:

однородные-дополнения ::= предлог однородные-дополнения
 | группа-дополнения группа-дополнения
 {группа-дополнения} сочинительный-союз

Прямая речь всегда находится в группе дополнения, поскольку она может быть заменена на соответствующее изъяснительное придаточное. Её синтаксическая группа состоит из отдельного предложения прямой речи и отделяющих его от главного предложения знаков препинания. Формально её структура описывается следующим правилом:

прямая-речь ::= предложение знаки-препинания

3.5. Группа определения. В состав группы определения, кроме самого определения, могут входить другие определения: *самый лучший*; прямые и не прямые дополнения: *читавший Пушкина*, *сбежавший из дома*; обстоятельства: *сдающие сейчас*. Структура синтаксической группы определения описывается следующим формальным правилом:

группа-определения ::= вводная-конструкция группа-определения
 | однородные-определения
 | группа-определения подчинительный-союз грамматическая-основа
 | группа-определения группа-непрямого-дополнения
 | предлог группа-определения
 | группа-определения группа-обстоятельства
 | группа-определения группа-определения
 | группа-определения группа-прямого-дополнения
 | группа-определения частица
 | определение

В рамках данной грамматики приложения не выделяются в отдельный вид членов предложения, а относятся к определениям. В словосочетаниях, в которых грань между главным и зависимым словом размыта, например, *инженер-конструктор*, *река Волга*, главным словом для упрощения всегда считается идущее первым: *инженер-конструктор*, *река Волга*, *Волга-река*.

Для отделения определений от определительных дополнений используется упрощённый критерий: слово считается определением, если оно может быть заменено на прилагательное без потери смысла. Например, *двор школы* = *школьный двор*, *колесо велосипеда* = *велосипедное колесо*, но *шип дорожки* ≠ *дорожный шип*.

3.6. Группа обстоятельства. В состав синтаксической группы обстоятельства, кроме непосредственно обстоятельства, могут входить прямые и не прямые дополнения: слуш^ая му^зыку, ме^ншая пре^по^да^вате^лю; определения: рабо^та^л це^ль^ный ве^че^р; другие обстоятельства: слу^ша^я нев^ни^мате^ль^но. Формально структура синтаксической группы обстоятельства описывается правилом:

группа-обстоятельства ::= вводная-конструкция группа-обстоятельства
 | однородные-обстоятельства
 | группа-обстоятельства подчинительный-союз грамматическая-основа
 | группа-обстоятельства группа-непрямого-дополнения
 | предлог группа-обстоятельства
 | группа-обстоятельства группа-обстоятельства
 | группа-обстоятельства группа-прямого-дополнения
 | группа-обстоятельства частица
 | группа-обстоятельства группа-определения
 | обстоятельство

Для отделения обстоятельств от обстоятельственных дополнений используется несколько упрощённых правил:

– Обстоятельством считается любое уточнение, относящееся к вопросам времени: бе^га^ть по у^тра^м, уе^ха^ть по^сле се^сси^и.

– Обстоятельством считается любая группа с одним характерных для обстоятельств предлогов: во^пре^ки, в^пло^ть до, в слу^ча^е, в ка^че^ст^ве, в свя^зи с, за с^чёт, из-за, на^зло, не^смо^тря на, по^сле, вс^ле^дст^вие, с по^мо^щью, при.

– Обстоятельством считается любое уточнение, относящееся к абстрактным понятиям, отвечающим на вопрос «как?»: хо^ди^ть бо^ко^м, жи^ть о^тще^ль^ни^ко^м.

– Во всех остальных случаях уточняющее слово считается дополнением.

4. Алгоритм. Предлагаемый алгоритм позволяет построить дерево синтаксических единиц русскоязычного предложения, соответствующее описанной выше грамматике, на основе его дерева синтаксических связей.

На вход алгоритма подаётся дерево синтаксических связей, узлами которого являются отдельные слова и знаки препинания предложения (токены), а дугами – синтаксические связи между ними. Для каждой дуги известен её тип связи, размеченный в соответствии с Universal Dependencies [31, 32]. Результатом работы алгоритма является дерево синтаксических единиц, корень которого имеет тип *предложение*.

4.1. Общая схема. Пусть $D = (V_D, E_D)$ обозначает дерево синтаксических связей; v_D – его узел; $C(v_D)$ – не включённые в дерево

синтаксических единиц прямые потомки узла v_D , отсортированные от наиболее близких в тексте предложения к v_D к наиболее удалённым; $e(v_i, v_j)$ – метка дуги между узлами v_i и v_j . Также $R_t = (r_{t1}, \dots, r_{tn})$ – набор продукций для синтаксических единиц типа t в порядке приоритета их применения (раздел 3); $N = (V_N, E_N)$ – дерево синтаксических единиц.

Общая схема построения дерева синтаксических единиц по данному дереву синтаксических связей описывается алгоритмом 1. Для построения дерева синтаксических единиц русскоязычного предложения по его дереву синтаксических связей $D_s = (V_{D_s}, E_{D_s})$ используется вызов *CreateConstituencyTree*(D_s , предложение).

Алгоритм 1. *CreateConstituencyTree* – создаёт дерево синтаксических единиц с корнем заданного типа по данному дереву синтаксических связей

Input: $D = (V_D, E_D)$ – дерево синтаксических связей, v_r – его корень; t – тип синтаксической единицы.

Output: дерево синтаксических единиц $N = (V_N, E_N)$, корень которого имеет тип t .

if t является терминалом **then**

 /* создать синтаксическую единицу типа t из токена v_r */

return ($\{v_r\}, \emptyset$)

end if

 /* выбор продукции для применения */

for $r \in R_t$ **do**

 /* попытка применить r к D */

 ($success, N$) := *ApplyProduction*(r, D)

if $success$ **then**

 /* если попытка была успешна, то её результат – искомый N */

return N

end if

end for

Для выбора продукции для построения синтаксических единиц используется вспомогательный алгоритм 2, который, в свою очередь, рекурсивно вызывает основной алгоритм. Рекурсия конечна, так как на каждом её шаге уменьшается мощность множества $C(v_D)$.

Алгоритм 2. *ApplyProduction* – пытается построить дерево синтаксических единиц по данным продукции и дереву синтаксических связей

Input: $D = (V_D, E_D)$ – дерево синтаксических связей, v_r – его корень;
 $r = (t_1 \dots t_n)$ – продукция, каждый её элемент – тип синтаксической единицы.

Output: *success* – успешность построения (**true** или **false**);

$N = (V_N, E_N)$ – дерево синтаксических единиц, построенное в соответствии с продукцией r .

$W = ()$

/* поиск токена, подходящего каждому элементу продукции */

for $t \in r$ **do**

for $w \in v_r \cup C(v_r)$ **do**

 /* каждому элементу продукции должен соответствовать свой токен */

if $w \in W$ **then**

continue

else if w соответствует критерию для t **then**

 /* токен для t найден */

$W := W + w$

break

end if

end for

end for

if $|W| \neq |r|$ **then**

 /* для какого-то из элементов продукции не удалось найти токен */

return (**false**, $(\emptyset, \emptyset)$)

end if

/* найдены токены для всех элементов продукции */

/* включим все найденные токены в дерево синтаксических связей */

/* сформируем искомое дерево синтаксических единиц */

$N = (\emptyset, \emptyset)$

for $i := 1 \dots |r|$ **do**

 /* создадим дерево синтаксических единиц для каждой пары из элемента r и соответствующего ему токена */

$D_i := (V_i, E_i)$ – поддерево D для W_i

$N_i := \text{CreateConstituencyTree}(D_i, r_i)$

 добавить N_i к N

end for

return (**true**, N)

Нужно отметить, что продукция r , в соответствии с которой можно построить v_N , существует всегда, поскольку в предложении на

русском языке всегда есть хотя бы один из главных членов предложения, среди наборов продукций для синтаксических групп каждого члена предложения есть продукция, состоящая из одного элемента – самого члена предложения, а в случае построения синтаксических групп однородных членов предложения и группы прямой речи заранее производится проверка возможности построения такой группы.

Далее описаны критерии для различных типов синтаксических единиц грамматики, используемые алгоритмом 2 (условие « w соответствует критерию для t »).

4.2. Критерии для элементов предложения в целом и его грамматической основы. В данной работе считается, что синтаксическая единица предложения в целом состоит из грамматической основы и, возможно, знака препинания. Токен, на основе которого создаётся синтаксическая единица знака препинания, должен находиться в конце предложения и быть знаком препинания. Синтаксическая единица грамматической основы предложения может быть создана на основе любого токена.

Синтаксическая единица вводной конструкции, предваряющей грамматическую основу, может быть построена на основе токена v из $C(v_r)$, находящегося ближе всего к началу предложения, такого, что выполняется хотя бы одно из условий:

- Метка синтаксической связи $e(v_r, v) = parataxis$. Большинство вводных конструкций имеет именно такой тип синтаксической связи: *Безусловно, программист знает математику. Кроме того, река замёрзла.*

- Метка синтаксической связи $e(v_r, v) = advcl$, среди $C(v)$ есть союз «как» и между v и v_r в предложении есть запятая. Это условие необходимо, чтобы обнаружить вводные конструкции, схожие с обстоятельствами, например: *Как выражаются моряки, ветер крепчает.*

При этом v не должно быть заключено в скобки или кавычки. Этот же критерий в данной работе используется и для построения вводных конструкций, входящих во все остальные синтаксические единицы.

Синтаксическая единица сложносочинённого предложения может быть построена на основе токена v_r , если среди $C(v_r)$ есть хотя бы один токен v , для которого выполняются следующие условия:

- метка синтаксической связи $e(v_r, v) = conj$;
- среди $C(v)$ есть хотя бы один токен, соответствующий критерию для группы подлежащего.

Первая грамматическая основа сложносочинённого предложения строится на основе самого v_r , вторая и последующие – на основе всех $v \in C(v_r)$, для которых выполняются приведённые выше условия.

Синтаксическая единица расчленённого сложноподчинённого предложения может быть построена на основе токена v_r , если среди $C(v_r)$ есть токен v , для которого выполняется любой из критериев грамматической основы подчинённого предложения:

- v присоединяется к v_r подчинительным союзом, а метка синтаксической связи $e(v_r, v)$ – *parataxis*, *advcl* или *conj*;
- метка синтаксической связи $e(v_r, v)$ – *parataxis* либо *advcl*, а также либо среди $C(v)$ есть хотя бы один токен, соответствующий критерию для группы подлежащего, либо v и v_r не являются глаголами одного времени.

При этом v не должно быть заключено в кавычки или относиться к прямой речи.

Синтаксическая единица группы подлежащего может быть построена на основе токена v_r , если выполняется одно из условий:

- v_r – именная часть речи, и либо $e(v_r, v) \in \{nsubj, nsubj:pass\}$, либо v_r – корень дерева синтаксических связей D ;
- v_r – глагол в начальной форме, и либо $e(v_r, v) \in \{csubj, csubj:pass\}$, либо v_r – корень дерева синтаксических связей D .

Синтаксическая единица группы сказуемого может быть построена на основе токена v_r , если он является корнем дерева синтаксических связей D , и выполняется любое из следующих условий:

- v_r – глагол;
- v_r удовлетворяет критериям для главной части составного именного сказуемого: v_r – это именная часть речи, а среди $C(v_r)$ есть хотя бы один токен, удовлетворяющий критерию для глагола-связки или группы подлежащего. Также главными частями составного именного сказуемого всегда считаются слова-категории состояния и краткие прилагательные.

4.3. Критерии для элементов групп подлежащего, сказуемого, определения, дополнения и обстоятельства. Синтаксическая единица однородных подлежащих может быть построена на основе токена v_r , если среди $C(v_r)$ есть хотя бы один токен v , для которого $e(v_r, v) = conj$. Первое из однородных подлежащих строится на основе самого v_r , второе и последующие – на основе каждого из $v \in C(v_r)$ с меткой синтаксической связи $e(v_r, v) = conj$. Такой же критерий используется и для всех остальных синтаксических единиц однородных членов предложения, если не указано иное.

Синтаксическая единица грамматической основы нерасчленённого придаточного предложения может быть построена на основе токена v_r , если выполняется хотя бы одно из условий:

- v_r отделено от v_d – корня дерева синтаксических связей D – знаком препинания, при этом либо синтаксическая связь $e(v_r, v_d)$ должна иметь метку acl , $acl:relcl$ или $ccomp$, либо v_r должно присоединяться к v_d подчинительным союзом;

- v_r отделено от v_d двоеточием и метка синтаксической связи $e(v_r, v_d) = appos$.

При этом синтаксическая единица подчинительного союза строится на основе токена $v \in C(v_r)$, являющегося подчинительным союзом, находящегося в предложении между v_r и v_d и такого, что $e(v, v_r) = mark$. Этот же критерий для нерасчленённого придаточного предложения используется в данной работе и при построении синтаксических единиц для остальных членов предложения.

Синтаксическая единица группы непрямого дополнения может быть построена на основе токена v_r такого, что выполняется хотя бы одно из условий:

- тип синтаксической связи между v_r и корнем дерева синтаксических связей $e(v_r, v_d) = iobj, nmod, obl$ или $xcomp$;
- при v_r есть предлог, и $e(v_r, v_d) = nummod:gov$.

При этом v_r не должно быть характерного для обстоятельств предлога, а среди $v_r \cup C(v_r)$ не должно быть слов, называющих время суток, день, время года или меру времени (раздел 3.6).

Синтаксическая единица группы обстоятельства может быть построена на основе токена v_r , если выполняется хотя бы одно из условий:

- тип синтаксической связи между v_r и корнем дерева синтаксических связей $e(v_r, v_d) = advmod$;
- v_r – наречие или прилагательное, и $e(v_r, v_d) = advcl$ или obl ;
- при v_r есть характерный для обстоятельств предлог (раздел 3.6);
- среди $v_r \cup C(v_r)$ есть хотя бы одно слово, называющие время суток, день недели, время года или меру времени.

Синтаксическая единица группы определения может быть построена на основе токена v_r , если выполняется хотя бы одно из условий:

- тип синтаксической связи между v_r и корнем дерева синтаксических связей $e(v_r, v_d) = amod, det$ или $appos$;
- v_r – полное прилагательное, и $e(v_r, v_d) = acl$;
- v_r – полное причастие.

Синтаксическая единица группы прямого дополнения может быть построена на основе токена v_r , если выполняется хотя бы одно из условий:

- тип синтаксической связи между v_r и корнем дерева синтаксических связей $e(v_r, v_d) = obj$;
- при v_r нет предлога, и $e(v_r, v_d) = nummod:gov$;
- v_r – именная часть речи в родительном падеже, и v_d – отглагольное существительное.

При построении синтаксической единицы группы однородных сказуемых на основе токенов $V = v_1, \dots, v_n$ синтаксическая единица их общего непрямого дополнения может быть построена при условии того, что V соединены сочинительным союзом, а не только знаками препинания, на основе токена $v_r \in \bigcup_{i=1}^n C(v_i)$ такого, что v_r находится в предложении либо сразу перед первым токеном из V , либо сразу после последнего из них.

Синтаксическая единица составного именного сказуемого может быть построена на основе токена v_r , являющегося корнем дерева синтаксических связей, если выполняется одно из условий:

- v_r входит в список глаголов-связок (раздел 3.3), а также существует $v \in C(v_r)$, являющееся именной частью речи, кратким причастием или категорией состояния. В этом случае синтаксическая единица именной части строится на основе v , а синтаксическая единица глагола-связки – на основе v_r .

– v_r является именной частью речи, кратким причастием или категорией состояния. В этом случае синтаксическая единица именной части строится на основе v_r . Если при этом существует $v \in C(v_r)$, входящий в список глаголов-связок, то на основе v строится синтаксическая единица глагола-связки; если же такого v не существует, то связка считается нулевой и синтаксическая единица для неё не строится.

Синтаксическая единица составного глагольного сказуемого может быть построена на основе токена v_r , являющегося корнем дерева синтаксических связей, если выполняется одно из условий:

- v_r входит в список глаголов-связок, и существует токен $v \in C(v_r)$ такой, что $e(v_r, v) = csubj$ или $xcomp$. В этом случае синтаксическая единица глагольной части строится на основе v , а синтаксическая единица глагола-связки – на основе v_r .

– v_r – глагол начальной формы, и существует токен $v \in C(v_r)$, являющийся глаголом-связкой. В этом случае синтаксическая единица

глагольной части строится на основе v_r , а синтаксическая единица глагола-связки – на основе v .

Синтаксическая единица группы прямой речи может быть построена на основе токена v_r такого, что выполняются все условия:

- первый по порядку в предложении элемент $C(v_r)$ является открывающей кавычкой, а последний – закрывающей;
- между v_r и корнем синтаксического дерева v_d есть тире, если v_r в предложении предшествует v_d , и двоеточие, если v_r находится после v_d .

5. Эксперименты

5.1. Корпус. При построении корпуса для экспериментов авторы работы исходили из следующих соображений. Во-первых, необходимо оценить способность предложенного алгоритма качественно строить деревья синтаксических единиц предложений, в которых присутствуют разнообразные синтаксические конструкции русского языка. Во-вторых, необходимо оценить пригодность предложенного алгоритма к использованию в рамках решения прикладных задач компьютерной лингвистики, для чего требуется оценить его способность качественно строить деревья синтаксических единиц предложений, используемых в таких задачах.

Для оценки способности предложенного алгоритма строить деревья синтаксических единиц предложений, в которых встречаются разнообразные синтаксические конструкции русского языка, был использован набор из 200 предложений, выбранных из раздела «Синтаксис» классической работы [13]. Предложения выбирались так, чтобы включать примеры применения как можно более разнообразных конструкций, относящихся к различным разделам синтаксиса, например, устройству простых предложений, различным видам сложных предложений, различным способам выражения членов предложения, последовательностям однородных членов. Разделы синтаксиса и количество подобранных для них предложений приведены в таблице 1.

Для оценки способности предложенного алгоритма строить деревья синтаксических единиц предложений, используемых в прикладных задачах компьютерной лингвистики, например, анализе тональности, было выбрано 100 случайных предложений из OpenCorpora – открытого корпуса русского языка, широко использующегося для решения разнообразных прикладных задач [33 – 35].

Экспертами-лингвистами была выполнена разметка всех 300 предложений по синтаксической структуре.

Таблица 1. Разделы синтаксиса и количество предложений, выбранных для них в качестве примеров

Раздел синтаксиса	Количество предложений
Простое двусоставное предложение	20
Односоставное предложение	10
Составное сказуемое	10
Определение	25
Приложение	5
Дополнение	25
Обстоятельство	25
Однородные члены предложения	10
Вводные конструкции	5
Обособленные члены предложения	10
Сложносочинённое предложение	20
Сложноподчинённое предложение	20
Бессоюзное сложное предложение	10
Прямая речь	5
Итого	200

5.2. Метрики. При проведении экспериментов для оценки качества построения деревьев синтаксических единиц используется несколько метрик.

Первая из них – доля предложений, для которых деревья синтаксических единиц были построены правильно, то есть дерево, полученное в результате работы алгоритма, в точности совпадает с деревом, построенным экспертами-лингвистами:

$$FullyCorrect = \frac{\text{число правильно построенных деревьев}}{\text{объём корпуса}}.$$

Существенный недостаток использования такой метрики – она не учитывает различия между полностью неверно построенными деревьями синтаксических единиц и деревьями синтаксических единиц, построенными с незначительными ошибками. Однако, например, различие между ошибкой, заключающейся в построении для двусоставного предложения дерева синтаксических единиц без группы подлежащего, и ошибкой, связанной с определением прямого дополнения как непрямого, достаточно существенна.

Для компенсации этого недостатка были использованы метрики точности, полноты и F_1 -меры построения деревьев синтаксических

единиц [36]. Эти метрики схожи с классическими, используемыми в задачах классификации, однако рассчитываются для всего множества входящих в деревья синтаксических единиц.

– Точность построения деревьев синтаксических единиц – отношение числа синтаксических единиц, входящих и в деревья, построенные алгоритмом, и в деревья, построенные экспертами-лингвистами, к общему числу синтаксических единиц в деревьях, построенных алгоритмом. Точность показывает, какая доля синтаксических единиц, построенных алгоритмом, существует в предложениях на самом деле.

– Полнота построения деревьев синтаксических единиц – отношение числа синтаксических единиц, входящих и в деревья, построенные алгоритмом, и в деревья, построенные экспертами-лингвистами, к общему числу синтаксических единиц в деревьях, построенных экспертами-лингвистами. Полнота характеризует способность алгоритма отыскивать реально существующие в предложении синтаксические единицы.

– F_1 -мера – среднее гармоническое точности и полноты.

Обозначив G_{TP} – синтаксические единицы, входящие и в деревья, построенные экспертами, и в деревья, построенные алгоритмом, G_{FP} – синтаксические единицы, входящие в деревья, построенные алгоритмом, но не в деревья, построенные экспертами, G_{FN} – синтаксические единицы, входящие в деревья, построенные экспертами, но не в деревья, построенные алгоритмом, можно записать формулы для точности, полноты и F_1 -меры построения деревьев синтаксических единиц:

$$Precision = \frac{G_{TP}}{G_{TP} + G_{FP}}; Recall = \frac{G_{TP}}{G_{TP} + G_{FN}}; F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}.$$

Помимо информации о внутренней синтаксической структуре предложения, дерево синтаксических единиц также содержит информацию о том, каким членом предложения является каждое из слов (листьев дерева). Для того, чтобы отдельно оценить точность этой информации, используется дополнительная метрика точности определения членов предложения, рассчитываемая по формуле:

$$Tagging = \frac{\text{число верно определённых членов предложений}}{\text{число слов в предложениях}}.$$

Поскольку деревья синтаксических связей редко бывают входными данными в практических задачах, авторам работы представляется

разумным при проведении экспериментов использовать в качестве входных данных сами предложения на естественном языке: текст предложения преобразуется анализатором синтаксических связей в дерево синтаксических связей; на основе этого дерева предложенный алгоритм строит дерево синтаксических единиц, которое сравнивается с эталоном, размеченным экспертами-лингвистами. При этом результаты работы анализатора синтаксических связей не валидируются – если для какого-то предложения он допустил ошибку, то это предложение не будет удалено из рассмотрения, и ошибки, допущенные из-за неверной работы анализатора синтаксических связей, не отделяются от ошибок, допущенных из-за неверной работы предложенного алгоритма. Конечно, такая оценка эффективности предложенного алгоритма может оказаться значительно ниже его реальной эффективности, однако, она позволяет сделать наиболее точный вывод о применимости предложенного алгоритма для решения практических задач.

Безусловно, при наличии достаточных ресурсов можно валидировать результаты работы анализатора синтаксических связей, и оценивать точность работы предложенного алгоритма только на тех предложениях, дерево синтаксических связей для которых было построено полностью верно. Однако, нужно отметить, что, валидация деревьев синтаксических связей – достаточно трудоёмкий процесс, и, при наличии достаточных для него ресурсов, возможно, целесообразнее будет потратить их напрямую на разметку банка деревьев для создания алгоритма анализа синтаксической структуры предложения, не использующего анализатор синтаксических связей.

5.3. Результаты. Для проведения экспериментов предложенный алгоритм был реализован на языке программирования Python; программный код реализации доступен в репозитории <https://github.com/yarfruct/constituency-from-dependency>. Для построения деревьев синтаксических связей, размеченных в соответствии с системой Universal Dependencies, использовались анализаторы из состава трёх библиотек: Stanza (обученный на наборе данных SynTagRus), SpaCy (версии `ru_core_news_lg`) и Natasha. Полученные в результате экспериментов метрики качества приведены в таблице 2.

На 200 предложениях из справочника Д.Э. Розенталя наилучшие результаты получаются при использовании анализатора синтаксических связей из состава библиотеки Stanza – для почти трёх четвертей предложений дерева синтаксических единиц строятся полностью верно; верно определяются практически все (95 %) члены предложения; точность, полнота и F_1 -мера также достаточно высоки и составляют

почти 0.9 каждая. Отсутствие существенной разницы между точностью и полнотой построения деревьев синтаксических единиц говорит о том, что алгоритм не склонен ни строить «лишние» синтаксические единицы, отсутствующие в построенных экспертами деревьях, ни упускать присутствующие в предложении синтаксические единицы. При использовании анализаторов синтаксических связей из состава библиотек SpaCy и Natasha качество значительно хуже: доля предложений, для которых деревья синтаксических единиц строятся полностью корректно, ниже на 0.16 для SpaCy и на 0.28 для Natasha; F_1 -мера – на 0.08 для SpaCy и на 0.15 для Natasha; точность определения членов предложения – на 0.05 для SpaCy и на 0.15 для Natasha. Нужно отметить, что для некоторых предложений анализатор деревьев синтаксических связей из Natasha строил деревья синтаксических связей, содержащие циклы, то есть заведомо некорректные.

Таблица 2. Метрики качества построения деревьев синтаксических единиц с помощью различных анализаторов синтаксических связей

Анализатор	<i>FullyCorrect</i>	<i>Precision</i>	<i>Recall</i>	F_1	<i>Tagging</i>
Для 200 предложений из справочника Д. Э. Розенталя					
Stanza	0.73	0.88	0.89	0.89	0.95
SpaCy	0.57	0.81	0.81	0.81	0.90
Natasha	0.45	0.78	0.70	0.74	0.80
Для 100 предложений из OpenCorpora					
Stanza	0.46	0.83	0.83	0.83	0.92
SpaCy	0.23	0.70	0.69	0.70	0.84
Natasha	0.12	0.68	0.49	0.57	0.62
Для всех 300 предложений					
Stanza	0.64	0.85	0.85	0.85	0.93
SpaCy	0.45	0.75	0.75	0.75	0.87
Natasha	0.34	0.73	0.58	0.65	0.70

На 100 предложениях из OpenCorpora наилучшие результаты также получаются при использовании анализатора из состава Stanza. При этом метрики качества ниже, чем для предложений из справочника – доля предложений с полностью правильно построенными деревьями снизилась существеннее всего, на 0.27, до 0.46; снижение F_1 -меры и точности определения членов предложения оказалось не таким существенным, всего на 0.06 и 0.03 соответственно. Для двух других анализаторов синтаксических связей снижение качества оказалось ещё более существенным – для анализатора из состава SpaCy доля полностью правильно построенных деревьев снизилась до 0.23, F_1 -мера – до 0.70,

а точность определения членов предложения – до 0.84; для анализатора из состава Natasha эти показатели составили 0.12, 0.57 и 0.62 соответственно, кроме того, по разнице между точностью, составившей 0.68 и полнотой, составившей 0.49, при использовании этого анализатора алгоритм склонен включать в дерево меньше синтаксических единиц, чем реально существует в предложении.

Скорее всего, такое снижение качества при использовании всех анализаторов связано с тем, что предложения из OpenCorpora в среднем более сложные и содержат в себе больше комбинаций различных синтаксических конструкций, что может затруднять работу анализаторов синтаксических связей. В пользу этой гипотезы говорит то, что для наиболее точного из анализаторов – из состава библиотеки Stanza – показатели качества снизились меньше, чем для анализаторов из состава SpaCy и Natasha.

Итоговые метрики качества, рассчитанные на всех 300 предложениях, позволяют говорить о том, что при использовании анализатора синтаксических связей из состава библиотеки Stanza предлагаемый алгоритм включает в деревья практически все синтаксические единицы, реально существующие в предложениях, и достаточно редко включает несуществующие (F_1 -мера – 0.85), а также достаточно точно определяет, каким членом предложения является то или иное слово (точность определения – 0.93), однако примерно в двух третях построенных деревьев синтаксических единиц присутствовали хотя бы минимальные ошибки. При использовании других синтаксических анализаторов построение деревьев синтаксических единиц происходит гораздо менее точно, с существенным снижением метрик качества. Тем не менее, поскольку во многих прикладных задачах, например, анализа тональности и извлечения информации, не требуется абсолютная точность синтаксического анализа, можно предположить, что предложенный алгоритм может успешно применяться на практике.

Закключение. В работе предложен алгоритм преобразования деревьев синтаксических связей в деревья синтаксических единиц для русскоязычных предложений. Он основан на рекурсивном применении к узлам дерева синтаксических связей продукции формальной грамматики, построенной на основе классического справочника Д. Э. Розенталя. Для оценки качества работы алгоритма был собран корпус из 300 предложений на русском языке, из которых 200 были выбраны из вышеупомянутого справочника, и 100 из открытого корпуса публицистических текстов OpenCorpora.

В ходе экспериментов предложения подавались на вход анализаторов синтаксических связей из состава библиотек Stanza, SpaCy и Natasha, после чего полученные деревья синтаксических связей обрабатывались предложенным алгоритмом. Полученные в результате обработки деревья синтаксических единиц сравнивались с размеченными вручную экспертами-филологами. Наилучшее качество было получено при использовании анализатора синтаксических связей из библиотеки Stanza: доля полностью правильно построенных деревьев синтаксических единиц составила 0.64, точность определения членов предложения составила 0.93, а F_1 -мера построения синтаксических единиц составила 0.85, чего должно быть достаточно для решения многих практических задач.

Дальнейшее повышение качества работы алгоритма может осуществляться с помощью расширения используемой формальной грамматики. Доработка грамматики и алгоритма, безусловно, потребуют проведения анализа ошибок и привлечения экспертов-лингвистов для уточнения используемых продукций.

Литература

1. Jurafsky D., Martin J.H. Speech and Language Processing. 2nd Edition. USA: Prentice-Hall, Inc., 2009. 1024 p.
2. Батура Т.В., Чаринцева М.В. Основы обработки текстовой информации: Учебное пособие. Новосибирск: Институт систем информатики им. А.П. Ершова СО РАН, 2016. 45 с.
3. Андреева С.В. Типология конструктивно-синтаксических единиц в русской речи // Вопросы языкознания. 2004. № 5. С. 32–45.
4. Онипенко Н.К. Об основаниях классификации синтаксических единиц // Труды института русского языка им. В.В. Виноградова. 2019. Т. 20. С. 189–201.
5. Percival W.K. On the historical source of immediate constituent analysis // Notes from the linguistics underground. 1976. pp. 229–242.
6. Waziri Z.Y., Safana M.I. Contrastive analysis of English and Hausa sentence structures and its pedagogical implications // Voices: A Journal of English Studies. 2021. vol. 5. pp. 15–27.
7. Dewi N.M.P., Putra I.G.W.N., Winarta I.B.G.N. Imperative Sentence in «The Guidance iPhone Support Website» // Elysian Journal: English Literature, Linguistics and Translation Studies. 2021. vol. 1. pp. 81–92.
8. Nguyen H.V., Tan N., Quan N.H., Huong T.T., Phat N.H. Building a Chatbot System to Analyze Opinions of English Comments // Informatics and Automation. 2023. vol. 22. no. 2. pp. 289–315.
9. Matchin W., Hickok G. The cortical organization of syntax // Cerebral Cortex. 2020. vol. 30. no. 3. pp. 1481–1498.
10. Ениколопов С.Н., Кузнецова Ю.М., Осипов С.Г., Смирнов И.В., Чудова Н.В. Метод реляционно-ситуационного анализа текста в психологических исследованиях // Психология. Журнал Высшей школы экономики. 2021. Т. 18. № 4. С. 748–769.

11. Zhang Y., Zhang Y. Tree communication models for sentiment analysis // Proceedings of the 57th annual meeting of the association for computational linguistics. 2019. pp. 3518–3527. DOI: 10.18653/v1/P19-1342.
12. Marcus M., Santorini B., Marcinkewicz M.A. Building a large annotated corpus of English: The Penn Treebank // Computational Linguistics. 1993. vol. 19 no. 2. pp. 313–330.
13. Розенталь Д.Э., Голуб И.Б., Теленкова М.А. Современный русский язык. 16-е изд. М.: АйРИС-пресс, 2018. 448 с.
14. Chomsky N. On certain formal properties of grammars // Information and control. 1959. vol. 2. no. 2. pp. 137–167.
15. Chomsky N. Some Puzzling Foundational Issues: the Reading Program // Catalan journal of linguistics. 2019. pp. 263–285. DOI: 10.5565/rev/catjl.287.
16. Muller S. Grammatical theory: From transformational grammar to constraint-based approaches. Fifth revised and extended edition. Berlin: Language Science Press, 2023. 889 p. DOI: 10.17169/langsci.b25.167.
17. Taylor A., Marcus M., Santorini B. The Penn Treebank: an overview // Treebanks: Building and using parsed corpora. Dordrecht: Springer Netherlands, 2003. 407 p. DOI: 10.1007/978-94-010-0201-1.
18. Zhou J., Zhao H. Head-Driven Phrase Structure Grammar Parsing on Penn Treebank // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. pp. 2396–2408.
19. Gaddy D., Stern M., Klein D. What's Going On in Neural Constituency Parsers? An Analysis // Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018. vol. 1. pp. 999–1010.
20. Zhang M.S. A survey of syntactic-semantic parsing based on constituent and dependency structures // Science China Technological Sciences. 2020. vol. 63. no. 10. pp. 1898–1920.
21. Yang S., Cui L., Ning R., Wu D., Zhang Y. Challenges to open-domain constituency parsing // Findings of the Association for Computational Linguistics: ACL 2022. 2022. pp. 112–127.
22. Гладкий А.В., Мельчук И.А. Элементы математической лингвистики. М.: Наука, 1969. 192 с.
23. Гладкий А.В. Синтаксические структуры естественного языка. Изд. 2-е. М.: УРСС, 2007. 146 с.
24. Коротаев Н.А. Синтаксические группы А.В. Гладкого: анализ конструкций с сочинением // Вестник РГТУ. Серия: Литературоведение. Языкознание. Культурология. 2013. № 8(109). С. 16–36.
25. Кагиров И.А., Леонтьева А.Б. Модуль синтаксического анализа для литературного русского языка // Труды СПИИРАН. 2008. Т. 6. С. 171–183.
26. Leontyeva A., Kagirov I. The module of morphological and syntactic analysis SMART // Text, Speech and Dialogue: 11th International Conference, TSD 2008. 2008. pp. 373–380.
27. Леонтьева Н.Н., Ермаков М.В., Крылов С.А., Семенова С.Ю., Соколова Е.Г. Прикладной семантический словарь РУСЛАН: основная концепция и обновленный подход // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог». 2020. С. 1049–1064.
28. Москвина А.Д., Орлова Д., Паничева П.В., Митрофанова О.А. Разработка ядра синтаксического анализатора для русского языка на основе библиотек NLTK // Сборник научных статей. Труды XIX Международной объединённой научной

- конференции «Интернет и современное общество». Санкт-Петербург: Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики. 2016. С. 44–54.
29. Shelmanov A., Pisarevskaya D., Chistova E., Toldova S., Kobozeva M., Smirnov I. Towards the data-driven system for rhetorical parsing of Russian texts // *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking*. 2019. pp. 82–87.
 30. Гаврилов Д.А. Сопоставительное изучение пунктуации в сетевом газетном заголовке: к постановке проблемы // *Вестник Чувашского государственного педагогического университета им. И.Я. Яковлева*. 2021. № 3(112). С. 3–8.
 31. De Marneffe M.C., Manning C.D., Nivre J., Zeman D. Universal Dependencies // *Computational Linguistics*. 2021. vol. 47. no. 2. pp. 255–308.
 32. Lyashevskaya O., Bocharov V., Sorokin A., Shavrina T., Granovsky D., Alexeeva S. Text collections for evaluation of Russian morphological taggers // *Journal of Linguistics / Jazykovedny Casopis*. 2017. vol. 68. no. 2. pp. 258–267.
 33. Kirillovich A., Loukachevitch N., Kulaev M., Bolshina A., Ilvovsky D. Sense-Annotated Corpus for Russian // *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*. 2022. pp. 130–136.
 34. Volkova L., Bocharov V. An approach to inter-annotation agreement evaluation for the named entities annotation task at OpenCorpora // *Communications in Computer and Information Science*. 2019. vol. 1119. pp. 33–44.
 35. Lagutina K. Topical Text Classification of Russian News: a comparison of BERT and Standard Models // *31st Conference of Open Innovations Association FRUCT*. 2022. pp. 160–166.
 36. Yang S., Tu K. Bottom-up constituency parsing and nested named entity recognition with pointer networks // *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. 2022. vol. 1. pp. 2403–2416.

Полетаев Анатолий Юрьевич – ассистент, кафедра компьютерных сетей, факультет информатики и вычислительной техники, Ярославский государственный университет им. П.Г. Демидова. Область научных интересов: анализ и моделирование естественного языка, математическая статистика. Число научных публикаций — 14. anatoliy-poletaev@mail.ru; улица Советская, 14, 150003, Ярославль, Россия; р.т.: +7(910)819-8325.

Парамонов Илья Вячеславович – канд. физ.-мат. наук, доцент, кафедра компьютерных сетей, факультет информатики и вычислительной техники, Ярославский государственный университет им. П.Г. Демидова; руководитель лаборатории, ярославская лаборатория Ассоциации открытых инноваций FRUCT. Область научных интересов: компьютерная лингвистика, искусственные нейронные сети, методология разработки программного обеспечения. Число научных публикаций — 55. ilya.paramonov@fruct.org; улица Советская, 14, 150003, Ярославль, Россия; р.т.: +7(905)633-3993.

Бойчук Елена Игоревна – д-р филол. наук, доцент, старший научный сотрудник, отдел управления наукой и инновациями, Ярославский государственный университет им. П.Г. Демидова. Область научных интересов: компьютерная лингвистика, филология, литературоведение, лингвистика текста, функциональная грамматика, теория языка и сравнительно-сопоставительные исследования. Число научных публикаций — 108. elena-boychouk@rambler.ru; улица Советская, 14, 150003, Ярославль, Россия; р.т.: +7(903)824-1797.

Поддержка исследований. Исследование выполнено за счет гранта Российского научного фонда № 23-21-00495 (<https://rscf.ru/project/23-21-00495/>).

A. POLETAEV, I. PARAMONOV, E. BOYCHUK
**ALGORITHM OF CONSTITUENCY TREE FROM DEPENDENCY
TREE CONSTRUCTION FOR A RUSSIAN-LANGUAGE SENTENCE**

Poletaev A., Paramonov I., Boychuk E. **Algorithm of Constituency Tree from Dependency Tree Construction for a Russian-Language Sentence.**

Abstract. Automatic syntactic analysis of a sentence is an important computational linguistics task. At present, there are no syntactic structure parsers for Russian that are publicly available and suitable for practical applications. Ground-up creation of such parsers requires building of a treebank annotated according to a given formal grammar, which is quite a cumbersome task. However, since there are several syntactic dependency parsers for Russian, it seems reasonable to employ dependency parsing results for syntactic structure analysis. The article introduces an algorithm that allows to construct the constituency tree of a Russian sentence by a syntactic dependency tree. The formal grammar used by the algorithm is based on the D.E. Rosenthal's classic reference. The algorithm was evaluated on 300 Russian-language sentences. 200 of them were selected from the aforementioned reference, and 100 from OpenCorpora, an open corpus of sentences extracted from Russian news and periodicals. During the evaluation, the sentences were passed to syntactic dependency parsers from Stanza, SpaCy, and Natasha packages, then the resulted dependency trees were processed by the proposed algorithm. The obtained constituency trees were compared with the trees manually annotated by experts in linguistics. The best performance was achieved using the Stanza parser: the constituency parsing F_1 -score was 0.85, and the sentence parts tagging accuracy was 0.93, that would be sufficient for many practical applications, such as event extraction, information retrieval and sentiment analysis.

Keywords: computational linguistics, natural language processing, syntactic parsing, constituency tree, dependency tree, formal grammar.

References

1. Jurafsky D., Martin J.H. Speech and Language Processing. 2nd Edition. USA: Prentice-Hall, Inc., 2009. 1024 p.
2. Batura T V., Charinceva M.V. Osnovy obrabotki tekstovoj informacii: Uchebnoe posobie. [Basics of textual information processing: Study Guide]. Novosibirsk: Institut sistem informatiki im. A.P. Ershova SO RAN, 2016. 45 p. (in Russ.).
3. Andrejeva S.V. [Typology of constructive-syntactic units in Russian speech]. Voprosy yazykoznaviya – Problems of linguistics. 2004. no. 5. pp. 32–45. (in Russ.).
4. Onipenko N.K. [About the grounds for the classification of syntactic units]. Trudy Instituta Russkogo Iazyka im. V.V. Vinogradova – Proceedings of the V.V. Vinogradov Russian Language Institute. 2019. vol. 20. pp. 189–201. (in Russ.).
5. Percival W.K. On the historical source of immediate constituent analysis. Notes from the linguistics underground. 1976. pp. 229–242.
6. Waziri Z.Y., Safana M.I. Contrastive analysis of English and Hausa sentence structures and its pedagogical implications. Voices: A Journal of English Studies. 2021. vol. 5. pp. 15–27.
7. Dewi N.M.P., Putra I.G.W.N., Winarta I.B.G.N. Imperative Sentence in «The Guidance iPhone Support Website». Elysian Journal: English Literature, Linguistics and Translation Studies. 2021. vol. 1. pp. 81–92.

8. Nguyen H.V., Tan N., Quan N.H., Huong T.T., Phat N.H. Building a Chatbot System to Analyze Opinions of English Comments. *Informatics and Automation*. 2023. vol. 22. no. 2. pp. 289–315.
9. Matchin W., Hickok G. The cortical organization of syntax. *Cerebral Cortex*. 2020. vol. 30. no. 3. pp. 1481–1498.
10. Enikolopov S.N., Kuznetsova Y.M., Osipov G.S., Smirnov I.V., Chudova N.V. [The Method of Relational-Situational Analysis of Text in Psychological Research]. *Psihologiya. Zhurnal vysshej shkoly ekonomiki – Psychology. Journal of the Higher School of Economics*]. 2021. vol. 18. no. 4. pp. 748–769. (in Russ.).
11. Zhang Y., Zhang Y. Tree communication models for sentiment analysis. *Proceedings of the 57th annual meeting of the association for computational linguistics*. 2019. pp. 3518–3527. DOI: 10.18653/v1/P19-1342.
12. Marcus M., Santorini B., Marcinkewicz M.A. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*. 1993. vol. 19 no. 2. pp. 313–330.
13. Rozental D.E., Golub I.B., Telenkova M.A. *Sovremennyy russkij jazyk [Modern Russian language (16th Edition)]*. Moscow: AJRIS-press, 2018. 448 p. (in Russ.).
14. Chomsky N. On certain formal properties of grammars. *Information and control*. 1959. vol. 2. no. 2. pp. 137–167.
15. Chomsky N. Some Puzzling Foundational Issues: the Reading Program. *Catalan journal of linguistics*. 2019. pp. 263–285. DOI: 10.5565/rev/catjl.287.
16. Muller S. *Grammatical theory: From transformational grammar to constraint-based approaches*. Fifth revised and extended edition. Berlin: Language Science Press, 2023. 889 p. DOI: 10.17169/langsci.b25.167.
17. Taylor A., Marcus M., Santorini B. *The Penn Treebank: an overview*. Treebanks: Building and using parsed corpora. Dordrecht: Springer Netherlands, 2003. 407 p. DOI: 10.1007/978-94-010-0201-1.
18. Zhou J., Zhao H. Head-Driven Phrase Structure Grammar Parsing on Penn Treebank. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019. pp. 2396–2408.
19. Gaddy D., Stern M., Klein D. What's Going On in Neural Constituency Parsers? An Analysis. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2018. vol. 1. pp. 999–1010.
20. Zhang M.S. A survey of syntactic-semantic parsing based on constituent and dependency structures. *Science China Technological Sciences*. 2020. vol. 63. no. 10. pp. 1898–1920.
21. Yang S., Cui L., Ning R., Wu D., Zhang Y. Challenges to open-domain constituency parsing. *Findings of the Association for Computational Linguistics: ACL 2022*. 2022. pp. 112–127.
22. Gladkij A.V., Melchuk I.A. *Jelementy matematicheskoy lingvistiki [Elements of mathematical linguistics]*. Moscow: Nauka, 1969. 192 p. (in Russ.).
23. Gladkij A.V. *Sintaksicheskie struktury estestvennogo jazyka [Syntactic structures of natural language (2nd Edition)]*. Moscow: URSS, 2007. 146 p. (in Russ.).
24. Korotaev N.A. [A.V. Gladkij syntactic groups: analysis of compound constructions]. *Vestnik RGGU. Seriya: Literaturovedenie. Yazykoznanie. Kulturologiya – RSUH Bulletin. «Literary theory. Linguistics. Cultural Studies.» Series.*. 2013. no. 8(109). pp. 16–36. (in Russ.).
25. Kagiroy I.A., Leontyeva A.B. [Module for syntax parsing of the literary Russian language]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2008. vol. 6. pp. 171–183. (in Russ.).
26. Leontyeva A., Kagiroy I. The module of morphological and syntactic analysis SMART. *Text, Speech and Dialogue: 11th International Conference, TSD 2008*. 2008. pp. 373–380.

27. Leontyeva N.N., Ermakov M.V., Krylov S.A., Semenova S.Yu., Sokolova E.G. [On traditional conception and upgrading of one applied semantic dictionary]. *Kompyuternaya lingvistika i intellektualnye tekhnologii: Po materialam ezhegodnoj mezhdunarodnoj konferencii «Dialog» – Computational linguistics and intellectual technologies: Papers from the annual international conference «Dialogue»*. 2020. pp. 1049–1064. (in Russ.).
28. Moskvina A.D., Orlova D., Panicheva P.V., Mitrofanova O.A. *Razrabotka yadra sintaksicheskogo analizatora dlya russkogo yazyka na osnove bibliotek NLTK [Development of the Core for Syntactic Parser for Russian based on NLTK libraries] Kompyuternaya lingvistika i vychislitelnye ontologii: Sbornik nauchnyh statej. Trudy XIX Mezhdunarodnoj ob'edinyonnoj nauchnoj konferencii [Computational linguistics and computational ontologies: Collection of scientific articles. Proceedings of the XIXth International joint scientific conference]*. St. Petersburg: Sankt-Peterburgskij nacional'nyj issledovatel'skij universitet informacionnyh tekhnologij, mekhaniki i optiki, 2016. pp. 44–54. (in Russ.).
29. Shelmanov A., Pisarevskaya D., Chistova E., Toldova S., Kobozeva M., Smirnov I. Towards the data-driven system for rhetorical parsing of Russian texts. *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking*. 2019. pp. 82–87.
30. Gavrilo D.A. [Comparative study of punctuation in an online newspaper headline: statement of the problem]. *Vestnik Chuvashskogo Gosudarstvennogo Pedagogicheskogo Universiteta im. I. Y. Yakovleva – I. Yakovlev Chuvash State Pedagogical University Bulletin*. 2021. no. 3(112). pp. 3–8. (in Russ.).
31. De Marneffe M.C., Manning C.D., Nivre J., Zeman D. Universal Dependencies. *Computational Linguistics*. 2021. vol. 47. no. 2. pp. 255–308.
32. Lyashevskaya O., Bocharov V., Sorokin A., Shavrina T., Granovsky D., Alexeeva S. Text collections for evaluation of Russian morphological taggers. *Journal of Linguistics / Jazykovedny Casopis*. 2017. vol. 68. no. 2. pp. 258–267.
33. Kirillovich A., Loukachevitch N., Kulaev M., Bolshina A., Ilvovsky D. Sense-Annotated Corpus for Russian. *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*. 2022. pp. 130–136.
34. Volkova L., Bocharov V. An approach to inter-annotation agreement evaluation for the named entities annotation task at OpenCorpora. *Communications in Computer and Information Science*. 2019. vol. 1119. pp. 33–44.
35. Lagutina K. Topical Text Classification of Russian News: a comparison of BERT and Standard Models. *31st Conference of Open Innovations Association FRUCT*. 2022. pp. 160–166.
36. Yang S., Tu K. Bottom-up constituency parsing and nested named entity recognition with pointer networks. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. 2022. vol. 1. pp. 2403–2416.

Poletaev Anatoliy – Assistant, Chair of computer networks, faculty of computer science, P.G. Demidov Yaroslavl State University. Research interests: natural language analysis and modeling, mathematical statistics. The number of publications — 14. anatoliy-poletaev@mail.ru; 14, Sovetskaya St., 150003, Yaroslavl, Russia; office phone: +7(910)819-8325.

Paramonov Ilya – Ph.D., Associate professor, Chair of computer networks, faculty of computer science, P.G. Demidov Yaroslavl State University; Head of the laboratory, Yaroslavl Laboratory of Open Innovations Association FRUCT. Research interests: computational linguistics, artificial neural networks, software development methods. The number of publications — 55. ilya.paramonov@fruct.org; 14, Sovetskaya St., 150003, Yaroslavl, Russia; office phone: +7(905)633-3993.

Boyчук Elena – Ph.D., Dr.Sci., Associate Professor, Senior researcher, Research and development department, P.G. Demidov Yaroslavl State University. Research interests: computational linguistics, philology, literary criticism, text linguistics, functional grammar, language theory and comparative studies. The number of publications — 108. elena-boychouk@rambler.ru; 14, Sovetskaya St., 150003, Yaroslavl, Russia; office phone: +7(903)824-1797.

Acknowledgements. The reported study was funded by the grant of Russian Science Foundation No. 23-21-00495 (<https://rscf.ru/en/project/23-21-00495/>).

В.А. ГОЛОВИН, К.С. ЯКОВЛЕВ
**ПРИМИТИВЫ ДВИЖЕНИЯ РОБОТА В ЗАДАЧЕ
ПЛАНИРОВАНИЯ ТРАЕКТОРИИ С КИНЕМАТИЧЕСКИМИ
ОГРАНИЧЕНИЯМИ**

Головин В.А., Яковлев К.С. Прimitives движения робота в задаче планирования траектории с кинематическими ограничениями.

Аннотация. Автоматическое планирование траектории – актуальная научно-техническая задача, решения которой востребованы во многих областях: беспилотный транспорт, роботизированная логистика, социальная робототехника и т.д. Зачастую при планировании траектории необходимо учитывать тот факт, что агент (робот, беспилотный автомобиль и др.) не может произвольно менять ориентацию при движении, другими словами – необходимо учитывать кинематические ограничения при планировании. Одним из широко-распространенных подходов к решению этой задачи является подход, опирающийся на конструирование траектории из заранее подготовленных фрагментов, примитивов движения, каждый из которых в свою очередь удовлетворяет кинематическим ограничениям. Зачастую, акцент при разработке методов, реализующих этот подход, делается на сокращении перебора вариантов при планировании (эвристический поиск), при этом сам набор доступных примитивов считается заданным извне. В этой же работе, мы наоборот ставим своей целью провести исследование и анализ влияния различных доступных примитивов движения на качество решения задачи планирования при фиксированном алгоритме поиска. В частности, рассматриваются 3 различных набора примитивов движения для колесного робота с дифференциальным приводом. В качестве алгоритма поиска используется известный в искусственном интеллекте и робототехнике алгоритм A*. Качество решения оценивается по 6 метрикам, включая время планирования, длину и кривизну результирующей траектории. На основании проведенного исследования делаются выводы о факторах, оказывающих наибольшее влияние на результат планирования, и даются рекомендации по построению примитивов движения, использование которых позволяет достичь баланса между скоростью работы алгоритма планирования и качеством отыскиваемых траекторий.

Ключевые слова: планирование траектории, кинематическое планирование, примитивы движения, эвристический поиск.

1. Введение. Одной из наиболее активно-развивающихся областей современной науки и техники является разработка мобильных роботов, способных к функционированию в сложно-структурированных средах без активного участия оператора. Отдельный интерес представляют полностью автономные мобильные роботы, которые могут функционировать и решать поставленные перед ними задачи самостоятельно. Использование подобных систем в таких приложениях как беспилотный транспорт, доставка, логистика, мониторинг позволяет повысить эффективность решаемых задач (например, сократить время обработки и маршрутизации посылки на складе) [1, 2]. Более того, в ряде случаев использование роботов – это единственный способ

решения задачи в принципе. Например, при устранении последствий чрезвычайных ситуаций, когда нахождение человека в определенных областях невозможно из-за высокого риска потери здоровья (например, при проведении спасательных работ в непосредственной близости от поврежденных объектов ядерной энергетики).

Системы управления мобильными роботами с высокой степенью автономности обычно построены по модульному принципу [3]. Так, зачастую, выделяют следующие модули высокого уровня, каждый из которых, обычно, разбивается на множество более специализированных модулей:

- модуль восприятия (в англоязычной терминологии – perception)
- этот модуль ответственен за обработку и анализ данных, поступающих от датчиков, построение карты, локализацию робота в ней, выделение объектов интереса и т.д.;
- модуль планирования (в англоязычной терминологии – planning)
- этот модуль ответственен за планирование траектории с учетом информации, предоставленной модулем восприятия (карта, текущее положение статических препятствий, прогнозные траектории движения динамических препятствий и пр.);
- модуль управления (в англоязычной терминологии – control) – этот модуль ответственен за следование по построенной траектории.

Все указанные модули, безусловно, важны для эффективного функционирования системы управления мобильным роботом, однако в этой работе мы концентрируем свое внимание на модуле планирования траектории. В простейшем виде этот модуль по известной карте строит геометрический путь – ломанную линию, соединяющую две точки пространства, не пересекающую ни одно из нанесенных на карту препятствий. Очевидно, что такой подход целесообразен лишь, когда речь идет о построении траектории для робототехнических систем, обладающих возможностью движения в произвольном направлении без смены ориентации. Примерами таких систем являются квадрокоптеры или мобильные роботы с омни-колесами. В более общей постановке при планировании необходимо учитывать дополнительные ограничения робототехнической системы, которые могут быть весьма специфичны в различных случаях [4 – 6]. Если же концентрироваться на планировании траектории с точки зрения последующей ее выполнимости (с помощью модуля управления), то целесообразно учитывать в первую очередь ограничения на смену направления движения. Например, беспилотный автомобиль не может повернуться на 90 градусов мгновенно, ему требуется совершить поворот для этого, причем радиус этого поворота

– ограниченная снизу величина. Такие ограничения носят названия кинематических и именно проблеме автоматического планирования траектории с учетом кинематических ограничений и посвящена эта работа. Примеры траекторий, учитывающей (и не учитывающей) кинематические ограничения изображены на рисунке 1.

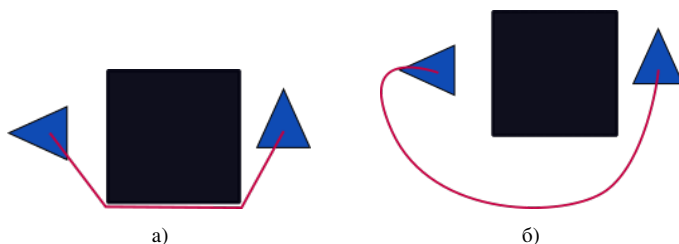


Рис. 1. Пример: а) траектории не учитывающей; б) траектории учитывающей кинематические ограничения

Одним из широко распространенных подходов к планированию траектории с учетом кинематических ограничений является декомпозиционный подход, основанный на построении траектории из небольших фрагментов, т.н. примитивов движения. Каждый примитив – это рассчитанный для заданной робототехнической системы короткий участок траектории, удовлетворяющий кинематическим ограничениям движения. Набор всех возможных примитивов называется сеткой или решеткой примитивов (в англо-язычной литературе – *lattice*). На основе этой решетки в процессе планирования происходит эвристический перебор вариантов последовательного комбинирования примитивов до построения полной траектории, соединяющий стартовое и целевое положения. Обычно в литературе, посвященной такому подходу к планированию, основное внимание уделяется вопросу сокращения перебора (разработке различных эвристик, способов комбинации множества эвристических функций и пр.), при этом сама решетка примитивов считается заданной. В данной работе, мы рассматриваем обратную задачу – как повысить эффективность планирования траектории (при фиксированном алгоритме поиска) за счет вариации решетки примитивов. Для этого в работе разрабатываются различные наборы примитивов, проводится обширное экспериментальное исследование наборов, результаты которого используются для определения свойств примитивов положительно (и отрицательно) влияющих на качество планирования по шести количественным метрикам (включающим,

скорость работы алгоритма поиска, длину и кривизну отыскиваемых траекторий и др.).

2. Обзор литературы. Задаче планирования с учетом кинематических ограничений посвящено достаточно большое количество работ. В целом, методы решения этой задачи можно разделить на следующие классы: методы на основе сэмплирования, методы на основе оптимизации, методы на основе машинного обучения, методы на основе примитивов движения. Рассмотрим указанные классы методов более подробно.

2.1. Методы на основе сэмплирования. Широко используемым классом подходов является планирование на основе сэмплирования (sampling-based). При таком подходе планирование представляет собой стохастический процесс. Новые состояния на пути к целевому генерируются случайным образом, постепенно заполняя доступное пространство, в надежде однажды достигнуть цели. Несмотря на кажущуюся наивность такого подхода, подобные алгоритмы успешно применяются в различных автономных системах и прочно зарекомендовали себя в области планирования траектории. Популярным примером подобного алгоритма является алгоритм RRT [7] – Rapidly-exploring Random Trees. Его идея заключается в следующем: из стартового положения в случайном направлении ставится точка. Далее производится попытка соединить новое и старое положение с помощью метода, обеспечивающего соблюдение кинематических ограничений (т.н. локальное планирование). Если соединение удастся, то новое состояние добавляется в дерево состояний, если – нет, то состояние отбрасывается. Далее процесс повторяется, и таким образом строится дерево из состояний, которое постепенно покрывает доступное пространство и в достигает желаемой цели (с вероятностью 1 при числе сэмплов, стремящемся к бесконечности). Соблюдение кинематических ограничений достигается за счет использования особых методов соединения состояний. В таких методах могут применяться, например, кривые Ридса-Шеппа или кривые Дюбинса, которые за счет параметризации позволяют учесть ограничения на максимальный радиус поворота. Также могут использоваться и более продвинутые методы, т.н. функции руления (англ. steering function). В таких методах заложена кинематическая модель агента, которая используется для генерации траектории, соединяющий состояния в дереве. Примером подобного подхода служит метод POSQ [8].

Популярны различные модификации RRT. К примеру, RRT-connect [9] позволяет растить дерево одновременно из стартовой и

конечной точки, и прекращать поиск тогда, когда будет раскрыто состояние, принадлежащее обоим деревьям. Существуют и другие модификации [10 – 12], различным образом улучшающие базовый алгоритм.

Преимуществом сэмплирующих подходов является их простота, а также теоретическая возможность работы в непрерывном пространстве без необходимости дискретизации. С другой стороны, в практических задачах дискретизация зачастую возникает достаточно естественно, т.к. обычно робот строит именно дискретную модель представления окружающего пространства, а не непрерывную. Поэтому указанное теоретическое преимущество теряется. Более того из-за опоры на случайный выбор, алгоритм не гарантирует повторяемости результата, т.е. выдает разные решения одной и той же задачи. Также на практике существуют проблемы с подбором основного параметра алгоритма – числа сэмплов. При низком значении траектории отыскиваются быстро, но имеют низкое качество (большую длину, наличие большого числа поворотов и др.). При высоком значении параметра на отыскание траектории тратится слишком много времени. В целом задача подбора подходящего значения этого параметра не имеет общего решения.

2.2. Методы на основе оптимизации. Задача планирования траектории может быть математически сформулирована как задача оптимизации некоторой (непрерывной) функции – стоимости пути. Таким образом становится возможным использование оптимизационных подходов для планирования траектории. Подобные подходы в большинстве своем устроены следующим образом: на вход оптимизационному алгоритму дается начальная догадка – некоторая кривая, соединяющая начальную и конечную точку, далее производится попытка деформировать данную кривую так, чтобы она была оптимальной по заданным критериям. Для соблюдения кинематических ограничений, последние выражаются в аналитическом виде, например, в виде уравнений, которые определяют дополнительные ограничения оптимизационного процесса. Примеры таких алгоритмов представлены в работах [13 – 15]. Подобные подходы часто применяются при движении в хорошо структурированных средах, например на автодорогах общего пользования, когда в качестве начальной догадки берется середина дорожной полосы. В неструктурированной же среде вопрос формирования начальной догадки стоит острее. Зачастую в качестве такой догадки берется отрезок прямой между стартом и финишем, однако при такой начальной догадке методы оптимизации способны качественно строить траектории лишь при наличии препятствий простых форм (выпуклые многоугольники), которые на практике

встречаются довольно редко в неструктурированных средах. При сложной конфигурации препятствий процесс оптимизации зачастую сходится к локальному минимуму и не способен сгенерировать траектории, обходящие сложные препятствия.

Одним из существующих способов частичного решения подобной проблемы являются оптимизационные подходы, которые в качестве начальной догадки используют геометрическую траекторию, соединяющую стартовую и финишную точку в обход статических препятствий. Подобные методы также носят название пост-оптимизационных. Так, популярен метод пост-оптимизации GRIPS [16], или гибридный подход, который для второго этапа использует алгоритм Timed-Elastic Band (TEB) [17]. Методы пост-оптимизации так же, как и методы оптимизации сильно зависят от начальной догадки и при неудачном ее выборе не в состоянии найти какое-либо решение, даже если оно существует. Также большинство оптимизационных подходов требуют аналитического описания препятствий, что значительно затрудняет их применение в робототехнических системах, которые в основном используют клеточное представление об окружающем пространстве.

2.3. Методы на основе машинного обучения. В последнее время набирают популярность также методы основанные на машинном обучении. Данные подходы весьма разнообразны по архитектуре нейросетей, входным данным и параметрам. Часто используются алгоритмы [18 – 20] обучения с подкреплением (RL – Reinforcement learning), суть которых заключается во взаимодействии агента со средой и зависимости награды от действий агента. Отдельным классом являются так называемые end-to-end методы [21, 22], которые в качестве входных данных принимают наблюдения среды, а в качестве выходных данных выдают траекторию, или сразу управляющие воздействия. Учет кинематических ограничений осуществляется на этапе обучения – действия, которые ведут к их нарушениям штрафуются функцией потерь. Методы машинного обучения многообразны, и зачастую адаптированы под некоторую конкретную задачу, поэтому, к примеру, алгоритм движения по шоссе неприменим к алгоритму движения по парковке, что приводит к недостатку универсальности и сильной зависимости от входных данных у этих подходов.

2.4. Методы на основе примитивов движения. Но широко распространен и другой подход, при котором мы не адаптируем новые состояния агента под кинематические ограничения, а сразу генерируем дерево с учетом всех ограничений. Таким является подход с использованием примитивов движения. Его суть заключается в

построении траектории из отдельных фрагментов, – примитивов движения. Каждый такой примитив представляет собой дугу или прямолинейный участок, разработанный индивидуально для каждого набора кинематических ограничений. Проследовав любому из примитивов движения агент гарантированно удовлетворит всем заданным кинематическим ограничениям. Этот подход интересен тем, что мы заранее определяем наши возможные шаги в пространстве поиска, в противовес другим, например оптимизационным, алгоритмам, которые в свою очередь действуют в непрерывном пространстве, где поиск вариантов выполнять значительно сложнее. Плюс, мы не ограничиваем себя необходимостью проверять сконструированную траекторию на выполнимость – она при любом составе из примитивов будет удовлетворять всем кинематическим ограничениям, поскольку каждый элементарный примитив им удовлетворяет, а на стыке между двух примитивов гарантируется соблюдение граничных условий на гладкость необходимой степени.

Главным преимуществом является то, что с использованием примитивов движения мы можем выбрать любой эвристический алгоритм поиска – A^* , Theta^* , JPS (с некоторыми дополнительными ограничениями), TBA* и многие другие. Это возможно, поскольку мы не меняем принцип алгоритма, а меняем только способ генерации новых состояний. При этом даже с использованием таких простых алгоритмов, как A^* , мы можем применить наш подход к решению современных и достаточно сложных задач [1]. Описанным ранее подходам присущи различные недостатки: случайность в сэмплирующих подходах, отсутствие гарантированного решения (при его существовании) в оптимизационных подходах, зависимость от входных данных в подходах на основе машинного обучения. Эвристические подходы лишены всех этих минусов, и напротив обладают собственными преимуществами – хорошо исследованное математическое обоснование, широкий выбор модификаций, возможность использования множества различных эвристических функций. Поэтому использование эвристических подходов открывает широкие возможности для решения задачи кинематического планирования.

2.4.1. Методы генерации примитивов. Ключевым вопросом является выбор подходящих примитивов движения. Именно от качества и приемлемости заданных примитивов движения и будет зависеть результирующая траектория. Так, например, довольно популярными в последнее время стали подходы, основанные на имитации траекторий человека-водителя [24, 25]. В таких работах обычно анализируются реальные данные, на основе которых происходит выделение и отбор

основных примитивов движения. Многие работы посвящены генерации примитивов на основе оптимизационных подходов [26, 27], что открывает новые возможности, ограниченные только выбором целевой функции. Также очень распространенными являются подходы основанные на динамических [28, 29] и адаптивных [30, 31] примитивах движения. Обе группы подходов отличает то, что примитивы в них могут изменяться от итерации к итерации, подстраиваясь под состояние агента и текущую задачу. Отличие между динамическими и адаптивными подходами заключается в том, что первые меняют примитивы пропорционально, т.е. один и тот же примитив может быть в два-три раза длиннее или короче на каждой итерации, а вторые же могут полностью менять форму примитива, а не только его размер.

Перечисленные методы могут достаточно подробно описывать способ генерации примитивов. Однако наблюдается явный недостаток исследований, направленных на сравнение различных примитивов между собой. Также неясным остается вопрос о качествах генерируемых примитивов – какими они должны быть? Длинными или короткими? Их должно быть много в наборе или мало? Данное исследование направлено на устранение этого пробела.

2.5. Результаты анализа литературы. В настоящее время существуют различные способы решения задачи планирования с кинематическими ограничениями, каждый из которых обладает определенными недостатками и преимуществами. Одним из перспективных является декомпозиционный подход на основе примитивов движения. При этом в литературе наблюдается недостаток исследований, направленных на анализ применимости различных примитивов движения к конкретной задаче и методики сравнения их между собой. Именно решению этой актуальной задачи и посвящена данная работа.

3. Планирование траектории с кинематическими ограничениями. Рассмотрим (точечного) мобильного робота, чья модель движения в общем случае имеет вид $\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t))$, где $\mathbf{x}$ – состояние, $\mathbf{u}$ – управление, f – вектор-функция. Здесь и далее будем акцентировать свое внимание на колесном роботе с дифференциальным приводом. В этом случае имеем: $\mathbf{x}(t) = (x, y, \theta)$, $\mathbf{u}(t) = (v, \omega)$, а уравнения кинематики материальной точки представляются как:

$$\dot{x} = v \cdot \cos(\theta) \quad \dot{y} = v \cdot \sin(\theta) \quad \dot{\theta} = \omega. \quad (1)$$

Примитивом движения $\mathbf{p}$ назовем кортеж функций $(x(t), y(t), \theta(t))$, $t \in [0, t_f]$, где $t_f > 0$ – продолжительность примитива, которые удовлетворяют модели движения (1). Пусть $\hat{t} \in [0, t_f]$ – некоторый момент времени, тогда состояние робота в этот момент времени при совершении примитива $\mathbf{p}$ обозначим как $\mathbf{p}(\hat{t})$: $\mathbf{p}(\hat{t}) = (\hat{x}, \hat{y}, \hat{\theta})$, где $\hat{x} = x(\hat{t})$, $\hat{y} = y(\hat{t})$, $\hat{\theta} = \theta(\hat{t})$. Геометрически каждый примитив представляется как ограниченный (по длине) сегмент некоторой кривой. Пример изображен на рисунке 2.

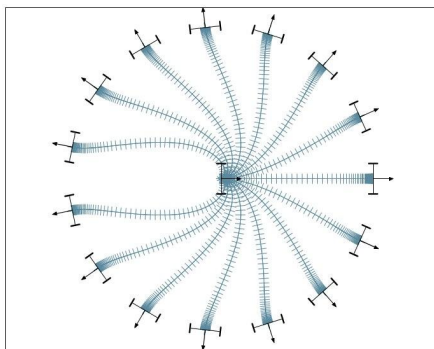


Рис. 2. Пример примитивов из работы [8]

С каждым примитивом движения будем ассоциировать действительное положительное число – стоимость (выполнения) примитива: $Cost(p) \in \mathbb{R}$. Чем меньше стоимость, тем предпочтительней примитив движения. В общем случае функция стоимости может быть произвольной и отражать любые пользовательские требования к задаче планирования. Самой распространенной функцией стоимостью является длина примитива. Также функция стоимости может учитывать кривизну, энергозатраты на выполнение примитива и пр.

Траекторией π назовем такую конечную последовательность примитивов, $\pi = \{p^i | i \in (0, N)\}$, $\pi[i] = \mathbf{p}^i$, для которой выполняются условия:

$$\mathbf{p}^i(t_f^i) = \mathbf{p}^{i+1}(0), \quad \dot{\mathbf{p}}^i(t_f^i) = \dot{\mathbf{p}}^{i+1}(0), \quad \forall i \in (0, N-1), \quad (2)$$

подразумевающие совпадение состояний и скоростей на стыке каждого двух последовательно идущих примитивов (другими словами, каждый

последующий примитив начинается из того же состояния, в котором заканчивается предыдущий).

Стоимостью траектории назовем сумму стоимостей входящих в нее примитивов $\text{Cost}(\pi) = \sum_i \text{Cost}(\mathbf{p}^i)$.

Определим далее рабочее пространство, в котором перемещается робот, как замкнутое подмножество на плоскости, $W \in \mathbb{R}^2$, состоящие из проходимых и непроходимых областей (препятствий): $W = W_{free} \cup W_{obst}$.

Задачей планирования траектории назовем кортеж $S = \langle W, \mathbf{x}_0, \mathbf{x}_f \rangle$, где $\mathbf{x}_0, \mathbf{x}_f$ – это начальное и целевое состояние робота соответственно. Без ограничения общности будем считать, что координаты начального и целевого положения принадлежат свободному пространству (в противном случае задача планирования считается некорректно поставленной).

Решением задачи планирования является траектория π , обладающая следующими свойствами:

$$\pi0 = \mathbf{p}^0(0) = \mathbf{x}_0, \quad (3)$$

$$\pi[N](t_f^N) = \mathbf{p}^N(t_f^N) = \mathbf{x}_f, \quad (4)$$

$$\forall i \in (0, N), \forall t \in (0, t_f^i) : \mathbf{p}^i(t) \in W_{free}, \quad (5)$$

то есть начало траектории совпадает с начальным состоянием, конец траектории с конечным состоянием, а вся траектория содержится полностью в области свободного пространства (не пересекает препятствия).

Заметим, что в общем случае решение задачи планирования может быть не единственным. Обозначим множество всех решений задачи планирования как Π . Оптимальным решением задачи планирования называется траектория π^* , для которой выполняется условие:

$$\forall \pi \in \Pi : \text{Cost}(\pi^*) \leq \text{Cost}(\pi).$$

Другими словами, ни одна другая траектория из заданного начального состояния в заданное конечное не имеет меньшей стоимости.

4. Исследуемая задача. Будем считать, что в нашем распоряжении имеется алгоритм, *Alg*, получения оптимального решения произвольной (корректно определенной) задачи планирования для заданной робототехнической системы. Формально алгоритм принимает на вход задачу планирования (описание рабочей области робота, начального

и целевого состояния), а также некоторый набор примитивов, из которых допускается конструирование траектории: $Alg(S, \mathbf{pr})$. Здесь $\mathbf{pr} = \{\mathbf{p}\}$ – это набор примитивов движения, удовлетворяющих кинематическим ограничениям (1). На выходе алгоритм возвращается оптимальную траекторию π^* .

Обычно работы по планированию траектории посвящены вопросам конструирования алгоритма Alg , исследования его теоретических свойств (доказательство сходимости, оптимальности), повышения вычислительной эффективности и так далее. С практической же точки зрения интерес представляет также изучение влияния второго аргумента, $\mathbf{pr}$ на эффективность работы Alg .

Для формализации этой задачи, которой и посвящена статья, рассмотрим конечную совокупность наборов примитивов: $\mathbf{pr}_1, \mathbf{pr}_2, \dots, \mathbf{pr}_n$, каждый из которых может быть использован для получения оптимального решения задачи планирования (с помощью алгоритма Alg , который считается фиксированным). Введем также в рассмотрение набор числовых функций (метрик), ставящих в соответствие траектории некоторое число: $M = \{m_i\}$, $m_i(\pi) \in \mathbb{R}$. Каждая такая функция количественно отражает определенное качество траектории – степень кривизны, среднее расстояние до препятствий и т.д. Без ограничения общности будем считать, что время получения траектории, т.е. время работы алгоритма Alg , также относится к метрикам. Теперь для произвольной задачи планирования S и фиксированного набора примитивов из рассматриваемой совокупности, $\mathbf{pr}_i$, мы можем получить количественную оценку качества решения задачи планирования в виде набора числовых значений (метрик). Проводя анализ этих значений для разных наборов примитивов и на разных задачах планирования (при фиксированном Alg), мы стремимся выявить зависимость между видом используемых примитивов (количество примитивов в наборе, их длина, форма и пр.) и качеством решения задачи планирования. Нашей конечной целью является выработка общих рекомендаций (по внешнему виду примитивов, их количеству, длине, кривизне и пр.), которые необходимо учитывать исследователям и разработчикам, использующим методы планирования на основе примитивов движения для решения задач планирования траектории движения колесного мобильного робота с дифференциальным приводом (т.е. робота, движение которого описывается (1)).

Таким образом, рассматриваемая в данной работе задача заключается в выявлении влияния характеристик примитивов движения на эффективность работы алгоритма планирования траектории

с кинематическими ограничениями, а также в выработке общих рекомендаций по конструированию примитивов движения.

5. Методы. В данном разделе рассмотрены методы и подходы, необходимые для решения поставленной выше задачи, а именно: алгоритм генерации примитивов, метод планирования траектории, а также методика проведения анализа зависимости эффективности работы алгоритма планирования траектории от характеристик примитивов движения.

5.1. Метод генерации примитивов. Для генерации примитивов, учитывающих кинематические ограничения (1), использовался метод из [32]. Будем на него ссылаться как на *GenPrim*. Этот метод основан на принципе накрытия для решения краевых задач для плоских систем управления (подробнее – в оригинальной публикации). Он подразумевает получение примитивов движения в виде кривых третьего порядка:

$$x(t) = \alpha_0 \cdot t^3 + \alpha_1 \cdot t^2 + \alpha_2 \cdot t + \alpha_3, \quad (6)$$

$$y(t) = \beta_0 \cdot t^3 + \beta_1 \cdot t^2 + \beta_2 \cdot t + \beta_3, \quad (7)$$

$$\theta(t) = \arctan\left(\frac{\dot{y}}{\dot{x}}\right), \quad (8)$$

$$t \in [0, t_f]. \quad (9)$$

Здесь коэффициенты α_i , β_i определяются *GenPrim*, а продолжительность примитива t_f задается пользователем.

Для создания примитивов использовалась программная реализация *GenPrim* на языке Matlab. В качестве входных данных передавалась желаемая финальная точка кривой (начальная точка всегда находится в начале координат), ориентация робота в начале и в конце движения, а также предельное значение радиуса кривизны. Все эти параметры позволяли учитывать модель робота при разработке примитивов.

Изначально при генерации примитивов стартовая ориентация робота считалась равной нулю градусов. После генерации примитивов из этой ориентации полученный набор поворачивался на плоскости с некоторым, заранее определенным, шагом. В итоге получалась роза примитивов, покрывающих определенные стартовые ориентации агента. Финальные ориентации примитивов подбираются так, чтобы в итоге выйти в одну из допустимых стартовых ориентаций, т.е. дискретизация шага финальных ориентаций совпадала с дискретизацией шага стартовых ориентаций. Поскольку в робототехнике наиболее часто используются графы-сетки (occupancy grid) для моделирования окружающего робота пространства, то мы в своей работе также опирались на это представление.

Таким образом при генерации примитивов их длины подбирались так, чтобы соединять центры клеток в графе-сетке. Пример набора примитивов изображен на рисунке 3.

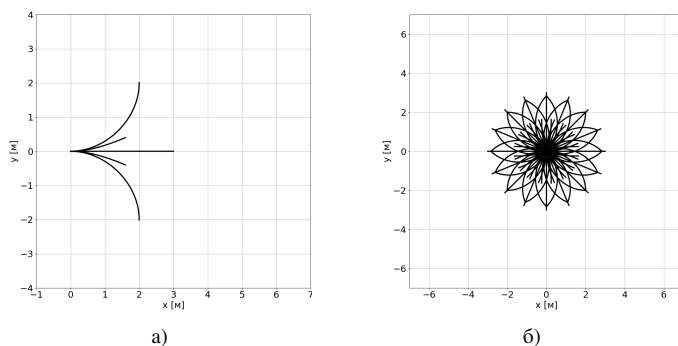


Рис. 3. Пример примитивов движения: а) примитивы для одной ориентации; б) роза примитивов

Для каждого построенного примитива сохранялась информация о клетках графа-сетки, которые накрывает робот при движении по данному примитиву. Это необходимо для дальнейшего определения столкновений с препятствиями при планировании всей траектории. Визуально пример такого набора клеток можно увидеть на рисунке 4.

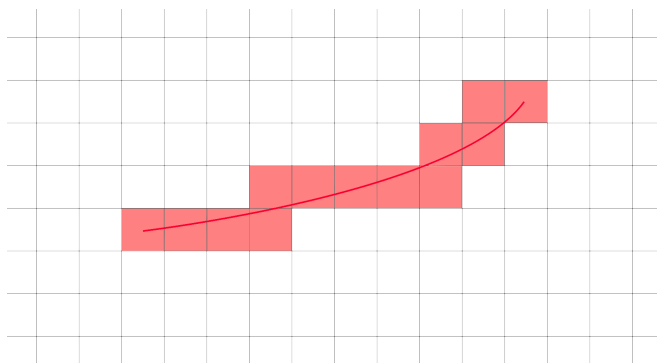


Рис. 4. Демонстрация множества клеток, накрываемого одним примитивом

Для расчета множества клеток использовался python-скрипт, код которого представлен в листинге 1. Здесь функция *calc_sweep_cells* принимает на вход отдельный примитив, далее увеличивая время t от 0 до t_f регистрируется положение агента вдоль кривой и вычисляется

текущая клетка пространства. Каждая клетка несет в себе целочисленные координаты, поэтому при разных значениях t агент может находиться в пределах одной и той же клетки. Далее обновляется список перекрываемых клеток, содержащихся в классе примитива.

```
def calcSweepCells(prim)
    t_init = t = 0
    t_final = primitive.getFinalTime()
    cells = []
    while t != t_final:
        cell = prim.getOccupiedCell(t)
        if cell not in cells:
            cells.append(cell)
        t = t + dt
    return cells
```

Листинг 1. Алгоритм расчета перекрываемых клеток

Возникает вопрос о расчете стоимости отдельного примитива. В качестве базового решения может использоваться только общая длина. Однако эксперименты показали, что в таком случае траектории могут быть «виляющими» – это объясняется тем, что планирование ведется исключительно из соображения приблизиться к цели за как можно меньшее число итераций. Поэтому было принято решение об использовании модифицированной функции стоимости. В нашем случае функция стоимости представляет собой линейную комбинацию из длины примитива и его средней кривизны. Длина примитива находится из параметрических уравнений на x и y с помощью численного интегрирования:

$$L(p) = \int_0^{t_f} \sqrt{\dot{x}^2(t) + \dot{y}^2(t)} dt.$$

Для нахождения средней кривизны примитив разбивается на множество точек с некоторым шагом по времени Δt . В нашем случае этот шаг составлял 0.05 с. Далее считается кривизна в каждой полученной точке, не считая крайних. Напомним, что математически кривизна считается как обратный радиус кривизны траектории. После нахождения кривизны в каждой точке, находится средняя кривизна как среднее арифметическое. Таким образом:

$$NC(p) = \frac{1}{I-2} \cdot \sum_1^{I-1} \kappa_i = \frac{1}{I-2} \cdot \sum_1^{I-1} \frac{1}{R_i},$$

где I – количество промежуточных точек примитива, κ_i – кривизна i -й точки, R_i – радиус кривизны траектории в i -й точке.

Стоимость примитива тогда находится как:

$$Cost(p) = \alpha_1 \cdot L(p) + \alpha_2 \cdot NC(p),$$

где α_1, α_2 – линейные коэффициенты. Изменяя величину коэффициентов можно менять вклад той или иной составляющей в общую стоимость примитива. В рамках данной работы предлагается больший вклад уделить длине и меньший кривизне траектории. Как будет показано далее, характерное значения длины примитива $2m$, а кривизны $0.5 m^{-1}$. Таким образом достаточно взять единичные коэффициенты, чтобы увеличить вклад линейной составляющей. Тогда стоимость примитива:

$$Cost(p) = L(p) + NC(p).$$

5.2. Метод планирования траектории. В качестве эвристического алгоритма планирования был выбран классический алгоритм A^* . Он прост в реализации, удобен в использовании и позволяет добавить различные модификации, об одной из которых пойдет речь в разделе . В случае использования примитивов движения требуется изменить алгоритм генерации новых состояний, в частности с использованием предрасчитанных перекрываемых клеток указанных выше. В листинге 2 приведен псевдокод алгоритма генерации новых состояний. Функция *get_successors* принимает на вход состояние, из которого нужно сгенерировать потомков, а на выходе выдает список допустимых потомков. Для этого состояние передается в функцию *get_prims*, которая возвращает список допустимых примитивов, которые могут следовать из этого состояния. Примитивы отсеиваются по стартовому углу (он должен совпадать с ориентации в состоянии) и по коллизиям с препятствиями, которые находятся на карте около этого состояния. После выполнения этой функции из допустимых примитивов генерируются новые допустимые состояния (потомки) и, если ранее они не были исследованы, добавляются в итоговый список.

```
def getSuccessors(node):
    successors = []
    prims = getPrims(node)
    for prim in prims:
        new_node = genNode(node, prim)
        if new_node not in Closed:
            successors.append(new_node)
    return successors
```

```
def getPrims(node):  
    prims = getPrimsByAngle(node.orientation)  
    for prim in prims:  
        if any prim.cell is in Map.obstacles:  
            delete prim from prims  
    return prims
```

Листинг 2. Алгоритм генерации новых состояний

Во всех остальных аспектах алгоритм полностью повторяет классический алгоритм А*. Результатом работы алгоритма является траектория π , состоящая из множества примитивов.

5.3. Методика проведения анализа. Задача выявления влияния характеристик примитивов движения на эффективность работы алгоритма планирования траектории может быть решена путем эмпирического исследования различных наборов примитивов движения и сравнения результатов работы алгоритма по совокупности количественных показателей. Такими показателями являются метрики качества – численные функции, отображающие траекторию на множество действительных чисел. Опишем их более подробно.

Метрики. Для оценки качества работы алгоритмов требуется рассчитать различные количественные метрики. Существуют разнообразные наборы данных и методик сравнения алгоритмов планирования [33 – 35], из которых можно выделить и использовать метрики качества. Нами был выбран широко известный и общепризнанный набор Bench-MR [36], который направлен на тестирование алгоритмов в неструктурированной среде, содержащей только препятствия и испытуемого агента. Как будет показано в следующем разделе, данные условия полностью соответствуют нашей постановке, что и обосновывает сделанные выбор.

Для расчета каждой метрики требуется представить траекторию в виде последовательности точек (состояний). Поскольку в нашей постановке траектория представляет собой последовательность примитивов, требуется разбить каждый примитив на последовательность точек. Так как примитивы представляют собой параметрические уравнения, сделать это можно элементарно с помощью подстановки различных значений времени в выражения соответствующего примитива. Тогда для всей траектории последовательность точек представляет собой объединение последовательностей точек каждого из включенных примитивов движения.

Для дальнейшего пояснения введем обозначения: s_i – i -й элемент траектории, содержащий состояние (координаты и ориентация) агента, $d(s_i, s_j)$ – евклидово расстояние между состояниями, $K(s_{i-1}, s_i, s_{i+1})$ – кривизна участка траектории между тремя точками, считается с помощью подгонки окружности по трем точкам и последующим делением единицы на ее радиус. Также I – это количество промежуточных точек во всей траектории. Тогда перечень метрик:

- Длина траектории (Path length, LEN) – считается как:

$$LEN = \sum_{i=0}^{I-1} d(s_i, s_{i+1}).$$

- Расстояние до препятствий (Clearing) – показывает среднее расстояние траектории до препятствий. Для расчетов используется предрасчитанная матрица расстояний, которая по размерам совпадает с тестовой картой, а в каждой клетке которой записано кратчайшее расстояние до препятствия из этой клетки. Тогда метрика считается как:

$$Clearing = \frac{\sum_{i=1}^I DistanceMatrix(s_i)}{I}.$$

- Угол-на-траекторию (Angle-over-length, AOL) – показывает в среднем насколько сильно меняется ориентация вдоль траектории, считается как:

$$AOL = \frac{\sum_{i=0}^{I-1} abs(s_i(\theta) - s_{i+1}(\theta))}{LEN}.$$

- Средняя кривизна (Normalized curvature, NormC) – показывает среднюю кривизну, т.е. средний радиус поворота вдоль траектории, считается как:

$$NormC = \frac{\sum_{i=1}^{I-1} K(s_{i-1}, s_i, s_{i+1})}{I}.$$

Помимо этих метрик вводилась также дополнительная метрика отвечающая времени работы алгоритма – *ComputationTime*.

Методика сравнения. Сами по себе метрики качества не позволяют выполнить сравнение между различными наборами примитивов движения. Требуется определить правила упорядочивания результатов сравнения в зависимости от значения показателя. В соответствии с использованными метриками существует единственное правило, по которому меньшее

значение метрики соответствует лучшему результату. Действительно, длина траектории, время работы алгоритма, средняя кривизна и т.д. – показатели, значение которых ожидается на минимальном уровне.

Сравнение различных примитивов движения выполнялось следующим образом:

1. На нескольких картах генерировалось множество различных заданий;
2. Для каждого сравниваемого набора примитивов проводился экспериментальный запуск алгоритма планирования для каждого из заданий;
3. По полученным траекториям выполняется подсчет метрик качества;
4. Для каждого набора примитивов на каждой карте подсчитывалось среднее из 100 значений метрик;
5. Полученные результаты заносились в таблицу и упорядочивались в зависимости от оцениваемого показателя.

Далее будет показано, что различные примитивы имеют преимущества по различным метрикам качества. Поэтому сравнивать наборы примитивов движения можно только по совокупности полученных результатов.

На основе выполненного сравнения строятся общие рекомендации ко свойствам генерируемых примитивов движения, следование которым должно привести исследователей к получению лучших результатов при решении задачи планирования траектории с кинематическими ограничениями.

6. Входные данные экспериментов. Для тестирования использовались две клеточные карты: первая (рисунок 5(а)) представляла собой свободное пространство, заполненное различными небольшими статическими препятствиями, вторая (рисунок 5(б)) моделировала реальную среду с препятствиями различного размера, в т.ч. зданиями. Каждая карта имела размер 600×600 клеток (пикселей), разрешение составляло 0.2 метра на клетку.

Было сгенерировано 100 различных заданий для каждой карты. Каждое задание представляло собой состояние (x, y, θ) . Гарантируется выполнение каждого из заданий по отдельности для точечного дифференциального робота. Также были проведены дополнительные тесты на пустой карте, где 45 заданий были равномерно распределены по окружности вокруг агента, с ориентациями направленными от центра наружу.

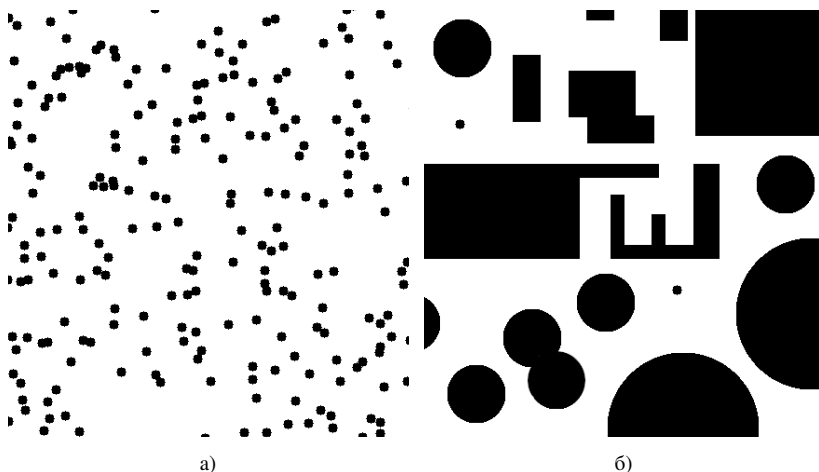


Рис. 5. Карты, используемые для тестирования: а) карта 1; б) карта 2

Для генерации примитивов был выбран шаг по ориентации в 22.5 градуса ($\frac{\pi}{8}$ радиан). Также для упрощения анализа скорость агента считается постоянной и равной 1 м/с. Для ускорения работы алгоритма была добавлена искусственная погрешность в 1 клетку для нахождения решения. Таким образом алгоритм может найти состояние не в точности совпадающее с заданным финальным, а отличающимся от него на эту погрешность. Также была добавлена погрешность в 11.25 градусов по ориентации, но это обусловлено только случайной генерацией ориентации финальных точек, а поскольку финальные ориентации примитивов дискретизованны, необходимо добавить погрешность размером 1/2 от шага в 22.5 градуса, чтобы гарантировать решение задачи.

Было сгенерировано 3 версии примитивов, на основе которых можно было делать выводы о связи параметров этих примитивов с качеством построенных траекторий. Все эти версии изображены на рисунке 6. Слева изображены примитивы для одной ориентации, справа – роза для всех ориентаций. Масштаб выбран единый для всех изображений, чтобы примитивы можно было сравнить визуально. Здесь и далее, если указывается угол у примитива, имеется ввиду его изменение по модулю от первоначального угла, т.е. если примитив стартует из ориентации $\frac{\pi}{8}$ и заканчивается с ориентацией $-\frac{\pi}{4}$, то соответствующее ему обозначение будет равно $abs(-\frac{\pi}{4} - (\frac{\pi}{8})) = \frac{3\pi}{8}$.

В качестве первоначальной догадки были выбраны примитивы, изображенные на рисунке 6(а). На картинке слева видны примитивы

для одной ориентации агента. Для каждой ориентации насчитывается 6 примитивов. Классификация примитивов ведется по их финальной ориентации, таким образом в этом наборе присутствуют:

- 2 прямолинейных примитива длинами 0.4 и 3 метра,
- 2 примитива с финальными ориентациями $\frac{\pi}{8}$,
- 2 примитива с финальными ориентациями $\frac{\pi}{2}$.

Роза примитивов изображена на рисунке 6(б). Всего в наборе $6 \times 16 = 96$ примитивов.

Дальнейшие версии примитивов получались путем модификации первой версии – увеличении длины примитивов, количества примитивов, формы и т.д.

На рисунке 6(в) изображены примитивы второй версии. Их основное отличие в дополнительных кривых, цель добавления которых заключается в проверке гипотезы о том, как вариативность влияет на качество траекторий. Таким образом в данном наборе добавлены примитивы:

- Для движения по прямой со смещением;
- С другими выходными ориентациями $\frac{\pi}{4}$ и $\frac{3\pi}{8}$;
- Выходящие в ту же точку, что и примитив с ориентацией $\frac{\pi}{2}$, но с другими ориентациями – $\frac{3\pi}{8}$ и $\frac{\pi}{4}$.

Роза примитивов изображена на рисунке 6(г). Всего в наборе $17 \times 16 = 272$ примитива.

Далее на рисунке 6(д) изображены примитивы третьей версии. Они отличаются от примитивов первой версии как увеличенными размерами, так и количеством примитивов и выходными ориентациями. Это позволит проверить, в первую очередь, как длина примитивов влияет на качество траекторий. Таким образом в наборе:

- Примитивы $\frac{\pi}{8}$ обладают большей длиной;
- Дополнительные прямолинейные примитивы длинами 1 и 6 м;
- Дополнительные примитивы с выходными ориентациями $\frac{\pi}{4}$ и $\frac{3\pi}{8}$.

Роза примитивов изображена на рисунке 6(е). Всего в наборе $14 \times 16 = 224$ примитива.

Также для наглядности была составлена таблица 1, содержащая информацию о примитивах в каждом наборе. В каждой строке таблицы указывается длина отдельного примитива, а в каждом столбце его финальная ориентация. На пересечении строки и столбца при наличии в наборе примитива с заданными параметрами ставится значок, соответствующий одному из трех наборов. Соответствие значков указано в описании к таблице.

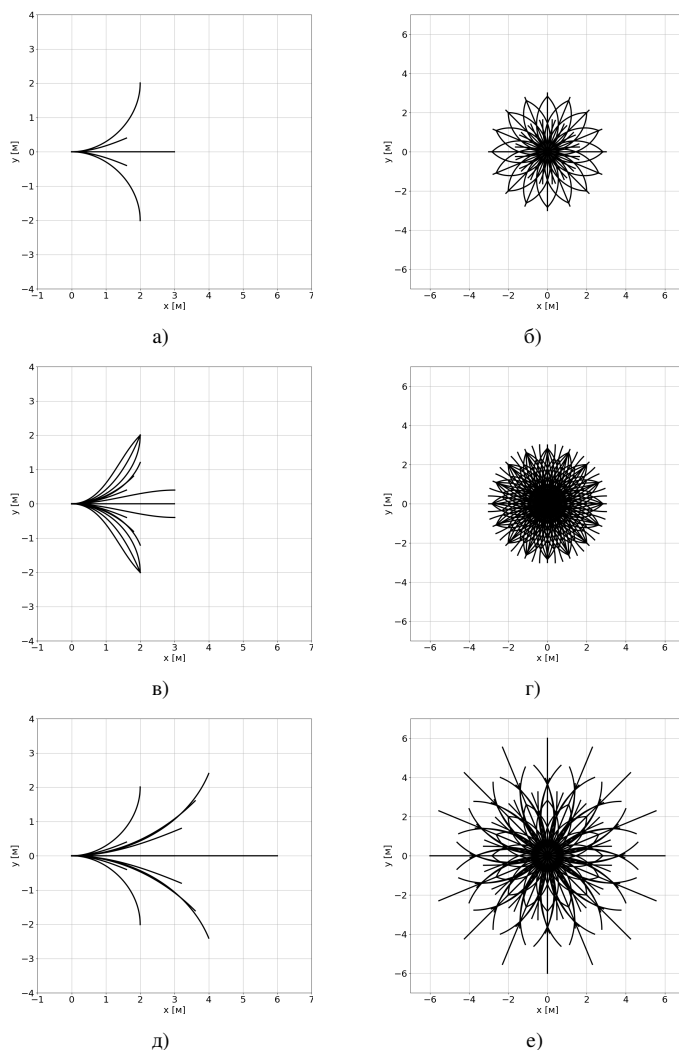


Рис. 6. Прimitives, используемые в экспериментах: а) primitives 1; б) роза primitives 1; в) primitives 2; г) роза primitives 2; д) primitives 3; е) роза primitives 3

Были проведены все тесты на всех картах и primitives, и по итогу были получены величины метрик, которые были сведены в таблицы и графики.

Таблица 1. Сводная таблица с характеристиками примитивов в используемых наборах. Здесь символ в каждой ячейке обозначает наличие примитива с заданной длиной и результирующей сменой угла в соответствующем наборе.

Введены обозначения: ▲ – 1-й набор, ■ – 2-й набор, ★ – 3-й набор

Угол \ Длина (м)	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$
0.400	▲■★				
1.000	■★				
1.660		▲■★			
2.020			■		
2.467				■	
2.935			■		
2.998				■	
3.000	▲■★				
3.032	■				
3.117					▲■★
3.320		★			
4.041			★		
4.934				★	
6.000	★				

7. Результаты экспериментов

7.1. Анализ результатов. В ходе экспериментов были получены численные значения метрик из раздела 5.3, но для наглядности предлагается рассмотреть значения относительно результатов первых примитивов.

Согласно полученным результатам (таблица 2) видно, что вторая и третья версия примитивов находят решение примерно в 3 раза дольше, чем первый примитивы.

Таблица 2. Результаты проведенных экспериментов на обеих картах с использованием трех видов примитивов. В каждой ячейке указано итоговое значение метрики относительно первой версии примитивов. Жирным шрифтом выделены лучшие результаты в каждой метрике на каждой карте

Номер примитивов	Карта 1			Карта 2		
	I	II	III	I	II	III
Длина пути	1.0	0.993	0.994	1.0	0.986	0.985
Время работы	1.0	3.157	2.829	1.0	3.149	2.531
Угол-на-длину	1.0	0.983	1.007	1.0	0.959	1.001
Расстояние до препятствий	1.0	1.015	1.011	1.0	1.165	1.196
Средняя кривизна	1.0	0.983	1.006	1.0	0.961	1.001

При этом отклонение по остальным метрикам на первой карте составило порядка 1%, и порядка 2-4% на второй карте. Такая большая разница во времени работы свидетельствует о том, что любое увеличение количества примитивов приводит к существенному повышению времени работы алгоритма. Это лишний раз подтверждается тем, что вторые примитивы оказались самыми ресурсозатратными, поскольку имели в своем наборе наибольшее число примитивов. Для проверки связи между временем работы и количеством примитивов были проведены дополнительные эксперименты на пустой карте без препятствий. Стартовое состояние находилось в центре карты с ориентацией вниз, а задания в количестве 45 штук были равномерно распределены по окружности радиусом 75 м с ориентациями направленными радиально наружу. Если посмотреть на траектории, построенные на пустой карте примитивами 1 и 2 (рисунок 7), то можно заметить, что итоговые траектории очень похожи, что свидетельствует о применении одинаковых примитивов. Таким образом примитивы, добавленные во второй версии лишь увеличивают фактор ветвления поиска, но по сути не изменяют траектории.

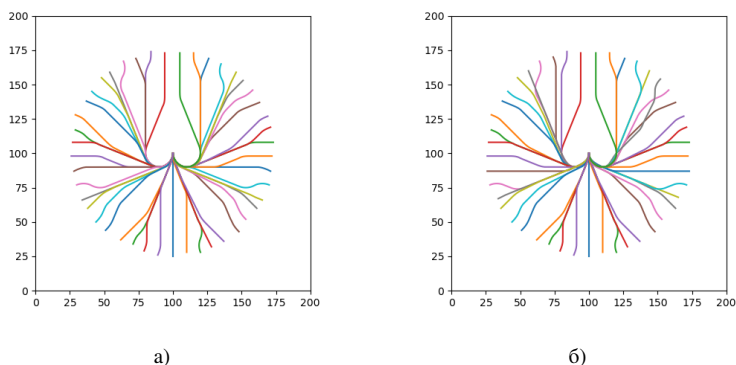


Рис. 7. Траектории, построенные с помощью первых и вторых примитивов на пустой карте: а) примитивы № 1; б) примитивы № 2

Рассмотрим значения остальных метрик. Результаты на карте 1 и карте 2 демонстрируют схожие тенденции, поэтому далее приводятся графики только для значений метрик на первой карте.

На графике распределения длин траекторий (рисунок 8(а)) можно заметить, что длины траекторий для всех трех версий практически совпадают, и лишь незначительно уменьшаются с увеличением средней длины примитивов.

Достаточно неожиданными оказались результаты по метрике Clearing. Если взглянуть на графики (рисунок 8(б)), то можно сделать вывод, что большее разнообразие примитивов приводит к менее аккуратным траекториям.

Метрики средней кривизны (рисунок 8(в)) и угла-на-длину (рисунок 8(г)) наоборот демонстрируют большую похожесть между примитивами. Для того, чтобы это объяснить, обратимся к статистике по примитивам.

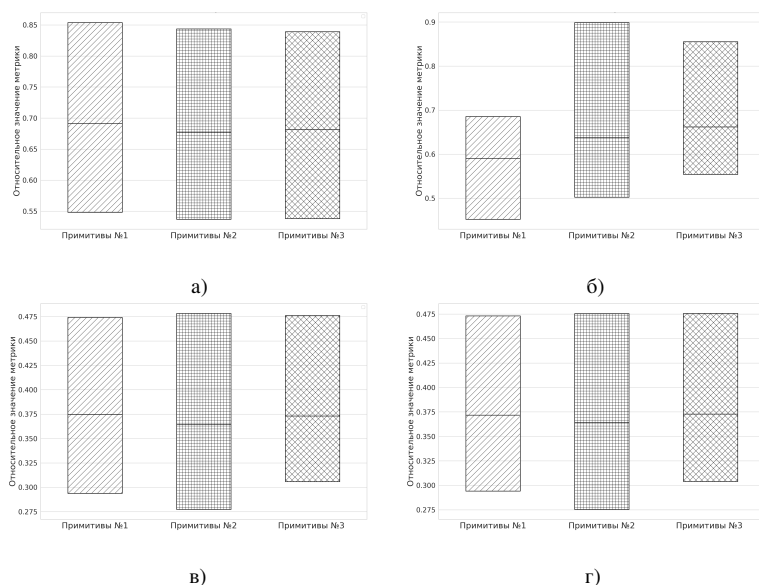


Рис. 8. Сводные графики метрик по результатам экспериментов на первой карте:
а) длина пути; б) среднее расстояние до препятствий; в) средняя кривизна;
г) угол-на-длину

Статистика использования примитивов состояла в трех сравнениях: соотношение между прямолинейными и криволинейными примитивами, соотношение между типами прямолинейных примитивов (по длине), и между типами криволинейных примитивов (по финальному углу). Полученные графики для обеих карт демонстрируют одинаковые тенденции, поэтому продемонстрируем рисунки только для первой карты (рисунок 9). Из графиков видно, что среди криволинейных примитивов всегда преобладает самый малый примитив с поворотом на $\frac{\pi}{8}$. Поэтому общая кривизна траектории равно как и изменения угла на единицу

длины обусловлены именно им. Так что нет ничего удивительного, что графики кривизны практически совпадают между первыми и вторыми примитивами, и лишь немного отличаются для третьих примитивов.

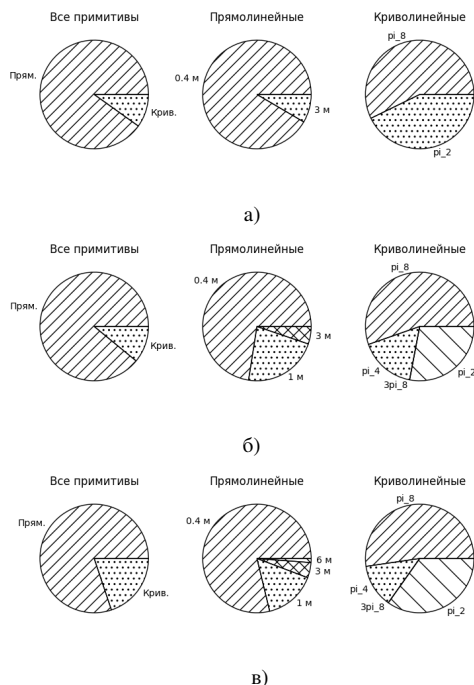


Рис. 9. Статистика использования примитивов: а) примитивы № 1; б) примитивы № 2; в) примитивы № 3

Также видно, что прямолинейные примитивы используются среди чаще криволинейных, причем предпочтение отдается опять же более коротким примитивам.

8. Заключение. В работе рассматривалась задача планирования траектории для мобильного робота с учетом кинематическим ограничений, в частности рассматривался случай колесного мобильного робота с дифференциальным приводом. Для решения этой задачи использовался подход планирования на основе примитивов движения. Основное внимание уделялось вопросу выявления взаимосвязей между характеристиками используемых примитивов и качеством решения задачи планирования траектории в неструктурированной среде. Качество решения оценивалось по шести различным метрикам. На основе более

двухсот заданий было произведено сравнение трех различных версий примитивов движения.

Анализ результатов проведенных экспериментальных исследований позволяет сделать следующие выводы:

1. Увеличение числа примитивов ведет к существенному (кратному) увеличению времени работы алгоритма, при этом улучшение качественных метрик (длина траектории, кривизна и пр.) – незначительно.

2. Увеличенная длина примитивов в наборе приводит к уменьшению общей длины траектории на 5–15%.

3. Прямолинейные примитивы используются гораздо чаще криволинейных, что объяснимо, поскольку поворотов в среднестатистической траектории гораздо меньше прямых участков.

4. Среди криволинейных примитивов самыми часто используемыми оказались кривые с малым углом поворота, при примерно равномерном распределении остальных типов.

Эти выводы позволяют сформировать следующий набор рекомендаций к набору примитивов для задач планирования с кинематическими ограничениями для колесного робота с дифференциальным приводом:

1. Количество примитивов на шаг угла составляет порядка 6–10 штук.

2. Преимущество при генерации отводится более коротким, маневренным примитивам.

3. Возможно добавление малого (1–2) числа удлинённых примитивов для получения более плавных траекторий на открытом пространстве.

Учет рекомендаций необходим для обеспечения наиболее качественного результата при планировании траектории с учетом кинематических ограничений.

Литература

1. Дудакова Д.С., Анохин В.М., Дудаков М.О., Ронжин А.Л. О теоретических основах аэролимнологии: изучение пресных водоемов и прибрежных территорий с применением воздушных робототехнических средств // Информатика и автоматизация. 2022. Т. 21. № 6. С. 1359–1393.
2. Балабанов А.Н., Безуглая А.Е., Шушляпин Е.А. Управление манипулятором подводного робота // Информатика и автоматизация. 2021. Т. 20. № 6. С. 1307–1332. DOI: 10.15622/ia.20.6.5.
3. Макаров Д.А., Панов А.И., Яковлев К.С. Архитектура многоуровневой интеллектуальной системы управления беспилотными летательными аппаратами // Искусственный интеллект и принятие решений. 2015. № 3. С. 18–33.
4. Otsu K., Matheron G., Ghosh S., Toupet O., Ono M. Fast approximate clearance evaluation for rovers with articulated suspension systems // Journal of Field Robotics.

2020. vol. 37. no. 5. pp. 768–785.
5. Al Mashhadany Y.I. Design and analysis of 7-DOF human-link manipulator based on hybrid intelligent controller // *Informatics and Automation*. 2020. vol. 19. no. 4. pp. 774–802.
6. Пшихопов В.Х., Медведев М.Ю. Планирование движения группы подвижных объектов в двумерной среде с препятствиями // *Известия Южного федерального университета. Технические науки*. 2016. Т. 2(175). С. 6–22.
7. LaValle S.M. Rapidly-exploring random trees: A new tool for path planning. *Research Report 9811*. 1998.
8. Kuffner J.J., LaValle S.M. RRT-connect: An efficient approach to single-query path planning // *Proceedings ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings*. 2000. vol. 2. pp. 995–1001.
9. Sharma P., Gupta A., Ghosh D., Honkote V., Nandakumar G., Ghose, D. PG-RRT: A Gaussian Mixture Model Driven, Kinematically Constrained Bi-directional RRT for Robot Path Planning // *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2021. pp. 3666–3673.
10. Webb D.J., Van Den Berg J. Kinodynamic RRT*: Asymptotically optimal motion planning for robots with linear dynamics // *IEEE international conference on robotics and automation*. IEEE. 2013. pp. 5054–5061.
11. Gammell J.D., Srinivasa S.S., Barfoot T.D. Informed RRT: Optimal sampling-based path planning focused via direct sampling of an admissible ellipsoidal heuristic // *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2014. pp. 2997–3004.
12. Zhang Y., Sun H., Zhou J., Pan J., Hu J., Miao J. Optimal vehicle path planning using quadratic optimization for baidu apollo open platform // *IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2020. pp. 978–984.
13. Li B., Wang K., Shao Z. Time-optimal maneuver planning in automatic parallel parking using a simultaneous dynamic optimization approach // *IEEE Transactions on Intelligent Transportation Systems*. 2016. vol. 17. no. 11. pp. 3263–3274.
14. Karlsson J., Murgovski N., Sjöberg J. Computationally efficient autonomous overtaking on highways // *IEEE Transactions on Intelligent Transportation Systems*. 2019. vol. 21. no. 8. pp. 3169–3183.
15. Heiden E., Palmieri L., Koenig S., Arras K.O., Sukhatme G.S. Gradient-informed path smoothing for wheeled mobile robots // *Proceedings of the IEEE International Conference on Robotics and Automation*. 2018. pp. 1710–1717.
16. Yongzhe Z., Ma B., Wai C.K. A practical study of time-elastic-band planning method for driverless vehicle for auto-parking // *International Conference on Intelligent Autonomous Systems (ICoIAS)*. IEEE. 2018. pp. 196–200.
17. Kicki P., Gawron T., Cwian K., Ozay M., Skrzypczynski P. Learning from experience for rapid generation of local car maneuvers // *Engineering Applications of Artificial Intelligence*. 2021. vol. 105. pp. 104399. DOI: 10.1016/j.engappai.2021.104399.
18. Vitelli M., Chang Y., Ye Y., Ferreira A., Wolczyk M., Osinski B., Niendorf M., Grimmett H., Huang Q., Jain A., Ondruska P. Safetynet: Safe planning for real-world self-driving vehicles using machine-learned policies // *International Conference on Robotics and Automation (ICRA)*. IEEE. 2022. pp. 897–904.
19. Nasiriany S., Pong V., Lin S., Levine S. Planning with goal-conditioned policies // *Advances in Neural Information Processing Systems*. 2019. vol. 32.
20. Chen L., Hu X., Tang B., Cheng Y. Conditional DQN-Based Motion Planning With Fuzzy Logic for Autonomous Driving // *IEEE Transactions on Intelligent Transportation Systems*. 2020. vol. 23. no. 4. pp. 2966–2977.

21. Wu K., Wang H., Esfahani M.A., Yuan S. Achieving Real-Time Path Planning in Unknown Environments Through Deep Neural Networks. *IEEE Transactions on Intelligent Transportation Systems*. 2022. vol. 23. no. 3. pp. 2093–2102.
22. Cohen B.J., Chitta S., Likhachev M. Search-based planning for manipulation with motion primitives // *IEEE International Conference on Robotics and Automation*. 2010. pp. 2902–2908. DOI: 10.1109/ROBOT.2010.5509685.
23. Low T., Bandyopadhyay T., Borges P.V. Identification of effective motion primitives for ground vehicles // *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020. pp. 2027–2034.
24. Wang B., Gong J., Chen H. Motion primitives representation, extraction and connection for automated vehicle motion planning applications // *IEEE Transactions on Intelligent Transportation Systems*. 2019. vol. 21. no. 9. pp. 3931–3945.
25. Jarin-Lipschitz L., Paulos J., Bjorkman R., Kumar V. Dispersion-minimizing motion primitives for search-based motion planning // *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021. pp. 12625–12631.
26. Bergman K., Ljungqvist O., Axehill D. Improved optimization of motion primitives for motion planning in state lattices // *IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2019. pp. 2307–2314.
27. Koutras L., Doulgeri Z. Dynamic movement primitives for moving goals with temporal scaling adaptation // *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2020. pp. 144–150.
28. Abu-Dakka F.J., Kyrki V. Geometry-aware dynamic movement primitives // *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2020. pp. 4421–4426.
29. Sood R., Vats S., Likhachev M. Learning to use adaptive motion primitives in search-based planning for navigation // *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020. pp. 6923–6929.
30. Chernik C., Tajvar P., Tumova J. Robust Feedback Motion Primitives for Exploration of Unknown Terrains // *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2021. pp. 8173–8179.
31. Palmieri L., Arras K.O. A novel RRT extend function for efficient and smooth mobile robot motion planning // *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2014. pp. 205–211.
32. Яковлев К.С., Белинская Ю.С., Макаров Д.А., Андрейчук А.А. Безопасно-интервальное планирование и метод накрытий для управления движением мобильного робота в среде со статическими и динамическими препятствиями // *Автоматика и телемеханика*. 2022. № 6. С. 96–117.
33. Wang X., Krasowski H., Althoff M. CommonRoad-RL: A configurable reinforcement learning environment for motion planning of autonomous vehicles // *IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE. 2021. pp. 466–472.
34. Ilievski M. Wisebench: A motion planning benchmarking framework for autonomous vehicles: MS thesis. Ontario: University of Waterloo, 2020. 129 p.
35. Xu C., Ding W., Lyu W., Liu Z., Wang S., He Y., Hu H., Zhao D., Li B. SafeBench: A Benchmarking Platform for Safety Evaluation of Autonomous Vehicles // *Advances in Neural Information Processing Systems*. 2022. vol. 35. pp. 25667–25682.
36. Heiden E., Palmieri L., Bruns L., Arras K.O., Sukhatme G.S., Koenig S. Bench-MR: A motion planning benchmark for wheeled mobile robots // *IEEE Robotics and Automation Letters*. 2021. vol. 6. no. 3. pp. 4536–4543.

Головин Владислав Андреевич — инженер, центр когнитивного моделирования, Московский физико-технический институт (МФТИ); студент, кафедра "научно-образовательный центр когнитивного моделирования", Московский физико-технический институт (МФТИ). Область научных интересов: искусственный интеллект, планирование траектории, эвристические алгоритмы. Число научных публикаций — 1. golovin.va@phystech.edu; Институтский переулок, 9, 141701, Долгопрудный, Россия; р.т.: +7(495)408-4554.

Яковлев Константин Сергеевич — канд. физ.-мат. наук, ведущий научный сотрудник, Федеральное государственное учреждение "Федеральный исследовательский центр "Информатика и управление" Российской академии наук" (ФИЦ ИУ РАН); ведущий научный сотрудник, Институт искусственного интеллекта AIRI. Область научных интересов: интеллектуальная робототехника, планирование траектории, эвристический поиск, много-агентные системы. Число научных публикаций — 134. yakovlev@isa.ru; проспект 60-летия Октября, 9, 141701, Москва, Россия; р.т.: +7(495)135-5457.

Поддержка исследований. Работа выполнена при поддержке Аналитического центра при Правительстве Российской Федерации в соответствии с договором о субсидии (идентификатор договора 000000D730321P5Q0002; грант № 70-2021-00138).

V. GOLOVIN, K. YAKOVLEV
**MOTION PRIMITIVES IN THE TRAJECTORY PLANNING
PROBLEM WITH KINEMATIC CONSTRAINTS**

Golovin V., Yakovlev K. Motion Primitives in the Trajectory Planning Problem with Kinematic Constraints.

Abstract. Automatic trajectory planning is an urgent scientific and technical problem, whose solutions are in demand in many fields: unmanned transportation, robotic logistics, social robotics, etc. Often, when planning a trajectory, it is necessary to consider the fact that the agent (robot, unmanned car, etc.) cannot arbitrarily change its orientation while moving, in other words, it is necessary to consider kinematic constraints when planning. One widespread approach to solving this problem is the approach that relies on the construction of a trajectory from prepared parts, motion primitives, each of which satisfies kinematic constraints. Often, the emphasis in the development of methods implementing this approach is on reducing the combinations of choices in planning (heuristic search), with the set of available primitives itself being regarded as externally defined. In this paper, on the contrary, we aim to investigate and analyze the effect of different available motion primitives on the quality of solving the planning problem with a fixed search algorithm. Specifically, we consider 3 different sets of motion primitives for a wheeled robot with differential drive. As a search algorithm, the A* algorithm well known in artificial intelligence and robotics is used. The solution quality is evaluated by 6 metrics, including planning time, length and curvature of the resulting trajectory. Based on the study, conclusions are made about the factors that have the strongest influence on the planning result, and recommendations are given on the construction of motion primitives, the use of which allows to achieve a balance between the speed of the planning algorithm and the quality of the trajectories found.

Keywords: path planning, kinematic constraints, motion primitives, heuristic search.

References

1. Dudakova D.S., Anohin V.M., Dudakov M.O., Ronzhin A.L. [On Theoretical Foundations of Aerolimnology: Study of Fresh Water Bodies and Coastal Territories Using Air Robot Equipment]. *Informatika i avtomatizaciya – Informatics and Automation*. 2022. vol. 21. no. 6. pp. 1359–1393.
2. Balabanov A.N., Bezuglaya A.E., Shushlyapin E.A. [Underwater Robot Manipulator Control]. *Informatika i avtomatizaciya – Informatics and Automation*. 2021. vol. 20. no. 6. pp. 1307–1332. DOI: 10.15622/ia.20.6.5.
3. Makarov D.A., Panov A.I., Jakovlev K.S. Architecture of a multi-level intelligent control system for unmanned aerial vehicles. *Iskusstvennyj intellekt i prinjatie reshenij – Artificial Intelligence and Decision Making*. 2015. no 3. pp. 18–33.
4. Otsu K., Matheron G., Ghosh S., Toupet O., Ono M. Fast approximate clearance evaluation for rovers with articulated suspension systems. *Journal of Field Robotics*. 2020. vol. 37. no. 5. pp. 768–785.
5. Al Mashhadany Y.I. Design and analysis of 7-DOF human-link manipulator based on hybrid intelligent controller. *Informatics and Automation*. 2020. vol. 19. no. 4. pp. 774–802.
6. Pshihopov V.H., Medvedev M.J. Planning the movement of a group of moving objects in a two-dimensional environment with obstacles. *Izvestija Juzhnogo federal' nogo universiteta*.

- Technicheskie nauki – News of the Southern Federal University. Technical science. 2016. vol. 2(175). pp. 6–22.
7. LaValle S.M. Rapidly-exploring random trees: A new tool for path planning. Research Report 9811. 1998.
8. Kuffner J.J., LaValle S.M. RRT-connect: An efficient approach to single-query path planning. Proceedings ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings. 2000. vol. 2. pp. 995–1001.
9. Sharma P., Gupta A., Ghosh D., Honkote V., Nandakumar G., Ghose, D. PG-RRT: A Gaussian Mixture Model Driven, Kinematically Constrained Bi-directional RRT for Robot Path Planning. Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. 2021. pp. 3666–3673.
10. Webb D.J., Van Den Berg J. Kinodynamic RRT*: Asymptotically optimal motion planning for robots with linear dynamics. IEEE international conference on robotics and automation. IEEE. 2013. pp. 5054–5061.
11. Gammell J.D., Srinivasa S.S., Barfoot T.D. Informed RRT: Optimal sampling-based path planning focused via direct sampling of an admissible ellipsoidal heuristic. IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE. 2014. pp. 2997–3004.
12. Zhang Y., Sun H., Zhou J., Pan J., Hu J., Miao J. Optimal vehicle path planning using quadratic optimization for baidu apollo open platform. IEEE Intelligent Vehicles Symposium (IV). IEEE. 2020. pp. 978–984.
13. Li B., Wang K., Shao Z. Time-optimal maneuver planning in automatic parallel parking using a simultaneous dynamic optimization approach. IEEE Transactions on Intelligent Transportation Systems. 2016. vol. 17. no. 11. pp. 3263–3274.
14. Karlsson J., Murgovski N., Sjöberg J. Computationally efficient autonomous overtaking on highways. IEEE Transactions on Intelligent Transportation Systems. 2019. vol. 21. no. 8. pp. 3169–3183.
15. Heiden E., Palmieri L., Koenig S., Arras K.O., Sukhatme G.S. Gradient-informed path smoothing for wheeled mobile robots. Proceedings of the IEEE International Conference on Robotics and Automation. 2018. pp. 1710–1717.
16. Yongzhe Z., Ma B., Wai C.K. A practical study of time-elastic-band planning method for driverless vehicle for auto-parking. International Conference on Intelligent Autonomous Systems (ICoIAS). IEEE. 2018. pp. 196–200.
17. Kicki P., Gawron T., Cwian K., Ozay M., Skrzypczynski P. Learning from experience for rapid generation of local car maneuvers. Engineering Applications of Artificial Intelligence. 2021. vol. 105. pp. 104399. DOI: 10.1016/j.engappai.2021.104399.
18. Vitelli M., Chang Y., Ye Y., Ferreira A., Wolczyk M., Osinski B., Niendorf M., Grimm H., Huang Q., Jain A., Ondruska P. Safetynet: Safe planning for real-world self-driving vehicles using machine-learned policies. International Conference on Robotics and Automation (ICRA). IEEE. 2022. pp. 897–904.
19. Nasiriany S., Pong V., Lin S., Levine S. Planning with goal-conditioned policies. Advances in Neural Information Processing Systems. 2019. vol. 32.
20. Chen L., Hu X., Tang B., Cheng Y. Conditional DQN-Based Motion Planning With Fuzzy Logic for Autonomous Driving. IEEE Transactions on Intelligent Transportation Systems. 2020. vol. 23. no. 4. pp. 2966–2977.
21. Wu K., Wang H., Esfahani M.A., Yuan S. Achieving Real-Time Path Planning in Unknown Environments through Deep Neural Networks. IEEE Transactions on Intelligent Transportation Systems. 2022. vol. 23. no. 3. pp. 2093–2102.
22. Cohen B.J., Chitta S., Likhachev M. Search-based planning for manipulation with motion primitives. IEEE International Conference on Robotics and Automation. 2010. pp. 2902–2908. DOI: 10.1109/ROBOT.2010.5509685.

23. Low T., Bandyopadhyay T., Borges P.V. Identification of effective motion primitives for ground vehicles. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020. pp. 2027–2034.
24. Wang B., Gong J., Chen H. Motion primitives representation, extraction and connection for automated vehicle motion planning applications. *IEEE Transactions on Intelligent Transportation Systems*. 2019. vol. 21. no. 9. pp. 3931–3945.
25. Jarin-Lipschitz L., Paulos J., Bjorkman R., Kumar V. Dispersion-minimizing motion primitives for search-based motion planning. *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021. pp. 12625–12631.
26. Bergman K., Ljungqvist O., Axehill D. Improved optimization of motion primitives for motion planning in state lattices. *IEEE Intelligent Vehicles Symposium (IV)*. IEEE. 2019. pp. 2307–2314.
27. Koutras L., Doulgeri Z. Dynamic movement primitives for moving goals with temporal scaling adaptation. *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2020. pp. 144–150.
28. Abu-Dakka F.J., Kyrki V. Geometry-aware dynamic movement primitives. *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2020. pp. 4421–4426.
29. Sood R., Vats S., Likhachev M. Learning to use adaptive motion primitives in search-based planning for navigation. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020. pp. 6923–6929.
30. Chernik C., Tajvar P., Tumova J. Robust Feedback Motion Primitives for Exploration of Unknown Terrains. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2021. pp. 8173–8179.
31. Palmieri L., Arras K.O. A novel RRT extend function for efficient and smooth mobile robot motion planning. *Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2014. pp. 205–211.
32. Yakovlev K.S., Belinskaya Yu.S., Makarov D.A., Andrejchuk A.A. Safe interval planning and covering method for controlling the movement of a mobile robot in an environment with static and dynamic obstacles. *Avtomatika i telemekhanika – Automation and telemechanics*. 2022. no. 6. pp. 96–117.
33. Wang X., Krasowski H., Althoff M. CommonRoad-RL: A configurable reinforcement learning environment for motion planning of autonomous vehicles. *IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE. 2021. pp. 466–472.
34. Ilievski M. Wisebench: A motion planning benchmarking framework for autonomous vehicles: MS thesis. Ontario: University of Waterloo, 2020. 129 p.
35. Xu C., Ding W., Lyu W., Liu Z., Wang S., He Y., Hu H., Zhao D., Li B. SafeBench: A Benchmarking Platform for Safety Evaluation of Autonomous Vehicles // *Advances in Neural Information Processing Systems*. 2022. vol. 35. pp. 25667–25682.
36. Heiden E., Palmieri L., Bruns L., Arras K.O., Sukhatme G.S., Koenig S. Bench-MR: A motion planning benchmark for wheeled mobile robots. *IEEE Robotics and Automation Letters*. 2021. vol. 6. no. 3. pp. 4536–4543.

Golovin Vladislav — Software engineer, Cognitive modeling center, Moscow Institute of Physics and Technology (MIPT); Student, Department "research and educational center for cognitive modeling", Moscow Institute of Physics and Technology (MIPT). Research interests: artificial intelligence, trajectory planning, heuristic algorithms. The number of publications — 1. golovin.va@phystech.edu; 9, Institutskiy Lane, 141701, Dolgoprudniy, Russia; office phone: +7(495)408-4554.

Yakovlev Konstantin — Ph.D., Leading researcher, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC CSC RAS); Leading researcher, Artificial Intelligence Research Institute (AIRI). Research interests: intelligent robotics, path planning, heuristic search, multi-agent systems. The number of publications — 134. yakovlev@isa.ru; 9, 60-letiya Oktyabrya Ave., 141701, Moscow, Russia; office phone: +7(495)135-5457.

Acknowledgements. This work was supported by the Analytical Center for the Government of the Russian Federation in accordance with a subsidy agreement (agreement identifier 000000D730321P5Q0002; grant no. 70-2021-00138).

Б.Г. МАЙОРОВ
**ПРИМЕНЕНИЕ ГАРМОНИЧЕСКИХ ПОЛУВОЛН ДЛЯ
АВТОМАТИЗАЦИИ УПРАВЛЕНИЯ ВЫСОКОСКОРОСТНЫМИ
ПОЕЗДАМИ**

Майоров Б.Г. Применение гармонических полуволн для автоматизации управления высокоскоростными поездами.

Аннотация. Процессы экстренного торможения в Европейской системе управления поездами (European Train Control System (ETCS)) связаны со ступенчатым регулированием ускорения (замедления) в зависимости от способности торможения поезда, данных рельефа и меняющейся погоды на маршруте движения. Эти процессы являются определяющими в ETCS. Процедура ступенчатого регулирования замедления осуществляется машинистом многократно в процессе торможения до полной остановки поезда. Начало экстренного торможения и его окончание, а так же сам процесс торможения сопровождается многократным импульсным срабатыванием тормозов, что приводит к скачкам замедления и, соответственно, к повышенному износу тормозной системы, снижению комфорта для пассажиров, из чего следует ограничение максимально допустимой скорости движения. В статье предложена новая концепция и методика построения математических моделей кривых экстренного торможения отличных от кривых ETCS и основанных на гармонических полуволнах. Показано, что кривые торможения ETCS описываются известными степенными полуволнами второго порядка. Совместное их исследование даёт основание утверждать, что применение этих кривых приводит к обязательному импульсному режиму срабатывания тормозов. Предложены два новых варианта моделей кривых экстренного торможения, описываемых гармоническими полуволнами. Первый вариант имеет одно импульсное срабатывание тормозов в конце интервала торможения. Второй вариант свободен от тормозных импульсов и позволяет использовать непрерывное регулирование. Эти модели объясняют особенности ETCS, содержат предложения по их устранению, применимы для разработки новых кривых экстренного торможения, которые позволяют плавно осуществлять экстренное торможение поездов. Работоспособность, отличия и преимущества перед кривыми торможения ETCS показаны на результатах математического моделирования процессов экстренного торможения.

Ключевые слова: высокоскоростные поезда, максимальная скорость движения, экстренное торможение, кривые экстренного торможения, плавное регулирование замедления поезда, гармонические полуволны, методика построения математических моделей кривых экстренного торможения.

1. Введение. В Европейской системе управления поездами (European Train Control System (ETCS)) применяют для автоматизации управления высокоскоростных поездов процедуру «Расчет и контроль кривых торможения» [1 – 3]. Основная задача ETCS и этой процедуры – контроль отсутствия превышения поездом допустимой скорости в реальном времени и остановка перед местом возможной аварии. Для этого текущая скорость постоянно сравнивается с допустимой скоростью, рассчитанной для текущего положения, которая определяется с помощью кривых экстренного торможения. Расчет

кривой учитывает зависимость скорости от мгновенного замедления (отрицательного ускорения) и характеризует способность выполнения поездом торможения. Кроме данных поезда, учитываются постоянно меняющиеся данные маршрута, например, уклон пути и условия сцепления колёс с рельсами, которые поступают от служб, обслуживающих железнодорожную инфраструктуру. Расчет кривой торможения ETCS описан в требованиях к системе ETCS [2, 3]. Эти требования поддерживаются европейскими фирмами разработчиками высокоскоростных систем железнодорожного движения (Alstom, Ansaldo, Bombardier, Invensys, Siemens, Thales), фирмами Китая и Японии, которые входят в союз производителей систем безопасности (Union of Signaling Industry – UNISIG). При скоростях более 500 км/ч применяют электромагнитные тормоза и не используют рельсы, используют принцип «маглев» – магнитной левитации [4]. Для проектирования и эксплуатации систем и средств высокоскоростного движения используются авиационный дизайн и технологии [5], методы обеспечения надежности и безопасности, хорошо развитые в трудах отечественных и зарубежных ученых, вводятся единые правила управления [6] и эргономического обеспечения [7].

Математические модели и алгоритмы являются собственностью упомянутых фирм, поэтому необходимо научно обосновывать и разрабатывать оригинальные способы и модели систем управления скоростным железнодорожным движением для развития управления.

Существующие и применяемые в ETCS кривые экстренного торможения математически записываются в виде зависимости скорости высокоскоростного поезда от расстояния, проходимого в процессе торможения. Математически это соответствует степенной функции от расстояния в степени 0.5. Рассмотрим примеры временных диаграмм экстренного торможения [1] и покажем полное совпадение этих диаграмм и их математических моделей с авторской моделью обобщённого критерия наибольшего отклонения [8]. Этот факт позволяет рассматривать и развивать новый подход к построению кривых экстренного торможения высокоскоростных поездов на основе процедур цифровой обработки сигналов, имеющих трудности в определении спектра, и применять идеи оптимизации [9] процессов управления. Подобные подходы применяют достаточно часто при обработке сигналов процессов, не имеющих частотного спектра, что является неопределённостью объекта управления и его информационного описания. Такие средства и системы реального времени относят к робастным системам [9 – 11] и применяют в разнообразных областях автоматизации процессов управления:

– *геоакустическая эмиссия*: одним из способов предсказания извержения вулканов является описание и систематизация геоимпульсов в амплитудно-временных координатах, выявление аномальных паттернов и их анализ для привязки к сейсмическим событиям [12];

– *робототехника*: для управления приводами манипулятора при копирующем типе управления, обеспечения перемещения схвата манипулятора из начального положения покоя в заданное конечное положение покоя за минимальное время [13, 14];

– *оптико-электронное наблюдение*: для перехвата движущейся по предписанной траектории цели машиной Дубинса [15 – 17];

– *биотехнология, металлургия*: для обеспечения качества продуктов питания и металлов в условиях априорной неопределённости технологических параметров процесса изготовления;

– *космические полёты*: для предсказания траекторий при резком изменении массы корабля из-за сброса пустых топливных баков [18].

Модели подобных процессов и их сигналов традиционно описывают в амплитудно-временной области: в автоматических системах график оптимального процесса управления задают в виде сопрягаемых парабол [9, 18]; моделируют равноускоренное движение манипулятора руки квадратичным трёхчленом [13, 14]; представляют сейсмические события в виде линейных степенных функций [12]. В качестве коэффициентов трёхчлена используют максимальные значения ускорения и скорости сигнала. Известны две формулы В.Н. Хлистунова [19] в виде линейного и квадратичного остаточного члена формулы Лагранжа в измерительной технике и в других применениях цифровых сигнальных процессоров [16, 20, 21]. Обобщением этих двух формул является формула [8], которая существенно повышает обоснованность и точность аппроксимации сигнала автоматизируемого процесса при его приближении по выборкам.

Общим недостатком применения степенных приближений является отсутствие математической связи с частотными характеристиками сигналов и процессов или с математическими моделями, использующими эти частотные характеристики.

В [22] рассмотрена модель в амплитудно-временной области, в виде гармонической полуволны (и её модификаций), которая устанавливает связь её параметров с амплитудно-частотными параметрами. Получены выражения, связывающие математически два типа описания сигналов. Гармоническая полуволна является половиной

волны частоты среза f спектра (повторяется с частотой $f_c = 2f$). Кроме наибольшего параметра f_c она характеризуется наибольшей скоростью сигнала $\max p^{(1)}(t) = V$, наибольшим ускорением сигнала $\max p^{(2)}(t) = a$. Амплитуда полуволны $\max p(t) = \Delta$ является максимально допустимой погрешностью аппроксимации сигнала ступенькой или отрезком линии по двум соседним выборкам. В статье [22] показано, что гармоническая полуволна является универсальной моделью для описания в амплитудно-частотной области и в амплитудно-временной области процессов и сигналов и связывает две группы максимальных величин параметров сигнала $\max\{\Delta, f_c\}$ и $\max\{\Delta, V, a, T_c\}$, где $T_c = 1/f_c$ – временной интервал существования полуволны.

Основными задачами настоящей статьи является выявление полезных особенностей гармонических полуволн для построения принципиально новых тормозных кривых по сравнению с кривыми торможения ETCS, которые описываются обобщённым критерием наибольшего отклонения (в виде степенных полуволн) и разработка методики проектирования гармонических тормозных кривых.

Для этого рассмотрены временные диаграммы экстренного торможения высокоскоростного состава [1–3] и обнаружены два основных недостатка использования кривых экстренного торможения ETCS:

- ведётся ступенчатое регулирование ускорения (замедления) в ручном режиме, в зависимости от способности торможения поезда и меняющихся данных о путях маршрута. Фактически ступенчатое регулирование торможения осуществляется машинистом многократно в процессе и до полного останова поезда, что усложняет процедуру экстренного торможения и ставит её в зависимость от машиниста;

- начало и окончание торможения сопровождается многократным ударным включением и выключением тормозов, что приводит к скачкам замедления и, соответственно, к повышенному износу тормозной системы и снижению комфорта для пассажиров (ударным воздействиям на организмы), особенно на скоростях около 500 км/ч [7].

Предложено и исследовано несколько вариантов кривых экстренного торможения, развивающих концепцию обобщённого критерия наибольшего отклонения [8, 22], которые позволяют исследовать указанные особенности кривых ETCS. Преимущества гармонических кривых торможения перед кривыми ETCS показаны на результатах моделирования процессов экстренного торможения. При этом процесс движения высокоскоростного состава сопровождается

постоянным определением его координат на линии по способу чтения RFID-меток в области железнодорожного полотна, в которые записаны коды их координат [23], где слова в названии способа «в закрытых помещениях» означают отсутствие сигналов спутников ГЛОНАСС или GPS.

2. Изучение диаграмм экстренного торможения ETCS.

Рассмотрим математическую модель экстренного торможения высокоскоростного поезда, применяемую в ETCS, для обеспечения безопасного движения попутно следующих поездов с минимальным временным интервалом следования и сравним её со степенной полуволевой наибольшего отклонения для установления их идентичности.

2.1. Пример диаграмм экстренного торможения ETCS. На рисунке 1 изображены временные диаграммы шести реальных экстренных торможений состава ETCS [1]. В [1] приведено выражение взаимосвязи параметров движения состава [в размерностях] при экстренном его торможении: расстояния торможения – $[s]=м$, начальной, максимальной скорости торможения – $[V]=км/ч$, максимального замедления (отрицательного ускорения) – $[a]=м/с^2$:

$$V^2 = 2sa . \quad (1)$$

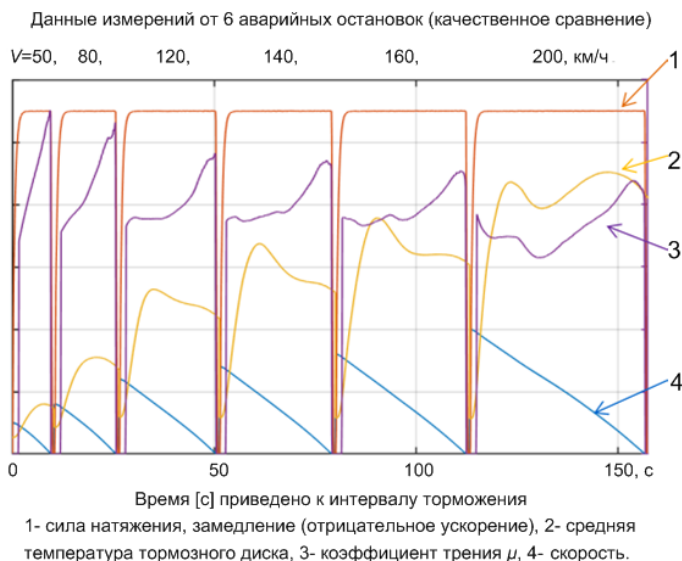


Рис. 1. Данные измерений шести аварийных остановок состава ETCS для различных значений начальной скорости торможения V

Как следует из рисунка 1, замедление в процессе торможения является постоянной величиной $a = \text{const}$, скорость уменьшается от V до нуля по линейному закону при аргументе: интервал торможения s . При этом *время торможения T приведено к интервалу торможения s* .

Выражение (1) полностью совпадает с условием согласования параметров $\{\Delta, V, a\}$ обобщённого критерия наибольшего отклонения $V=2\Delta a$ [8], если произвести замены обозначений $V = V$, $s = \Delta$. Исходя из этого совпадения, содержания рисунка 1 и для конкретизации постановки задачи статьи поясним критерий наибольшего отклонения процесса торможения по степенной полуволне второго порядка.

2.2. Степенная полуволна второго порядка и её особенности.

Покажем, что рисунок 1 и выражение (1) получаются из обобщённого критерия наибольшего отклонения [8], из чего следует основание для применения новых вариантов критерия для математического моделирования процессов торможения поездов. Используем изображение степенной полуволны на рисунке 2 и формулу степенной полуволны второго порядка, как основную часть обобщённого критерия:

$$p(t) = \Delta - \frac{a}{2}(t - 0.5T_2)^2, \quad (2)$$

где Δ – амплитуда степенной полуволны; a – максимальное значение ускорения (замедления); T_2 – интервал времени между двумя значениями сигнала, когда $p(t) = 0$. Производные от (2) по времени t :

$$p^{(1)}(t) = -a(t - 0.5T_2) = V(t), \quad p^{(2)}(t) = -a = \text{const}. \quad (3)$$

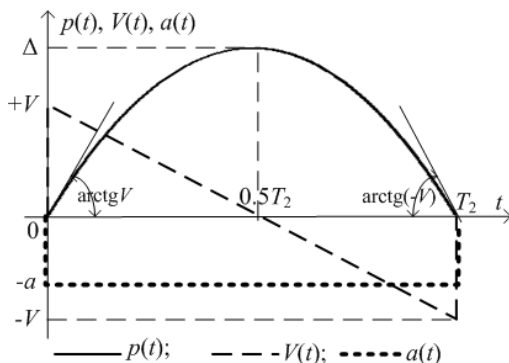


Рис. 2. Степенная полуволна второго порядка $p(t)$ (2) и две производные от неё по времени (3)

Определим особенности степенной полуволны из её описания.

При $t = 0$ и $t = T_2$ из (2): $p(0) = p(T_2) = 0 \rightarrow T_2 = 2(2\Delta/a)^{1/2}$.

При $t = 0$ и $t = T_2$ из (3): $p^{(1)}(0) = \max V(t) = +V$,

$p^{(1)}(T_2) = \min V(t) = -V \rightarrow T_2 = 2(V/a)$.

Обозначим $T_{\text{ст}}$ – интервал времени торможения по степенной полуволне при $0 \leq t \leq 0.5T_2 = T_{\text{ст}}$. Физический смысл этого обозначения виден из (2) и рисунка 2, где при $t = 0$, $p(0) = 0$ – начало времени и пути торможения; из (3) $V(0) = +V$ – максимальная скорость состава в начале торможения; при $t = T_{\text{ст}}$ из (2) и (3) имеем, соответственно, $p(T_{\text{ст}}) = \Delta$ – величина тормозного пути и $V(T_{\text{ст}}) = 0$ – остановка состава.

Несколько равных формул для T_2 позволяют записать цепь равных элементов:

$$T_{\text{ст}} = 0.5T_2 = \frac{2\Delta}{V} = \frac{V}{a} = \left(\frac{2\Delta}{a}\right)^{1/2}. \quad (4)$$

Из (4) следует, что справедливость равных элементов наступает при соблюдении условия согласованности максимальных параметров $\{\Delta, V, a\}$ модели степенной полуволны процесса $p(t)$. Это условие запишем путём преобразования любых пар элементов цепи равных элементов (4) в одно равенство:

$$V^2 = 2\Delta a. \quad (5)$$

Как упомянуто выше, а теперь можно утверждать, что (5) совпадает с (1). Ещё это указывает на возможность моделирования операции *приведения времени торможения* $0 \leq t \leq 0.5T_2$ к *интервалу торможения* s или $0 \leq p(t) \leq \Delta$ (как в диаграммах торможения ETCS [1]) на примере исследования модели степенной полуволны.

2.3. Приведение времени торможения к расстоянию торможения по степенной полуволне второго порядка. Применим степенную полуволну для описания кривой экстренного торможения высокоскоростного состава по протоколу ETCS [1], по предложенной в статье методике:

1) Так как в конце тормозного пути $p(t) = \Delta$, $V \geq V(t) = 0$, то интервал времени экстренного торможения (рисунок 2) будет $T_{\text{ст}} = 0.5T_2$ и находится из любого элемента цепи равных элементов (4).

2) Кривая экстренного торможения определяет зависимость скорости состава от расстояния, поэтому из выражения (2) выразим

время через расстояние: $(t - 0.5T_2) = [2/a(\Delta - p(t))]^{1/2}$ и подставим в первое выражение (3), получим:

$$V(t) = -[2a(\Delta - p(t))]^{1/2}, \quad (6)$$

где $0 \leq V(t) \leq V$; $0 \leq p(t) \leq \Delta$.

3) В (6) учтём (5): $2a = V^2/\Delta$, получим выражение для зависимости скорости от расстояния торможения: $V(t) = V[1 - p(t)/\Delta]^{1/2}$, которое переведём в относительную, безразмерную форму:

$$\frac{V(t)}{V} = \left[1 - \frac{p(t)}{\Delta}\right]^{1/2}. \quad (7)$$

На рисунке 3 приведена зависимость относительной скорости $V(t)/V$ от относительного тормозного пути $p(t)/\Delta$ (7) и выражение относительного замедления $-a(t) = -V^2/2\Delta = -a = \text{const}$ из (5), которое не зависит от $p(t)$: $a(t)/a = 1$.

Примечание 1: «Относительная величина скорости $V(t)/V$ (рисунок 3) по выражению (7) является степенной функцией, в которой используем только положительную ветвь $0 \leq V(t) \leq V$ (рисунок 2), так как она технически соответствует процессу торможения состава».

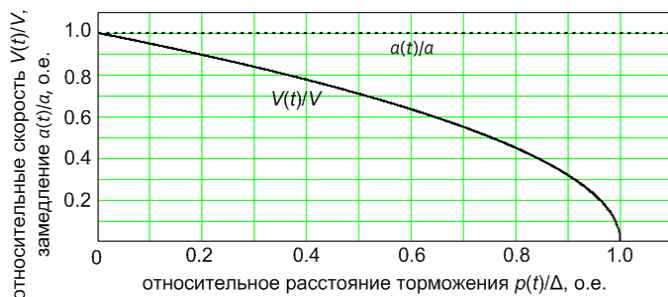


Рис. 3. Относительные зависимости скорости и замедления от тормозного пути по степенной полуовне

Примечание 2: «Относительная величина замедления $a(t)/a \geq 0$ соответствует процессу торможения состава и является отношением функции $a(t) \leq 0$ и константы $a \leq 0$, поэтому на рисунке 3 величина изображена в виде положительной константы».

Для примера расчёта примем предельные *исходные значения*: тормозной путь $\Delta = 1500$ м, скорость в начале торможения: $V = 500$ км/ч = 138.8889 м/с. Определим *параметры торможения*: найдём необходимое замедление высокоскоростного состава (5): $a = -V^2/2\Delta = -6.43$ м/с², время торможения (4): $T_{ст} = 2\Delta/V = 21.60$ с.

2.4. Особенности реальной кривой экстренного торможения второго порядка. Кривая взята из ETCS [1, 3], представлена на рисунке 4 и состоит из семи участков в виде квадратичных парабол с различными постоянными замедлениями, описывает снижение скорости из-за экстренного торможения. В составе ETCS она называется кривой EBD (Emergency Brake Deceleration). Кривая EBD является полностью детерминированной, зависит как от характеристик поезда, так и от характеристик пути. Эти характеристики задаются двумя группами поправок: A_T или «A_brake_safe» – изменяет величину замедления состава в зависимости от скорости состава и погодных условий, $A_{п/с}$ или «A_gradient» – изменяет замедление от подъёма/спуска пути. Поправки задают в виде двух ступенчатых функций (рисунок 4).

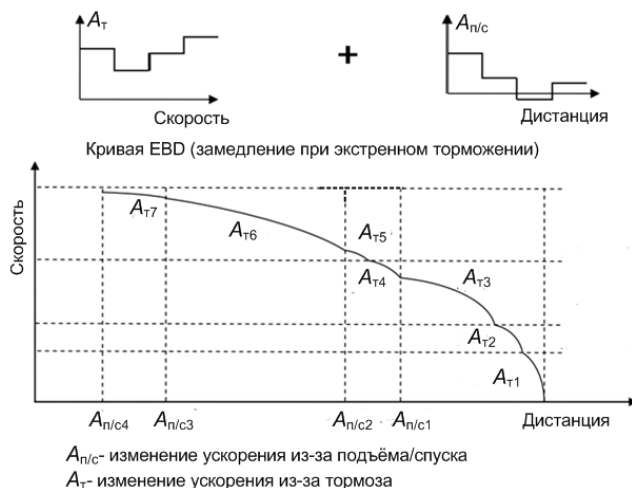


Рис. 4. Кривая экстренного торможения Emergency Brake Deceleration (EBD)

Кривая EBD имеет сложный состав и постоянно сравнивается в компьютере локомотива с кривой, подобной идеальной кривой торможения (рисунок 3). Кроме этого, за её поведением постоянно следит машинист локомотива. При пересечении или опасном

сближении двух кривых включается автоматическое управление торможением состава и включается звуковой сигнал предупреждения машиниста об опасности столкновения. Выступы на кривой EBD обусловлены тем, что в некоторые моменты времени меняются величины замедления от меняющихся поправок, а парабола (7) порядка 0.5 является гладкой (рисунок 3), если замедление (ускорение) постоянно $a = \text{const}$ в соответствии с (3) и (5).

Выступы провоцируют пересечение двух кривых, кроме этого величина многократных ударных замедлений до $a = -6.43 \text{ м/с}^2$ (пример расчёта в пункте 2.3) приводят к снижению комфорта для пассажиров, особенно на максимальных скоростях.

Предложим несколько кривых экстренного торможения, свободных от описанных особенностей. Кривые торможения построим на основе гармонических полувольт [22] и покажем, что они имеют необходимые положительные свойства для использования в управлении.

3. Модель гармонической полувольты и применение её для описания кривых экстренного торможения. Отметим свойства гармонической полувольты, которые можно использовать для устранения некоторых особенностей степенной полувольты или эквивалентной ей кривой EBD без импульсных поправок замедления A_T ($A_{\text{brake_safe}}$) и $A_{n/c}$ ($A_{\text{gradient_safe}}$). Получим параметры экстренного торможения, зависящие от времени, и методику приведения времени торможения к интервалу торможения (аналогично пункту 2.3).

3.1. Основные параметры и некоторые особенности гармонической полувольты. Параметры получены в [22] и приведены кратко для целостности материала. Уравнение гармонической полувольты:

$$p(t) = \Delta \sin \omega t, \quad (8)$$

где $0 \leq t \leq T_T$; $T_T = T/2 = 1/2f$ – интервал Найквиста; f – частота среза в спектре сигнала $p(t)$ после фильтрации; T – период частоты среза; $\omega = 2\pi f$ – круговая частота; Δ – амплитуда гармонической полувольты или максимальная погрешность линейной аппроксимации двух выборок из полувольты в моменты времени $t = 0$ и $t = T_T$. Первая производная от этой полувольты определяет скорость $V(t)$ её изменения:

$$p^{(1)}(t) = \omega \Delta \cos \omega t = V(t). \quad (9)$$

Вторая и третья производные от полуволны определяют, соответственно, ускорение или замедление $a(t)$ и резкость торможения $R(t)$:

$$p^{(2)}(t) = -\omega^2 \Delta \sin \omega t = a(t), \quad p^{(3)}(t) = -\omega^3 \Delta \cos \omega t = R(t). \quad (10)$$

Гармоническая полуволна и три производные от неё по времени изображены на рисунке 5.

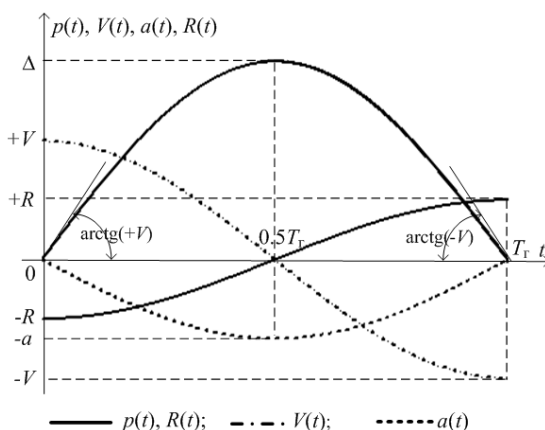


Рис. 5. Гармоническая полуволна $p(t)$ (8) и три производных от неё по времени (9), (10)

Выражение (8) описывает половину $T_r = T/2 = 1/2f$ периода T частоты среза f в спектре сигнала $p(t)$. Частота Найквиста $2f$ широко используется в теории связи, управления и в информатике [22] и в этом исследовании сделана попытка ещё одного её применения.

Первая выборка из сигнала $p(t)$ осуществлена в момент времени $t = 0$, вторая – в момент времени $t = T_r = T/2$. Линия, интерполирующая сигнал между двумя выборками лежит на оси абсцисс $0t$.

Погрешность такой интерполяции Δ достигает максимального значения в точке с координатами $(0.5T_r, \Delta)$, которая является вершиной гармонической полуволны. Проходящая через вершину вертикальная линия $t = 0.5T_r = 0.25T = T_{\text{гт}}$ служит её осью симметрии.

Величину интервала времени экстренного гармонического торможения обозначим $T_{\text{гт}}$, она соответствует двум точкам гармонической полуволны: точке начала полуволны $t = 0$ с

максимальной скоростью V изменения процесса и точке на вершине полуволны $t = 0.5T_r = T_{гт}$, в которой скорость процесса равна нулю, а отклонение или тормозной путь равен Δ . Таким образом, половина гармонической полуволны полностью соответствует физическим требованиям к кривой экстренного торможения. Связь круговой частоты ω гармонической полуволны и интервала торможения $T_{гт}$ запишем в виде выражения:

$$\omega = 2\pi f = \frac{2\pi}{T} = \frac{\pi}{T_r} = \frac{\pi}{2T_{гт}}. \quad (11)$$

Определим интервал гармонической полуволны T_r через максимальные параметры сигнала $p(t)$ процесса управления. Эти параметры должны быть получены при обработке результатов обследования объекта управления, так как процесс торможения поезда по пути $p(t)$ можно отнести к процессу с трудно определяемым спектром сигналов [22].

В выражение первой производной (9) подставим моменты времени выборки из сигнала $p(t)$: $p(t=0) = 0$ и $p(t=T_r) = 0$ и приравняем двум её максимальным значениям: $\max p^{(1)}(t=0, t=T_r) = |\pm V|$, учтём (11), $\cos(0)=1$, $\cos(\pi)=-1$, получим:

$$T_{гт} = \frac{\pi}{2} \frac{\Delta}{V}. \quad (12)$$

В выражение второй производной (10) подставим момент времени ($t = 0.5T_r$), при котором наступает максимум ускорения сигнала $p(t)$ процесса торможения и приравняем её максимальному значению $\max p^{(2)}(t = 0.5T_r) = |-a|$. Используем (11), получим цепь равенств интервалов гармонической полуволны:

$$T_{гт} = \frac{\pi}{2} \frac{\Delta}{V} = \frac{\pi}{2} \frac{V}{a} = \frac{\pi}{2} \left(\frac{\Delta}{a} \right)^{1/2}, \quad (13)$$

где условие согласования параметров $\{\Delta, V, a\}$ процесса: $V^2 = \Delta a$. Аналогично (13), при измеренном значении $\max p^{(3)}(t=0, t=T_r) = |\pm R|$, запишем более длинную цепь равенств интервалов:

$$T_{гт} = \frac{\pi}{2} \frac{\Delta}{V} = \frac{\pi}{2} \frac{V}{a} = \frac{\pi}{2} \left(\frac{\Delta}{a} \right)^{1/2} = \frac{\pi}{2} \left(\frac{V}{R} \right)^{1/2} = \frac{\pi}{2} \frac{a}{R} = \frac{\pi}{2} \left(\frac{\Delta}{R} \right)^{1/3}, \quad (14)$$

где два частных условия согласования параметров $\{\Delta, V, a, R\}$ процесса торможения: $V^2 = \Delta a$, $a^2 = VR$ или общее условие: $Va = \Delta R$. С учётом (11) запишем (14) через круговую частоту ω :

$$\omega = \frac{V}{\Delta} = \frac{a}{V} = \left(\frac{a}{\Delta}\right)^{1/2} = \left(\frac{R}{V}\right)^{1/2} = \frac{R}{a} = \left(\frac{R}{\Delta}\right)^{1/3}. \quad (15)$$

3.2. Сравнение свойств гармонической полуволны и полуволны EBD или модели степенной полуволны. Совместное рассмотрение выражений (13), (14) и (4) показывает, что:

1) У гармонической полуволны продолжительность и форма не зависят от количества производных. В степенной полуволне количество производных, форма и продолжительность определяются величиной степени степенной функции.

2) Гармонические полуволны позволяют определить частоту эквивалентной граничной частоты сигналов процессов с трудно определяемым спектром [22], напротив, степенные полуволны имеют спектры, граничные частоты которых существенно выше частоты самих степенных волн.

3) В соответствии с рисунком 4 [1, 3] при работе с тормозами состава машинист осуществляет ступенчатое регулирование ускорения (замедления) в ручном режиме, в зависимости от способности торможения поезда и меняющихся условий внешней среды на пути следования. В дальнейшем покажем, что гармонические полуволны позволяют осуществлять плавное регулирование замедления высокоскоростного состава.

4) Ударное включение и выключение тормозов системы EBD приводит к скачкам замедления и, соответственно, к повышенному износу тормозной системы и снижению комфорта для пассажиров.

3.3. Приведение времени торможения к интервалу (расстоянию) торможения гармонической полуволны. Применим гармоническую полуволну для описания кривой экстренного торможения высокоскоростного состава по протоколу ETCS [2, 3] подобно предложенной методике (пункт 2.3).

1) Так как в конце тормозного пути $p(t) = \Delta$, $V(t) = 0$, то время экстренного торможения (рисунок 5) будет $T_{\text{т}} = 0,5T_r$ и находится из любого элемента двух цепей равных элементов (13) или (14).

2) Кривая экстренного торможения определяет зависимость скорости состава от расстояния, поэтому из выражения (8) выразим безразмерное произведение ωt через относительное безразмерное расстояние $p(t)/\Delta$: $\omega t = \arcsin(p(t)/\Delta)$ и подставим в (9), преобразуем,

используем $V = \omega \Delta$ (15), получим: $V(t) = V[1 - (p(t)/\Delta)^2]^{1/2}$, которое переведем в относительную, безразмерную форму (рисунок 6 и Примечание 1):

$$\frac{V(t)}{V} = \left[1 - \left(\frac{p(t)}{\Delta} \right)^2 \right]^{1/2}. \quad (16)$$

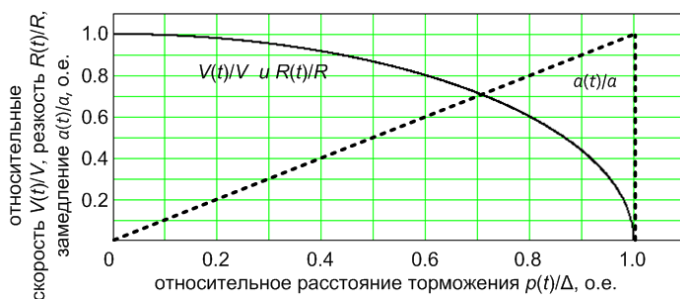


Рис. 6. Зависимости относительной скорости, замедления и резкости движения от относительного тормозного пути

3) В первую часть (10) подставим $\omega t = \arcsin(p(t)/\Delta)$ и $\omega^2 = a/\Delta$ из (15), получим: $a(t) = a \cdot p(t)/\Delta$, которую переведем в относительную, безразмерную форму (рисунок 6 и Примечание 2):

$$\frac{a(t)}{a} = \frac{p(t)}{\Delta}. \quad (17)$$

Во вторую часть (10) подставим $\omega t = \arcsin(p(t)/\Delta)$, $\omega^3 = R/\Delta$ из (15), получим: $R(t) = R[1 - (p(t)/\Delta)^2]^{1/2}$, которую переведем в относительную, безразмерную форму (рисунок 6 и Примечание 1):

$$\frac{R(t)}{R} = \left[1 - \left(\frac{p(t)}{\Delta} \right)^2 \right]^{1/2}. \quad (18)$$

Для примера расчёта примем предельные *исходные значения*: тормозной путь $\Delta = 1500$ м, скорость в начале торможения: $V = 500$ км/ч = 138.8889 м/с. Определим *параметры*

торможения: необходимое максимальное замедление состава по (13) и условию $p(t=T_{\text{т}}) = \Delta$: $a = -V^2/\Delta = -12.86 \text{ м/с}^2$, время торможения (14): $T_{\text{т}} = \pi\Delta/2V = 16.96 \text{ с}$, $R = a^2/V = 1.19 \text{ м/с}^3$.

3.4. Связь величины торможения (замедления) и разрешаемой скорости состава. Процесс экстренного торможения состава можно представить математически, как зависимость контролируемой (разрешаемой) скорости перемещения состава $V(t)$ по тормозному участку пути под действием силы инерции и усилия тормозов состава, которое определяется величиной замедления $a(t)$. Для чего (17) подставим в (16), получим два эквивалентных выражения:

$$\frac{V(t)}{V} = \left[1 - \left(\frac{a(t)}{a} \right)^2 \right]^{1/2}; \quad \frac{a(t)}{a} = \left[1 - \left(\frac{V(t)}{V} \right)^2 \right]^{1/2}. \quad (19)$$

На рисунке 7 приведена зависимость по выражениям (19), с учётом Примечания 1.

Тормозное усилие пропорционально замедлению [1], поэтому из (19) следует Правило 1: «Если замедление $a(t)$ изменять плавно по формуле (19), то скорость $V(t)$ будет изменяться плавно, так как функция (19) не имеет разрывов, как в функциях поправок $A_{\text{т}}$ ($A_{\text{brake_safe}}$) и $A_{\text{вс}}$ (A_{gradient}), характерных для кривой EBD ETCS».



Рис. 7. Зависимость относительной скорости состава от замедления и обратная зависимость

Основной недостаток гармонической кривой торможения по выражению (19) следует из рисунков 6 и 7, где наблюдается положительный скачок замедления от $a = -12.86 \text{ м/с}^2$ до $a = 0 \text{ м/с}^2$ в

конце тормозного пути $p(T_{\text{бр}}) = \Delta = 1500$ м при начальной скорости торможения $V = 500$ км/ч = 138.8889 м/с. При этом цель торможения (остановка состава) достигается: $V(T_{\text{бр}}) = 0$.

Покажем, что предлагаемая ниже гармоническая тормозная кривая позволяет устранить эффект скачка замедления.

4. Методика разработки модели гармонической полуволны и применение её для описания безударных кривых экстренного торможения. Методика проектирования включает в себя два основных этапа: этап получения кривой экстренного торможения в виде функции времени для определения временных параметров процесса торможения и этап получения кривой экстренного торможения в виде функции расстояния для безударного управления торможением.

4.1. Определение параметров торможения, в виде функций времени.

1) Математическое описание изменения скорости состава $V(t)$ во время торможения запишем, в соответствии с Правилom 1, как гармоническую полуволну в виде косинусоиды (рисунок 8), которая при $t = 0$ принимает максимальное положительное значение $V(0) = V$, а при $t = T_{\text{бр}}$ равна нулю $V(T_{\text{бр}}) = 0$:

$$V(t) = 0.5V \cos \omega t + 0.5V, \quad (20)$$

где $0 \leq t \leq T_{\text{бр}}$ – интервал времени безударного, экстренного, гармонического торможения, $T_{\text{бр}} = 0.5T$; T – период частоты среза или период гармонической волны, круговая частота которой $\omega = 2\pi/T = \pi/T_{\text{бр}}$.

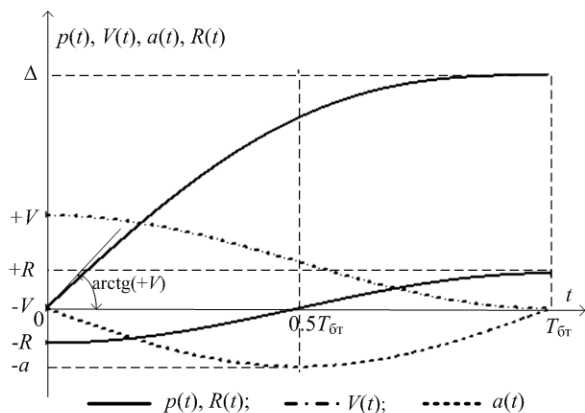


Рис. 8. Безударная кривая экстренного торможения $p(t)$ (23) и три производных от неё по времени (20), (21), (25)

2) Определим замедление состава $a(t)$ (или отрицательное ускорение) торможения, как производную от скорости торможения (20) по времени: $[V(t)]^{(1)} = a(t) = -0.5\omega V \sin \omega t$, в которую подставим $\omega = \pi/T_{\text{бр}}$ и $t = 0.5T_{\text{бр}}$, при этом учтём, что минимальное значение $a(0.5T_{\text{бр}}) = -a$ определяется максимальным тормозным усилием состава, получим:

$$a(t) = -a \sin \omega t, \quad (21)$$

где $a = \pi V/2T_{\text{бр}}$ – максимальное по величине значение замедления состава в момент времени $t = 0.5T_{\text{бр}}$. Из (20) и (21), при $t = 0.5T_{\text{бр}}$, запишем величину интервала времени экстренного, безударного торможения:

$$T_{\text{бр}} = \frac{\pi}{2} \frac{V}{a}. \quad (22)$$

3) Определим траекторию торможения, как неопределённый интеграл от скорости $V(t)$ (20), при этом учтём, что постоянная интегрирования C определяется из условия $p(t=0) = 0$ – величина пути в момент начала торможения, получим:

$$p(t) = \int V(t)dt = 0.5 \frac{V}{\omega} \sin \omega t + 0.5Vt. \quad (23)$$

Из (23) и условия $p(t = T_{\text{бр}}) = \Delta$ – максимальный путь в момент конца торможения определим время экстренного торможения:

$$T_{\text{бр}} = 2 \frac{\Delta}{V}. \quad (24)$$

Резкость торможения определим, как производную от закона замедления состава (20), получим:

$$R(t) = (a(t))^{(1)} = -0.5\omega^2 V \cos \omega t. \quad (25)$$

Из (25) и условия $R(t = T_{\text{бр}}, t = 0) = \pm R$ – максимальная скорость изменения замедления (резкость) в конце и начале торможения определим время экстренного, безударного торможения:

$$T_{\text{бр}} = \pi \left(\frac{V}{2R} \right)^{1/2}. \quad (26)$$

4) Из (22), (24) и (22), (26) для обеспечения однозначного значения интервала экстренного торможения $T_{\text{бр}}$, получим два частных условия согласования части параметров процесса торможения:

$$V^2 = \frac{4}{\pi} \Delta a, \quad a^2 = \frac{1}{2} VR. \quad (27)$$

Перемножение правых и левых частей выражений (27) даёт общее условие согласования всех параметров процесса торможения:

$$2\Delta R = \pi Va. \quad (28)$$

5) Совместное преобразование выражений (22), (24), (25) и (26) позволяет записать цепь равных элементов для величины интервала времени экстренного торможения по безударной траектории:

$$T_{\text{бр}} = 2 \frac{\Delta}{V} = \frac{\pi}{2} \frac{V}{a} = \left(\pi \frac{\Delta}{a} \right)^{1/2} = \pi \left(\frac{V}{2R} \right)^{1/2} = \pi \frac{a}{R} = \left(\pi^2 \frac{\Delta}{R} \right)^{1/3}. \quad (29)$$

С учётом круговой частоты $\omega = 2\pi/T = \pi/T_{\text{бр}}$ из (29) запишем:

$$\omega = \frac{\pi}{2} \frac{V}{\Delta} = 2 \frac{a}{V} = \left(\pi \frac{a}{\Delta} \right)^{1/2} = \left(\frac{2R}{V} \right)^{1/2} = \frac{R}{a} = \left(\pi \frac{R}{\Delta} \right)^{1/3}. \quad (30)$$

6) Для примера расчёта примем предельные исходные значения подраздела (2.3): тормозной путь $\Delta = 1500$ м, скорость в начале торможения: $V = 500$ км/ч = 138.8889 м/с. Определим *параметры безударного торможения*: период круговой частоты торможения из выражений (24) или (26) $T = 2T_{\text{бр}} = 4\Delta/V = 43.20$ с; максимальное замедление состава из (20) в середине интервала торможения $a(0.25T) = -\pi V^2/4\Delta = -10.1003$ м/с²; время торможения состава (26) $T_{\text{бр}} = 2\Delta/V = 21.60$ с; максимальная резкость торможения из (25) в конце и начале интервала торможения $R = -\pi^2 V^3/8\Delta^2 = \pm 1.469$ м/с³.

4.2. Методика приведения времени торможения к расстоянию торможения гармонической безударной полуволны.

Кривая экстренного торможения определяет зависимость скорости состава от расстояния, поэтому из выражения (23) необходимо выразить произведение ωt через относительное расстояние $p(t)$, но это действие затруднено, так как (23) является трансцендентным уравнением относительно переменных ωt и t . Для выхода из этой ситуации определим две обратных зависимости: расстояния экстренного торможения от скорости $p(V(t))$ и от ускорения $p(a(t))$.

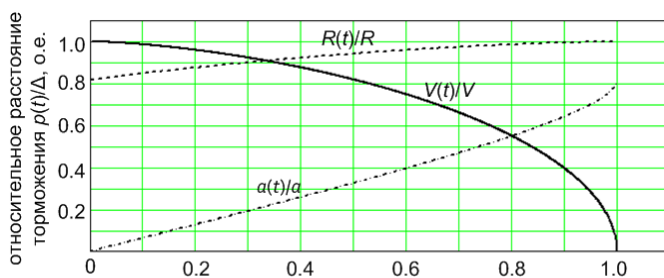
1) Из выражения (21) выделим время торможения состава $t = (1/\omega) \arccos(2V(t)/V) - 1$ и подставим в (23), применим $\omega = \pi V/2\Delta$ из (30), получим:

$$\frac{p(t)}{\Delta} = \frac{2}{\pi} \frac{V(t)}{V} \left(\frac{V}{V(t)} - 1 \right)^{1/2} + \frac{1}{\pi} \arccos \left(2 \frac{V(t)}{V} - 1 \right). \quad (31)$$

На рисунке 9 изображены зависимости относительного расстояния $p(t)/\Delta$ траектории торможения от относительной скорости торможения, замедления, резкости торможения.

2) Из выражения (20) выделим время торможения скоростного состава $t = (-1/\omega) \arccos(2a(t)/\omega V)$ и подставим в (23), применим $\omega^2 = \pi a/\Delta$ из (30), учтём что $a(t) \leq 0$, $a \leq 0$, при $0 \leq t \leq 0.5T_{\text{эт}}$, получим:

$$\frac{p(t)}{\Delta} = \frac{1}{\pi} \frac{a(t)}{a} + \frac{1}{\pi} \arcsin \left(\frac{a(t)}{a} \right). \quad (32)$$



относительные скорость $V(t)/V$, резкость $R(t)/R$, замедление $a(t)/a$, о.е.

Рис. 9. Зависимости тормозного пути от скорости, замедления и резкости движения скоростного состава

3) Из выражения (25) выделим время торможения скоростного состава $t = (1/\omega) \arccos(-2R(t)/R)$ и подставим в (23), применим $\omega = \pi V/2\Delta$ из (30), получим:

$$\frac{p(t)}{\Delta} = \frac{1}{\pi} \left[1 - \left(\frac{R(t)}{R} \right)^2 \right]^{1/2} + \frac{1}{\pi} \arccos \left(-\frac{R(t)}{R} \right). \quad (33)$$

На рисунке 9 изображены зависимости относительного расстояния $p(t)/\Delta$ траектории торможения от относительной скорости торможения $V(t)/V$ (31), от относительной величины замедления $a(t)/a$ (32) и от относительной величины резкости торможения $R(t)/R$ (33). При построении рисунка 9 учтены Примечания 1 и 2 при выборе знака перед выражением корня.

4.3. Связь величин замедления и скорости состава при торможении по гармонической безударной полуволне. Определим время торможения через скорость торможения состава $V(t)$ из выражения (21) $t = (1/\omega) \arccos(2V(t)/V-1)$ и $\omega = 2a/V$ из (30) для преобразования (20):

$$\frac{a(t)}{a} = -2 \frac{V(t)}{V} \left(\frac{V}{V(t)} - 1 \right)^{1/2}. \quad (34)$$

Обратную зависимость получим путём преобразования (34) или подстановкой $\omega t = -\arcsin(a(t)/a)$ из (20) в (21):

$$\frac{V(t)}{V} = 0.5 \pm 0.5 \left[1 - \left(\frac{a(t)}{a} \right)^2 \right]^{1/2}. \quad (35)$$

На рисунке 10 приведены зависимости по выражениям (34), (35). При построении рисунка 10 учтены Примечания 1 и 2 при выборе знака перед выражением корня.

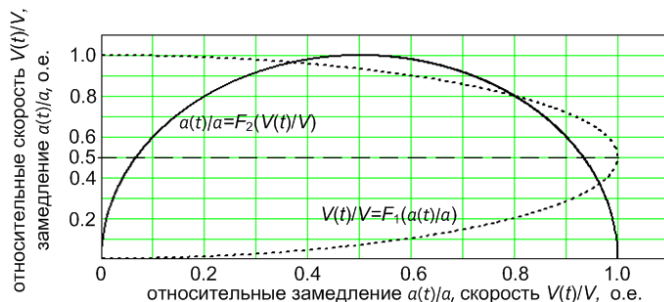


Рис. 10. Зависимость относительной скорости состава от замедления (отрицательного ускорения) F_1 и обратная зависимость F_2

5. Эффективность использования моделей. Результаты исследования можно использовать не только для экономии технического ресурса тормозной системы состава и повышения комфорта пассажиров, но и для сравнения технических возможностей (характеристик) высокоскоростных поездов, использующих различные тормозные кривые. Предположим, что поезда оснащены *одинаковыми* пневматическими и (или) электромагнитными тормозами, на которые задают тормозные усилия в одинаковых диапазонах изменения, но различные по законам изменения (для трёх кривых торможения, рассмотренных выше) в соответствии с полученными тремя группами выражений. При этом выделим две ситуации.

5.1. Сравнение моделей (тормозных кривых) при общем ограничении величины замедления.

1) Для тормозной кривой EBD (Emergency Brake Deceleration) или степенной полуволны второго порядка используем (4), перепишем (5) с присвоением индекса $V_{ст}$: $T_{ст} = (2\Delta/a)^{1/2}$; $V_{ст} = (2\Delta a)^{1/2}$.

2) Для тормозной гармонической кривой с единственным положительным скачком ускорения в конце процесса торможения используем (13), перепишем ограничительную часть выражения (13) с присвоением индекса $V_{гт}$: $T_{гт} = (\pi/2)(\Delta/a)^{1/2}$; $V_{гт} = (\Delta a)^{1/2}$.

3) Для тормозной гармонической кривой без скачков (ударов) замедления в процессе торможения используем (29), перепишем выражение (27) с присвоением индекса $V_{бт}$: $T_{бт} = (\pi\Delta/a)^{1/2}$; $V_{бт} = 2(\Delta a/\pi)^{1/2}$.

В соответствии с записанными и переиндексированными выражениями скоростей приведём зависимости на рисунке 11.

Из зависимостей на рисунке 11 следует, что степенная полуволна, воздействуя на тормоза состава с меньшей амплитудой замедления, останавливает состав быстрее или останавливает за то же

время останова при большей начальной скорости в начале торможения. Для нахождения количественных оценок перейдём к относительным единицам для трёх рассматриваемых кривых торможения.

Запишем соответствующие количественные соотношения выражений из пунктов 1) – 3):

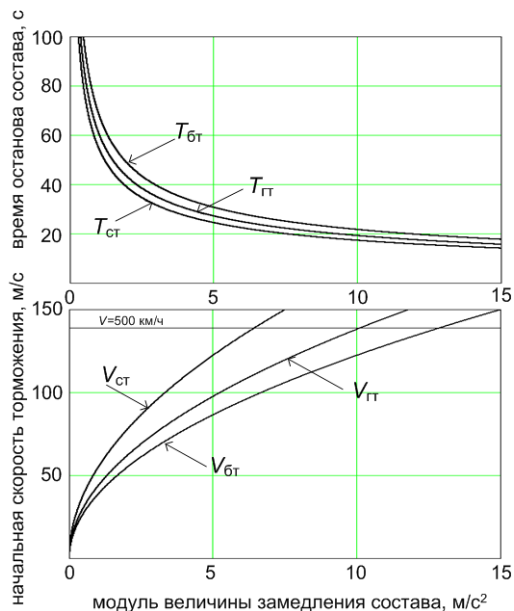


Рис. 11. Зависимости начальной скорости торможения и времени на останов состава от величины замедления

$$\frac{V_{гт}}{V_{ст}} = \frac{(\Delta a)^{1/2}}{(2\Delta a)^{1/2}} = 0.707; \quad \frac{V_{бт}}{V_{ст}} = \frac{2(\Delta a/\pi)^{1/2}}{(2\Delta a)^{1/2}} = 0.798.$$

$$\frac{T_{гт}}{T_{ст}} = \frac{(\pi/2)(\Delta a)^{1/2}}{(2\Delta a)^{1/2}} = 1.111; \quad \frac{T_{бт}}{T_{ст}} = \frac{2(\pi\Delta a)^{1/2}}{(2\Delta a)^{1/2}} = 1.253.$$

Из количественных соотношений видно, что ограничение на величину замедления приводит к необходимости сокращения максимальных скоростей вхождения в режим торможения для гармонических кривых замедления состава. При этом соответственно увеличиваются временные интервалы останова состава.

5.2. Сравнение моделей (тормозных кривых) при пропорциональном уменьшении максимальных величин замедлений. Для примера используем максимальные значения замедлений для трёх типов тормозных кривых, полученных ранее в примерах расчётов параметров торможения: в пункте статьи 2.3 раздела 2: $a_{\text{ст}} = -6.43 \text{ м/с}^2$; в пункте 4.3: $a_{\text{гт}} = -12.86 \text{ м/с}^2$; в пункте 5.1.: $a_{\text{бт}} = -10.10 \text{ м/с}^2$. Выражения для расчёта заимствуем из пункта 6.1, в которых используем индексированные, в части написания обозначения замедлений, выражения.

$$\frac{V_{\text{гт}}}{V_{\text{ст}}} = \frac{(a_{\text{гт}})^{1/2}}{(2a_{\text{ст}})^{1/2}} = \frac{(12.86)^{1/2}}{(2 \cdot 6.43)^{1/2}} = 1; \quad \frac{V_{\text{бт}}}{V_{\text{ст}}} = \frac{(2a_{\text{бт}})^{1/2}}{(\pi a_{\text{ст}})^{1/2}} = \frac{(2 \cdot 10.1)^{1/2}}{(\pi \cdot 6.43)^{1/2}} = 1;$$

$$\frac{T_{\text{гт}}}{T_{\text{ст}}} = \frac{\pi(a_{\text{ст}})^{1/2}}{2(2a_{\text{гт}})^{1/2}} = \frac{\pi(6.43)^{1/2}}{(8 \cdot 12.86)^{1/2}} = 0.785; \quad \frac{T_{\text{бт}}}{T_{\text{ст}}} = \frac{(\pi a_{\text{ст}})^{1/2}}{(2a_{\text{бт}})^{1/2}} = \frac{(\pi \cdot 6.43)^{1/2}}{(2 \cdot 10.1)^{1/2}} = 1.$$

Из последних количественных соотношений видно, что отсутствие ограничений на величины замедлений приводят к более оптимистичным выводам, чем в пункте 6.1:

- отсутствует уменьшение максимальных скоростей вхождения в режим торможения для гармонических кривых замедления состава;

- временные интервалы останова состава сохраняются, а в случае использования гармонической кривой торможения с импульсом замедления в конце интервала торможения, временной интервал останова меньше, чем у кривой EBD ETCS $T_{\text{гт}} < T_{\text{ст}}$.

При этом если понизить величины максимальных замедлений в процессах торможения состава для всех трёх рассмотренных моделей пропорционально коэффициенту $k \leq 1$, то положительные результаты не исчезнут, так как в последних количественных соотношениях после подстановки коэффициенты k взаимно сократятся.

6. Заключение. Математические модели на основе гармонических полуволн для совершенствования процессов торможения высокоскоростных поездов позволяют исследовать недостатки Европейской системы управления поездами (European Train Control System (ETCS)): импульсное, многократное срабатывание тормозов при экстренном торможении, что приводит к износу тормозной системы и снижает комфорт пассажиров многократным повторением перегрузок.

Доказано, что математической моделью кривых торможения ETCS является степенная полуволна или авторский обобщённый критерий наибольшего отклонения (2005 год) [8]. Результаты доказательства показывают, что математическая модель обобщённого критерия наибольшего отклонения является актуальной и применима для моделирования процесса торможения высокоскоростных поездов. Эта модель определяет новые свойства и методы построения изменения скорости и замедления высокоскоростного состава, полностью отвечающие требованиям ETCS. Сверх того позволяет осуществлять плавное регулирование процесса торможения, что актуально для новейших поездов типа «маглев». Описанная методика, в свою очередь, позволила разработать две новые модели на основе гармонических полувольт, которые свободны от многократного, импульсного срабатывания тормозов, обеспечивают непрерывное регулирование скорости, приводят время торможения к интервалу (расстоянию) торможения гармонической полуволны. Первая из двух новых гармонических полувольт характеризуется однократным импульсным отпусканием тормоза в конце торможения и равномерным нарастанием в процессе торможения. Этот результат можно использовать, например, в системах управления высокоскоростными лифтами высотных зданий. При этом можно обеспечить постепенное линейное увеличение ускорения. Вторая новая гармоническая полуволна имеет замедление по закону отрицательной синусоиды. Из этого закона путём операций интегрирования и дифференцирования получены законы пути торможения, скорости и резкости торможения свободные от необходимости импульсных тормозных воздействий, что повышает комфорт пассажиров и сохраняет ресурс тормозных систем поездов типа «маглев» и обычных высокоскоростных поездов с магнитными тормозами.

Таким образом, результаты статьи могут быть использованы для моделирования режима реверса оптико-электронных систем при наблюдении и сопровождении подвижных объектов, дистанционного и автономного управления плавным перемещением роботов и их элементов с сохранением максимальных скоростей, ускорений и резкостей движения. Применение общего подхода к моделям открывает перспективы управления движениями роботов и механизмов с необходимой точностью, плавностью, обеспечивает их работу в реальном времени.

Литература

1. Ehret M. Virtual Train Brakes // Eisenbahnwesen-Seminar, TU-Berlin. 2020. 49 p.
2. ERTMS/ETCS. System Requirements Specification. Chapter 6. Management of older System Versions. 2016. no. 3.6.0. 46 p.
3. Introduction to ETCS Braking Curves. Version 1.5. ERA ERTMS Unit. 2020. 28 p.
4. «СИНКАНСЭН»: как работают скоростные поезда в Японии. URL: <https://varlamov.ru/3259684.html> (дата обращения: 26.07.2023).
5. Performance and Financing Agreement – Infrastructure Condition and Development Report. Deutsche Bahn AG Group, 2019. Available at: https://www.eba.bund.de/DE/Themen/Finanzierung/LuFV/IZB/izb_node.html;jssessionid=9FA5FFD4733F0B2522224C879CABE1A9.live11292 (accessed: 10.09.2023).
6. Furness N., van Houten H., Arenas L., Bartholomeus M. ERTMS Level 3: the Game-Chander. IRSE NEWS. 2017. vol. 232. pp. 2–9.
7. Thales consortium to install ECTS Level 1 ATR system in Thailand. Railway Technology. 2019. Available at: <https://www.railway-technology.com/news/thales-consortium-to-install-ets-level-1-atp-system-in-thailand/?cf-view> (accessed: 10.09.2023).
8. Майоров Б.Г. Обобщенный критерий наибольшего отклонения входных сигналов систем управления // Автоматика и Телемеханика. 2005. № 10. pp. 148–157.
9. Фельдбаум А.А. Оптимальные процессы в системах автоматического регулирования // Автоматика и телемеханика. 1953. Т. 14. № 6. С. 712–728.
10. Никифоров В.В., Баранов С.Н. Статическая проверка корректности разделения ресурсов в системах реального времени // Труды СПИИРАН. 2017. Т. 3. № 52. С. 137–156.
11. Ровелли К. Срок времени / Пер. с итал. Д. Баюка // М.: ООО «Издательство АСТ»: CORPUS. 2020. 224 с.
12. Сенкевич Ю.И., Марапулец Ю.В., Луковенкова О.О., Солодчук А.А. Методика выделения информативных признаков в сигналах геоакустической эмиссии // Труды СПИИРАН. 2019. Т. 18. № 5. С. 1066–1092.
13. Петренко В.И., Тебужева Ф.Б., Гурчинский М.М., Антонов В.О., Павлов А.С. Прогнозная оценка траектории руки оператора для решения обратной задачи динамики при копирующем управлении // Труды СПИИРАН. 2019. Т. 18. № 1. С. 123–147.
14. Demydyuk M.V., Hoshovs'ka N. Parametric optimization of the transport operations of a two-link manipulator // Journal of Mathematical Sciences. 2019. vol. 238. pp. 174–188.
15. Manyam S.G. Optimal dubins paths to intercept a moving target on a circle // Proc. of the American Control Conference. 2019. pp. 828–834.
16. Poulsen D.R., Davis J.M., Gravagne I.A. Optimal Control on Stochastic Time Scales // IFAC-PapersOnLine. 2017. vol. 50. no. 1. pp. 14861–14866.
17. Майоров Б.Г. Способ непрерывной оптической связи с низколетящей целью // Патент RU №2 715 499 C1. 28.02.2020.
18. Григорьев И.С., Данилина И.А. Оптимизация траектории перелётов космических аппаратов с дополнительным топливным баком. II // Автоматика и Телемеханика. 2018. № 2. С. 135–153.
19. Хлистунов В.Н. О погрешности аппроксимации дискретных методов измерения // Приборостроение. 1960. № 5. С. 3–5.
20. Vavilov S.A., Lytaev M.S. Calibration and verification of models defining radar visibility zones in marine geoinformation systems // Proceedings of the 8th international Symposium on Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17). 2018. pp. 115–125.
21. Uneyama T., Miyaguchi T., Akimoto T. Relaxation Functions of the Ornstein Uhlenbeck Process with Fluctuating Diffusivity. Physical Review E. 2019. vol. 99. no. 3. DOI:10.1103/PhysRevE.99.032127.

22. Майоров Б.Г. Свойства гармонических и составных полувольт, определение интервала равномерной дискретизации времени цифровых сигнальных процессоров // Информатика и автоматизация. 2022. Т. 21. № 1. С. 95–125. DOI: 10.15622/ia.2022.21.4.
23. Майоров Б.Г. Способ определения координат подвижного объекта в закрытых помещениях // Патент RU 2 284 542 C2. 30.08.2004.

Майоров Борис Геннадьевич — канд. техн. наук, ведущий научный сотрудник, центр специальных систем, АО "Научно-производственное предприятие "Рубин". Область научных интересов: цифровая обработка сигналов, системы сбора и отображения информации в реальном времени. Число научных публикаций — 87. bgmayorov@yandex.ru; улица Байдукова, 2, 440015, Пенза, Россия; р.т.: 7(8412)204-777.

Поддержка исследований. Работа выполнена по плану научно-технического развития предприятия 2023 года.

B. MAYOROV

**APPLICATION OF HARMONIC HALF-WAVES FOR
AUTOMATION OF HIGH-SPEED TRAINS CONTROL**

Mayorov B. Application of Harmonic Half-Waves for Automation of High-Speed Trains Control.

Abstract. The emergency braking processes in the European Train Control System (ETCS) of high-speed trains are associated with stepwise regulation of acceleration (deceleration) depending on the braking ability of the train, terrain data and changing weather on the route. These processes are defined in ETCS. The procedure for stepwise regulation of deceleration is carried out by the driver repeatedly in the process of braking until the train stops completely. The beginning of emergency braking and its end, as well as the braking process itself, is accompanied by repeated pulsed operation of the brakes, which leads to jumps in deceleration and, accordingly, to increased wear of the brake system, a decrease in comfort for passengers, which results in the limitation of the maximum allowable speed. The article proposes a new concept and technique for constructing mathematical models of emergency braking curves different from ETCS curves and based on harmonic half-waves. It is shown that the ETCS deceleration curves are described by known second-order power half-waves. Their joint study gives grounds to assert that the application of these curves leads to the obligatory pulsed mode of brake operation. Two new variants of models of emergency braking curves described by harmonic half-waves are proposed. The first option has one pulsed brake application at the end of the braking interval. The second option is free from braking impulses and allows the use of continuous regulation. These models explain the features of ETCS, contain proposals for their elimination, and are applicable to the development of new emergency braking curves that allow smooth control of emergency braking of trains. Efficiency, differences and advantages over ETCS braking curves are shown on the results of mathematical modeling of emergency braking processes.

Keywords: high-speed trains, maximum speed, emergency braking, emergency braking curves, smooth train deceleration control, harmonic half-waves, method of constructing mathematical models of emergency braking curves.

References

1. Ehret M. Virtual Train Brakes. Eisenbahnwesen-Seminar, TU-Berlin. 2020. 49 p.
2. ERTMS/ETCS. System Requirements Specification. Chapter 6. Management of older System Versions. 2016. no. 3.6.0. 46 p.
3. Introduction to ETCS Braking Curves. Version 1.5. ERA ERTMS Unit. 2020. 28 p.
4. “SINKANSEN”: kak rabotayut poezda v Yaponii [“SHINKANSEN”: how trains work in Japan]. Available at: [https:// varlamov.ru/3259684. html](https://varlamov.ru/3259684.html) (accessed: 26.07.2023). (In Russ.).
5. Performance and Financing Agreement – Infrastructure Condition and Development Report. Deutsche Bahn AG Group, 2019. Available at: https://www.eba.bund.de/DE/Themen/Finanzierung/LuFV/IZB/izb_node.html;jsessionid=9FA5FFD4733F0B252224C879CABE1A9.live!1292 (accessed: 10.09.2023).
6. Furness N., van Houten H., Arenas L., Bartholomeus M. ERTMS Level 3: the Game-Changer. IRSE NEWS. 2017. vol. 232. pp. 2–9.
7. Thales consortium to install ECTS Level 1 ATR system in Thailand. Railway Technology. 2019. Available at: <https://www.railway-technology.com/news/thales-consortium-to-install-etcs-level-1-atp-system-in-thailand/?cf-view> (accessed: 26.07.2023).

8. Maierov B.G. The generalized maximum deviation criterion for input signals of control systems. *Autom. Remote Control*. 2005. vol. 66. pp. 1666–1672. DOI: 10.1007/s10513-005-0199-6.
9. Feldbaum A.A. [Optimal processes in automatic control systems]. *Avtomatika i telemekhanika – Automation and telemechanics*. 1953. vol. 14. no. 6. pp. 712–728.
10. Nikiforov V.V., Baranov S.N. [Static verification of the correctness of resource sharing in real-time systems]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2017. vol. 3. no. 52. pp. 137–156.
11. Rovelli K. *Srok vremeni [Time period]*. Moscow: Publishing house ACT: CORPUS, 2020. 224 p. (In Russ.).
12. Senkevich Yu.I., Marapulez Yu.V., Lukovenkova O.O., Solodchuk A.A. [Technique for identifying informative features in geoacoustic emission signals]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2019. vol. 18. no. 5. pp. 1066–1092.
13. Petrenko V., Tebueva F., Gurchinsky M., Antonov V., Pavlov A. [Predictive estimate of the trajectory of the operator's hand for solving the inverse problem of dynamics with copying control]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2019. vol. 18. no. 1. pp. 123–147.
14. Demydyuk M.V., Hoshovs'ka N. Parametric optimization of the transport operations of a two-link manipulator. *Journal of Mathematical Sciences*. 2019. vol. 238. pp. 174–188.
15. Manyam S.G. Optimal dubins paths to intercept a moving target on a circle. *Proc. of the American Control Conference*. 2019. pp. 828–834.
16. Poulsen D.R., Davis J.M., Gravagne I.A. Optimal Control on Stochastic Time Scales. *IFAC-PapersOnLine*. 2017. vol. 50. no. 1. pp. 14861–14866.
17. Mayorov B.G. [Method of continuous optical communication with low-flying target] Patent RU. No. 2 715 499 C1. 02.28.2020. (In Russ.).
18. Grigoriev I.S., Danilina L.A. Optimising flight for space vehicles with an additional fuel tank. II. *Autom. Remote Control*. 2018. vol. 79. no. 2. pp. 311–326.
19. Khlistunov V.N. [On the error of approximation of discrete measurement methods]. *Priborostroenie – Instrumentation*. 1960. no. 5. pp. 3–5.
20. Vavilov S.A., Lytaev M.S. Calibration and verification of models defining radar visibility zones in marine geoinformation systems. *Proceedings of the 8th international Symposium on Information Fusion and Intelligent Geographic Information Systems (IF&IGIS'17)*. 2018. pp. 115–125.
21. Uneyama T., Miyaguchi T., Akimoto T. Relaxation Functions of the Ornstein Uhlenbeck Process with Fluctuating Diffusivity. *Physical Review E*. 2019. vol. 99. no. 3. DOI:10.1103/PhysRevE.99.032127.
22. Mayorov B.G. [Properties of harmonic and composite half-waves, determination of the uniform time sampling interval of digital signal processors]. *Informatika i avtomatizaciya – Informatics and Automation*. 2022. vol. 21. no. 1. pp. 95–125. DOI: 10.15622/ia.2022.21.4. (In Russ.).
23. Mayorov B.G. [Method for determining the coordinates of a moving object in enclosed spaces] Patent RU 2 284 542 C2. 30.08.2004. (In Russ.).

Mayorov Boris — Ph.D., Leading researcher, Center for special systems, Joint Stock Company Scientific and Production Enterprise «Rubin». Research interests: digital signal processing, systems for collecting and displaying information in real time. The number of publications — 87. bgmayorov@yandex.ru; 2, Baidukova St., 440015, Penza, Russia; office phone: 7(8412)204-777.

Acknowledgements. The work was carried out according to the plan of scientific and technical development of the enterprise in 2023.

К.В. КРОТОВ

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ
ВЫПОЛНЕНИЯ ПАКЕТОВ ЗАДАНИЙ В КОНВЕЙЕРНЫХ
СИСТЕМАХ С ПРОМЕЖУТОЧНЫМИ БУФЕРАМИ
ОГРАНИЧЕННЫХ РАЗМЕРОВ**

Кротов К.В. Математическое моделирование процессов выполнения пакетов заданий в конвейерных системах с промежуточными буферами ограниченных размеров.

Аннотация. Современные методы планирования процессов в конвейерных системах с буферами определенного размера между обрабатывающими приборами позволяют оптимизировать расписания выполнения единичных заданий либо фиксированных пакетов заданий при их ограниченном количестве и ограниченном количестве приборов. Применение математических моделей процессов выполнения единичных заданий (пакетов заданий), используемых этими методами, при оптимизации составов пакетов и расписаний их выполнения в системах с произвольным количеством пакетов и приборов является невозможным. В то же время математические модели процессов выполнения пакетов заданий в конвейерных системах при наличии буферов ограниченных размеров между приборами являются основой для разработки методов оптимизации их (пакетов) составов и расписаний реализации действий с ними на приборах конвейерных систем. В связи с этим в статье разработаны математические модели многостадийных процессов выполнения произвольного количества пакетов заданий в конвейерных системах при наличии промежуточных буферов ограниченных размеров для двух и трех приборов, а также для произвольного количества приборов. Использование этих моделей позволяет определять моменты времени начала выполнения пакетов заданий на приборах конвейерных систем с учетом ограниченных размеров промежуточных буферов, а также длительности интервалов времени использования этих ресурсов и эффективность их использования в течение времени. Также разработан алгоритм математического моделирования процессов выполнения пакетов заданий в конвейерных системах при наличии промежуточных буферов ограниченных размеров, осуществляющий на основе заданного порядка реализации действий с пакетами заданий на приборах конвейерных систем вычисление временных характеристик этих процессов. Осуществлена разработка приложения, реализующего синтезированные математические модели процессов выполнения пакетов заданий в конвейерных системах с промежуточными буферами ограниченных размеров и соответствующий метод моделирования этих процессов. Разностороннее тестирование разработанного приложения показало, что полученные математические модели и метод моделирования адекватно описывают ход многостадийных процессов выполнения пакетов заданий в конвейерных системах, задаваемый с использованием различных значений их (процессов) параметров.

Ключевые слова: конвейерные системы, обрабатывающие приборы, промежуточные буферы, пакеты заданий, многостадийные процессы, расписания, математическое моделирование.

1. Введение. Обрабатывающая система конвейерного типа представляет собой последовательность приборов, выполняющих задания в соответствии со строго заданным маршрутом их перемещения между приборами. Процесс выполнения заданий в конвейерных системах (КС), характеризуемый наличием стадий

(этапов) реализации действий с ними на приборах этих систем, называется многостадийным процессом. Выполнение заданий в конвейерных системах характеризуется длительностями: 1) выполнения заданий на приборах КС; 2) длительностями переналадки приборов КС на выполнение заданий. С учетом указанных параметров задания могут быть типизированы следующим образом: к одному типу относятся задания, имеющие равные значения длительностей их выполнения на приборах КС и длительностей переналадки приборов на реализацию действий с ними. Повышение эффективности выполнения заданий разных типов в КС связано с формированием пакетов. Под пакетом заданий (ПЗ) подразумевается совокупность заданий одного типа, реализация действий с которыми на приборах КС не связана с их переналадкой. Автором предложен метод многоуровневой оптимизации, позволивший реализовать в составе иерархической системы решение задач: 1) оптимизации расписаний многостадийных процессов выполнения ПЗ в КС [1] посредством поиска локально оптимальных решений по их составам на верхнем уровне и поиска локально оптимальных решений по порядкам их выполнения на приборах на нижнем уровне; 2) оптимизации расписаний многостадийных процессов выполнения ПЗ в КС при формировании комплектов из результатов [2 – 3] посредством поиска локально оптимальных решений по составам ПЗ и по порядкам реализации действий с ними на приборах на соответствующих уровнях иерархии; 3) оптимизации расписаний процессов выполнения ПЗ в КС при задании ограничения на длительности интервалов времени их функционирования [4] посредством поиска локально оптимальных решений по: составам ПЗ на первом уровне, составам групп пакетов, выполняемых в течение интервалов времени функционирования КС ограниченной длительности, на втором уровне, порядкам выполнения пакетов из групп на приборах на нижнем (третьем) уровне.

Развитие иерархического подхода связано с решением задачи оптимизации расписаний многостадийных процессов выполнения ПЗ в КС при задании ограничения на другой вид ресурсов, используемых при реализации действий с пакетами в КС – размеров буферов, находящихся между обрабатывающими приборами. В них размещаются задания, входящие в пакеты, после завершения каждой стадии многостадийных процессов реализации действий с ними на приборах КС.

С использованием иерархического подхода решение задачи оптимизации расписаний многостадийных процессов выполнения ПЗ в

КС при ограничении на размеры промежуточных буферов осуществляется путем оптимизации решений по составам пакетов на верхнем уровне и оптимизации расписаний реализации действий с ними на приборах указанных систем на нижнем уровне. Оптимизация порядков реализации действий с заданиями на приборах КС осуществляется на основе математической модели многостадийных процессов выполнения ПЗ в КС при учете ограничения на размеры буферов. Задача разработки математической модели многостадийных процессов выполнения ПЗ в КС при ограниченных размерах промежуточных буферов в настоящее время не решена требуемым образом с точки зрения возможности ее использования при оптимизации составов пакетов и расписаний. В связи с этим разработка математической модели процессов выполнения ПЗ в КС при наличии между обрабатываемыми приборами буферов ограниченных размеров является актуальной.

Одной из первых работ по планированию процессов выполнения единичных заданий (ЕЗ) в системе Flow Shop (FS) при ограничении на размер b промежуточного буфера является работа [5]. В ней рассматривается задача построения расписаний выполнения ЕЗ в системе с двумя приборами и одним промежуточным буфером единичного размера. Доказывается, что задача является NP-трудной, а также исследуются свойства процесса выполнения заданий, связывающие моменты времени начала/окончания выполнения i -х заданий на первом приборе и $(i-b-1)$ -х заданий на втором приборе. В [6] введена в рассмотрение модель процесса выполнения ЕЗ на приборах КС при ограниченных в b единиц размерах буферов, позволяющая формировать моменты времени начала реализации действий с i -ми заданиями на l -х приборах, учитывающая моменты времени окончания выполнения этих заданий на $(l-1)$ -х приборах, предшествующих им $(i-1)$ -х заданий на l -х приборах, а также моменты времени начала выполнения $(i-b-1)$ -х заданий на $(l+1)$ -х приборах. При этом способы вычисления моментов времени начала выполнения i -х заданий на первом и последнем приборах этой моделью не определяются. Использование предложенных в [6] эвристических правил позволяет формировать порядок выполнения ЕЗ на приборах КС с учетом промежуточных буферов ограниченных размеров. В работе [7] рассматривается система Flexible Flow Shop (FFS), представляющая собой последовательность из машин с параллельно функционирующими приборами (Batch Processing Machine (BPM)) и промежуточными буферами между ними. Приборы в BPM разной производительности. В связи с этим возникает задача выбора для каждого задания

обрабатывающего прибора, входящего в BPM, таким образом, чтобы минимизировать общее время реализации их выполнения. Разработана модель MILP (mixed-integer linear programming), в которой процессу выполнения ЕЗ на приборах КС при заданных размерах буферов соответствует система ограничений на множество допустимых решений. В [8] рассматривается обрабатывающая система, представляющая собой последовательность BPM и промежуточных буферов между ними. Постановка задачи интерпретирует рассматриваемую систему как FFS, что предусматривает выбор в каждой BPM приборов, реализующих выполнение ЕЗ. Для оптимизации маршрутов выполнения ЕЗ в FFS и расписаний проведения действий с ними на приборах разработан алгоритм имитации поведения стаи китов (Whale Optimization Algorithm (WOA)), для реализации которого разработана математическая модель рассматриваемого процесса в системе указанного типа. В [9] для решения рассматриваемой задачи оптимизации расписаний выполнения ЕЗ в КС применены метаэвристические алгоритмы (в частности, алгоритм оценки распределения – Estimation of Distribution Algorithm (EDA)). При этом используется математическая модель определения моментов времени начала выполнения каждого i -го ЕЗ на l -х приборах КС ($l \neq 1$, $l \neq L$, где L – идентификатор последнего прибора в КС), аналогичная [6]. Построению модели MILP оптимизации расписаний выполнения ЕЗ в системах FS на примере металлургического производства посвящена работа [10]. Наличие в системе буферов ограниченных размеров учитывается в ограничениях разработанной модели MILP. Для понижения вычислительной сложности задачи в [10] реализуется релаксация ограничений модели MILP. Рассматриваемая в [11] система выполнения ЕЗ представляет собой последовательность из одной BPM и одной Job Processing Machine (JPM), в которой задания из пакетов выполняются последовательно. Размер промежуточного буфера кратен количеству приборов в BPM. Для оптимизации расписаний выполнения ЕЗ в КС в [11] введена модель MILP, в которой наличие промежуточного буфера ограниченного размера и заданная конфигурация системы (в виде последовательности BPM и JPM) определяются видом ограничений на множество допустимых решений. В [12] рассматривается задача оптимизации расписаний выполнения ЕЗ в КС, состоящей из двух приборов и цифрового промежуточного буфера между ними. Первый прибор обеспечивает получение файла для воспроизведения и размещение его в буфере, второй – воспроизведение файла. В [12] сопоставляется длительность выполнения i -й работы по получению файла и его объем (единица объема данных загружается в единицу времени). В выражениях для расчета моментов времени начала

выполнения заданий по получению файла его объем интерпретируется как длительность закачки. Аналогично интерпретируется и размер буфера – максимальное количество данных в буфере сопоставляется с длительностью интервала их получения. Также размер данных в буфере интерпретируется как длительность интервала времени воспроизведения находящегося в нем файла. В силу указанных особенностей введенная в [12] модель имеет ограничения на ее использование, вытекающее из особенностей интерпретации входных параметров. Рассматриваемая в [13] система представляет собой две параллельно функционирующие системы Flow Shop с общими рабочими станциями и промежуточными буферами ограниченного размера между приборами. Для оптимизации расписаний выполнения ЕЗ в системе указанного вида разработана модель MILP, учитывающая наличие единичных промежуточных буферов между приборами в параллельных потоках выполнения заданий и общих рабочих станций, которые разделяются между заданиями. Оптимизация расписаний реализуется с использованием модифицированного алгоритма NEH (авторы алгоритма Nawaz, Enscore, Ham – NEH). Задача, рассматриваемая в [14], аналогична задаче с промежуточным буфером ограниченного размера. В ней буфер является выходным и интерпретируется как склад готовой продукции. Обработка заданий может приводить к заполнению склада, продукция из которого извлекается в директивные сроки, что приводит к блокированию обрабатывающего прибора. С целью решения указанной задачи синтезирована модель MILP с нелинейными ограничениями. В качестве целевой функции выступают суммарные штрафы за хранение продукции на складе и за нарушение директивных сроков. В [15] рассматривается система FS гибридного типа с повторной обработкой на приборах и буферами ограниченных размеров (задания повторно выполняются на одном и том же приборе в соответствии с технологическими маршрутами). Сформирована оптимизационная модель построения расписаний, в которой целевая функция имеет вид взвешенного аддитивного критерия, учитывающего время окончания выполнения всех заданий в системе и среднее время выполнения заданий, а дополнительные условия определяют: 1) необходимость повторного направления заданий на приборы в соответствии с технологическими маршрутами; 2) запрет на прерывание выполнения задания на приборе; 3) ограничения на размеры промежуточных буферов. В [16] рассматривается задача оптимизации расписаний выполнения ЕЗ в FS, содержащей 3 прибора, с нулевым буфером между приборами, ограниченным количеством ЕЗ (не более 9) и требованием минимизации блокирования предшествующих машин при

невозможности последующих машин приступить к выполнению заданий, действия с которыми завершены на предыдущем этапе. Представлены модели MILP, различающиеся способами вычисления момента времени окончания выполнения всех заданий в системе. Результаты, полученные с использованием указанных моделей, сравниваются с результатами, полученными с использованием жадного алгоритма, предложенного в работе. В [17] рассматривается задача оптимизации расписаний выполнения ЕЗ в FS с двумя альтернативными маршрутами. Разработана модель процесса выполнения заданий в FS с отсутствующими операциями (операциями с нулевой длительностью). Для оптимизации решений разработан алгоритм дискретной китовой стаи (DWSA). Используется модель процесса выполнения заданий в FS, аналогичная моделям из [6, 9].

В [18] решается задача оптимизации расписаний в системах Flexible Job Shop (FJS) с промежуточными буферами ограниченных размеров при использовании сетей Петри в качестве моделей производственных процессов реализации действий с ЕЗ на приборах указанных систем. Оптимизация расписаний выполнения ЕЗ в системах FJS реализуется с использованием предложенного алгоритма Hybrid Estimation of Distribution Algorithm (HEDA), который воспроизводит новую популяцию потомков (новую совокупность решений) на основе вероятностных моделей. Предложенный в [18] алгоритм HEDA реализует локальную оптимизацию расписаний выполнения ЕЗ в системах FJS посредством формирования новых решений на основе вероятностных моделей. Формируемые с использованием HEDA решения интерпретируются с помощью предложенного в [18] подхода на основе сетей Петри.

В [19] задача оптимизации расписаний процессов выполнения ЕЗ в системах FFS с промежуточными буферами ограниченных размеров решается с использованием многоцелевого алгоритма имитации отжига (Multi Objective Simulated Annealing – MOSA) и многоцелевого эволюционного алгоритма, основывающегося на декомпозиции (Multi Objective Evolutionary Algorithm Based on Decomposition – MOEA/D). В качестве критериев заданы: 1) взвешенное время окончания выполнения заданий; 2) штраф за время ожидания заданиями в буфере и стоимость производства. В модели введены бюджетные ограничения на техническое обслуживание и приостановку работ. Применение алгоритма MOEA/D для решения рассматриваемой проблемы предполагает разложение многоцелевой задачи на ряд одноцелевых задач с помощью скалярной функции и введения весовых векторов.

Использование метода MOEA/D при решении задач оптимизации расписаний выполнения ЕЗ в системах Hybrid Flow Shop (HFS–аналог FFS) рассмотрено в [20]. В качестве критериев введены: 1) энергопотребление процесса выполнения ЕЗ в HFS в соответствии с расписанием; 2) суммарное время запаздывания относительно задаваемых для ЕЗ директивных сроков. В [20] буферы рассматриваются как машины с нулевым временем обработки – этапы буферизации интерпретируются как этапы обработки с нулевой длительностью.

В [21] задача оптимизации расписаний выполнения ЕЗ в системах FS решается с использованием алгоритма Hybrid Shuffled Frog Leaping Algorithm – HSFLA, сочетающего в себе поиск по переменным окрестностям (Variable Neighborhood Search – VNS) и алгоритм «прыжок лягушки» (Shuffled Frog Leaping Algorithm – SFLA). Оптимизация расписаний предусматривает генерацию начальной совокупности решений с использованием алгоритма NEH и их оптимизацию посредством HSFLA. С целью исключения преждевременной сходимости алгоритма введен адаптивный оператор перемещения для улучшения размера шага, оператор пересечения для обновления подгрупп решений, а также используются три вида окрестностей.

В [22] рассматривается трех стадийная система, в которой на первом этапе реализуется производство деталей на параллельно функционирующих приборах, на втором этапе реализуется сборка на одном приборе из обработанных деталей готовых изделий, на третьем этапе в соответствии с типом изделий реализуется дифференциация их обработки на параллельных приборах. Предложенный в [22] Dual-Space Co-Evolutionary Memetic algorithm (DCMA) является разновидностью алгоритмов степенной эволюции (метаэвристический алгоритм), предусматривает поиск лучших расписаний как в непрерывном пространстве, так и в дискретном пространстве решений.

В [23] определение расписаний выполнения ЕЗ в системах FFS осуществляется путем построения бинарных деревьев выбора решений, реализующих определенные правила по формированию последовательностей их (заданий) выполнения на приборах указанных систем. Алгоритм построения деревьев предусматривает использование обучающих примеров – решений, полученных с использованием программирования в ограничениях. Дерево позволяет реализовать выбор с точки зрения анализа текущих ситуаций (состояний системы), соответствующих решениям, сформированным с использованием программирования в ограничениях. Формирование нескольких деревьев

позволяет идентифицировать лучший способ выполнения заданий в FFS.

В [24] рассмотрена система FFS выполнения ЕЗ с индивидуальными буферами для каждого прибора и общедоступным буфером ограниченного размера. В общедоступный буфер задания перемещаются в случае заполнения индивидуальных буферов соответствующих приборов и извлекаются из него при освобождении позиций в индивидуальных буферах. Для оптимизации расписаний выполнения ЕЗ в системах указанного вида применены нейронные сети Хопфилда, использующие для расширения пространства поиска алгоритм имитации отжига. В дополнении к правилам использования индивидуальных и общего буферов в модели и алгоритме оптимизации учтены правила управления транспортными устройствами, перемещающими детали между накопителями. Количество итераций оптимизации решений с использованием нейронных сетей Хопфилда является ограниченным.

В [25] решается задача оптимизации расписаний выполнения ЕЗ в системе FS с BPM в качестве обрабатывающих устройств, промежуточными буферами ограниченного размера между ними, при профилактическом обслуживании (ПО) приборов в BPM. Особенностью является стохастический характер задачи, связанный с исправностью отдельных приборов BPM, в целом BPM, а также с доступностью всей системы. С учетом статистических характеристик процесса функционирования приборов в BPM оптимизируются расписания выполнения ЕЗ и периодичность операций ПО при учете ограничений на доступность приборов системы с точки зрения работоспособности. В [26] рассмотрена задача оптимизации расписаний выполнения ЕЗ в системах FS с BPM и буферами при наличии транспортеров, которые размещают задания в накопителях при окончании их выполнения на приборах и перемещают задания из накопителей на последующие приборы в случае их готовности. При этом реализуется ПО транспортеров. Синтезирована двухкритериальная модель оптимизации решений по назначению заданий на приборы BPM, порядкам выполнения заданий на приборах, выбору операций ПО транспортеров и периодичности реализации ПО. Статистические характеристики процесса функционирования транспортеров, связанные с отказом и восстановлением, учитываются при определении уровня недоступности системы в критерии и в ограничениях. Рассмотренные работы [5 – 26] ориентированы на оптимизацию расписаний выполнения ЕЗ в системах FS, FFS, FJS при ограниченных размерах буферов. Методы, в них изложенные, не позволяют решать задачу оптимизации составов пакетов

и расписаний их выполнения в КС при учете ограниченных ресурсов для их хранения.

Одной из первых работ, посвященных оптимизации расписаний выполнения пакетов заданий (ПЗ) в КС (Batch Scheduling Processing – BSP), состоящей из двух приборов, при введении между приборами промежуточного буфера ограниченного размера, является работа [27]. Доказывается, что задача BSP в указанной постановке является NP-трудной при размере буфера $b > 0$. Модель процесса выполнения пакетов позволяет вычислять интервал времени от начала выполнения заданий в каждом пакете на первом приборе до окончания выполнения заданий в пакете на этом же приборе. Критерий оптимизации расписаний имеет вид суммы интервалов времени выполнения всех пакетов в последовательности на первом приборе и дополнительного интервала времени обработки оставшихся заданий в последнем пакете на втором приборе. Введено понятие устойчивого состояния пакета, а также условий, при выполнении которых пакет переходит в устойчивое состояние. Определена нижняя граница количества заданий в пакетах, для которого пакет перейдет в указанное состояние, а сама задача BSP может быть проинтерпретирована как частный случай задачи коммивояжера, для решения которой используются полиномиальные алгоритмы. Аналогичным образом в работе [28] формулируется подход к решению задачи BSP для последовательности из двух приборов и промежуточного буфера ограниченного размера между ними. Работа [28] отличается от работы [27] способом вычисления интервала времени выполнения ПЗ на первом приборе. В [28] способ вычисления значения указанного интервала дополнительно учитывает время наладки первого прибора на выполнение заданий в пакете, а также время удаления заданий из системы после окончания их выполнения на первом приборе. Вид критерия оптимизации расписаний выполнения ПЗ в КС, введенный в [28], аналогичен работе [27]. Аналогичны условия устойчивого состояния ПЗ и условия, определяющие возможность интерпретации рассматриваемой задачи BSP как частного случая задачи коммивояжера с целью применения полиномиальных алгоритмов.

В [29] также рассматривается система потокового типа, состоящая из двух приборов с одним промежуточными буфером между ними. Модель многостадийного процесса строиться с учетом предположения, что расписание формируется только для двух ПЗ, выполняемых на приборах (два пакета размещаются в последовательностях реализации действий с ними на двух приборах). Составы пакетов являются заданными и не оптимизируются (задача

оптимизации расписаний выполнения двух фиксированных ПЗ в двух приборной системе потокового типа с буфером ограниченного размера). Предложенная модель обеспечивает определение интервалов времени, затрачиваемых на выполнение каждого из пакетов на приборах КС (между моментом времени начала настройки первого прибора на выполнение заданий в пакете и моментом времени удаления пакета со второй машины).

Рассмотренные в [27 – 29] способы вычисления интервалов времени выполнения ПЗ могут быть использованы только в двух приборных КС. В тоже время эти способы не позволяют определять моменты времени начала выполнения отдельных заданий, включенных в пакеты, на соответствующих приборах и моменты времени начала/окончания использования ограниченных ресурсов хранения. Это делает невозможным формирование критериев, позволяющих определять эффективность использования ограниченных ресурсов хранения заданий. В силу сказанного использование представленных в [27 – 29] подходов не позволяет решать задачу оптимизации составов ПЗ и расписаний их выполнения в КС. По этой причине разработка новых моделей процессов выполнения ПЗ на приборах КС при наличии промежуточных буферов ограниченного размера является актуальной.

2. Математическая модель многостадийных процессов выполнения пакетов заданий в конвейерных системах при задании ограничения на размеры промежуточных буферов. Постановка задачи предполагает, что в КС выполняются наборы однотипных заданий (количество типов заданий равно N). Переход от выполнения заданий одного типа к выполнению заданий другого типа требует переналадки приборов КС. Для повышения эффективности использования ресурсов приборов КС, а также уменьшения длительности выполнения всех заданий, входящих в наборы, формируются пакеты. Выполнение однотипных заданий, входящих в пакеты, осуществляется без переналадки приборов КС на реализацию действий с ними. Особенностью КС является наличие буферов между обрабатывающими приборами, в каждый из которых задания поступают после окончания реализации действий с ними на предшествующих им приборах. При этом размеры буферов являются ограниченными. В том случае, если в буфере находится количество заданий, равное его размеру, то выполнение заданий на предшествующем приборе прекращается (прибор блокируется). Предшествующий буферу обрабатывающий прибор блокируется до тех пор, пока следующий за буфером прибор не начнет выполнение задания из буфера. Таким образом, в КС реализуется выполнение ПЗ

разных типов, между обрабатываемыми приборами имеются буферы ограниченных размеров, выполнение заданий, входящих в пакеты, на обрабатываемых приборах, предшествующих буферам, блокируется в случае их заполнения до тех пор, пока следующий за буфером прибор не начнет выполнения задания из буфера.

В работе реализуется синтез математических моделей процессов выполнения ПЗ в КС при ограничении на размеры промежуточных буферов для следующих задач: 1) двух приборная задача с одним промежуточным буфером; 2) трех приборная задача с двумя промежуточными буферами; 3) обобщенная задача моделирования процессов выполнения ПЗ в КС, содержащих L приборов и $(L-1)$ промежуточный буфер. Построение математической модели процессов выполнения ПЗ в КС обеспечивается введением обозначений: N – количество типов заданий, пакеты которых выполняются в КС; i – идентификатор типа заданий, пакеты которых выполняются в КС ($i = \overline{1, N}$); l – идентификаторы приборов КС ($l = \overline{1, L}$); π^l – последовательности выполнения ПЗ на l -х приборах КС, соответствующие порядкам реализации действий с ними; $\mathbf{M} = (m_1, m_2, \dots, m_N)^T$ – вектор количества ПЗ i -ых типов ($i = \overline{1, N}$); N_p –

количество ПЗ, размещаемых в последовательностях π^l ($N_p = \sum_{i=1}^N m_i$);

$\mathbf{T} = \|t_{li}\|_{L \times N}$ – матрица длительностей выполнения заданий разных типов на приборах КС; $\mathbf{T}^l = \|t_{li}^l\|_{N \times N}$ ($l = \overline{1, L}$) – матрицы длительностей

переналадок l -х приборов с выполнения заданий i -х на выполнение заданий i' -х типов ($t_{ii}^l = 0$); j – номера позиций, которые занимают ПЗ в последовательностях π^l ($l = \overline{1, L}$); $\mathbf{P} = \|p_{ij}\|_{N \times N_p}$ – матрица порядка

выполнения ПЗ в последовательностях π^l ($p_{ij} = 1$, если ПЗ i -го типа занимает в π^l j -ю позицию; $p_{ij} = 0$, если ПЗ i -го типа не занимает π^l j -ю позицию); $\mathbf{R} = \|r_{ij}\|_{N \times N_p}$ – матрица количества заданий i -х

типов в пакетах, занимающих в π^l j -е позиции; N_j – количество заданий в пакетах, занимающих в π^l j -е позиции ($N_j = \sum_{h=1}^N r_{hj}$;

$j = \overline{1, N_p}$); $\mathbf{T}^{ol} = \|t_{jq}^{ol}\|_{N_p \times Q}$ – матрицы моментов времени начала

выполнения q -х заданий ($q = \overline{1, N_j}$), входящих в пакеты, занимающих j -е позиции в последовательностях π^l ($Q = \max_j(N_j)$); b – размер буферов.

В соответствии с введенными обозначениями постановка задачи математического моделирования многостадийных процессов выполнения ПЗ в КС при ограничении на размеры промежуточных буферов формулируется следующим образом. Заданными являются: вектор **М** количества ПЗ i -х типов ($i = \overline{1, N}$); матрица **Т** длительностей выполнения заданий на приборах КС; матрицы **T'** ($l = \overline{1, L}$) длительностей переналадок l -х приборов; матрица **Р** порядка выполнения ПЗ в π^l ; матрица **Р** количества заданий i -х типов в пакетах, занимающих в π^l j -е позиции; размеры b промежуточных буферов. Требуется определить значения элементов t_{jq}^{ol} ($j = \overline{1, N_p}$; $q = \overline{1, N_j}$; $l = \overline{1, L}$) матриц **T''** моментов времени начала выполнения заданий в пакетах, занимающих j -е позиции в последовательностях π^l ($l = \overline{1, L}$).

Для задачи с двумя приборами и промежуточным буфером между ними рассуждения с целью построения математической модели многостадийных процессов выполнения ПЗ в КС реализуются в соответствии с заданными видами последовательностей, представленными на рисунке 1. Рассматривается задача моделирования выполнения двух ПЗ, занимающих в последовательностях π^l ($l = \overline{1, 2}$) соответственно $(j=1)$ -ю и $(j=2)$ -ю позиции. Типы заданий в пакетах, занимающих в π^l указанные позиции, обозначены как i_1 и i_2 . Виды последовательностей и соответствующие им выражения модели формируются с учетом значений параметров t_{li} ($l = \overline{1, 2}$) для заданий в пакетах занимающих j -х позиции в π^l (i_1 -го и i_2 -го типов). Рассматриваются виды последовательностей и формируются выражения модели для следующих соотношений указанных параметров: 1) $t_{1i_1} \leq t_{2i_1}$, $t_{1i_2} \leq t_{2i_2}$; 2) $t_{1i_1} \leq t_{2i_1}$, $t_{1i_2} > t_{2i_2}$; 3) $t_{1i_1} > t_{2i_1}$, $t_{1i_2} \leq t_{2i_2}$; 4) $t_{1i_1} > t_{2i_1}$, $t_{1i_2} > t_{2i_2}$. Количество заданий в пакетах равно 5 ($N_1 = 5$, $N_2 = 5$), размер буфера b равен 3 ($b=3$). При $t_{1i_1} \leq t_{2i_1}$, $t_{1i_2} \leq t_{2i_2}$ (рисунок 1(а)) задание с $q=1$ в пакете в $(j=1)$ -й позиции в π^1 после выполнения на $(l=1)$ -м приборе перемещается в буфер, а затем из буфера перемещается на обработку на $(l=2)$ -м приборе. Следующие b заданий после обработки на $(l=1)$ -м

приборе размещаются в буфере (заполнение буфера) и $(l=1)$ -й прибор блокируется в ожидании освобождения позиции в буфере. В момент времени $t_{12}^{02} = t_{11}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1}$ задание с $q=2$ в пакете в $(j=1)$ -й позиции в

π^1 поступает для выполнения на $(l=2)$ -й прибор (позиция в буфере освобождается) и $(l=1)$ -й прибор приступает к выполнению задания с $q=5$ (тогда $t_{15}^{01} = t_{12}^{02}$). Аналогичные рассуждения строятся для всех заданий с q , удовлетворяющих условию: $b+1 < q \leq N_1$. С учетом выполненных рассуждений выражения для определения значений t_{1q}^{01} ($q = \overline{1, N_1}$) для $(j=1)$ -го пакета на $(l=1)$ -м приборе получены в виде:

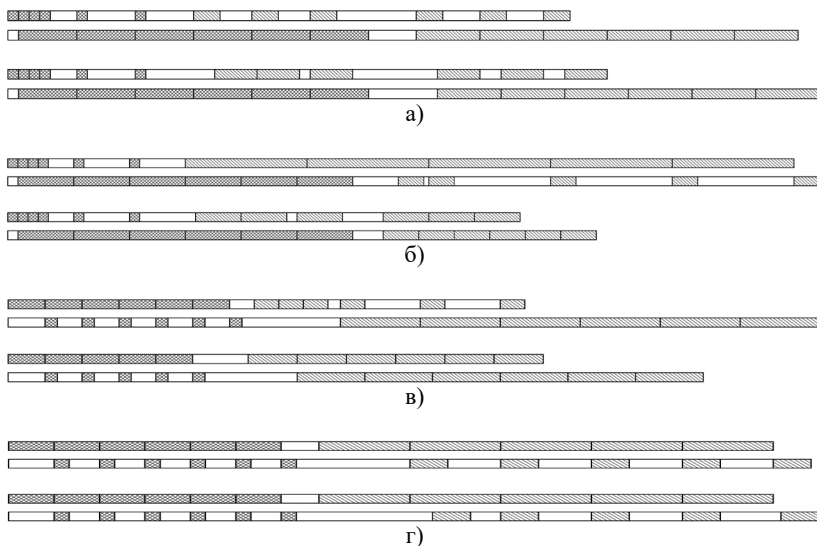


Рис. 1. Вид последовательностей реализации действий с ПЗ, используемых при построении математической модели многостадийных процессов их выполнения в КС, состоящей из двух приборов, при:

а) $t_{1i_1} \leq t_{2i_1}$, $t_{1i_2} \leq t_{2i_2}$; б) $t_{1i_1} \leq t_{2i_1}$, $t_{1i_2} > t_{2i_2}$; в) $t_{1i_1} > t_{2i_1}$, $t_{1i_2} \leq t_{2i_2}$;

г) $t_{1i_1} > t_{2i_1}$, $t_{1i_2} > t_{2i_2}$

$$q=l: t_{11}^{0l} = 0, \quad (1)$$

$$1 < q \leq b+1: t_{1q}^{0l} = t_{1,(q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}, \quad (2)$$

$$b+1 < q \leq N_1: t_{1q}^{0l} = t_{1,(q-b)}^{02}, \quad (3)$$

где $(q-b)$ – идентификатор задания в $(j=1)$ -м ПЗ на $(l=2)$ -м приборе.

Для q -х заданий ($q = \overline{2, N_1}$) в $(j=1)$ -м пакете в π^2 на $(l=2)$ -м приборе моменты времени начала их выполнения равны моментам времени окончания реализации действий на этом приборе с предшествующими им заданиями (рисунок 1(а)), тогда значения t_{1q}^{02} ($q = \overline{1, N_1}$) вычисляются в соответствии с выражениями вида:

$$q=1: t_{11}^{02} = t_{11}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}, \quad (4)$$

$$1 < q \leq N_1: t_{1q}^{02} = t_{1,(q-1)}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1}. \quad (5)$$

При $t_{1i_1} \leq t_{2i_1}$, $t_{1i_2} \leq t_{2i_2}$ (рисунок 1(а)) $(l=1)$ -й прибор приступает в выполнении $(q=1)$ -го задания в пакете в $(j=2)$ -й позиции в π^1 в случае начала выполнения $(q=4)$ -го ($(N_1 - b + 1)$ -го) задания из пакета в $(j=1)$ -й позиции на $(l=2)$ -м приборе либо при условии окончания переналадки $(l=1)$ -го прибора на выполнение заданий i_2 -го типа (в этом пакете). Аналогично, $(l=1)$ -й прибор может приступить в выполнении q -х заданий ($1 < q \leq b$) в пакете в $(j=2)$ -й позиции в π^1 в случае начала выполнения $(N_1 - b + q)$ -го задания из пакета в $(j=1)$ -й позиции на $(l=2)$ -м приборе либо при окончании выполнения $(q-1)$ -го задания в пакете в $(j=2)$ -й позиции в π^1 (окончания выполнения задания в $(j=2)$ -м пакете, предшествующего рассматриваемому q -му заданию из этого же пакета). Также, $(l=1)$ -й прибор может приступить к выполнению q -х заданий ($b+1 \leq q \leq N_2$) в пакете в $(j=2)$ -й позиции в π^1 в случае начала выполнения $(q-b)$ -го задания из пакета в $(j=2)$ -й позиции на $(l=2)$ -м приборе. Тогда выражения для вычисления значений t_{2q}^{0l} ($q = \overline{1, N_2}$) для q -х заданий из $(j=2)$ -го пакета на $(l=1)$ -м приборе имеют следующий вид:

$$q=l: t_{21}^{0l} = \max \{ t_{1,N_1}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1} + t_{i_1, i_2}^l; t_{1, (N_1 - b + 1)}^{02} \}, \quad (6)$$

$$1 < q \leq b: t_{2q}^{0l} = \max \{ t_{2, (q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s2}; t_{1, (N_1 - b + q)}^{02} \}, \quad (7)$$

$$b+1 \leq q \leq N_2: t_{2q}^{0l} = t_{2, (q-b)}^{02}. \quad (8)$$

Для q -х заданий из $(j=2)$ -го пакета в последовательности π^2 на $(l=2)$ -м приборе значения t_{2q}^{02} ($q = \overline{1, N_2}$) определяются в соответствии с выражениями вида (рисунки 1(a)):

$$q=l: t_{21}^{02} = t_{1, N_1}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1} + t_{i_1, i_2}^2, \quad (9)$$

$$1 < q \leq N_2: t_{2q}^{02} = t_{2, (q-1)}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s2}. \quad (10)$$

Выражение (9) позволяет идентифицировать момент времени окончания переналадки $(l=2)$ -го прибора на выполнение заданий i_2 -го типа, выражение (10) – момент времени окончания выполнения предшествующего $(q-1)$ -го задания в $(j=2)$ -м пакете в π^2 .

При $t_{1i_1} \leq t_{2i_1}$, $t_{1i_2} > t_{2i_2}$ (рисунок 1(б)) выражения для вычисления значений t_{1q}^{0l} и t_{2q}^{0l} для q -х заданий ($q = \overline{1, N_1}$), входящих в пакет в $(j=1)$ -й позиции в π^1 ($l = \overline{1, 2}$), аналогичны выражениям (1) – (3), (4), (5). Выражения для вычисления значений t_{2q}^{0l} для q -х заданий ($1 \leq q \leq b$), входящих в пакет в $(j=2)$ -й позиции в π^1 на $(l=1)$ -м приборе аналогичны выражениям (6), (7). Для q -х заданий ($b+1 \leq q \leq N_2$), входящих в пакет в $(j=2)$ -й позиции в π^1 на $(l=1)$ -м приборе, выражение для вычисления t_{2q}^{0l} имеет вид $t_{2q}^{0l} = t_{2, (q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s2}$. Выражение для вычисления t_{21}^{02} аналогично выражению (9), а выражение для вычисления t_{2q}^{02} ($1 < q \leq N_2$) имеет вид:

$$t_{2q}^{02} = \max \{ t_{2,q}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s2}; t_{2,(q-1)}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s2} \}. \quad (11)$$

При $t_{1i_1} > t_{2i_1}$, $t_{1i_2} \leq t_{2i_2}$ (рисунок 1(в)) выражения для вычисления значений t_{1q}^{01} для q -х заданий ($q = 1, 1 < q \leq b+1$), входящих в пакет в $(j=1)$ -й позиции в π^1 , аналогичны выражениям (1), (2) соответственно, выражение для вычисления значений t_{1q}^{01} ($b+1 < q \leq N_1$) аналогично выражению (2). Для q -х заданий ($q = \overline{1, N_1}$) в $(j=1)$ -м пакете в π^2 на $(l=2)$ -м приборе выражения для определения значений t_{1q}^{02} имеют вид:

$$t_{1q}^{02} = t_{1q}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s1}. \quad (12)$$

При $t_{1i_1} > t_{2i_1}$, $t_{1i_2} \leq t_{2i_2}$ (рисунок 1(в)) выражение для вычисления значения t_{21}^{01} для $(q=1)$ -го задания, входящего в пакет в $(j=2)$ -й позиции в π^1 имеет вид:

$$t_{21}^{01} = t_{1,N_1}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s1} + t_{i_1,i_2}^1. \quad (13)$$

Выражение для вычисления значений t_{2q}^{01} для q -х заданий ($1 < q \leq b+1$), входящих в этот же пакет, получено в виде:

$$t_{2q}^{01} = t_{2,(q-1)}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s2}, \quad (14)$$

а выражение для вычисления значений t_{2q}^{01} для q -х заданий ($b+1 < q \leq N_2$) в этом же пакете имеет вид:

$$t_{2q}^{01} = \max \{ t_{2,(q-1)}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s2}; t_{2,(q-b)}^{02} \}. \quad (15)$$

Также при $t_{li} > t_{2i}$, $t_{li} \leq t_{2i}$ выражение для вычисления значения t_{21}^{02} для $(q=1)$ -го задания, входящего в пакет в $(j=2)$ -й позиции в π^2 , получено в виде (рисунок 1(в)):

$$t_{21}^{02} = \max \left\{ t_{21}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s2} \cdot t_{1,N_1}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1} + t_{i_1, i_2}^2 \right\}, \quad (16)$$

а выражение для вычисления значений t_{2q}^{02} для q -х заданий ($1 < q \leq N_2$), входящих в этот же пакет, аналогично выражению (11).

При $t_{li} > t_{2i}$, $t_{li} > t_{2i}$ (рисунок 1(г)) выражение для определения значения t_{11}^{01} для $(q=1)$ -го задания в пакете в $(j=1)$ -й позиции в π^1 аналогично выражению (1), выражение для вычисления значений t_{1q}^{01} для q -х заданий в этом же пакете ($1 < q \leq N_1$) аналогично выражению (2). Выражения для вычисления t_{1q}^{02} для $(q=1)$ -го задания и q -х заданий ($1 < q \leq N_1$), входящих в $(j=1)$ -й пакет в π^2 на $(l=2)$ -м приборе аналогичны (12). Выражение для определения значения t_{21}^{01} для $(q=1)$ -го задания в пакете в $(j=2)$ -й позиции в π^1 аналогично (13), а выражение для вычисления значений t_{2q}^{01} для q -х заданий в этом же пакете ($1 < q \leq N_2$) на $(l=1)$ -м приборе аналогично (14). Выражение для вычисления значения t_{21}^{02} для $(q=1)$ -го задания в пакете в $(j=2)$ -й позиции в π^2 на $(l=2)$ -м приборе аналогично выражению (16), а выражение для вычисления значений t_{2q}^{02} для q -х заданий в этом же пакете ($1 < q \leq N_2$) на $(l=2)$ -м приборе аналогично (10).

Обобщение полученных выражений при $N_p \geq 2$ позволило сформировать математическую модель многостадийных процессов выполнения ПЗ в КС, состоящей из двух приборов, при ограниченном размере b буфера в следующем виде (выражения для вычисления значений t_{jq}^{0l} для q -х заданий ($q = \overline{1, N_j}$), входящих в пакеты, занимающие j -е позиции ($j = \overline{1, N_p}$) в последовательностях π^l их выполнения на l -х приборах ($l = \overline{1, 2}$)):

1) выражения для вычисления значений t_{1q}^{0l} ($q = \overline{1, N_1}$) для ПЗ в $(j=1)$ -й позиции в последовательности π^1 :

$$q=1: t_{11}^{0l} = 0, \quad (17)$$

$$1 < q \leq b+1: t_{1q}^{0l} = t_{1,(q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}, \quad (18)$$

$$b+1 < q \leq N_1: t_{1q}^{0l} = \max\{t_{1,(q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}; t_{1,(q-b)}^{02}\}; \quad (19)$$

2) выражения для вычисления значений t_{1q}^{02} ($q = \overline{1, N_1}$) для ПЗ в $(j=1)$ -й позиции в последовательности π^2 :

$$q=1: t_{11}^{02} = t_{11}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}, \quad (20)$$

$$1 < q \leq N_1: t_{1q}^{02} = \max\{t_{1q}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}; t_{1,(q-1)}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1}\}; \quad (21)$$

3) выражения для вычисления значений t_{jq}^{0l} ($q = \overline{1, N_j}$) для ПЗ в j -х позициях ($j = \overline{2, N_p}$; $N_p \geq 2$) в последовательности π^1 :

$$q=1: t_{j1}^{0l} = \max\{t_{(j-1), N_{j-1}}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s, j-1} + t_{j-1, i_j}^l; t_{(j-1), (N_{j-1}-b+1)}^{02}\}, \quad (22)$$

$$1 < q \leq b: t_{jq}^{0l} = \max\{t_{j,(q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{sj}; t_{(j-1), (N_{j-1}-b+q)}^{02}\}, \quad (23)$$

$$b+1 \leq q \leq N_2: t_{jq}^{0l} = \max\{t_{j,(q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{sj}; t_{j,(q-b)}^{02}\}; \quad (24)$$

4) выражения для вычисления значений t_{jq}^{02} для ПЗ в j -х позициях ($j = \overline{2, N_p}$; $N_p \geq 2$) в последовательности π^2 :

$$q=1: t_{j1}^{02} = \max \{ t_{j1}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{sj}; t_{(j-1), N_{j-1}}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s, (j-1)} + t_{i_{j-1}, i_j}^2 \}, \quad (25)$$

$$1 < q \leq N_2: t_{jq}^{02} = \max \{ t_{jq}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{sj}; t_{j, (q-1)}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s, j} \}. \quad (26)$$

Выражения (17) – (26) являются математической моделью в общем виде многостадийных процессов выполнения ПЗ в двух приборной КС при наличии промежуточного буфера ограниченного размера b и количестве пакетов $N_p \geq 2$ в последовательностях π^l ($l = \overline{1, 2}$).

Рассмотренный подход использован при построении математической модели многостадийных процессов выполнения ПЗ в КС, содержащей три прибора, которая затем обобщена для произвольного количества приборов. Построение математической модели многостадийных процессов выполнено при условиях: $L=3$; $b=3$; $N_p = 2$; $N_j = 8$ ($j = \overline{1, 2}$). Синтез выражений математической модели реализован с использованием заданных видов последовательностей π^l ($l = \overline{1, 3}$) выполнения ПЗ в КС, представленных на рисунке 2 ((а) и (б), которые различаются значениями t_{i_1, i_2}^1). Выражения математической модели процессов выполнения ПЗ в КС получены в следующем виде:

1) выражения для вычисления значений t_{1q}^{01} ($q = \overline{1, N_1}$) для q -х заданий в пакете в $(j=1)$ -й позиции в последовательности π^1 :

$$- \quad q=1: t_{11}^{01} = 0, \quad (27)$$

- $1 < q \leq b+1$ (реализуется заполнение буфера между первым и вторым приборами при условии, что $(q=1)$ -е задание в $(j=1)$ -м пакете после окончания выполнения на $(l=1)$ -м приборе начинает выполнение на $(l=2)$ -м приборе, рисунок 2(а, б, в)):

$$t_{1q}^{01} = t_{1, (q-1)}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s1}, \quad (28)$$

- $b+1 < q \leq N_1$ (выполнение q -го задания начинается либо в момент времени окончания действий с предшествующим ему $(q-1)$ -м

заданием в этом пакете, либо в момент времени освобождения позиции в буфере, связанного с началом выполнения $(q-b)$ -го задания из этого же пакета на $(l=2)$ -м приборе, рисунок 2(а, б, в)):

$$t_{1q}^{0l} = \max \{ t_{1,(q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}; t_{1,(q-b)}^{02} \}; \quad (29)$$

2) выражения для вычисления значений t_{1q}^{02} ($q = \overline{1, N_1}$) для q -х заданий в пакете в $(j=l)$ -й позиции в последовательности π^2 :

$$- \quad q=l: t_{11}^{02} = t_{11}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}, \quad (30)$$

– $1 < q \leq b+1$ (заполнение буфера между вторым и третьим приборами при условии, что $(q=l)$ -е задание в $(j=l)$ -м пакете после окончания выполнения на $(l=2)$ -м приборе начинает выполнение на $(l=3)$ -м приборе, рисунок 2 (а, в)):

$$t_{1q}^{02} = \max \{ t_{1q}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}; t_{1,(q-1)}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1} \}, \quad (31)$$

– $b+1 < q \leq N_1$ (выполнения q -го задания начинается либо в момент времени окончания его выполнения на $(l=1)$ -м приборе, либо в момент времени окончания выполнения $(q-1)$ -го задания в этом пакете на $(l=2)$ -м приборе, либо в момент времени освобождения позиции в буфере между $(l=2)$ -м и $(l=3)$ -м приборами, рисунок 2(а, в, г, д)):

$$t_{1q}^{02} = \max \{ t_{1q}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}; t_{1,(q-1)}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1}; t_{1,(q-b)}^{03} \}; \quad (32)$$

3) выражения для вычисления значений t_{1q}^{03} ($q = \overline{1, N_1}$) для q -х заданий в пакете в $(j=l)$ -й позиции в последовательности π^3 :

$$- \quad q=l: t_{11}^{03} = t_{11}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1} = \sum_{l=1}^2 \sum_{s=1}^N t_{ls} \cdot p_{s1}, \quad (33)$$

– $1 < q \leq N_1$ (начало выполнения q -го задания возможно либо в момент времени окончания его выполнения на $(l=2)$ -м приборе, либо в момент времени окончания выполнения $(q-1)$ -го задания в этом же пакете на $(l=3)$ -м приборе, рисунок 2(а, д)):

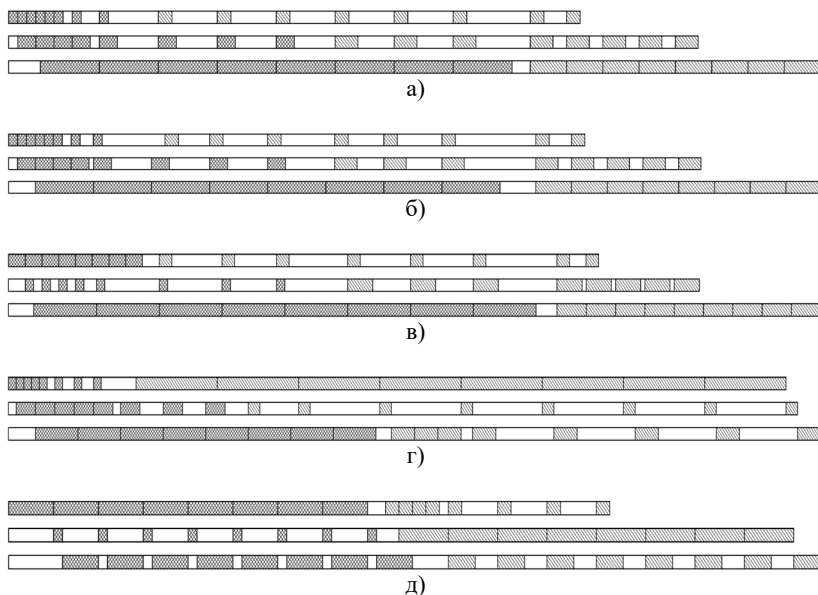


Рис. 2. Вид последовательностей реализации действий с ПЗ, используемых при построении математической модели многостадийных процессов их выполнения в КС, состоящей из трех приборов, при:

- а) $t_{1l_1} \leq t_{2l_1}, t_{2l_1} \leq t_{3l_1}, t_{1l_2} \leq t_{2l_2}, t_{2l_2} \leq t_{3l_2}$; б) $t_{1l_1} \leq t_{2l_1}, t_{2l_1} \leq t_{3l_1}, t_{1l_2} \leq t_{2l_2}, t_{2l_2} \leq t_{3l_2}$; в) $t_{1l_1} > t_{2l_1}, t_{2l_1} \leq t_{3l_1}, t_{1l_2} \leq t_{2l_2}, t_{2l_2} \leq t_{3l_2}$; г) $t_{1l_1} \leq t_{2l_1}, t_{2l_1} \leq t_{3l_1}, t_{1l_2} > t_{2l_2}, t_{2l_2} \leq t_{3l_2}$; д) $t_{1l_1} > t_{2l_1}, t_{2l_1} \leq t_{3l_1}, t_{1l_2} > t_{2l_2}, t_{2l_2} \leq t_{3l_2}$

$$t_{1q}^{03} = \max \{ t_{1q}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1}; t_{1,(q-1)}^{03} + \sum_{s=1}^N t_{3s} \cdot p_{s1} \}; \quad (34)$$

4) выражения для вычисления значений $t_{2q}^{01} (q = \overline{1, N_2})$ для q -х заданий в $(j=2)$ -м пакете в последовательности π^1 на $(l=1)$ -м приборе:

– $q=l$ (выполнение $(q=l)$ -го задания начинается в момент времени окончания переналадки $(l=l)$ -го прибора на выполнение

заданий i_2 -го типа в пакете в $(j=2)$ -й позиции в π^1 , либо в момент времени освобождения одной позиции в буфере между $(l=1)$ -м и $(l=2)$ -м приборами, вызванного началом выполнения $(N_1 - b + 1)$ -го задания из пакета в $(j=1)$ -й позиции на $(l=2)$ -м приборе, рисунок 2(а, б, г)):

$$t_{21}^{0l} = \max \{ t_{1,N_1}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1} + t_{i_1, i_2}^1; t_{1, (N_1 - b + 1)}^{02} \}, \quad (35)$$

– $1 < q \leq b$ (начало выполнения q -го задания в пакете в $(j=2)$ -й позиции в π^1 возможно в момент времени окончания выполнения предшествующего $(q-1)$ -го задания в этом же пакете, либо в момент времени освобождения позиции в буфере между $(l=1)$ -м и $(l=2)$ -м приборами, вызванного началом выполнения $(N_1 - b + q)$ -го задания из пакета в $(j=1)$ -й позиции на $(l=2)$ -м приборе, рисунок 2(а, г, д)):

$$t_{2q}^{0l} = \max \{ t_{2, (q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s2}; t_{1, (N_1 - b + q)}^{02} \}, \quad (36)$$

– $b+1 \leq q \leq N_2$ (начало выполнения q -го задания в пакете в $(j=2)$ -й позиции в π^1 возможно в момент времени окончания выполнения предшествующего $(q-1)$ -го задания в этом же пакете, либо в момент времени освобождения одной позиции в буфере между $(l=1)$ -м и $(l=2)$ -м приборами, связанного с началом выполнения $(q-b)$ -го задания из этого же пакета на $(l=2)$ -м приборе, рисунок 2(а, г)):

$$t_{2q}^{0l} = \max \{ t_{2, (q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s2}; t_{2, (q-b)}^{02} \}; \quad (37)$$

5) выражения для вычисления значений $t_{2q}^{02} (q = \overline{1, N_2})$ для q -х заданий в пакете в $(j=2)$ -й позиции в последовательности π^2 :

– $q=1$ (выполнение $(q=1)$ -го задания начинается в момент времени окончания выполнения этого задания на $(l=1)$ -м приборе, либо в момент времени окончания переналадки $(l=2)$ -го прибора на выполнение заданий i_2 -го типа, либо в момент времени освобождения позиции в буфере между $(l=2)$ -м и $(l=3)$ -м приборами, связанном с

началом реализации действий с $(N_1 - b + 1)$ -м заданием из пакета в $(j=1)$ -й позиции в π^3 на $(l=3)$ -м приборе, рисунок 2 (а, б, д)):

$$t_{21}^{02} = \max \left\{ t_{2q}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s2}; t_{1,N_1}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s1} + t_{i_1, i_2}^2; t_{1, (N_1 - b + 1)}^{03} \right\}, \quad (38)$$

– $1 < q \leq b$ (выполнение q -го задания начинается в момент времени окончания выполнения этого же задания на $(l=1)$ -м приборе, либо в момент времени окончания выполнения предшествующего $(q-1)$ -го задания из этого же пакета на $(l=2)$ -м приборе, либо в момент времени освобождения позиции в буфере между $(l=2)$ -м и $(l=3)$ -м приборами, связанном с началом реализации действий с $(N_1 - b + q)$ -м заданием из пакета в $(j=1)$ -й позиции в π^3 :

$$t_{2q}^{02} = \max \left\{ t_{2q}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s2}; t_{2, (q-1)}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s2}; t_{1, (N_1 - b + q)}^{03} \right\}, \quad (39)$$

– $b + 1 \leq q \leq N_2$ (выполнение q -го задания начинается в момент времени окончания выполнения этого же задания на $(l=1)$ -м приборе, либо в момент времени окончания выполнения предшествующего $(q-1)$ -го задания из этого же пакета на $(l=2)$ -м приборе, либо в момент времени освобождения одной позиции в буфере между $(l=2)$ -м и $(l=3)$ -м приборами, связанном с началом реализации действий с $(q - b)$ -м заданием из этого же пакета в $(j=2)$ -й позиции в π^3 , рисунок 2 (а, г, д)):

$$t_{2q}^{02} = \max \left\{ t_{2q}^{01} + \sum_{s=1}^N t_{1s} \cdot p_{s2}; t_{2, (q-1)}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s2}; t_{1, (q-b)}^{03} \right\}; \quad (40)$$

б) выражения для вычисления значений $t_{2q}^{03} (q = \overline{1, N_2})$ для q -х заданий в пакете в $(j=2)$ -й позиции в последовательности π^3 :

– $q=1$ (выполнение $(q=1)$ -го задания начинается в момент времени окончания выполнения этого же задания на $(l=2)$ -м приборе, либо в момент времени окончания переналадки $(l=3)$ -го прибора на выполнение заданий i_2 -го типа, рисунок 2 (в, г, д)):

$$t_{21}^{03} = \max \left\{ t_{21}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s2}; t_{1,N_1}^{03} + \sum_{s=1}^N t_{3s} \cdot p_{s1} + t_{h,i_2}^3 \right\}, \quad (41)$$

– $1 < q \leq N_2$ (выполнение q -го задания начинается в момент времени окончания выполнения этого задания на $(l=2)$ -м приборе, либо в момент времени окончания выполнения предшествующего $(q-1)$ -го задания из этого же пакета на $(l=3)$ -м приборе, рисунок 2 (а, г, д)):

$$t_{2q}^{03} = \max \left\{ t_{2q}^{02} + \sum_{s=1}^N t_{2s} \cdot p_{s2}; t_{2,(q-1)}^{03} + \sum_{s=1}^N t_{2s} \cdot p_{s2} \right\}. \quad (42)$$

Выражения (27) – (42) представляют собой математическую модель многостадийных процессов выполнения ПЗ в КС, содержащей три прибора с промежуточными буферами между ними размером b . На основе модели (27) – (42) синтезирована математическая модель многостадийных процессов выполнения ПЗ в КС в общем виде (при произвольном количестве приборов L ($L \geq 3$), произвольном количестве пакетов $N_p \geq 2$ в последовательностях π^l ($l = \overline{1, L}$) их выполнения на приборах КС). Выражения математической модели в общем виде многостадийных процессов выполнения ПЗ в КС при количестве приборов $L \geq 3$, ограничении на размеры b буферов между приборами, произвольном количестве $N_p \geq 2$ ПЗ в последовательностях π^l ($l = \overline{1, L}$) их выполнения на приборах КС имеют следующий вид:

1) выражения для вычисления значений t_{1q}^{0l} ($q = \overline{1, N_1}$) для q -х заданий в пакете в $(j=l)$ -й позиции в последовательности π^l :

$$- \quad q=l: t_{11}^{0l} = 0, \quad (43)$$

$$- \quad 1 < q \leq b+1: t_{1q}^{0l} = t_{1,(q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}, \quad (44)$$

$$- \quad b+1 < q \leq N_1: t_{1q}^{0l} = \max \left\{ t_{1,(q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}; t_{1,(q-b)}^{02} \right\}; \quad (45)$$

2) выражения для вычисления значений t_{jq}^{0l} ($q = \overline{1, N_j}$) для q -х заданий в пакетах в j -х позициях ($j = \overline{2, N_p}$) в последовательности π^1 на $(l=1)$ -м приборе:

– $q=1$:

$$t_{j1}^{0l} = \max \{ t_{(j-1), N_{j-1}}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s, (j-1)} + t_{i_{j-1}, i_j}^1; t_{(j-1), (N_{j-1}-b+1)}^{02} \}, \quad (46)$$

– $1 < q \leq b$:

$$t_{jq}^{0l} = \max \{ t_{j, (q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s, j}; t_{(j-1), (N_{j-1}-b+q)}^{02} \}, \quad (47)$$

– $b+1 \leq q \leq N_j$:

$$t_{jq}^{0l} = \max \{ t_{j, (q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{sj}; t_{j, (q-b)}^{02} \}; \quad (48)$$

3) выражения для вычисления значений t_{1q}^{0l} ($q = \overline{1, N_1}$) для q -х заданий в $(j=1)$ -м пакете в последовательностях π^l ($l = \overline{2, L-1}$):

– $q=1$:

$$t_{11}^{0l} = t_{11}^{0l-1} + \sum_{s=1}^N t_{(l-1), s} \cdot p_{s1}, \quad (49)$$

– $1 < q \leq b+1$:

$$t_{1q}^{0l} = \max \{ t_{1q}^{0l-1} + \sum_{s=1}^N t_{(l-1), s} \cdot p_{s1}; t_{1, (q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1} \}, \quad (50)$$

– $b+1 < q \leq N_1$:

$$t_{1q}^{0l} = \max \{ t_{1q}^{0l-1} + \sum_{s=1}^N t_{(l-1), s} \cdot p_{s1}; t_{1, (q-1)}^{0l} + \sum_{s=1}^N t_{1s} \cdot p_{s1}; t_{1, (q-b)}^{0l+1} \}; \quad (51)$$

4) выражения для вычисления значений t_{jq}^{0l} ($q = \overline{1, N_j}$) для q -х заданий в пакетах в j -х позициях ($j = \overline{2, N_p}$) в последовательностях π^l на l -х приборах ($l = \overline{2, L-1}$):

– $q=l$:

$$t_{j1}^{0l} = \max \{ t_{j1}^{0l-1} + \sum_{s=1}^N t_{(l-1),s} \cdot p_{sj}; t_{(j-1),N_{j-1}}^{0l} + \sum_{s=1}^N t_{ls} \cdot p_{s,(j-1)} + t_{l_{j-1},i_j}^l; t_{(j-1),(N_{j-1}-b+1)}^{0l+1} \}, \quad (52)$$

– $1 < q \leq b$:

$$t_{jq}^{0l} = \max \{ t_{jq}^{0l-1} + \sum_{s=1}^N t_{(l-1),s} \cdot p_{sj}; t_{j,(q-1)}^{0l} + \sum_{s=1}^N t_{ls} \cdot p_{sj}; t_{(j-1),(N_{j-1}-b+q)}^{0l+1} \}, \quad (53)$$

– $b+1 \leq q \leq N_j$:

$$t_{jq}^{0l} = \max \{ t_{jq}^{0l-1} + \sum_{s=1}^N t_{(l-1),s} \cdot p_{sj}; t_{j,(q-1)}^{0l} + \sum_{s=1}^N t_{ls} \cdot p_{sj}; t_{j,(q-b)}^{0l+1} \}; \quad (54)$$

5) выражения для вычисления значений t_{1q}^{0L} ($q = \overline{1, N_1}$) для q -х заданий в пакете в $(j=1)$ -й позиции в последовательности π^L на L -м приборе:

$$– \quad q=1: t_{11}^{0L} = t_{11}^{0L-1} + \sum_{s=1}^N t_{(L-1),s} \cdot p_{s1} = \sum_{l=1}^{L-1} \sum_{s=1}^N t_{ls} \cdot p_{s1}, \quad (55)$$

– $1 < q \leq N_1$:

$$t_{1q}^{0L} = \max \{ t_{1q}^{0L-1} + \sum_{s=1}^N t_{(L-1),s} \cdot p_{s1}; t_{1,(q-1)}^{0L} + \sum_{s=1}^N t_{Ls} \cdot p_{s1} \}; \quad (56)$$

6) выражения для вычисления значений t_{jq}^{0L} ($q = \overline{1, N_j}$, $j = \overline{2, N_p}$) для q -х заданий в пакетах в j -х позициях в последовательности π^L):

– $q=l$:

$$t_{j1}^{0L} = \max \{ t_{jq}^{0L-l} + \sum_{s=1}^N t_{(L-1),s} \cdot p_{sj}; t_{(j-1),N_{j-1}}^{0L} + \sum_{s=1}^N t_{Ls} \cdot p_{s,(j-1)} + t_{i_{j-1},i_j}^L \}, \quad (57)$$

$$- \quad 1 < q \leq N_j :$$

$$t_{jq}^{0L} = \max \{ t_{jq}^{0L-l} + \sum_{s=1}^N t_{(L-1),s} \cdot p_{sj}; t_{j,(q-1)}^{0L} + \sum_{s=1}^N t_{Ls} \cdot p_{sl} \}. \quad (58)$$

Сформированные выражения (43) – (58) представляют собой математическую модель в общем виде многостадийных процессов выполнения ПЗ в КС при наличии между приборами промежуточных буферов ограниченного размера для произвольного количества L обрабатываемых приборов и произвольного их (пакетов) количества в последовательностях π^l ($l = \overline{1, L}$) реализации действий с ними.

3. Алгоритм вычисления значений t_{jq}^{0l} ($q = \overline{1, N_j}$, $j = \overline{1, N_p}$, $l = \overline{1, L}$) моментов времени начала выполнения q -х заданий, входящих в пакеты, занимающие j -е позиции в последовательностях π^l их выполнения на приборах КС. С учетом вида полученных выражений (43) – (58) разработан алгоритм вычисления значений элементов t_{jq}^{0l} ($q = \overline{1, N_j}$, $j = \overline{1, N_p}$, $l = \overline{1, L}$) матриц T^{0l} моментов времени начала выполнения q -х заданий, входящих в пакеты, занимающие j -е позиции в последовательностях π^l их выполнения на приборах КС, которые являются результатом вычислений с использованием рассматриваемой модели. Алгоритм вычисления значений t_{jq}^{0l} ($q = \overline{1, N_j}$, $j = \overline{1, N_p}$, $l = \overline{1, L}$) на основе выражений (43) – (58) имеет следующий порядок шагов:

1. Задать номер позиции j , в которой размещен в последовательностях π^l ($l = \overline{1, L}$) рассматриваемый ПЗ, равным 0 ($j=0$). Вычислить значение $N_p: N_p = \sum_{i=1}^N m_i$.

2. Задать номер l прибора, для пакетов в последовательности π^l которого реализуется вычисление значений t_{jq}^{0l} моментов времени начала выполнения заданий, равным 1.

3. Модифицировать номер j позиции в последовательности π^l , в которой размещен рассматриваемый ПЗ: $j=j+1$. Определить

количество заданий N_j в пакете, занимающем j -ю позицию в последовательностях π^l ($l = \overline{1, L}$). Задать значение номера q задания в рассматриваемом пакете, занимающем j -ю позицию в последовательностях π^l , равным 1: $q=1$.

4. Если $l=1$ и $j=1$, то выполнить:

- а. инициализацию значения t_{11}^{0l} по формуле (43) при $q=1$;
- б. вычисление значения t_{1q}^{0l} по формуле (44) при $1 < q \leq b+1$;
- в. вычисление значения t_{1q}^{0l} по формуле (45) при $b+1 < q \leq N_1$.

Перейти на пункт 10.

5. Если $l=1$ и $2 \leq j \leq N_p$, то выполнить:

- а. вычисление значения t_{j1}^{0l} по формуле (46) при $q=1$;
- б. вычисление значения t_{1q}^{0l} по формуле (47) при $1 < q \leq b$;
- в. вычисление значения t_{1q}^{0l} по формуле (48) при $b+1 \leq q \leq N_1$.

Перейти на пункт 10.

6. Если $2 \leq l \leq L-1$ и $j=1$, то выполнить:

- а. вычисление значения t_{11}^{0l} по формуле (49) при $q=1$;
- б. вычисление значения t_{1q}^{0l} по формуле (50) при $1 < q \leq b+1$;
- в. вычисление значения t_{1q}^{0l} по формуле (51) при $b+1 < q \leq N_1$.

Перейти на пункт 10.

7. Если $2 \leq l \leq L-1$ и $2 \leq j \leq N_p$, то выполнить:

- а. вычисление значения t_{j1}^{0l} по формуле (52) при $q=1$;
- б. вычисление значения t_{jq}^{0l} по формуле (53) при $1 < q \leq b$;
- в. вычисление значения t_{jq}^{0l} по формуле (54) при $b+1 \leq q \leq N_j$.

Перейти на пункт 10.

8. Если $l=L$ и $j=1$, то выполнить:

- а. вычисление значения t_{11}^{0L} по формуле (55) при $q=1$;
- б. вычисление значения t_{1q}^{0L} по формуле (56) при $1 < q \leq N_1$.

Перейти на пункт 10.

9. Если $l=L$ и $2 \leq j \leq N_p$, то выполнить:

- а. вычисление значения t_{j1}^{0L} по формуле (57) при $q=1$;
- б. вычисление значения t_{jq}^{0L} по формуле (58) при $1 < q \leq N_j$.

10. Модифицировать значение номера прибора l : $l=l+1$. Если $l \leq L$, то перейти на пункт 6. Модифицировать значение q : $q=q+1$. Если $q \leq N_j$, то $l=l$, перейти на пункт 4. Если $q > N_j$, то перейти на пункт 11.

11. Модифицировать значение номера позиции j , которую занимает ПЗ: $j=j+1$. Если $j \leq N_p$, то инициализировать номер l прибора КС: $l=1$, перейти на пункт 3. Если $j > N_p$, то перейти на пункт 12.

12. Останов алгоритма.

Разработанная математическая модель многостадийных процессов выполнения ПЗ в КС при количестве приборов $L \geq 3$ и количестве пакетов $N_p \geq 2$, а также разработанный алгоритм моделирования указанных процессов, осуществляющий вычисление значений t_{jq}^{ol} на основе выражений модели, реализованы программно.

На основе вычисленных значений t_{jq}^{ol} формируются диаграммы Ганта, которые отображают процесс выполнения заданий, входящих в пакеты, приборами КС, включающий как интервалы блокирования и простоев приборов в ожидании начала реализации действий с заданиями, так и интервалы непосредственно функционирования при осуществлении действий с ними. На диаграммах Ганта отображаются моменты времени начала выполнения заданий, входящих в пакеты. С целью формирования диаграмм Ганта и отображения значений t_{jq}^{ol} , соответствующих процессу выполнения заданий, входящих в пакеты, разработанное приложение использует листы в программе Microsoft Excel. Вид формируемых приложением на основе модели и алгоритма расчета значений t_{jq}^{ol} диаграмм Ганта представлен на рисунке 3. Анализ формируемых приложением для различных параметров многостадийных процессов диаграмм Ганта показал корректность синтезированных моделей и алгоритма вычисления значений t_{jq}^{ol} .

Разработанные модели и алгоритм вычисления значений t_{jq}^{ol} в дальнейшем будут положены в основу применения разработанного иерархического подхода к оптимизации решений по составам ПЗ и расписаниям их выполнения на приборах КС для решения рассматриваемой задачи планирования многостадийных процессов при учете промежуточных буферов ограниченного размера.

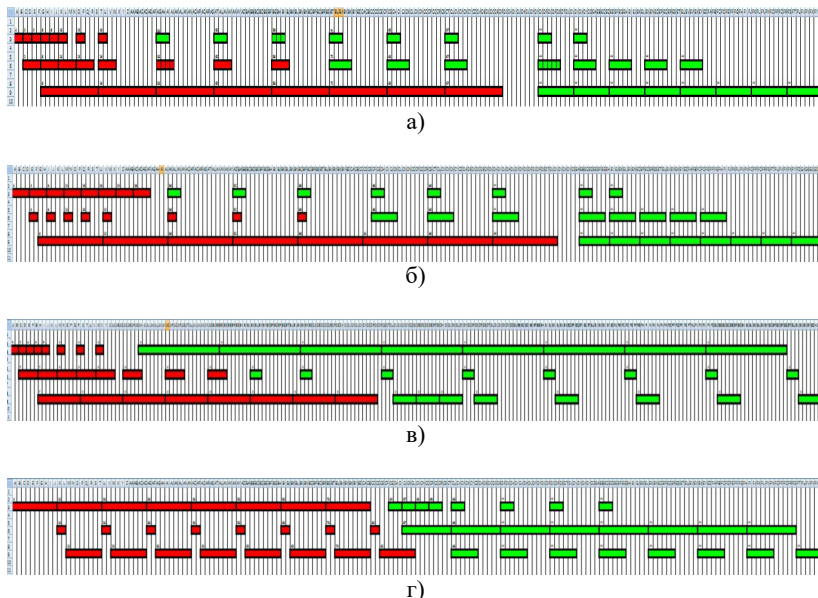


Рис. 3. Виды последовательностей выполнения ПЗ на трех приборах КС и значения моментов времени начала реализации действий с ними, сформированные программой моделирования многостадийных процессов

4. Заключение. Применение современных методов математического моделирования и оптимизации расписаний многостадийных процессов выполнения ПЗ в КС, включающих буферы заданных размеров, ограничено размерностью решаемых задач (ограниченное количество пакетов заданий и приборов в КС). В тоже время применение предложенного автором иерархического подхода к оптимизации составов ПЗ и расписаний их выполнения в КС (при наличии буферов ограниченных размеров) основывается на математических моделях процессов выполнения пакетов в КС. В связи с этим выполнены исследования, позволившие получить следующие новые научные результаты: 1) синтезирована математическая модель процессов выполнения ПЗ в КС, состоящей из двух приборов при произвольном (не ограниченном, в отличие от существующих методов) их (пакетов) количестве $N_p \geq 2$; 2) синтезирована математическая модель процессов выполнения ПЗ в КС, состоящей из трех приборов при произвольном (не ограниченном) количестве пакетов $N_p \geq 2$; 3) синтезирована математическая модель процессов

выполнения ПЗ в КС, состоящей из $L \geq 3$ приборов при количестве пакетов $N_p \geq 2$; 4) разработан алгоритм вычисления значений моментов времени начала выполнения заданий, входящих в пакеты, на приборах КС. Также разработана программа, интерпретирующая полученные математические модели многостадийных процессов и алгоритм вычисления моментов времени начала выполнения заданий, входящих в пакеты, на приборах КС. Результатом выполнения программы являются формируемые ею диаграммы Ганта, отображающие ход моделируемых многостадийных процессов выполнения ПЗ в КС. Использование разработанных математических моделей процессов выполнения ПЗ в КС, включающих промежуточные буферы ограниченных размеров, позволяет получать не только моменты времени окончания выполнения всех ПЗ в КС, но и интервалы простоя приборов, связанные с их блокированием по причине заполнения буферов, а также эффективность использования буферов при выполнении ПЗ. Это делает возможным при синтезе математических моделей оптимизации решений по составам ПЗ и расписаний их выполнения в КС формировать критерии, позволяющие учесть эффективность использования различных ресурсов, используемых системой при выполнении пакетов заданий. Дальнейшие исследования в рассматриваемом направлении связаны с: 1) построением модели двухуровневой иерархической игры оптимизации решений, позволяющей определять составы ПЗ на верхнем уровне и расписаниям выполнения пакетов на нижнем уровне; 2) синтезом методов локальной оптимизации решений на каждом из уровней; 3) получением результатов моделирования и оптимизации, которые позволяют определить особенности многостадийных процессов выполнения ПЗ в КС при наличии между приборами буферов ограниченных размеров между приборами.

Литература

1. Кротов К.В. Комплексный метод определения эффективных решений по составам партий данных и расписаниям их обработки в конвейерных системах // Вычислительные технологии. 2018. Т. 23. № 3. С. 58–76.
2. Кротов К.В. Обоснование методов построения комплексных расписаний обработки партий данных при условии оперативного формирования комплектов из результатов // Вестник Воронежского Государственного Университета. Серия: Системный анализ и информационные технологии. 2018. № 4. С. 58–72.
3. Кротов К.В., Скاتков А.В. Построение комплексных расписаний выполнения пакетов заданий при формировании комплектов в заданные директивные сроки // Информатика и автоматизация. 2021. Т. 20. № 3. С. 654–689.
4. Кротов К.В. Построение комплексных расписаний обработки пакетов данных в конвейерной системе при задании ограничений на длительность интервалов

- времени ее функционирования // Труды учебных заведений связи. 2020. Т. 6. № 3. С. 75–89.
5. Papadimitriou C.H., Kanellakis P.C. Flowshop scheduling with limited temporary storage // Journal of Association for Computing Machinery. 1980. vol. 27. no. 3. pp. 533–549.
 6. Leisten R. Flowshop sequencing problems with limited buffer storage // International Journal of Production Research. 1990. vol. 28. no. 11. pp. 2085–2100. DOI: 10.1080/00207549008942855.
 7. Crowder B. Minimizing the makespan in a flexible flowshop with sequence dependent setup times, uniform machines and limited buffers // Graduate Theses, Dissertations and Problem Reports. Morgantown: West Virginia University, 2006. 145 p.
 8. Han Zh., Zhang Q., Shi H., Qi Yu., Sun L. Research on limited buffer scheduling problems in flexible flow shops with setup times // International Journal of Modelling. Identification and Control. 2019. vol. 32. no. 2. pp. 93–104.
 9. Eddaly M., Jarboui B., Siarry P., Rebaï A. An Estimation of Distribution Algorithm for Flowshop Scheduling with Limited Buffers // Natural Intelligence for Scheduling, Planning and Packing Problems. Part of the Studies in Computational Intelligence book series. 2009. vol. 250. pp. 89–110.
 10. Frasc J.V., Krumke S.O., Westphal S. MIP Formulations for Flowshop Scheduling with Limited Buffers // Proceedings of the First International ICST Conference «Theory and Practice of Algorithms in (Computer) Systems» (TAPAS). 2011. pp. 127–138.
 11. Fu Q., Sivakumar A.I., Li K. Optimisation of flow-shop scheduling with batch processor and limited buffer // International Journal of Production Research. 2012. vol. 50. no. 8. pp. 2267–2285.
 12. Кононова П.А. Алгоритмы построения расписаний для цеховых задач потокового типа с цифровым буфером // Дисс. канд. физ.-мат. наук. Новосибирск: Институт вычислительной математики и математической геофизики, 2012. 104 с.
 13. Cakici M.K. Parallel flow shop scheduling with common workstations // MS thesis. Middle East Technical University. 2019. 138 p.
 14. Кононова П.А., Кочетов Ю.А. Алгоритм локального поиска для построения расписаний работы одного станка с переналадкой оборудования и складом // Дискретный анализ и исследование операций. 2019. Т. 26. № 2. С. 60–78. DOI: 10.33048/daio.2019.26.634.
 15. Lin C.C., Liu W.-Y., Chen Y.-H. Considering Stockers in Reentrant Hybrid Flow Shop Scheduling with Limited Buffer Capacity // Computers and Industrial Engineering. 2020. vol. 139. no. 106154. DOI: 10.1016/j.cie.2019.106154.
 16. Takano M.I., Nagano M.S. Solving the permutation flow shop problem with blocking and setup time constraints // International Journal of Industrial Engineering Computations. 2020. no. 11. pp. 469–480. DOI: 10.5267/j.ijiec.2019.11.002.
 17. Zhang C., Tan J., Peng K., Gao L., Shen W., Lian K. A discrete whale swarm algorithm for hybrid flow-shop scheduling problem with limited buffers // Robotics and Computer-Integrated Manufacturing. 2021. no. 68. DOI: 10.1016/j.rcim.2020.102081.
 18. Gao Z., Feng Ya., Xing K. A hybrid estimation-of-distribution algorithm for scheduling flexible job shop with limited buffers based on Petri Nets // IEEE Access. 2020. vol. 8. pp. 165396–165408. DOI: 10.1109/ACCESS.2020.3022655.
 19. Kazemi Esfeh M., Shojaie A.A., Javanshir H., Khalili-Damghani K. Flexible flow shop scheduling problem with reliable transporters and intermediate limited buffers via considering learning effects and budget constraint // Hindawi journal. Complexity. 2022. vol. 2022. no. 1253336. DOI: 10.1155/2022/1253336.

20. Jiang S.-L., Zhang L. Energy-oriented scheduling for hybrid flow shop with limited buffers through efficient multi-objective optimization // IEEE Access. 2019. vol. 7. pp. 34477–34487. DOI: 10.1109/ACCESS.2019.2904848.
21. Liang X., Wang P., Huang M. Flow shop scheduling problem with limited buffer based on hybrid shuffled frog leaping algorithm // 7th International Conference on Computer Science and Network Technology (ICCSNT). 2019. pp. 87–93. DOI: 10.1109/iccsnt47585.2019.8962427.
22. Zhang G., Wang L., Xing K. Dual-space co-evolutionary memetic algorithm for scheduling hybrid differentiation flow shop with limited buffer constraints // IEEE Transactions on systems, man and cybernetics systems. 2021. vol. 52. no. 11. pp. 6822–6836. DOI: 10.1109/TSMC.2021.3102658.
23. Benda F., Braune R., Doerner K.F., Hartl R.F. A machine learning approach for flow shop scheduling problems with alternative resources, sequence-dependent setup times and blocking // OR Spectrum. 2019. vol. 41. pp. 871–893. DOI: 10.1007/s00291-019-00567-8.
24. Han Zh., Han Ch., Lin Sh., Dong X., Shi H. Flexible flow shop scheduling method with public Buffer // Processes. 2019. vol. 7(10). no. 681. DOI: 10.3390/pr7100681.
25. Rooeinfar R., Raissi S., Ghezavati V.R. Stochastic flexible flow shop scheduling problem with limited buffers and fixed interval preventive maintenance: a hybrid approach of simulation and metaheuristic algorithms // Simulation. 2019. vol. 95(6). pp. 509–528.
26. Kazemi Esfeh M., Shojaei A.A., Javanshir H., Khalili Damghani K. Solving a bi-objective flexible flow shop problem with transporter preventive maintenance planning and limited buffers by NSGA-II and MOPSO // International Journal Nonlinear Analysis and Applications. 2022. vol. 13. no. 1. pp. 217–246.
27. Agnetis A., Pacciarelli D., Rossi F. Batch scheduling in a two-machine flow shop with limited buffer // Discrete Applied Mathematics. 1997. vol. 72. no. 3. pp. 243–260.
28. Pranzo M. Batch scheduling in a two-machine flow shop with limited buffer and sequence independent setup times and removal times // European Journal of Operational Research. 2004. vol. 153. no. 3. pp. 581–592.
29. Dai J. Batch Scheduling of Two-machine Limited-buffer Flow Shop with Setup and Removal Times // Georgia Institute of Technology, 2003. 108 p.

Кротов Кирилл Викторович — д-р техн. наук, доцент, профессор кафедры «информационные системы», ФГАОУ ВО «Севастопольский государственный университет» (СевГУ). Область научных интересов: теория расписаний, многоуровневая оптимизация, методы дискретной оптимизации. Число научных публикаций — 120. krotov_kl@mail.ru; улица Университетская, 33, 299053, Севастополь, Россия; п.т.: +7(978)730-3819.

K. KROTOV

**MATHEMATICAL MODELING OF THE PROCESSES OF
EXECUTING PACKAGES OF TASKS IN CONVEYOR SYSTEMS
WITH INTERMEDIATE BUFFERS OF LIMITED SIZE**

Krotov K. Mathematical Modeling of the Processes of Executing Packages of Tasks in Conveyor Systems with Intermediate Buffers of Limited Size.

Abstract. Modern methods of process planning in conveyor systems with buffers of a certain size between processing devices allow optimizing schedules for single tasks or fixed task packages with a limited number of them and a limited number of devices. The use of mathematical models of the processes of performing single tasks (task packages) used by these methods in optimizing the composition of packages and schedules for their execution in systems with an arbitrary number of packages and devices is impossible. At the same time, mathematical models of the processes of executing task packages in conveyor systems in the presence of buffers of limited sizes between devices are the basis for the development of methods for optimizing their (package) compositions and schedules for the implementation of actions with them on the devices of conveyor systems. In this regard, the article develops mathematical models of multi-stage processes of performing an arbitrary number of task packages in conveyor systems in the presence of intermediate buffers of limited sizes for two and three devices, as well as for an arbitrary number of devices. The use of these models makes it possible to determine the time points of the start of the execution of task packages on the devices of conveyor systems, taking into account the limited size of intermediate buffers, as well as the duration of time intervals for the use of these resources and the efficiency of their use over time. An algorithm has also been developed for mathematical modeling of the processes of executing task packages in conveyor systems in the presence of intermediate buffers of limited size, which calculates the time characteristics of these processes based on a given order of implementation of actions with task packages on the devices of conveyor systems. An application has been developed that implements synthesized mathematical models of the processes of executing task packages in conveyor systems with intermediate buffers of limited sizes and an appropriate method for modeling these processes. Versatile testing of the developed application has shown that the obtained mathematical models and the modeling method adequately describe the course of multi-stage processes of task packages in pipeline systems, set using different values of their (processes) parameters.

Keywords: pipeline systems, schedules, buffers of limited size, mathematical modeling of multi-stage processes for the execution of task packages.

References

1. Krotov K.V. [A complex method for determining effective solutions for data batch compositions and processing schedules in conveyor systems]. *Vychislitel'nye tekhnologii*, Izd-vo Instituta vychislitel'nyh tekhnologij SO RAN – Computational Technologies, Publishing House of the Institute of Computational Technologies SB RAS. 2018. vol. 23. no. 3. pp. 58–76. (In Russ).
2. Krotov K.V. [Substantiation of methods for constructing complex schedules for processing batches of data under the condition of prompt formation of sets from the results]. *Vestnik Voronezhskogo Gosudarstvennogo Universiteta. Seriya: Sistemnyj analiz i informacionnye tekhnologii – Bulletin of the Voronezh State University. Series: System analysis and Information Technologies*. 2018. no. 4. pp. 58–72. (In Russ).

3. Krotov K.V., Skatkov A.V. [Construction of complex schedules for the execution of task packages when forming sets within the specified directive deadlines]. *Informatika i avtomatizacija – Informatics and Automation*. 2021. vol. 20. no. 3. pp. 654–689. (In Russ).
4. Krotov K.V. [Building complex schedules for processing data packets in a pipeline system when setting limits on the duration of time intervals of its operation] *Trudy uchebnyh zavedenij svyazi – Proceedings of educational institutions of communication*. 2020. vol. 6. no. 3. pp. 75–89. (In Russ).
5. Papadimitriou C.H., Kanellakis P.C. Flowshop scheduling with limited temporary storage. *Journal of Association for Computing Machinery*. 1980. vol. 27. no. 3. pp. 533–549.
6. Leisten R. Flowshop sequencing problems with limited buffer storage. *International Journal of Production Research*. 1990. vol. 28. no. 11. pp. 2085–2100. DOI: 10.1080/00207549008942855.
7. Crowder B. Minimizing the makespan in a flexible flowshop with sequence dependent setup times, uniform machines and limited buffers. *Graduate Theses, Dissertations and Problem Reports*. Morgantown: West Virginia University, 2006. 145 p.
8. Han Zh., Zhang Q., Shi H., Qi Yu., Sun L. Research on limited buffer scheduling problems in flexible flow shops with setup times. *International Journal of Modelling. Identification and Control*. 2019. vol. 32. no. 2. pp. 93–104.
9. Eddaly M., Jarbouli B., Siarry P., Rebaï A. An Estimation of Distribution Algorithm for Flowshop Scheduling with Limited Buffers. *Natural Intelligence for Scheduling, Planning and Packing Problems. Part of the Studies in Computational Intelligence book series*. 2009. vol. 250. pp. 89–110.
10. Frasc J.V., Krumke S.O., Westphal S. MIP Formulations for Flowshop Scheduling with Limited Buffers. *Proceedings of the First International ICST Conference «Theory and Practice of Algorithms in (Computer) Systems» (TAPAS)*. 2011. pp. 127–138.
11. Fu Q., Sivakumar A.I., Li K. Optimisation of flow-shop scheduling with batch processor and limited buffer. *International Journal of Production Research*. 2012. vol. 50. no. 8. pp. 2267–2285.
12. Kononova P.A. [Algorithms for building schedules for shop tasks of a streaming type with a digital buffer. Dissertation of the Candidate of Physical and Mathematical Sciences]. Novosibirsk: Institute of Computational Mathematics and Mathematical Geophysics, 2012. 104 p.
13. Cakici M.K. Parallel flow shop scheduling with common workstations. MS thesis. Middle East Technical University. 2019. 138 p.
14. Kononova P.A., Kochetov Yu.A. [Local search algorithm for building schedules for the operation of one machine with equipment changeover and warehouse]. *Diskretnyj analiz i issledovanie operacij – Discrete analysis and operations research*. 2019. vol. 26. no. 2. pp. 60–78. DOI: 10.33048/daio.2019.26.634.
15. Lin C.C., Liu W.-Y., Chen Y.-H. Considering Stockers in Reentrant Hybrid Flow Shop Scheduling with Limited Buffer Capacity. *Computers and Industrial Engineering*. 2020. vol. 139. no. 106154. DOI: 10.1016/j.cie.2019.106154.
16. Takano M.I., Nagano M.S. Solving the permutation flow shop problem with blocking and setup time constraints. *International Journal of Industrial Engineering Computations*. 2020. no. 11. pp. 469–480. DOI: 10.5267/j.ijiec.2019.11.002.
17. Zhang C., Tan J., Peng K., Gao L., Shen W., Lian K. A discrete whale swarm algorithm for hybrid flow-shop scheduling problem with limited buffers. *Robotics and Computer-Integrated Manufacturing*. 2021. no. 68. DOI: 10.1016/j.rcim.2020.102081.

18. Gao Z., Feng Ya., Xing K. A hybrid estimation-of-distribution algorithm for scheduling flexible job shop with limited buffers based on Petri Nets. *IEEE Access*. 2020. vol. 8. pp. 165396–165408. DOI: 10.1109/ACCESS.2020.3022655.
19. Kazemi Esfeh M., Shojaie A.A., Javanshir H., Khalili-Damghani K. Flexible flow shop scheduling problem with reliable transporters and intermediate limited buffers via considering learning effects and budget constraint. *Hindawi journal. Complexity*. 2022. vol. 2022. no. 1253336. DOI: 10.1155/2022/1253336.
20. Jiang S.-L., Zhang L. Energy-oriented scheduling for hybrid flow shop with limited buffers through efficient multi-objective optimization. *IEEE Access*. 2019. vol. 7. pp. 34477–34487. DOI: 10.1109/ACCESS.2019.2904848.
21. Liang X., Wang P., Huang M. Flow shop scheduling problem with limited buffer based on hybrid shuffled frog leaping algorithm. *7th International Conference on Computer Science and Network Technology (ICCSNT)*. 2019. pp. 87–93. DOI: 10.1109/iccsnt47585.2019.8962427.
22. Zhang G., Wang L., Xing K. Dual-space co-evolutionary memetic algorithm for scheduling hybrid differentiation flow shop with limited buffer constraints. *IEEE Transactions on systems, man and cybernetics systems*. 2021. vol. 52. no. 11. pp. 6822–6836. DOI: 10.1109/TSMC.2021.3102658.
23. Benda F., Braune R., Doerner K.F., Hartl R.F. A machine learning approach for flow shop scheduling problems with alternative resources, sequence-dependent setup times and blocking. *OR Spectrum*. 2019. vol. 41. pp. 871–893. DOI: 10.1007/s00291-019-00567-8.
24. Han Zh., Han Ch., Lin Sh., Dong X., Shi H. Flexible flow shop scheduling method with public Buffer. *Processes*. 2019. vol. 7(10). no. 681. DOI: 10.3390/pr7100681.
25. Rooeinfar R., Raissi S., Ghezavati V.R. Stochastic flexible flow shop scheduling problem with limited buffers and fixed interval preventive maintenance: a hybrid approach of simulation and metaheuristic algorithms. *Simulation*. 2019. vol. 95(6). pp. 509–528.
26. Kazemi Esfeh M., Shojaei A.A., Javanshir H., Khalili Damghani K. Solving a bi-objective flexible flow shop problem with transporter preventive maintenance planning and limited buffers by NSGA-II and MOPSO. *International Journal Nonlinear Analysis and Applications*. 2022. vol. 13. no. 1. pp. 217–246.
27. Agnetis A., Pacciarelli D., Rossi F. Batch scheduling in a two-machine flow shop with limited buffer. *Discrete Applied Mathematics*. 1997. vol. 72. no. 3. pp. 243–260.
28. Pranzo M. Batch scheduling in a two-machine flow shop with limited buffer and sequence independent setup times and removal times. *European Journal of Operational Research*. 2004. vol. 153. no. 3. pp. 581–592.
29. Dai J. Batch Scheduling of Two-machine Limited-buffer Flow Shop with Setup and Removal Times. *Georgia Institute of Technology*, 2003. 108 p.

Krotov Kirill — Ph.D., Dr.Sci., Associate Professor, Professor of the department, Department of "information systems", Sevastopol State University. Research interests: theory of schedules, multilevel optimization, methods of discrete optimization. The number of publications — 120. krotov_k1@mail.ru; 33, Universitetskaya St., 299053, Sevastopol, Russia; office phone: +7(978)730-3819.

В.И. ВОРОТНИКОВ
**О ЧАСТИЧНОЙ УСТОЙЧИВОСТИ НЕЛИНЕЙНЫХ
ДИСКРЕТНЫХ СИСТЕМ С ЗАПАЗДЫВАНИЕМ**

Воротников В.И. О частичной устойчивости нелинейных дискретных систем с запаздыванием.

Аннотация. Рассматривается система нелинейных дискретных (конечно-разностных) уравнений общего вида с ограниченным запаздыванием. Интерес к задачам устойчивости таких систем в последние годы значительно возрос; в частности, это связано с актуальными проблемами управления через сеть. В основном анализируется задача устойчивости по всем переменным нулевого положения равновесия, поскольку заменой переменных к такой задаче сводится задача устойчивости по всем переменным любого решения рассматриваемой системы. Одним из основных методов исследования является дискретно-функциональный вариант прямого метода Ляпунова, получивший существенное развитие в теоретическом и прикладном аспектах. В данной статье предполагается, что рассматриваемая система уравнений допускает «частичное» (нулевое) положение равновесия, и ставится задача устойчивости по отношению к части определяющих это положение равновесия переменных. Такая задача относится к более общим задачам частичной устойчивости, которые исследуются для нелинейных динамических систем различной формы математического описания. Предложенная постановка задачи частичной устойчивости дополняет круг указанных исследований применительно к классу рассматриваемых систем. Для решения поставленной задачи применяется метод функционалов Ляпунова – Красовского в пространстве дискретных функций при соответствующей конкретизации требований к функционалам. Ослабления таких требований можно добиться введением дополнительных дискретных функций, посредством которых: 1) проводится корректировка области функционального пространства, где строятся функционалы Ляпунова – Красовского; 2) находятся оценки функционалов и их разностей (приращений) в силу рассматриваемой системы. В результате используемые функционалы и их разности (приращения) могут быть знакопеременными в области функционального пространства, обычно рассматриваемой при анализе частичной устойчивости. На основе предложенного подхода получены достаточные условия частичной устойчивости (асимптотической устойчивости) указанного вида. Особенности подхода показаны на примере двух классов нелинейных систем заданной структуры, для которых частичная устойчивость анализируется в пространстве параметров. При этом обращается внимание на целесообразность использования семейства функционалов.

Ключевые слова: система нелинейных дискретных уравнений с запаздыванием, частичная устойчивость, метод функционалов Ляпунова – Красовского в пространстве дискретных функций, однопараметрическое семейство функционалов.

1. Введение. Нелинейные системы дискретных (конечно-разностных) уравнений широко используются в современной теории управления, математическом моделировании в различных областях, а также в вычислительной математике и ее приложениях.

Теории и методам качественного исследования таких систем к настоящему времени посвящена обширная литература, в которой значительное внимание уделяется важным в прикладном смысле

задачам устойчивости; укажем монографии [1 – 4], наиболее близкие по тематике к данной статье. Наряду с наиболее часто изучаемой задачей устойчивости по всем переменным нулевого положения равновесия также рассматриваются естественным образом возникающие в приложениях более общие задачи *частичной устойчивости* [5, 6]: устойчивости по части переменных нулевого положения равновесия, а также устойчивости по всем и по части переменных «частичного» (нулевого) положения равновесия. Применительно к нелинейным стандартным одношаговым системам дискретных уравнений:

$$\mathbf{x}(k+1) = \mathbf{F}(k, \mathbf{x}(k)), \quad (1)$$

указанные задачи частичной устойчивости изучались в контексте прямого метода Ляпунова [2, 4, 6, 7 – 11].

В последние 30 лет стали интенсивно разрабатываться подходы к качественному анализу (включая анализ устойчивости) более общих систем *дискретных уравнений с запаздыванием* [12 – 14]. В случае *постоянного* запаздывания такие системы являются системами обычных дискретных уравнений порядка выше первого ($m \geq 1$):

$$\mathbf{x}(k+1) = \mathbf{F}(k, \mathbf{x}(k), \mathbf{x}(k-1), \dots, \mathbf{x}(k-m)), \quad (2)$$

которые можно свести к стандартным дискретным системам вида (1) введением новых переменных и расширением пространства состояний.

Системы дискретных уравнений с *переменным* запаздыванием:

$$\mathbf{x}(k+1) = \mathbf{F}(k, \mathbf{x}(k), \mathbf{x}(k-\tau_1(k)), \dots, \mathbf{x}(k-\tau_m(k))), \quad (3)$$

где функции $\tau_i(k)$ принимают одно из целочисленных значений из промежутка $0 < \tau_i(k) \leq m$, возникают в задачах управления через сеть (networked control). Системы уравнений (3) можно рассматривать [15] как системы с переключениями (switched systems) между набором систем вида (2), соответствующих всем допустимым значениям запаздываний $\tau_i(k)$ и их комбинациям в системе (3).

Возможности анализа устойчивости систем вида (2) и (3) можно расширить, используя дискретно-функциональную трактовку этих систем и соответствующий аппарат прямого метода Ляпунова. К настоящему времени в данном направлении получен ряд теорем для

общего класса систем нелинейных функционально-разностных уравнений с запаздыванием [16 – 24], а также найдены конструктивно проверяемые условия устойчивости линейных дискретных систем с переменным запаздыванием на основе дискретных аналогов функционалов Ляпунова – Красовского и матричных неравенств [15, 16, 25 – 30].

Имеются также другие эффективные подходы к этим задачам, рассмотрение которых выходит за рамки данной статьи; укажем, например, работы [31, 32], где можно найти библиографию.

Круг указанных исследований небезынтересно дополнить за счет постановки задачи *частичной* устойчивости для общего класса нелинейных дискретных уравнений с ограниченным запаздыванием. А именно, далее предполагается, что система допускает «частичное» (по части переменных) нулевое положение равновесия и устойчивость рассматривается также только по части определяющих это положение равновесия переменных. Анализируется возможность решения поставленной задачи на основе вспомогательных функционалов, являющихся дискретными аналогами функционалов Ляпунова – Красовского, которые эффективно используются при анализе функционально-дифференциальных уравнений. Для расширения возможностей построения функционалов предлагается проводить корректировку области пространства дискретных функций, в которой функционалы строятся, а также оценивать функционалы и их разности (приращения) в силу рассматриваемой системы, используя подходящие дополнительные вспомогательные функции.

2. Постановка задачи. Пусть $\mathbb{R}^n$ – линейное n -мерное пространство векторов $\mathbf{x}$ с евклидовой нормой $|\mathbf{x}|$; $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ – множество целых неотрицательных чисел; $\mathbb{Z}_0 = \{-m, -m+1, \dots, 0\}$, m – заданное целое положительное число.

Будем рассматривать общий класс систем нелинейных дискретных уравнений с ограниченным запаздыванием, определяемый системой функционально-разностных уравнений вида [17 – 24]:

$$\mathbf{x}(k+1) = \mathbf{F}(k, \mathbf{x}_k), \quad (4)$$

состояние системы (4) на каждом шаге (в каждый дискретный момент времени $k \in \mathbb{Z}_+$) определяется дискретной вектор-функцией $\mathbf{x}_k = \mathbf{x}(k+j)$, $j \in \mathbb{Z}_0$, а начальное состояние $\mathbf{x}_0$ – дискретным набором

значений $\mathbf{x}_0 = \{\mathbf{x}(k_0), \mathbf{x}(k_0 - 1), \dots, \mathbf{x}(k_0 - m)\}$, образующих матрицу размера $n \times (m + 1)$.

Пусть оператор $\mathbf{F}(k, \boldsymbol{\psi}): \mathbb{Z}_+ \times \mathcal{H} \rightarrow \mathbb{R}^n$, определяющий правую часть системы (4) в пространстве $\mathcal{H}$ дискретных (сеточных) вектор-функций $\boldsymbol{\psi}: \mathbb{Z}_0 \rightarrow \mathbb{R}^n$ с нормой $\|\boldsymbol{\psi}\| = \max\{|\boldsymbol{\psi}(0)|, |\boldsymbol{\psi}(-1)|, \dots, |\boldsymbol{\psi}(-m)|\}$, при каждом $k \in \mathbb{Z}_+$ непрерывен по $\boldsymbol{\psi}$ в области $\|\boldsymbol{\psi}\| < \infty$. Тогда для любых $k_0 \geq 0$, $\mathbf{x}_0$ существует единственный дискретный процесс $\mathbf{x}(k, \mathbf{x}_0)$, при всех $k \geq k_0$ определяющий решение этой системы с начальным состоянием $\mathbf{x}_0$. Обозначим $\mathbf{x}(k) = \mathbf{x}(k; k_0, \mathbf{x}_0)$ значения вектор-функции $\mathbf{x}(k_0, \mathbf{x}_0)$ на k -ом шаге процесса.

Проводя разбиение $\mathbf{x} = (\mathbf{y}^T, \mathbf{z}^T)^T$ фазового вектора системы (4) на две группы переменных, представим систему (4) в виде:

$$\mathbf{y}(k+1) = \mathbf{F}_y(k, \mathbf{y}_k, \mathbf{z}_k), \quad \mathbf{z}(k+1) = \mathbf{F}_z(k, \mathbf{y}_k, \mathbf{z}_k).$$

Если выполнено условие $\mathbf{F}_y(k, \mathbf{0}, \mathbf{z}_k) \equiv \mathbf{0}$, то множество $M = \{\mathbf{x}(k): \mathbf{y}_k = \mathbf{0}\}$ определяет «частичное» положение равновесия системы (4), являющееся инвариантным множеством этой системы. Предположения о существовании «полного» положения равновесия $\mathbf{x}_k = \mathbf{0}$ системы (4) не требуется.

Имея в виду анализ устойчивости «частичного» положения равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) не по всем определяющим его переменным, а только по отношению к их некоторой заданной части [6], предположим, что $\mathbf{y} = (\mathbf{y}_1^T, \mathbf{y}_2^T)^T$ и устойчивость рассматривается по переменным, входящим в вектор $\mathbf{y}_1$. Для расширения круга понятий $\mathbf{y}_1$ -устойчивости введем также произвольным образом разбиение $\mathbf{z} = (\mathbf{z}_1^T, \mathbf{z}_2^T)^T$ вектора $\mathbf{z}$ на две группы переменных.

Обозначим через D_δ область значений $\mathbf{x}_0$ таких, что $\|\mathbf{y}_0\| < \delta$, $\|\mathbf{z}_{10}\| \leq L, \|\mathbf{z}_{20}\| < \infty$; область D_Δ получается заменой δ на Δ . Здесь $\|\mathbf{y}_0\| = \max |\mathbf{y}(k_0 + j)|$, $\|\mathbf{z}_{0i}\| = \max |\mathbf{z}_i(k_0 + j)|$ ($i = 1, 2$) при $j \in \mathbb{Z}_0$.

Определения. «Частичное» положение равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) *при больших значениях $\mathbf{z}_{10}$ в целом по $\mathbf{z}_{20}$* :

1) *$\mathbf{y}_1$ -устойчиво*, если для каждого $k_0 \in \mathbb{Z}_+$, а также для произвольного числа $\varepsilon > 0$, как бы мало оно не было, и для любого

наперед заданного числа $L > 0$ найдется $\delta(\varepsilon, k_0, L) > 0$ такое, что неравенство $|\mathbf{y}_1(k, k_0, \mathbf{x}_0)| < \varepsilon$ имеет место для всех $k \geq k_0$, если $\mathbf{x}_0 \in D_\delta$;

2) *равномерно $\mathbf{y}_1$ -устойчиво*, если $\delta = \delta(\varepsilon, L)$;

3) *равномерно асимптотически $\mathbf{y}_1$ -устойчиво*, если оно равномерно $\mathbf{y}_1$ -устойчиво и существует $\Delta(L) > 0$ такое, что для произвольного решения системы (4), для которого $\mathbf{x}_0 \in D_\Delta$, предельное соотношение $|\mathbf{y}_1(k; k_0, \mathbf{x}_0)| = 0$, $k \rightarrow \infty$ выполняется равномерно по $k_0, \mathbf{x}_0$ из области $k_0 \geq 0, \mathbf{x}_0 \in D_\Delta$.

Задача. Найти условия $\mathbf{y}_1$ -устойчивости и равномерной асимптотической $\mathbf{y}_1$ -устойчивости при больших значениях $\mathbf{z}_{10}$ в целом по $\mathbf{z}_{20}$ «частичного» положения равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) в контексте дискретного варианта метода функционалов Ляпунова – Красовского.

Данную задачу можно рассматривать и как вспомогательную при анализе устойчивости по *всем* переменным «частичного» положения равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4), конкретизируя при этом использующуюся в нелинейной теории управления концепцию детектируемости динамических систем [33]. Если систему (4) рассматривать как замкнутую систему управления, то возникает соответствующая задача частичной стабилизации для рассматриваемого в статье класса систем посредством управлений вида $\mathbf{u} = \mathbf{u}(k, \mathbf{x}_k)$.

Замечание 1. Наиболее близкими к введенным являются понятия устойчивости по всем [4] и по части переменных [11, 34, 35] «частичного» положения равновесия систем дискретных, а также функционально-дифференциальных уравнений [36, 37]. Сделанное разбиение $\mathbf{z} = (\mathbf{z}_1^T, \mathbf{z}_2^T)^T$ можно конкретизировать с целью минимизации требований к используемым для решения поставленной задачи дискретным аналогам функционалов Ляпунова – Красовского.

3. Условия частичной устойчивости. В пространстве $\mathcal{H}$ дискретных вектор-функций $\boldsymbol{\psi}: \mathbb{Z}_0 \rightarrow \mathbb{R}^n$ будем рассматривать однозначные, непрерывные по $\boldsymbol{\psi}$ при каждом $k \in \mathbb{Z}_+$ скалярные функционалы $V = V(k, \boldsymbol{\psi})$, $V(k, \mathbf{0}) \equiv 0$, определенные в области:

$$k \geq 0, \|\boldsymbol{\psi}_{y_1}\| < h, \|\boldsymbol{\psi}_{y_2}\| + \|\boldsymbol{\psi}_z\| < \infty. \quad (5)$$

Разбиение $\Psi = (\Psi_{y_1}^T, \Psi_{y_2}^T, \Psi_z^T)^T$ соответствует сделанному разбиению $\mathbf{x} = (\mathbf{y}_1^T, \mathbf{y}_2^T, \mathbf{z}^T)^T$ фазового вектора $\mathbf{x}$ системы (4); $\|\Psi_{y_i}\| = \max |\Psi_{y_i}(\theta)|$ ($i=1, 2$), $\|\Psi_z\| = \max |\Psi_z(\theta)|$ при $\theta \in \mathbb{Z}_0$.

Аналогом производной рассматриваемых функционалов в силу системы (4) являются их *разности* (приращения) в силу этой системы $\Delta V(k, \Psi) = V(k+1, \mathbf{X}(k, \Psi)) - V(k, \Psi)$. Имеет место соотношение:

$$V(k, \Psi) = V(k_0, \mathbf{x}_0) + \sum_{j=k_0}^{k-1} \Delta V(j, \Psi). \quad (6)$$

Для анализа y_1 -устойчивости «частичного» положения равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) дополнительно будут использоваться следующие вспомогательные функционалы и функции.

1) Скалярные непрерывные в области (5) функционалы $V^*(k, \Psi_y, \Psi_{z_1})$, $V^*(\Psi_y, \Psi_{z_1})$ для указания *оценки сверху* для основного V -функционала, необходимость которой обусловлена спецификой поставленной задачи частичной устойчивости. Вектор-функция Ψ_{z_1} определяется разбиением $\Psi_z = (\Psi_{z_1}^T, \Psi_{z_2}^T)^T$, соответствующим разбиению $\mathbf{z} = (\mathbf{z}_1^T, \mathbf{z}_2^T)^T$.

2) Непрерывная в области (5) вектор-функция $\mu(k, \Psi)$, $\mu(k, \mathbf{0}) \equiv \mathbf{0}$, посредством которой корректируется область, где строится V -функционал. Определим $\|\mu(k, \Psi)\| = \sup |\mu(k, \Psi(\theta))|$ при $k \in \mathbb{Z}_+$, $\theta \in \mathbb{Z}_0$.

3) Непрерывные монотонно возрастающие по $r > 0$ скалярные функции $a_i(r)$, $a_i(0) = 0$ ($i = 1, \dots, 4$) функции типа Хана, позволяющие указать для основного V -функционала необходимую *оценку снизу*.

Введение вспомогательной $\mu(t, \Psi)$ -функции связано с тем, что анализ y_1 -устойчивости «частичного» положения равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) в обычно рассматриваемой области:

$$k \geq 0, \|\Psi_{y_1}\| < h_1 < h, \|\Psi_{y_2}\| + \|\Psi_z\| < \infty, \quad (7)$$

пространства $\mathcal{H}$ не всегда позволяет выявить желаемые свойства используемого V -функционала или наделить его этими свойствами.

Поэтому целесообразно перейти к рассмотрению V -функционала в более узкой области:

$$k \geq 0, \|\Psi_{y_1}\| + \|\mu(k, \Psi)\| < h_1 < h, \|\Psi_{y_2}\| + \|\Psi_z\| < \infty, \quad (8)$$

учитывая, что фактически y_1 -устойчивость «частичного» положения равновесия $y_k = \mathbf{0}$ системы (4) означает выполнение соответствующих оценок не только для компонент вектора y_1 , но и для компонент некоторой векторной, вообще говоря, $\mu(k, \mathbf{x})$ - функции фазовых переменных системы (4).

Однако указанную $\mu(k, \mathbf{x})$ - функцию не всегда возможно указать заранее и соответствующую ей $\mu(k, \Psi)$ - функцию в рассматриваемом пространстве $\mathcal{H}$ дискретных вектор-функций естественно трактовать как дополнительную вектор-функцию, которая (как и сам подходящий V -функционал) определяется в процессе решения исходной задачи y_1 -устойчивости «частичного» положения равновесия $y_k = \mathbf{0}$ системы (4).

Указанные обстоятельства приводят к целесообразности корректировки области (7) функционального пространства $\mathcal{H}$, в которой происходит построение V -функционалов (являющихся дискретными аналогами функционалов Ляпунова – Красовского), посредством дополнительной вспомогательной $\mu(k, \Psi)$ - функции.

Теорема 1. Пусть для системы (4) найдется V -функционал, а также вектор-функция $\mu(k, \Psi)$, $\mu(k, \mathbf{0}) \equiv \mathbf{0}$, для которых при достаточно малом $h_1 > 0$ в области (8) выполняются условия:

$$V(k, \Psi) \geq a_1(|\Psi_{y_1}(0)| + |\mu(k, \Psi(0))|), \quad (9)$$

$$V(k, \Psi) \leq V^*(k, \Psi_y, \Psi_{z_1}), \quad V^*(k, \mathbf{0}, \Psi_{z_1}) \equiv 0, \quad (10)$$

$$V(k, \Psi) \leq 0. \quad (11)$$

Тогда «частичное» положение равновесия $y_k = \mathbf{0}$ системы (4) y_1 -устойчиво при больших значениях z_{10} в целом по z_{20} .

Доказательство. Для любых $\varepsilon > 0$, $k_0 \in \mathbb{Z}_+$ и любого заданного числа $L > 0$ в силу непрерывности функционалов $V(k, \Psi)$ и $V^*(k, \Psi_y, \Psi_{z_1})$, условия $V(k, \mathbf{0}) \equiv 0$ и условий (10) можно найти $\delta(\varepsilon, k_0, L) > 0$ такое, что из $\mathbf{x}_0 \in D_\delta$ следует $V(k_0, \mathbf{x}_0) < a_1(\varepsilon)$.

Принимая во внимание соотношение (6), при выполнении условия (11) имеем $V(k, \mathbf{x}_k) \leq V(k_0, \mathbf{x}_0)$. Поэтому, учитывая неравенство

(9), для произвольного решения $\mathbf{x}(k_0, \mathbf{x}_0)$ системы (4) с $\mathbf{x}_0 \in D_\delta$ при всех $k \geq k_0$ справедливы неравенства:

$$a_1(|\mathbf{y}_1(k; k_0, \mathbf{x}_0)| + |\boldsymbol{\mu}(k, \mathbf{x}(k; k_0, \mathbf{x}_0))|) \leq V(k, \mathbf{x}_k) < a_1(\varepsilon). \quad (12)$$

В силу свойств функции Хана $a_1(r)$, на основании полученных неравенств (12) заключаем, что $|\mathbf{y}_1(k; k_0, \mathbf{x}_0)| < \varepsilon$ при всех $k \geq k_0$ для произвольного решения системы (4), если $\mathbf{x}_0 \in D_\delta$. Теорема доказана.

Следствие. Если условия (10) теоремы 1 заменить условиями:

$$V(k, \boldsymbol{\Psi}) \leq V^*(\boldsymbol{\Psi}_y, \boldsymbol{\Psi}_{z1}), \quad V^*(\mathbf{0}, \boldsymbol{\Psi}_{z1}) \equiv 0, \quad (13)$$

то «частичное» положение равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) равномерно $\mathbf{y}_1$ -устойчиво при больших значениях $\mathbf{z}_{10}$ в целом по $\mathbf{z}_{20}$.

Доказательство. При выполнении условия (13) для любых $\varepsilon > 0$, $k_0 \geq 0$ и для любого наперед заданного числа $L > 0$ найдется $\delta(\varepsilon, L) > 0$ такое, что $V(k_0, \mathbf{x}_0) < a_1(\varepsilon)$ при $\mathbf{x}_0 \in D_\delta$.

Тогда из соотношений (12) следует, что $|\mathbf{y}_1(k; k_0, \mathbf{x}_0)| < \varepsilon$ при всех $k \geq k_0$, если $\mathbf{x}_0 \in D_\delta$.

Замечание 2. Особенность условий теоремы 1 и следствия из нее в том, что рассматриваемый V -функционал и его разность (приращение) ΔV в силу системы (4), вообще говоря, *знакопеременны* в области (7). При этом наряду с V -функционалом дополнительная вспомогательная $\boldsymbol{\mu}$ -функция подбирается с целью наиболее рациональной замены области (7) областью (8). Условия (10), (13) выделяют *допустимую структуру* V -функционала, которая определяется спецификой поставленной задачи частичной устойчивости; допускается *произвольный* непрерывный V^* -функционал, для которого $V^*(t, \mathbf{0}, \boldsymbol{\Psi}_{z1}) \equiv 0$ или $V^*(\mathbf{0}, \boldsymbol{\Psi}_{z1}) \equiv 0$, и ограничивающий V -функционал сверху.

Замечание 3. В качестве допустимых V -функционалов можно использовать знакоопределенные (по всем переменным) *квадратичные* функционалы $V(k, \boldsymbol{\Psi}) \equiv V^*(k, \boldsymbol{\Psi}_{y1}, \boldsymbol{\mu}(k, \boldsymbol{\Psi}))$, если выбор $\boldsymbol{\mu}$ -функций согласовать с условиями (10), (13): возможны, например, $\boldsymbol{\mu}$ -функции вида $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\Psi}_{y2}, \boldsymbol{\Psi}_{z1})$, $\boldsymbol{\mu}(\mathbf{0}, \boldsymbol{\Psi}_{z1}) \equiv \mathbf{0}$. Также допускается использование соответствующих знакоопределенных по всем переменным функционалов более высокого порядка.

Замечание 4. Пусть система (4) допускает «полное» положение равновесия $\mathbf{x}_k = \mathbf{0}$. Если $\boldsymbol{\mu}(k, \boldsymbol{\Psi}) \equiv 0$ (область (7) не корректируется), а

требование $\mathbf{x}_0 \in D_\delta$ заменяется требованием $\|\mathbf{x}_0\| < \delta$, то при выполнении условий (9), (11) имеем дискретно-функциональный вариант теоремы В.В. Румянцева [5] об устойчивости по отношению к части переменных (y_1 -устойчивости). В данном случае выполнения условий (8) или (13) не требуется.

4. Условия частичной асимптотической устойчивости. Для анализа асимптотической y_1 -устойчивости «частичного» положения равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) дополнительно будут использоваться:

- 1) в пространстве $\mathcal{H}$ дискретных вектор-функций $\boldsymbol{\psi}$ норма:

$$\|\boldsymbol{\psi}\|_m = \sum_{j=-m}^0 |\boldsymbol{\psi}(j)|^2;$$

- 2) непрерывная в области (5) вектор-функция $\mathbf{w}(\boldsymbol{\psi})$, $\mathbf{w}(\mathbf{0}) \equiv \mathbf{0}$, посредством которой задаются необходимые оценки V -функционала и его разности (приращения) в систему системы (4).

Теорема 2. Пусть для системы (4) найдется V -функционал, а также непрерывные вектор-функции $\boldsymbol{\mu}(k, \boldsymbol{\psi})$, $\boldsymbol{\mu}(k, \mathbf{0}) \equiv \mathbf{0}$ и $\mathbf{w}(\boldsymbol{\psi})$, $\mathbf{w}(\mathbf{0}) = \mathbf{0}$ такие, что в области (8) наряду с условиями (13) выполнены условия:

$$a_1(|\boldsymbol{\psi}_{y_1}(0)| + |\boldsymbol{\mu}(k, \boldsymbol{\psi}(0))|) \leq V(k, \boldsymbol{\psi}) \leq a_3(|\mathbf{u}(0)|) + a_3(\|\mathbf{u}\|_m), \quad (14)$$

$$V(k, \boldsymbol{\psi}) \leq -a_4(|\mathbf{u}(0)|), \quad (15)$$

$$\mathbf{u} = [\boldsymbol{\psi}_y, \mathbf{w}(\boldsymbol{\psi})], \quad \mathbf{u}(0) = [\boldsymbol{\psi}_y(0), \mathbf{w}(\boldsymbol{\psi}(0))].$$

Тогда «частичное» положение равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) равномерно асимптотически y_1 -устойчиво при больших значениях $\mathbf{z}_{10}$ в целом по $\mathbf{z}_{20}$.

Доказательство. Равномерная y_1 -устойчивость при больших значениях $\mathbf{z}_{10}$ в целом по $\mathbf{z}_{20}$ «частичного» положения равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) имеет место на основании следствия из теоремы 1.

Равномерная асимптотическая y_1 -устойчивость будет иметь место, если для любых $\varepsilon > 0$, $k_0 \in \mathbb{Z}_+$ найдется целое число $N = N(\varepsilon, L) > 0$ такое, что $|\mathbf{y}_1(k; k_0, \mathbf{x}_0)| < \varepsilon$ при всех $k > k_0 + N(\varepsilon, L)$, если $\mathbf{x}_0 \in D_\delta$; здесь $\delta(\varepsilon, L)$ находится в силу равномерной y -устойчивости.

Пусть $\varepsilon > 0$ задано. Обозначим:

$$\begin{aligned}\mathbf{u}_k &= [\mathbf{y}_k(k_0, \mathbf{x}_0), \mathbf{w}(\mathbf{x}_k(k_0, \mathbf{x}_0))], \\ \mathbf{u}(k) &= [\mathbf{y}(k; k_0, \mathbf{x}_0), \mathbf{w}(\mathbf{x}(k; k_0, \mathbf{x}_0))].\end{aligned}$$

Выберем $\varepsilon_1 < \varepsilon$ так, что $a_2(\varepsilon_1) < \frac{1}{2}a_1(\varepsilon)$. Поскольку при $\|\mathbf{u}_k\|_m < \varepsilon_2$ справедливо неравенство $a_3(\|\mathbf{u}_k\|_m) \leq a_3(\varepsilon_2(m+1))$, то также выберем такое $\varepsilon_2 < \varepsilon$, что $a_3(\|\mathbf{u}_k\|_m) < \frac{1}{2}a_1(\varepsilon)$.

Положим $\varepsilon_3 = \min(\varepsilon_1, \varepsilon_2)$.

Покажем, что существует целое число $N_1(\varepsilon, L) > m$ такое, что неравенство $|\mathbf{u}(k; k_0, \mathbf{x}_0)| \geq \varepsilon_3$, где $\mathbf{x}_0 \in D_\delta$, нарушается для некоторого значения k в каждом интервале длины N_1 . Действительно, при $\mathbf{x}_0 \in D_\delta$ значение $V(k_0, \mathbf{x}_0)$ в силу условий (13) ограничено числом $r(\delta, L) > 0$. Поэтому, если $|\mathbf{u}(k; k_0, \mathbf{x}_0)| \geq \varepsilon_3$ на всем указанном интервале, то на этом интервале в силу условия (15) будут выполнены соотношения:

$$V(k+1, \mathbf{x}_k) = V(k_0, \mathbf{x}_0) + \sum_{j=k_0}^k \Delta V(j, \mathbf{x}_j) \leq r - N_1 a_4(\varepsilon_3),$$

и $V(k+1, \mathbf{x}_{k+1}) < 0$ при $N_1 > r/a_4(\varepsilon_3)$, что противоречит условию (14).

В результате можно указать монотонную последовательность значений $\{k_i\} \rightarrow \infty$ такую, что $|\mathbf{u}(k; k_0, \mathbf{x}_0)| < \varepsilon_3$. Выберем из этой последовательности подпоследовательность значений $\{k_i\}$ так, чтобы $k_{i+1} - k_i > m$. Таким образом, если $\mathbf{x}_0 \in D_\delta$, то найдется целое число $N_1(\varepsilon, L) > m$ такое, что $|\mathbf{u}(k; k_0, \mathbf{x}_0)| < \varepsilon_3$ при $k_i \in [k_0 + iN_1, k_0 + (i+1)N_1]$.

Рассмотрим последовательность функций $\{\mathbf{u}_k\}$, соответствующую выбранной подпоследовательности $\{k_i\}$. Покажем, что найдется целое число $p > 0$ такое, что $a_3(\|\mathbf{u}_k\|_m) < \frac{1}{2}a_1(\varepsilon)$ при $k = k_i (1 \leq i \leq p)$.

Предполагая противное, при любом $p > 0$ рассмотрим те значения $k = k_i$, для которых $a_3(\|\mathbf{u}_k\|_m) \geq \frac{1}{2}a_1(\varepsilon)$ и можно подобрать такое число $\gamma > 0$, что $a_4(|\mathbf{u}(k_i)|) \geq \gamma$.

В данном случае в силу условия (15) при $k \geq k_i$ имеем:

$$V(k+1, \mathbf{x}_k) = V(k_0, \mathbf{x}_0) + \sum_{j=k_0}^k \Delta V(j, \mathbf{x}_j) \leq r - (k_i - k_0)\gamma,$$

и при $k = k_l > k_0 + r/\gamma$ получаем $V(k + 1, \mathbf{x}_{k+1}) < 0$, что противоречит условию (14).

Таким образом, найдется целое число $p > 0$ такое, что:

$$a_1(|\mathbf{y}_1(k; k_0, \mathbf{x}_0)| + |\boldsymbol{\mu}(k, \mathbf{x}(k; k_0, \mathbf{x}_0))|) \leq V(k, \mathbf{x}_k) \leq V(k_l, \mathbf{x}_k) \leq a_1(\varepsilon),$$

при $k \geq k_p$ и, следовательно, $|\mathbf{y}_1(k; k_0, \mathbf{x}_0)| < \varepsilon$ для всех $k \geq k_p$.

Учитывая, что $k_p < k_0 + 2N_1p$, положим $N = 2N_1p$. Тогда при всех значениях $k > k_0 + N$ справедливо неравенство $|\mathbf{y}_1(k; k_0, \mathbf{x}_0)| < \varepsilon$ для произвольного решения системы (4), если $\mathbf{x}_0 \in D_\delta$.

Таким образом, «частичное» положение равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4) равномерно асимптотически $\mathbf{y}_1$ -устойчиво при больших значениях $\mathbf{z}_{10}$ в целом по $\mathbf{z}_{20}$. Теорема доказана.

Замечание 5. Вспомогательный V -функционал и его разность (приращение) в силу системы (4) в теореме 2, вообще говоря, *знакопеременны* в области (7). Выбор дополнительных $\mathbf{w}$ -функций (как и дополнительных $\boldsymbol{\mu}$ -функций) должен быть согласован с условиями (13). Условия теоремы 2 являются дискретно-функциональным вариантом ранее полученных условий [6, 36] для систем обыкновенных дифференциальных, а также функционально-дифференциальных уравнений с последствием (запаздыванием).

Замечание 6. Наличие не только $\mathbf{y}_1$ -устойчивости, но и «расширенной» $(\mathbf{y}_1, \boldsymbol{\mu})$ -устойчивости «частичного» положения равновесия $\mathbf{y}_k = \mathbf{0}$ системы (4), фактически имеющей место при выполнении условий теорем 1 и 2, обеспечивает правомерность применяемого перехода от области (7) к области (8).

5. Примеры. Покажем особенности предложенного подхода к анализу частичной устойчивости системы (4), а также целесообразность использования однопараметрического семейства функционалов для расширения возможностей такого анализа.

Пример 1. Пусть система (4) состоит из уравнений:

$$\begin{aligned} y_1(k+1) &= ay_1(k) + a^*y_1(k-1) + ly_2(k-1)z_1(k-1), \\ y_2(k+1) &= [b + dy_1(k-1)]y_2(k), \\ z_1(k+1) &= [c + ey_1(k-1)]z_1(k), \quad z_2(k+1) = Z_2(k, \mathbf{x}_k), \end{aligned} \quad (16)$$

в которых a, a^*, b, c, d, e, l – постоянные параметры; оператор Z_2 удовлетворяет только общим требованиям к системе (4).

Система (16) допускает «частичное» положение равновесия:

$$y_{1k} = y_{2k} = 0. \quad (17)$$

Для анализа задачи y_1 -устойчивости данного положения равновесия в пространстве $\mathcal{H}$ дискретных вектор-функций Ψ рассмотрим функционал вида $(M, \beta_1, \beta_2 = \text{const} > 0)$:

$$V(\Psi) = \Psi_{y_1}^2(0) + M\Psi_{y_2}^2(0)\Psi_{z_1}^2(0) + \beta_1\Psi_{y_1}^2(\theta) + \beta_2\Psi_{y_2}^2(\theta)\Psi_{z_1}^2(\theta), \quad (18)$$

являющийся дискретным аналогом функционала Ляпунова – Красовского, а также две вспомогательные скалярные дискретные функции:

$$\mu_1(\Psi) = w_1(\Psi) = \Psi_{y_2}(\theta)\Psi_{z_1}(\theta), \mathbb{Z}_0 = \{k = -1, 0\}. \quad (19)$$

Имеют место соотношения:

$$\begin{aligned} \Psi_{y_1}^2(0) + M\mu_1^2(0) &\leq V(\Psi) \leq \Psi_{y_1}^2(0) + M\mu_1^2(0) + \beta \|\mathbf{u}\|_1, \\ V(\Psi) &= V^*(\Psi_{y_1}, \Psi_{y_2}, \Psi_{z_1}), \quad V^*(0, 0, \Psi_{z_1}) \equiv 0, \\ \mathbf{u} &= [\Psi_{y_1}, w_1(\Psi)], \quad \beta = \max(\beta_1, \beta_2), \end{aligned}$$

и, следовательно, V -функционал (18) удовлетворяет условиям (13) и (14), а его разность (приращение) ΔV в силу системы (16) при всех $k \in \mathbb{Z}_+ \cup \mathbb{Z}_0$ определяется следующим образом:

$$\begin{aligned} \Delta V &= \{[a\Psi_{y_1}(0) + a^*\Psi_{y_1}(-1) + l\Psi_{y_2}(-1)\Psi_{z_1}(-1)]^2 + \\ &+ M\Psi_{y_2}^2(0)\Psi_{z_1}^2(0)[b + d\Psi_{y_1}(-1)]^2[c + e\Psi_{y_1}(-1)]^2\} - \\ &- \Psi_{y_1}^2(0) - M\Psi_{y_2}^2(0)\Psi_{z_1}^2(0) + \\ &+ \beta_1[\Psi_{y_1}^2(0) - \Psi_{y_1}^2(-1)] + \beta_2[\Psi_{y_2}^2(0)\Psi_{z_1}^2(0) - \Psi_{y_2}^2(-1)\Psi_{z_1}^2(-1)] = \\ &= (a^2 - 1 + \beta_1)\Psi_{y_1}^2(0) + 2aa^*\Psi_{y_1}(0)\Psi_{y_1}(-1) + (a^{*2} - \beta_1)\Psi_{y_1}^2(-1) + \\ &+ 2a^*l\Psi_{y_1}(-1)\mu_1(-1) + 2al\Psi_{y_1}(0)\mu_1(-1) + \\ &+ (Mb^2c^2 - M + \beta_2)\mu_1^2(0) + (l^2 - \beta_2)\mu_1^2(-1) + r_1\Psi_{y_1}(-1)\mu_1^2(0) + \\ &+ r_2\Psi_{y_1}^2(-1)\mu_1^2(0) + r_3\Psi_{y_1}^3(-1)\mu_1^2(0) + Md^2e^2\Psi_{y_1}^4(-1)\mu_1^2(0), \\ &\mu_1(0) = \Psi_{y_2}(0)\Psi_{z_1}(0), \quad \mu_1(-1) = \Psi_{y_2}(-1)\Psi_{z_1}(-1), \end{aligned}$$

где r_i ($i = 1, 2, 3$) – постоянные, зависящие от значения M и параметров b, c, d, e системы (16).

Используя для упрощения последующего анализа y_1 -устойчивости «частичного» положения равновесия (17) неравенства:

$$\begin{aligned} 2aa^*\psi_{y_1}(0)\psi_{y_1}(-1) &\leq |aa^*| \cdot [\psi_{y_1}^2(0) + \psi_{y_1}^2(-1)], \\ 2a^*h\psi_{y_1}(-1)\mu_1(-1) &\leq |a^*l| \cdot [\psi_{y_1}^2(-1) + \mu_1^2(-1)], \end{aligned}$$

для квадратичной части $(\Delta V)_2$ полученного выражения для ΔV можно получить следующую оценку:

$$\begin{aligned} (\Delta V)_2 &\leq (a^2 + |aa^*| - 1 + \beta_1)\psi_{y_1}^2(0) + \\ &+ 2al\psi_{y_1}(0)\mu_1(-1) + (l^2 + |a^*l| - \beta_2)\mu_1^2(-1) + \\ &+ (a^{*2} + |aa^*| + |a^*l| - \beta_1)\psi_{y_1}^2(-1) + (Mb^2c^2 - M + \beta_2)\mu_1^2(0). \end{aligned}$$

При выполнении условий:

$$\begin{aligned} a^2 + |aa^*| - 1 + \beta_1 &< 0, \\ (a^2 + |aa^*| - 1 + \beta_1) \cdot (l^2 + |a^*l| - \beta_2) &> a^2l^2, \\ a^{*2} + |aa^*| + |a^*l| - \beta_1 &< 0, \quad Mb^2c^2 - M + \beta_2 < 0. \end{aligned}$$

$(\Delta V)_2$ является определенно отрицательной функцией переменных $\psi_{y_1}(0), \psi_{y_1}(-1), \mu_1(0), \mu_1(-1)$ на основании критерия Сильвестра.

Поэтому в данном случае при достаточно малом $h_1 > 0$ в области (8) для разности (приращения) ΔV функционала (18) будет иметь место следующая оценка:

$$\begin{aligned} \Delta V &\leq -\alpha[\psi_{y_1}^2(0) + \psi_{y_1}^2(-1) + \mu_1^2(0) + \mu_1^2(-1)] \leq \\ &\leq -\alpha[\psi_{y_1}^2(0) + \mu_1^2(0)] \quad (\alpha = \text{const} > 0). \end{aligned}$$

Пусть параметры системы (16) удовлетворяют условиям:

$$\begin{aligned} (|a| + |a^*|)^2 + |a^*l| &< 1, \\ [(|a| + |a^*|)^2 + |a^*l| - 1] \cdot [l^2 + |a^*l| + M(b^2c^2 - 1)] &> a^2l^2, \end{aligned} \quad (20)$$

а параметры β_1, β_2 в функционале (18) выберем следующим образом:

$$\beta_1 = a^{*2} + |aa^*| + |a^*l| + \varepsilon_1, \quad \beta_2 = M(1 - b^2c^2) - \varepsilon_2$$

$$(\varepsilon_1, \varepsilon_2 = \text{const} > 0).$$

Учитывая, что $\mu_1(\psi) = w_1(\psi)$, заключаем, что при достаточно малых значениях $\varepsilon_1, \varepsilon_2$ и $h_1 > 0$ для разности (приращения) ΔV функционала в области (8) при любых значениях параметров d, e имеет место оценка ($\alpha = \text{const} > 0$):

$$\Delta V \leq -\alpha(|\mathbf{u}(0)|), \quad \mathbf{u}(0) = [\psi_{y1}(0), w_1(0)]. \quad (21)$$

Значит для V -функционала (18) в области (8), помимо условий (13) и (14), также выполняется условие (15).

На основании теоремы 2 заключаем, что при выполнении условий (20) и при любых значениях параметров d, e «частичное» положение равновесия (17) системы (16) равномерно асимптотически y_1 -устойчиво при больших значениях z_{10} в целом по z_{20} . Отметим, что поскольку условие (21) выполняется в области (8), но не в области (7), то разность (приращение) ΔV рассматриваемого функционала (18) в силу системы (16) *знакопеременна* в области (7).

Равенство (18) определяет *семейство функционалов*, зависящее от параметра $M > 0$, что, в свою очередь, приводит к зависимости от параметра M и найденной области устойчивости (20). В результате, за счет выбора подходящего значения M , в область устойчивости можно включить (или исключить) некоторые заданные комбинации параметров системы (4). Так, например, если $l^2 + |a^*l| = 1$, то для казалось бы «естественного» выбора значения $M = 1$ в функционале (18), при любых значениях параметров a, b, c, d, e область устойчивости (20) оказывается пустым множеством, поскольку в данном случае условия (20) предполагают требование $b^2c^2 < 0$. Однако можно избежать указанной ситуации, полагая $M = 2$.

Пусть $a^* = 0$. Покажем, что при $M = 1$ область равномерной асимптотической y_1 -устойчивости (20) можно изменить за счет изменения оценки ΔV . Действительно, при достаточно малом $h_1 > 0$ в области (8) неравенство (21) также будет справедливо, если выполнены условия $a^2 + |al| - 1 < 0$, $l^2 - \beta_2 + |al| < 0$, $b^2c^2 - 1 < 0$. Поэтому область (20) можно заменить областью:

$$(|a| + |l|)^2 + b^2c^2 < 2, \quad b^2c^2 < 1. \quad (22)$$

В отличие от области (20), при $M = l^2 = 1$ область (22) не является пустым множеством. Например, при $M = 1$, $b^2 c^2 = 0$ области устойчивости (20) и (22) имеют соответственно вид $a^2 + l^2 < 1$ и $|a| + |l| < \sqrt{2}$, причем область (22) охватывает случай $l^2 = 1$.

Отметим, что при $a^* = 0$ аналогичная задача y_1 -устойчивости анализировалась [35] для системы (16) при отсутствии эффекта запаздывания посредством функции Ляпунова, имеющей вид (18) при $\beta_1 = \beta_2 = 0$, $M = 2$. Сравнение показывает, что полученная область y_1 -устойчивости совпадает с областью (20) при $a^* = 0$, $M = 2$.

Для численной характеристики результатов в таблице 1 приводятся результаты вычислений по рекуррентным соотношениям (16) на отрезке $k \in [0, 30]$ при начальных данных $y_i(-1) = y_i(0) = 0,1$ ($i = 1, 2$) и $z_1(-1) = z_1(0) = 1$, и при значениях параметров $a = 1/2$, $a^* = 0$, $b = 3/2$, $c = 1/3$, $d = e = l = 1$.

Таблица 1. Результаты вычислений

k	$y_1(k)$	$y_2(k)$	$z_1(k)$
-1	0,1	0,1	1
0	0,1	0,1	1
1	0,15	0,16	0,4333
2	0,1750	0,2560	0,1877
3	0,1568	0,4224	0,2094
4	0,1264	0,4288	0,0954
5	0,1517	0,7104	0,0468
6	0,1168	1,1554	0,0215
7	0,0916	1,9084	0,0104
8	0,0706	3,0854	0,0060
9	0,0551	4,9107	0,0025
10	0,0461	7,7127	0,0004
...	...	...	...
15	0,0050	65,557	$5,5 \times 10^{-6}$
...	...	...	...
20	0,00039	504,13	$2,4 \times 10^{-8}$
...	...	...	...
30	$1,9 \times 10^{-5}$	3832,1	$4,7 \times 10^{-10}$

Пример 2. Рассмотрим систему дискретных уравнений:

$$\begin{aligned}
y_1(k+1) &= [a + ly_2(k-1)z_1(k-1)]y_1(k), \\
y_2(k+1) &= [b + dy_1(k-1)]y_2(k), \\
z_1(k+1) &= [c + ey_1(k-1)]z_1(k), \quad z_2(k+1) = Z_2(k, \mathbf{x}_k),
\end{aligned} \tag{23}$$

которые представляют структурно измененный вариант системы (16).

Для анализа асимптотической y_1 -устойчивости «частичного» положения равновесия (17) системы (23) используем V -функционал (18), в котором $\beta_1 = 0$, и две вспомогательные функции (19).

Квадратичная часть $(\Delta V)_2$ разности (приращения) ΔV выбранного функционала в силу системы (23) имеет вид:

$$(\Delta V)_2 = (a^2 - 1)\psi_{y_1}^2(0) + (Mb^2c^2 - M + \beta_2)\mu_1^2(0) - \beta_2\mu_1^2(-1),$$

а члены более высокого порядка в выражении ΔV являются формами переменных $\psi_{y_1}(0)$, $\mu_1(0)$, $\mu_1(-1)$.

Пусть выполняются условия $a^2 < 1$, $M(b^2c^2 - 1) + \beta_2 < 0$. Тогда при достаточно малом $h_1 > 0$ в области (8) при любых значениях параметров l, d, e для разности (приращения) ΔV функционала (18), в котором $\beta_1 = 0$, $\beta_2 = M(1 - b^2c^2) - \varepsilon_2$ (где $\varepsilon_2 = \text{const} > 0$ – достаточно малое число) в силу системы (23) имеет место оценка вида (21).

В результате, при выполнении условий $a^2 < 1$, $b^2c^2 < 1$ «частичное» положение равновесия (17) системы (23) равномерно асимптотически y_1 -устойчиво при больших значениях z_{10} в целом по z_{20} на основании теоремы 2. Разность (приращение) ΔV рассматриваемого функционала (18) в силу системы (23) *знакопеременна* в области (7).

При отсутствии эффекта запаздывания в системе (23) аналогичная задача y_1 -устойчивости анализировалась [11] посредством функции Ляпунова, имеющей вид (18) при $\beta_1 = \beta_2 = 0$, $M = 2$. Полученная область y_1 -устойчивости совпадает с областью $a^2 < 1$, $b^2c^2 < 1$.

6. Заключение. Для системы нелинейных дискретных (конечно-разностных) уравнений общего вида с ограниченным запаздыванием дана постановка задачи частичной устойчивости: устойчивости «частичного» (нулевого) положения равновесия по отношению к части определяющих его переменных.

Найдены достаточные условия частичной устойчивости (асимптотической устойчивости) указанного вида, основанные на выборе подходящего V -функционала, являющегося дискретным аналогом функционала Ляпунова – Красовского, в сочетании с двумя

дополнительными (векторными, вообще говоря) функциями. Эти функции вводятся для корректировки области пространства дискретных функций, в которой строится V -функционал, а также для нахождения требуемых оценок самого функционала и его разности (приращения) в силу изучаемой системы. В результате как сам V -функционал, так и его разность (приращение) могут быть знакопеременными в области пространства дискретных функций, которая обычно рассматривается при анализе частичной устойчивости.

Результаты статьи являются развитием идей и результатов [6, 11, 34 – 36, 38] по решению задач частичной устойчивости применительно к рассматриваемому в статье классу дискретных систем с запаздыванием. Найденные условия частичной устойчивости могут использоваться при анализе как собственно систем дискретных уравнений высокого порядка, так и более общих систем дискретных уравнений с переменным запаздыванием.

Литература

1. Халанай А., Векслер Д. Качественная теория импульсных систем. М.: Мир, 1971. 309 с.
2. Фурасов В.Д. Устойчивость и стабилизация дискретных процессов. М.: Наука, 1982. 192 с.
3. Elaydi S. An Introduction to Difference Equations, 3-ed. N.Y.: Springer, 2005. 540 p. DOI: 10.1007/0-387-27602-5.
4. Александров А.Ю., Жабко А.П., Платонов А.В. Устойчивость движений дискретных динамических систем. СПб.: Изд. Дом Федоровой Г.В., 2015. 154 с.
5. Румянцев В.В. Об устойчивости движения по отношению к части переменных // Вестн. МГУ. Сер. Матем., Механика, Физика, Астрономия, Химия. 1957. № 4. С. 9–16.
6. Воротников В.И. Частичная устойчивость и управление: состояние проблемы и перспективы развития // Автоматика и телемеханика. 2005. № 4. С. 3–59.
7. Haddad W.M., Chellaboina V. Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach. Princeton: Princeton Univ. Press, 2008. 976 p.
8. Ramirez-Llanos E., Martínez S. Distributed Discrete-Time Optimization Algorithms with Applications to Resource Allocation in Epidemics Control // Optimal Control Appl. Meth. 2018. vol. 39. no. 1. pp. 160–180.
9. Shafiei M.H., Vazirpour N. The Approach of Partial Stabilization in Design of Discrete-Time Robust Guidance Laws against Maneuvering Targets // Aeronautical J. 2020. vol. 124. no. 1277. pp. 1114–1127.
10. Игнатьев А.О. Метод функций Ляпунова в системах разностных уравнений: устойчивость относительно части переменных // Дифференц. уравнения. 2022. Т. 58. № 3. С. 407–415.
11. Воротников В.И., Мартышенко Ю.Г. Об одном подходе к анализу устойчивости «частичных» положений равновесия нелинейных дискретных систем // Известия РАН. Теория и системы управления. 2022. Т. 63. № 3. С. 57–68.
12. Shaikhet L. Lyapunov Functionals and Stability of Stochastic Difference Equations. N.Y.: Springer, 2011. 370 p.
13. Astrom K.J., Wittenmark B. Computer Controlled Systems: Theory and Design. N.Y.: Dover Publ, Inc., 2013. 576 p.

14. Fridman E. Introduction to Time-Delay Systems: Analysis and Control. Boston: Birkhauser, 2014. 362 p. DOI: 10.1007/978-3-319-09393-2.
15. Hetel L., Daafouz J., Jung C. Equivalence between the Lyapunov–Krasovskii Functionals Approach for Discrete Delay Systems and that of the Stability Conditions for Switched Systems // *Nonlinear Analysis: Hybrid Systems*. 2008. vol. 2. no. 3. pp. 697–705.
16. Родионов А.М. Некоторые модификации теорем второго метода Ляпунова для дискретных уравнений // *Автоматика и телемеханика*. 1992. № 9. С. 86–93.
17. Elaydi S., Zhang S. Stability and Periodicity of Difference Equations with Finite Delay // *Funkcialaj Ekvacioj*. 1994. vol. 37. no. 3. pp. 401–413.
18. Анашкин О.В. Функции Ляпунова в теории устойчивости нелинейных разностных уравнений с запаздыванием // *Дифференц. уравнения*. 2002. Т. 38. № 7. С. 976–978.
19. Pepe P., Pola G., Di Benedetto M.D. On Lyapunov–Krasovskii Characterizations of Stability Notions for Discrete-Time Systems with Uncertain Time-Varying Time Delays // *IEEE Trans. Automatic Control*. 2017. vol. 63. no. 6. pp. 1603–1617.
20. Aleksandrov A.Y., Aleksandrova E.B. Delay-Independent Stability Conditions for a Class of Nonlinear Difference Systems // *J. of the Franklin Institute*. 2018. vol. 355. no. 7. pp. 3367–3380.
21. Zhou B. Improved Razumikhin and Krasovskii Approaches for Discrete-Time Time-Varying Time-Delay Systems // *Automatica*. 2018. vol. 91. pp. 256–269.
22. Li X., Wang R., Du S., Li T. An Improved Exponential Stability Analysis Method for Discrete-Time Systems with a Time-Varying Delay // *Intern. J. Robust Nonlin. Control*. 2022. vol. 32. no. 2. pp. 669–681.
23. Guo Y., Xu X., Liu L., Wang Y., Feng G. New Results on Stability of Discrete-Time Systems with Infinite Delays // *Automatica*. 2022. vol. 136. no. 110043.
24. Zhang K., Braverman E., Gharesifard B. Event-Triggered Control for Discrete-Time Delay Systems // *Automatica*. 2023. vol. 147. no. 110688.
25. Seuret A., Gouaisbaut F., Fridman E. Stability of Discrete-Time Systems with Time-Varying Delays via a Novel Summation Inequality // *IEEE Trans. Automatic Control*. 2015. vol. 60. no. 10. pp. 2740–2745.
26. Lin H., Zeng H., Wang W. New Lyapunov-Krasovskii Functional for Stability Analysis of Linear Systems with Time-Varying Delay // *J. Systems Science and Complexity*. 2021. vol. 34. no. 2. pp. 632–641.
27. Zhu L., Zhu C. Enhanced Stability Criteria for Discrete-Time Systems with Time-Varying Delay // *Intern. J. Control, Autom. Systems*. 2021. vol. 19. no. 7. pp. 2385–2394.
28. Demidenko G.V., Matveeva I.I. The Second Lyapunov Method for Time-Delay Systems // *International workshop on Functional Differential Equations and Applications: FDEA-2019*. Singapore: Springer, 2022. pp. 145–167.
29. Zhang X.M., Han Q.L., Ge X., Peng C. Stability Analysis of Delayed Discrete-Time Systems based on a Delay-Square-Dependent Lyapunov Functional // *Automatica*. 2023. vol. 147. no. 110592.
30. Diblík J. Exponential Stability of Linear Discrete Systems with Multiple Delays by Degenerated Lyapunov–Krasovskii Functionals // *Applied Mathematics Letters*. 2023. vol. 142. no. 110592.
31. Liz E. Stability of Non-Autonomous Difference Equations: Simple Ideas Leading to Useful Results // *J. Difference Equat.* 2011. vol. 17. no. 2. pp. 203–220.
32. Мальгина В.В. Асимптотические свойства решений линейных дифференциальных и разностных уравнений с последействием. Дисс. на соиск. уч. степени докт. физ.-мат. наук. Пермь: Пермский национальный исследовательский политехнический университет, 2021. 279 с.

33. Дашковский С.Н., Ефимов Д.В., Сонгаг Э.Д. Устойчивость от входа к состоянию и смежные свойства систем // Автоматика и телемеханика. 2011. № 8. С. 3–40.
34. Воротников В.И., Мартышенко Ю.Г. К задаче частичной устойчивости нелинейных дискретных систем // Мехатроника. Автоматизация. Управление. 2017. Т. 18. № 6. С. 371–375.
35. Воротников В.И., Мартышенко Ю.Г. К задаче частичной устойчивости нелинейных дискретных стохастических систем // Автоматика и телемеханика. 2021. № 9. С. 116–132.
36. Воротников В.И., Мартышенко Ю.Г. Об устойчивости по части переменных «частичных» положений равновесия систем с последствием // Матем. заметки. 2014. Т. 96. № 4. С. 496–503.
37. Воротников В.И. К частичной устойчивости и детектируемости функционально-дифференциальных систем с последствием // Автоматика и телемеханика. 2020. № 2. С. 3–17.
38. Vorotnikov V.I. Partial Stability and Control. Boston: Birkhauser, 1998. 448 p.

Воротников Владимир Ильич — д-р физ.-мат. наук, профессор кафедры, кафедра математики и информационных технологий, Сочинский институт Российского университета дружбы народов. Область научных интересов: устойчивость динамических систем, частичная устойчивость и стабилизация, теория управления, динамика управляемого твердого тела (космического аппарата). Число научных публикаций — 210. vorotnikov-vi@rambler.ru; улица Куйбышева, 32, 354340, Сочи, Россия; р.т.: +7(862)241-1270.

V. VOROTNIKOV
**ON THE PARTIAL STABILITY OF NONLINEAR DISCRETE-
TIME SYSTEMS WITH DELAY**

Vorotnikov V. On the Partial Stability of Nonlinear Discrete-Time Systems with Delay.

Abstract. A system of nonlinear discrete (finite-difference) of a general form with a bounded delay is considered. Interest in the tasks of qualitative analysis of such systems has increased significantly in recent years. At the same time, the problem of stability with respect to all variables of the zero equilibrium position, which has a great generality, is mainly analyzed in domestic and foreign literature. The main research method is a discrete-functional analogue of the direct Lyapunov method. In this article, it is assumed that the system under consideration admits a “partial” (in some part of the state variables) zero equilibrium position. The problem of stability of a given equilibrium position is posed, and stability is considered not in all, but only in relation to a part of the variables that determine this equilibrium position. Such a problem belongs to the class of problems of partial stability, which are actively studied for systems of various forms of mathematical description. The proposed statement of the problem complements the scope of the indicated studies in relation to the system under consideration. To solve this problem, a discrete version of the Lyapunov–Krasovskii functionals method is used in the space of discrete functions with appropriate specification of the functional requirements. To expand the capabilities of this method, it is proposed to use two types of additional auxiliary (vector, generally speaking) discrete functions in order to: 1) adjustments of the phase space region of the system in which the Lyapunov–Krasovskii functional is constructed; 2) finding the necessary estimates of the functionals and their differences (increment) due to the system under consideration, on the basis of which conclusions about partial stability are made. The expediency of this approach lies in the fact that as a result, the Lyapunov–Krasovskii functional, as well as its difference due to the system under consideration, can be alternating in the domain that is usually considered when analyzing partial stability. Sufficient conditions of partial stability, partial uniform stability, and partial uniform asymptotic stability of the specified type are obtained. The features of the proposed approach are shown on the example of two classes of nonlinear systems of a given structure, for which partial stability is analyzed in parameter space. Attention is drawn to the expediency of using a one-parameter family of functionals.

Keywords: nonlinear delay discrete-time system, partial stability, Lyapunov–Krasovskii functional, one-parameter family of functionals.

References

1. Halanay A., Wexler D. Qualitative Theory of Impulsive Systems. Bucharest: Ed. Acad. RPR, 1968. 312 p.
2. Furasov V.D. Ustojchivost' i stabilizaciya diskretnyh processov [Stability and stabilization of discrete processes]. Moscow: Nauka, 1982. 192 p. (In Russ.).
3. Elaydi S. An Introduction to Difference Equations, 3-ed. N.Y.: Springer, 2005. 540 p. DOI: 10.1007/0-387-27602-5.
4. Aleksandrov A.Yu., Zhabko A.P., Platonov A.V. Ustojchivost' dvizhenij diskretnyh dinamicheskikh sistem [Stability of motions of discrete dynamic systems]. SPb.: Izd. Dom Fedorovoj G.V., 2015. 154 p. (In Russ.).
5. Rumyantsev V.V. [On Stability of Motion with Respect to a Part of the Variables]. Vestn. MGU. Ser. Matematiki, mekhaniki, fiziki, astronomii, himii – Gerald of

- Moscow State University. Ser. Math., Mech., Phys., Astron., Chem. 1957. no. 4. pp. 9–16. (In Russ.).
6. Vorotnikov V.I. Partial Stability and Control: the State of the Art and Developing Prospects. *Autom. Remote Control*. 2005. vol. 66. no. 4. pp. 511–561.
7. Haddad W.M., Chellaboina V. *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. Princeton: Princeton Univ. Press, 2008. 976 p.
8. Ramírez-Llanos E., Martínez S. Distributed Discrete-Time Optimization Algorithms with Applications to Resource Allocation in Epidemics Control. *Optimal Control Appl. Meth.* 2018. vol. 39. no. 1. pp. 160–180.
9. Shafiei M.H., Vazirpour N. The Approach of Partial Stabilization in Design of Discrete-Time Robust Guidance Laws against Maneuvering Targets. *Aeronautical J.* 2020. vol. 124. no. 1277. pp. 1114–1127.
10. Ignatyev A.O. Lyapunov Function Method for Systems of Difference Equations: Stability with Respect to Part of the Variables. *Differential Equations*. 2022. vol. 58. no. pp. 405–414.
11. Vorotnikov V.I., Martyshenko Y.G. Approach to the Stability Analysis of Partial Equilibrium States of Nonlinear Discrete Systems. *Journal of Computer and Systems Sciences International. Theory and control systems*. 2022. vol. 61. no. 3. pp. 348–359.
12. Shaikhet L. *Lyapunov Functionals and Stability of Stochastic Difference Equations*. N.Y.: Springer, 2011. 370 p.
13. Astrom K.J., Wittenmark B. *Computer Controlled Systems: Theory and Design*. N.Y.: Dover Publ, Inc., 2013. 576 p.
14. Fridman E. *Introduction to Time-Delay Systems: Analysis and Control*. Boston: Birkhauser, 2014. 362 p. DOI: 10.1007/978-3-319-09393-2.
15. Hetel L., Daafouz J., Jung C. Equivalence between the Lyapunov–Krasovskii Functionals Approach for Discrete Delay Systems and that of the Stability Conditions for Switched Systems. *Nonlinear Analysis: Hybrid Systems*. 2008. vol. 2. no. 3. pp. 697–705.
16. Rodionov A.M. Certain Modifications of Theorems of the Second Lyapunov Method for Discrete Equations. *Autom. Remote Control*. 1992. vol. 53. no. 9. pp. 1381–1386.
17. Elaydi S., Zhang S. Stability and Periodicity of Difference Equations with Finite Delay. *Funkcialaj Ekvacioj*. 1994. vol. 37. no. 3. pp. 401–413.
18. Anashkin O.V. Lyapunov Functions in Stability Theory of Nonlinear Difference Delay Equations. *Differential Equations*. 2002. vol. 38. no. 7. pp. 1038–1041.
19. Pepe P., Pola G., Di Benedetto M.D. On Lyapunov–Krasovskii Characterizations of Stability Notions for Discrete-Time Systems with Uncertain Time-Varying Time Delays. *IEEE Trans. Automatic Control*. 2017. vol. 63. no. 6. pp. 1603–1617.
20. Aleksandrov A.Y., Aleksandrova E.B. Delay-Independent Stability Conditions for a Class of Nonlinear Difference Systems. *J. of the Franklin Institute*. 2018. vol. 355. no. 7. pp. 3367–3380.
21. Zhou B. Improved Razumikhin and Krasovskii Approaches for Discrete-Time Time-Varying Time-Delay Systems. *Automatica*. 2018. vol. 91. pp. 256–269.
22. Li X., Wang R., Du S., Li T. An Improved Exponential Stability Analysis Method for Discrete-Time Systems with a Time-Varying Delay. *Intern. J. Robust Nonlin. Control*. 2022. vol. 32. no. 2. pp. 669–681.
23. Guo Y., Xu X., Liu L., Wang Y., Feng G. New Results on Stability of Discrete-Time Systems with Infinite Delays. *Automatica*. 2022. vol. 136. no. 110043.
24. Zhang K., Braverman E., Gharesifard B. Event-Triggered Control for Discrete-Time Delay Systems. *Automatica*. 2023. vol. 147. no. 110688.
25. Seuret A., Gouaisbaut F., Fridman E. Stability of Discrete-Time Systems with Time-Varying Delays via a Novel Summation Inequality. *IEEE Trans. Automatic Control*. 2015. vol. 60. no. 10. pp. 2740–2745.

26. Lin H., Zeng H., Wang W. New Lyapunov-Krasovskii Functional for Stability Analysis of Linear Systems with Time-Varying Delay. *J. Systems Science and Complexity*. 2021. vol. 34. no. 2. pp. 632–641.
27. Zhu L., Zhu C. Enhanced Stability Criteria for Discrete-Time Systems with Time-Varying Delay. *Intern. J. Control, Autom. Systems*. 2021. vol. 19. no. 7. pp. 2385–2394.
28. Demidenko G.V., Matveeva I.I. The Second Lyapunov Method for Time-Delay Systems. *International workshop on Functional Differential Equations and Applications: FDEA-2019*. Singapore: Springer, 2022. pp. 145–167.
29. Zhang X.M., Han Q.L., Ge X., Peng C. Stability Analysis of Delayed Discrete-Time Systems based on a Delay-Square-Dependent Lyapunov Functional. *Automatica*. 2023. vol. 147. no. 110592.
30. Diblík J. Exponential Stability of Linear Discrete Systems with Multiple Delays by Degenerated Lyapunov–Krasovskii Functionals. *Applied Mathematics Letters*. 2023. vol. 142. no. 110592.
31. Liz E. Stability of Non-Autonomous Difference Equations: Simple Ideas Leading to Useful Results. *J. Difference Equat.* 2011. vol. 17. no. 2. pp. 203–220.
32. Malygina V.V. Asimptoticheskie svoistva reshenii lineinykh differentsialnykh i raznostnykh uravnenii s posledствием [Asymptotic Properties of Solutions of Linear Differential and Difference Equations with Aftereffect]. *Dr. Sci. Thesis*. Perm: Perm Univ., 2021. 279 p. (In Russ.).
33. Dashkovskiy S.N., Efimov D.V., Sontag E.D. Input to State Stability and Allied System Properties. *Autom. Remote Control*. 2011. vol. 72. no. 8. pp. 1579–1614.
34. Vorotnikov V.I., Martyshenko Yu.G. [To Problem of Partial Stability for Nonlinear Discrete-Time Systems]. *Mekhatronika. Avtomatizaciya. Upravlenie – Mechatronics. Automation and Control*. 2017. vol. 18. no. 6. pp. 371–375. (In Russ.).
35. Vorotnikov V.I., Martyshenko Yu.G. On the Partial Stability Problem for Nonlinear Discrete-Time Stochastic Systems. *Autom. Remote Control*. 2021. vol. 82. no. 9. pp. 1554–1567.
36. Vorotnikov V.I., Martyshenko Yu.G. Stability in a Part of Variables of “Partial” Equilibria of Systems with Aftereffect. *Math. Notes*. 2014. vol. 96. no. 4. pp. 477–483.
37. Vorotnikov V.I. On Partial Stability and Detectability of Functional Differential Systems with Aftereffect. *Autom. Remote Control*. 2020. vol. 81. no. 2. pp. 199–211.
38. Vorotnikov V.I. *Partial Stability and Control*. Boston: Birkhauser, 1998. 448 p.

Vorotnikov Vladimir — Ph.D., Dr.Sci., Professor of the department, Mathematics and information technology department, Sochi Institute of the Peoples’ Friendship University of Russia. Research interests: stability of dynamical systems, partial stability and stabilization, control theory, dynamics of controlled solid (spacecraft). The number of publications — 210. vorotnikov-vi@rambler.ru; 32, Kuibysheva St., 354340, Sochi, Russia; office phone: +7(862)241-1270.

А.С. СТЕПАНОВ, Е.А. ФОМИНА, Л.В. ИЛЛАРИОНОВА, К.Н. ДУБРОВИН,
Д.В. ФЕДОСЕЕВ

АППРОКСИМАЦИЯ ВРЕМЕННЫХ РЯДОВ ИНДЕКСОВ ВЕГЕТАЦИИ (NDVI И EVI) ДЛЯ МОНИТОРИНГА СЕЛЬХОЗКУЛЬТУР (ПОСЕВОВ) ХАБАРОВСКОГО КРАЯ

Степанов А.С., Фомина Е.А., Илларионова Л.В., Дубровин К.Н., Федосеев Д.В.
Аппроксимация временных рядов индексов вегетации (NDVI и EVI) для мониторинга сельхозкультур (посевов) Хабаровского края.

Аннотация. Аппроксимация рядов сезонного хода индексов вегетации является основой для эффективного мониторинга сельскохозяйственных культур, их идентификации и автоматизированной классификации пахотных земель. Для пахотных земель Хабаровского края в период с мая по октябрь 2021 года по мультиспектральным снимкам Sentinel-2A (20 м) с использованием маски облачности были построены временные ряды NDVI и EVI. Для приближения временных рядов были использованы пять видов аппроксимирующих функций: функция Гаусса; двойная гауссиана; двойная синусоида; ряд Фурье; двойная логистическая. Были построены и рассчитаны характеристики экстремумов аппроксимированных временных рядов для разных типов пахотных земель: гречихи, многолетних трав, сои, залежи и пара. Было показано, что для каждой сельхозкультуры аппроксимированные кривые сезонного хода имели характерный вид. Как было достоверно установлено ($p < 0,05$), наиболее высокую точность аппроксимации рядов NDVI и EVI показал ряд Фурье (средняя ошибка составила, соответственно, 8,5% и 16,0%). Аппроксимация рядов NDVI с использованием двойной синусоиды, двойной гауссианы и двойной логистической функции приводила к увеличению ошибки до 8,9-10,6%. Аппроксимация рядов EVI на основе двойной гауссианы и двойной синусоиды способствовала росту средней ошибки до 18,3-18,5%. Проведенный апостериорный анализ с использованием критерия Тьюки показал, что для полей с соей, парующих и залежных земель для приближения индексов вегетации достоверно лучше использовать ряд Фурье, двойную гауссиану или двойную синусоиду, для полей с гречихой целесообразно применять ряд Фурье или двойную гауссиану. В целом, средняя ошибка аппроксимации сезонных временных рядов NDVI в 1,5-4 раза меньше, чем ошибка аппроксимации рядов EVI.

Ключевые слова: индекс вегетации, Хабаровский край, аппроксимация, пахотные земли, сельскохозяйственная культура, временной ряд.

1. Введение. Эффективный мониторинг сельскохозяйственных культур с использованием данных дистанционного зондирования Земли является основой успешного перехода к точному и цифровому земледелию. Одно из важнейших направлений мониторинга связано с разработкой подходов по применению вегетационных индексов (VI) для автоматизированной классификации пахотных земель, выявления залежных земель, оценки состояния посевов, прогнозирования урожайности [1 – 4].

К наиболее популярным индексам вегетации, расчет которых базируется на анализе мультиспектральных изображений, относятся

Normalized Difference Vegetation Index (NDVI) и Enhanced Vegetation Index (EVI) [5 – 8].

Отмечено, что для решения задач автоматизированной классификации пахотных земель важнейшую роль имеет исследование графиков временных рядов VI разных культур, предварительное выявление их сходства между собой, а также выбор параметра – типа VI, который продемонстрирует самую высокую точность классификации [9 – 10]. Так, например, было показано, что для верной идентификации полей с рисом в провинции Хубей (Китай) целесообразно применять не NDVI и EVI, а специальную комбинацию красного, ближнего инфракрасного, и коротковолнового инфракрасного каналов [9]. В другой работе, посвященной картографированию пахотных земель северных провинций Китая, было установлено, что NDVI превзошел EVI в точности распознавания хлопка, пшеницы, кукурузы, поскольку NDVI обладал лучшей разделяемостью для различения культур на этапах созревания [10]. На примере сельскохозяйственных земель индийского штата Одиша были исследованы временные ряды NDVI и EVI для зернобобовых, хлопка, риса, а также парующих земель [11]. Были определены характерные параметры каждой кривой – значения максимумов, продолжительность фенологических циклов, а также рассчитана степень сходства между временными рядами разных видов сельхозкультур, а также вариации внутри одного класса. Моделирование фенологических циклов культур с помощью временных рядов VI также имеет большое значение при прогнозировании урожайности. В частности, рассматривались показатели, основанные на значениях NDVI и EVI для разных фаз роста сои в долине Миссисипи [12]. Модели прогнозирования урожайности сои, основанные как на NDVI, так и на EVI, имели точность около 95%. Утверждается, что разработанные подходы к моделированию урожайности полезны для мониторинга различных сельскохозяйственных регионов мира и могут быть применены к различным культурам для составления карт урожайности или моделирования жизненного цикла культур. В другом исследовании, при изучении полей кукурузы и сои в штате Небраска (США), была построена система, описывающая связь между фенологическими показателями, полученными с помощью дистанционного зондирования – на основе индекса NDVI, и ключевыми стадиями роста сельскохозяйственных культур [13]. В целом, это исследование продемонстрировало, что разработанная система обладает большим потенциалом для оценки сроков различных стадий роста кукурузы и

сои с приемлемой точностью (ошибка модели не превышала 8%). Прогнозирование дат наступления разных стадий роста сельскохозяйственных культур с опережением по времени весьма полезно для управления ростом культур и мониторинга в течение сезона [14]. Для создания общеевропейской базы фенологических данных по разным сельскохозяйственным культурам также тестировалась возможность использования NDVI и EVI как временных показателей, характеризующих наступление и продолжительность фенологических фаз. Было установлено, что EVI имел более тесную корреляционную связь с фенологическими циклами и меньшую ошибку, в сравнении с NDVI для большинства культур.

В качестве важнейшей задачи, связанной с обработкой значений VI, полученных в ходе одного сезона, часто рассматривается получение гладкой формы временного ряда индексов и подавление шума [15 – 16]. Это дает возможность рассчитывать параметры временного ряда и использовать их на практике (например, определить точку наступления и завершения, длительность вегетационного периода и прочие фенологические характеристики) [15, 17]. Также при отсутствии некоторых значений временного ряда применяется аппроксимация при помощи разных функций [16]. Применение аппроксимирующих функций для обработки временных рядов может быть использовано в дальнейшем для построения композитов и аналитического описания изменения VI в ходе одного сезона.

Наиболее простым способом является сглаживание скользящим средним; достаточно часто применяется фильтр Савицкого-Голлея, являющийся по сути развитием метода скользящего среднего с использованием полиномов разных степеней; также используется вейвлет-преобразование и фильтр Уиттакера [18]. При этом применение алгоритма скользящего окна с заполнением пропусков одновременно решает задачи интерполяции и фильтрации [19]. Другим подходом к сглаживанию является использование аппроксимирующих функций (асимметричная гауссова, двойная логистическая (DL) и полиномиальные функции) [20 – 23]. Например, сравнение точности классификации сельхозкультур при предварительном сглаживании временных рядов NDVI дискретным преобразованием Фурье, асимметричной функцией Гаусса, двойной логистической функцией (DL), а также фильтрами Савицкого-Голлея и Уиттакера показало, что наименьшая ошибка классификации была достигнута с помощью преобразования Фурье и фильтра Уиттакера (точность классификации выросла на 2-6%) [18]. Вместе с тем, использование асимметричной

функции Гаусса и DL напротив, способствовало даже некоторому снижению точности классификации пахотных земель. В схожем исследовании, выполненном в Индии, проводилась оценка применимости четырех методов (преобразования Фурье, асимметричной гауссовой модели, DL-модели и фильтра Уиттакера) для моделирования сезонных колебаний VI [20]. Асимметричная функция Гаусса и DL показали хорошие результаты только для кривых с одним выраженным экстремумом (в том числе для регионов, где в год получают только один урожай). С другой стороны, некоторыми авторами отмечалось преимущество использования логистических функций для моделирования EVI и NDVI не только сельскохозяйственных культур, но и естественных лугов, а также залежных земель [21, 22]. Так, например, для моделирования временных рядов NDVI, полученных для полей кукурузы и сои в США применялись две логистические кривые, одна для начального периода, а другая последующего периода роста [22]. Для залежей предлагался итеративный метод логистической подгонки для моделирования EVI. Было установлено, что именно устойчивость этого метода к изменению уровня шума обеспечивала ему преимущество [21]. Для временных рядов значений NDVI, рассчитанных по данным MODIS для основных сельхозкультур Самарской области, проводилась оценка точности аппроксимации при помощи кусочно-линейных функций, асимметричных функций Гаусса и DL, рядов Фурье, многочленов и кубического сплайна [23]. Было показано, что наиболее высокая точность аппроксимации достигалась при использовании кубического сплайна. В другом исследовании, для подавления шума в дополнение к асимметричным функциям Гаусса и DL использовали фильтр Савицкого-Голлея, а также три модифицированных фильтра для аппроксимации временных рядов NDVI в канадской провинции Альберта (при этом учитывался тип поверхности – горы, леса, пашни, луга и т.д.) [24]. В среднем по региону использование асимметричных функций Гаусса и DL наилучшим образом снижало уровень шума при сохранении соответствующей целостности сигнала NDVI. Однако в некоторых особых случаях (например, в горных регионах) альтернативные фильтры обеспечивали более высокую точность аппроксимации. Вместе с тем, другими авторами было отмечено, что применение асимметричной функции Гаусса и DL, а также гармонического анализа и некоторых прочих методов восстановления данных временного ряда NDVI гарантировало хорошую точность для малооблачных территорий [25]. В то же время моделирование рядов NDVI с использованием разных аппроксимирующих функций в

районах, подверженных облачности, в том числе южной части российского Дальнего Востока, требует дополнительных исследований. В частности, ранее были проведены исследования, связанные с оценкой возможности использования функции Гаусса, DL и полиномов второй и третьей степени для решения задачи раннего прогнозирования урожайности сои для Амурской области, Хабаровского и Приморского краев, а также Еврейской Автономной области (ЕАО) [26, 27]. Было установлено, что точность прогнозирования непосредственно связана с ошибкой аппроксимации временных рядов, а полиномиальные функции продемонстрировали недостаточную точность. Вместе с тем, в связи с особенностями севооборота, такими как посев овса с подсевом многолетних трав, поздним севом гречихи, обеспечивающими более одного максимума на кривых сезонного хода VI, целесообразно рассмотреть аппроксимацию временных рядов в модельном регионе Дальнего Востока на основе периодических функций (ряд Фурье, двойная синусоида), а также двойной логистической функции и двойной функции Гаусса.

Таким образом, основной целью исследования явилась оценка возможности аппроксимации временных рядов индексов вегетации пахотных земель Хабаровского края с использованием нелинейных функций. Для достижения этой цели в рамках настоящей работы решались следующие задачи: построение временных рядов NDVI и EVI для разных сельскохозяйственных культур Хабаровского края; описание основных характеристик сезонного хода VI для каждой культуры; расчет аппроксимированных значений временных рядов NDVI и EVI с помощью нелинейных функций; оценка ошибки аппроксимации для разных сельскохозяйственных культур и функций, а также определение статистических различий способов аппроксимации; выбор оптимальной модели аппроксимации для пахотных земель Хабаровского края; оценка применимости модели в других сельскохозяйственных регионах южной части Дальнего Востока.

2. Материалы и методы. В качестве области исследования рассматривались пахотные земли Хабаровского района (2021 г): более 3500 га (155 отдельных полей), в том числе посевы сои (1260 га), гречихи (238 га), многолетних трав (114 га), а также парующие (496 га) и залежные земли (1613 га). В период с 1 мая по 31 октября 2021 г для исследуемой области были получены 37 снимков (Sentinel-2A, размер пикселя 20 м). Выбранный временной интервал является достаточным для исследования изменений в фенологических циклах

сельскохозяйственных культур Хабаровского края – первые всходы появляются не раньше первой декады мая, уборка обычно заканчивается не позже середины октября. Для каждого пикселя формировались временные ряды NDVI и EVI, где значения VI определялись по формулам [28]:

$$NDVI = \frac{NIR-RED}{NIR+RED}, \quad (1)$$

$$EVI = 2,5 \frac{NIR-RED}{NIR+6RED-7,5BLUE+1}, \quad (2)$$

где NIR – отражение в ближней инфракрасной области спектра;

RED , $BLUE$ – отражения в красной и синей областях спектра.

С использованием маски облачности осуществлялась фильтрация полученных временных рядов NDVI и EVI. На следующем этапе рассчитывались средние значения VI для 155 отдельных полей.

Для аппроксимации временных рядов NDVI и EVI были использованы следующие функции: Гаусса (G); двойная гауссиана (DG); двойная синусоида (DS); ряд Фурье (DF); двойная логистическая (DL). Аппроксимация проводилась нелинейным методом наименьших квадратов, на основе алгоритма Левенберга-Марквардта [29].

Функция Гаусса представляет выражение следующего вида [30]:

$$G = a e^{\frac{-(x-b)^2}{2c^2}}, \quad (3)$$

где a , b , c – параметры.

DG представляет сумму двух гауссиан:

$$DG = a_1 e^{-\left(\frac{(x-b_1)}{c_1}\right)^2} + a_2 e^{-\left(\frac{(x-b_2)}{c_2}\right)^2}, \quad (4)$$

где a_1 , b_1 , c_1 , a_2 , b_2 , c_2 – параметры.

Значения функции DS рассчитываются по следующей формуле:

$$DS = a_1 \sin(b_1 x + c_1) + a_2 \sin(b_2 x + c_2), \quad (5)$$

где $a_1, b_1, c_1, a_2, b_2, c_2$ – параметры.

DF вычисляется по формуле:

$$DF = a_0 + a_1 \cos(xw) + b_1 \sin(xw) + a_2 \cos(2xw) + b_2 \sin(2xw), \quad (6)$$

где $a_0, a_1, w, b_1, a_2, b_2$ – параметры.

DL может быть задана следующей формулой:

$$DL = c_1 + c_2 \left(\frac{1}{1 + e^{\frac{a_1 - x}{a_2}}} - \frac{1}{1 + e^{\frac{a_3 - x}{a_4}}} \right), \quad (7)$$

где $c_1, c_2, a_1, a_2, a_3, a_4$ – параметры.

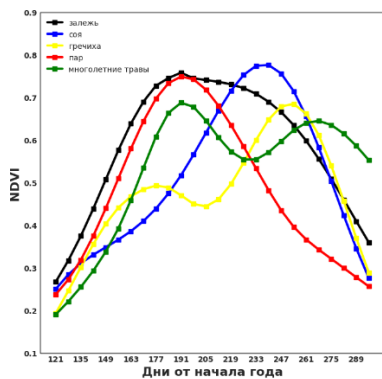
Для сравнительной оценки точности аппроксимации разными функциями рассчитывалась средняя абсолютная ошибка модели, MAPE, % (Mean absolute percentage error) по формуле:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y^i - Y_m^i|}{Y^i} 100, \quad (8)$$

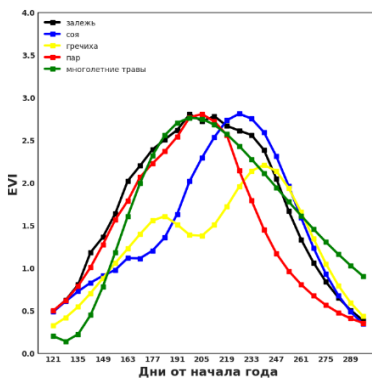
где n – количество членов временного ряда, Y^i – реальное значение VI, Y_m^i – аппроксимированное (смоделированное) значение VI.

В дальнейшем, с использованием двухфакторного дисперсионного анализа с апостериорным попарным сравнением на основе критерия Тьюки ($p=0,05$) по рассчитанным для 155 полей MAPE для NDVI и EVI, определялся оптимальный тип аппроксимирующей функции для каждого вида пахотных земель, а также тип функции, наилучшим образом моделирующий все типы сельскохозяйственных культур.

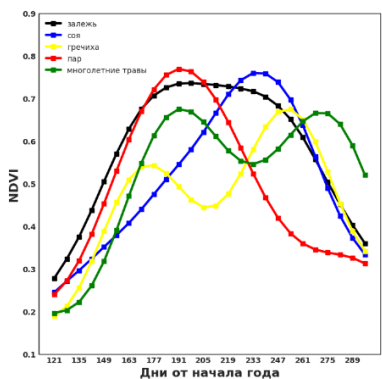
3. Результаты. Для каждого поля Хабаровского района в 2021 году были сформированы ряды средних значений NDVI и EVI. Далее для каждой культуры по усредненным временным рядам VI проводилась аппроксимация функциями DG, DS, DF, DL, G (рисунки 1).



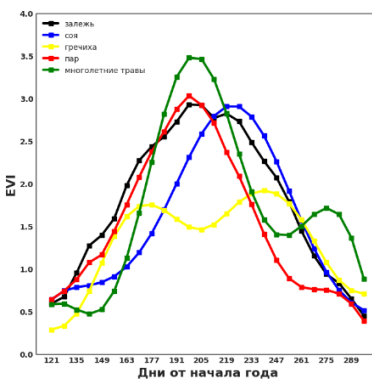
а)



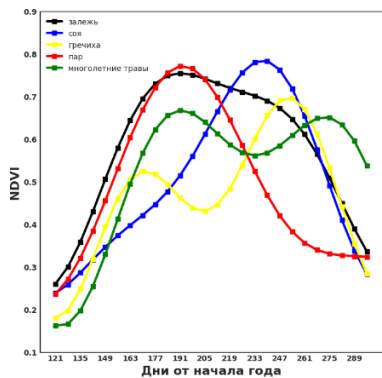
б)



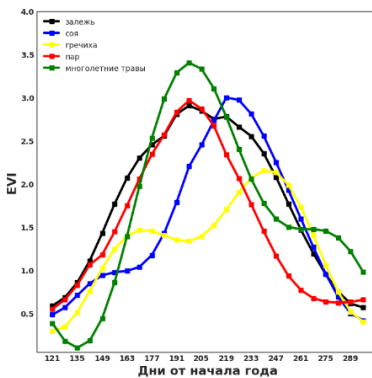
в)



г)



д)



е)

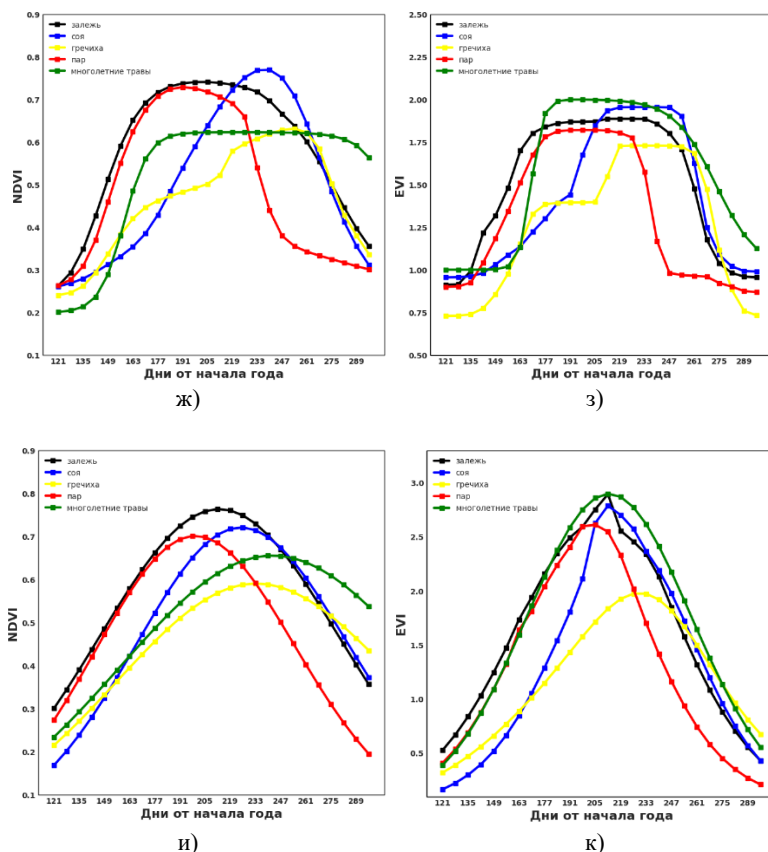


Рис. 1. Аппроксимация временных рядов VI гречихи, залежи, многолетних трав, пара, сои (2021 г, Хабаровский район) разными функциями: а), б) DG; в), г) DS; д), е) DF; ж), з) DL; и), к) G

Как видно из рисунков 1(а) и (б), для сои характерен один пик на аппроксимированной кривой NDVI и один пик на аппроксимированной кривой EVI (при использовании функции DG). День наступления максимума для сои на графике сезонного хода NDVI – 233 календарный день с начала года, EVI – 226, а численные значения $NDVI_{max}$ и EVI_{max} составили, соответственно, 0,78 и 2,81. Также наличие выраженного пика было характерно для парующих земель, при этом наступление максимума являлось достаточно ранним, в сравнение с соей: максимум NDVI приходился на 191 календарный день и достигал 0,75, а максимум EVI соответствовал 205

календарному дню и превышал 2,80. Сезонный ход EVI залежных земель, аппроксимированных DG, в целом схож с динамикой EVI растительного покрова парующих полей: максимум EVI пришелся на 198 день с начала года, но, при этом, в течение 5–7 календарных недель наблюдалось плато. Еще более выраженное плато в период со 184 по 247 календарный день соответствовало сезонному ходу NDVI. Сезонный ход NDVI для гречихи и многолетних трав (рисунок 1(а)) имел двухпиковый характер, что в первом случае отражало поздний сев гречихи (конец июня – начало июля), а во втором – укос многолетних трав, в частности, тимофеевки луговой, в конце июля.

Необходимо отметить, что график аппроксимированного ряда сезонного хода EVI для многолетних трав при помощи DG имел однопиковый вид (рисунок 1(б)), в то время как применение DS и DF (рисунок 1(г, е)) для аппроксимации приводило к двухпиковой форме, что более точно отражало особенности культивирования многолетних трав в Хабаровском крае. Вместе с тем, аппроксимация функциями DS и DF сезонного хода индексов вегетации soi показала, что вид кривых и характеристики экстремумов вполне соответствовали показателям для DG (рисунок 1(в, г, д, е)). Наступление максимума NDVI soi при аппроксимации DS и DF также пришлось на 233 календарный день (численные значения составили 0,76 и 0,78), а максимума EVI soi – на 226 день ($EVI_{\max}$, соответственно, 2,90 и 2,98). Также схожие результаты наблюдались при аппроксимации функциями DG, DS, DF индексов вегетации гречихи и парующих полей. Второй максимум NDVI гречихи приходился на 254 календарный день, значения $NDVI_{\max}$ находились в диапазоне 0,67–0,69 при использовании DG, DS, DF. Второй максимум EVI соответствовал 240 дню от начала года, при этом разброс значений $EVI_{\max}$ был более значителен – от 1,91 при аппроксимации DS до 2,20 при аппроксимации DG. Как видно из рисунка 1(в, д), для графика сезонного хода NDVI залежи, аппроксимированного функциями DS и DF, также характерно наличие плато в период со 177 по 247 календарный день. Значения NDVI на плато при аппроксимации DS находились в интервале 0,68–0,74, а при аппроксимации DF – в диапазоне 0,67–0,75. Как следует из рисунка 1(ж), аппроксимация с использованием DL для рядов сезонного хода NDVI парующих земель, многолетних трав и гречихи способствовала образованию плато на графике. Аналогично, применение DL для аппроксимации рядов сезонного хода EVI приводило к потере выраженного максимума для всех изучаемых типов пахотных земель Хабаровского района (рисунок 1(з)). В результате аппроксимации сезонного хода NDVI soi с помощью DL среднее значение максимума, пришедшегося на 233

календарный день, также составило 0,77, как и в случае DG, DS, DF. Использование для сглаживания сезонного хода NDVI и EVI функции Гаусса приводило к потере одного экстремума для гречихи и многолетних трав, а также к исчезновению плато для залежи (рисунок 1(и, к)). Аппроксимация сезонного хода NDVI сои на основе G привела к снижению значения максимума до 0,72 в сравнении с двойными функциями; аналогично, численное значение $NDVI_{max}$ для парующих полей составило 0,70, при этом максимум достигался к 205 дню с начала года.

Как видно из диаграммы размаха, ошибка аппроксимации сезонного хода NDVI выбранными функциями зависит от вида произрастающей культуры (рисунок 2). Так, при использовании DG, DS, DF и DL ошибка аппроксимации сезонного хода NDVI гречихи находилась в диапазоне от 10 до 25%, залежи и многолетних трав – от 4 до 13%, сои и пара – от 5 до 18%. MAPE при применении функции Гаусса находилась в пределах 10-30% для сезонного хода NDVI сои, 15-30% для гречихи, 12-28% для пара. Усреднение проводилось по полям с разными культурами, поэтому для многолетних трав, представленных двумя полями большого размера, на диаграмме размаха значения MAPE отображались точками. Максимальная вариативность MAPE наблюдалась при аппроксимации сезонного хода NDVI гречихи функциями DS, DF – размах средних квартилей превышал 5%.

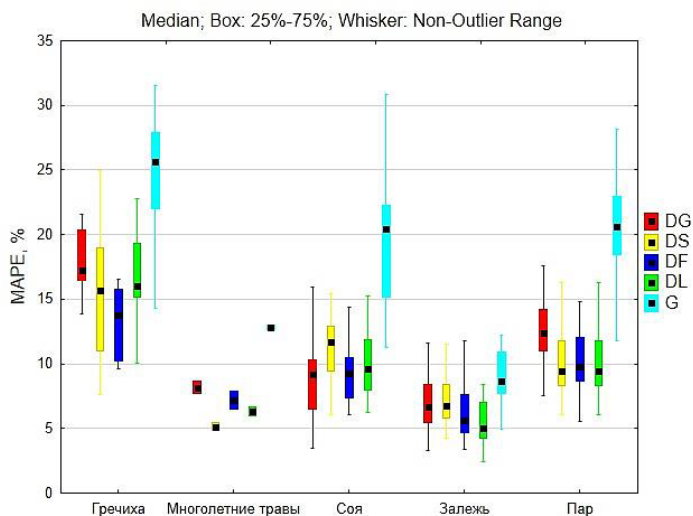


Рис. 2. Диаграмма размаха ошибки аппроксимации временных рядов сезонного хода NDVI для пахотных полей Хабаровского края в 2021 г.

Как видно из рисунка 3, в среднем значения MAPE аппроксимации сезонного хода EVI были существенно выше, в сравнении с NDVI. Ошибка аппроксимации временных рядов EVI полей с залежью составила от 8 до 18%, а для многолетних трав – от 15 до 24% при использовании DG, DS, DF и G. Применение DL для обработки исходных рядов в среднем способствовало увеличению средней ошибки, при этом можно отметить достаточно большую вариативность MAPE по отдельным полям – от 22 до 54% для залежи, от 28 до 48% для гречихи, от 33 до 58% для сои, и от 28 до 82% для парующих земель. Аппроксимация сезонного хода EVI с использованием функции Гаусса также приводила к достаточно большой средней ошибке и вариативности этого показателя: для гречихи в интервале от 18 до 52%, для сои – от 21 до 43%, а для пара – от 20 до 48%. При этом MAPE аппроксимации временных рядов EVI находилась в диапазоне 11-33% при применении DG, DS и DF (поля с соей), 11-40% (пар), 13-48% (гречиха).

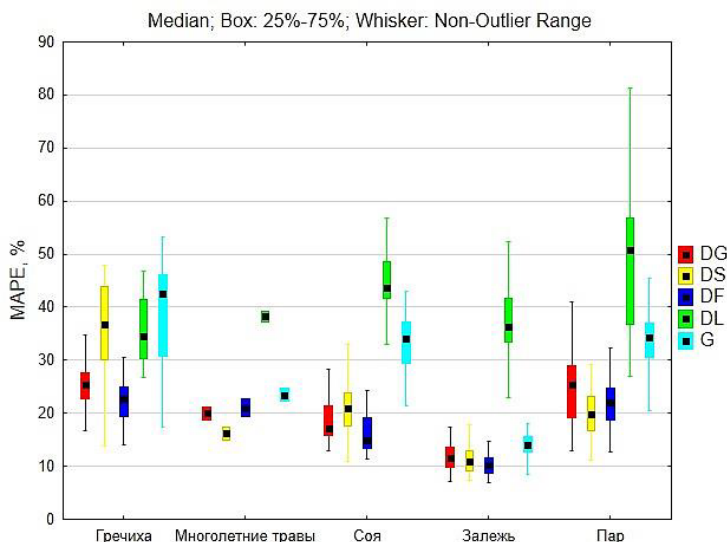


Рис. 3. Диаграмма размаха ошибки аппроксимации временных рядов сезонного хода EVI для пахотных полей Хабаровского края в 2021 г.

На следующем этапе для каждого вида произрастающей культуры проводилась оценка влияния типа функции в сочетании с NDVI или EVI на точность аппроксимации. Методами двухфакторного дисперсионного анализа было установлено, что значимо на ошибку

аппроксимации для каждой культуры влияют и тип функции, и показатель VI (таблица 1). При этом, как видно из таблицы, вклад фактора VI в точность аппроксимации в 2-4 раза выше, чем фактор типа функции для гречихи, многолетних трав и пара, и в 1,1-1,4 раза выше для залежи и сои.

Таблица 1. Оценка значимости факторов (VI и тип функции) на точность аппроксимации сезонного хода VI для разных культур

Вид культуры	Фактор	Характеристики двухфакторного анализа			
		SS	MS	F	p
Гречиха	VI	6402,0	6402,0	135,1	<0,05
	Функция	3012,4	753,1	15,9	<0,05
Многолетние травы	VI	1254,6	1254,6	50,6	<0,05
	Функция	319,6	79,9	3,2	<0,05
Соя	VI	23660,9	23660,9	444,3	<0,05
	Функция	17027,3	4256,8	79,9	<0,05
Залежь	VI	17658,5	17658,5	377,4	<0,05
	Функция	15665,8	3916,4	83,7	<0,05
Пар	VI	33669,0	33669,0	501,0	<0,05
	Функция	15621,6	3905,4	58,1	<0,05

Апостериорный анализ на основе критерия Тьюки для полей с гречихой показал, что использование DF давало существенно меньшую ошибку аппроксимации в сравнении с DS, DL и G (таблица 2). В целом, для аппроксимации сезонного хода VI гречихи целесообразно было использовать DG (ошибка 21,5%) или DF (ошибка 17,5%).

Таблица 2. Апостериорный анализ (критерий Тьюки) по аппроксимирующим функциям для полей с гречихой

	DG	DS	DF	DL	G
DG		p>0,05	p>0,05	p>0,05	p<0,05
DS	p>0,05		p<0,05	p>0,05	p>0,05
DF	p>0,05	p<0,05		p<0,05	p<0,05
DL	p>0,05	p>0,05	p<0,05		p>0,05
G	p<0,05	p>0,05	p<0,05	p>0,05	
MAPE, %	21,5	25,5	17,5	26,3	31,9

Как следует из таблицы 3, значимых различий между аппроксимацией сезонного хода VI многолетних трав пятью различными функциями выявлено не было. Очевидно, что в связи с тем, что в среднем ошибка аппроксимации NDVI многолетних трав в 2-5 раза меньше, чем EVI, необходимо оценить значимость различий отдельно для NDVI и EVI.

Таблица 3. Апостериорный анализ (критерий Тьюки) по аппроксимирующим функциям для полей с многолетними травами

	DG	DS	DF	DL	G
DG		p>0,05	p>0,05	p>0,05	p>0,05
DS	p>0,05		p>0,05	p>0,05	p>0,05
DF	p>0,05	p>0,05		p>0,05	p>0,05
DL	p>0,05	p>0,05	p>0,05		p>0,05
G	p>0,05	p>0,05	p>0,05	p>0,05	
MAPE, %	14,1	10,7	14,1	22,3	18,1

С помощью критерия Тьюки было установлено, что для полей с соей статистически значимы различия точности аппроксимации VI с применением функций G и DL (MAPE составила 26,3% и 28,2%), с одной стороны, и DG, DS, DF (ошибка от 12,7% до 16,4%) (таблица 4).

Таблица 4. Апостериорный анализ (критерий Тьюки) по аппроксимирующим функциям для полей с соей

	DG	DS	DF	DL	G
DG		p>0,05	p>0,05	<i>p<0,05</i>	<i>p<0,05</i>
DS	p>0,05		p>0,05	<i>p<0,05</i>	<i>p<0,05</i>
DF	p>0,05	p>0,05		<i>p<0,05</i>	<i>p<0,05</i>
DL	<i>p<0,05</i>	<i>p<0,05</i>	<i>p<0,05</i>		p>0,05
G	<i>p<0,05</i>	<i>p<0,05</i>	<i>p<0,05</i>	p>0,05	
MAPE, %	14,7	16,4	12,7	28,2	26,3

В тоже время для сезонного хода VI залежных земель ошибка аппроксимации функцией DF была равной 8,4%, функцией DS – 9,5%, функцией DG – 10,0%, а гауссианой – 12,5%, что достоверно ниже, чем при использовании DL (таблица 5).

Таблица 5. Апостериорный анализ (критерий Тьюки) по аппроксимирующим функциям для залежи

	DG	DS	DF	DL	G
DG		$p>0,05$	$p>0,05$	$p<0,05$	$p>0,05$
DS	$p>0,05$		$p>0,05$	$p<0,05$	$p>0,05$
DF	$p>0,05$	$p>0,05$		$p<0,05$	$p<0,05$
DL	$p<0,05$	$p<0,05$	$p<0,05$		$p<0,05$
G	$p>0,05$	$p>0,05$	$p<0,05$	$p<0,05$	
MAPE, %	10,0	9,5	8,4	21,7	12,1

Схожая ситуация была отмечена и для парующих земель (таблица 6). Точность аппроксимации VI была достоверно выше при применении функций DS и DF (ошибка, соответственно 15,4% и 15,9%) в сравнении с DL и G (ошибка 29,5% и 27,0%).

Таблица 6. Апостериорный анализ (критерий Тьюки) по аппроксимирующим функциям для пара

	DG	DS	DF	DL	G
DG		$p>0,05$	$p>0,05$	$p<0,05$	$p<0,05$
DS	$p>0,05$		$p>0,05$	$p<0,05$	$p<0,05$
DF	$p>0,05$	$p>0,05$		$p<0,05$	$p<0,05$
DL	$p<0,05$	$p<0,05$	$p<0,05$		$p>0,05$
G	$p<0,05$	$p<0,05$	$p<0,05$	$p>0,05$	
MAPE, %	19,1	15,4	15,9	29,5	27,0

В связи с тем, что аппроксимация сезонного хода индексов вегетации может применяться в задачах автоматизированной классификации пахотных земель, зачастую априори неизвестно, какой тип культуры произрастает на поле. Таким образом, с практической точки зрения, целесообразно найти метод, наилучшим образом аппроксимирующий сезонный ход каждого индекса для всех культур. С помощью двухфакторного дисперсионного анализа было установлено, что на точность аппроксимации NDVI значимо влияет и вид культуры, и аппроксимирующая функция (таблица 7). Как видно из таблицы, при аппроксимации временных рядов NDVI гречихи и сои наименьшая ошибка была характерна при использовании DF, для многолетних трав оптимальным методом явился DS, а для залежи и пара – DL. При этом различия в точности при применении этих трех функций для многолетних трав, залежи, пара и сои не являлось

достоверным ($p > 0,05$). Аналогичным образом было выявлено значимое влияние обоих факторов на сезонный ход EVI (таблица 8). В ходе попарного сравнения было установлено, что функция DF являлась оптимальным выбором для приближения временного ряда EVI гречихи и сои ($p < 0,05$), а для залежи, пара и многолетних трав возможно использование как DF, так и DS и DG ($p > 0,05$).

Таблица 7. Средние значения MAPE (%) для разных типов функций при аппроксимации сезонного хода NDVI гречихи, многолетних трав, сои, залежи и пара (Хабаровский край, 2021)

Вид культуры	DG	DS	DF	DL	G	p
Гречиха	17,5± 3,4	15,9 ±6,1	12,8 ±2,7	16,7± 3,3	24,7± 4,9	p<0,05
Многолетние травы	8,2 ± 0,6	5,2 ± 0,4	7,2 ± 1,0	6,3 ± 0,5	12,8± 0,2	
Соя	8,7 ± 2,6	11,8 ±3,4	9,3 ± 2,1	10,3± 3,4	19,3± 4,7	
Залежь	7,5 ± 3,5	7,2 ± 1,9	6,1 ± 1,8	5,7 ± 2,4	9,5 ± 3,2	
Пар	13,3± 4,0	10,1 ± 2,8	10,1± 2,3	10,1± 2,8	20,4± 3,3	
p	p<0,05					

Таблица 8. Средние значения MAPE (%) для разных типов функций при аппроксимации сезонного хода EVI гречихи, многолетних трав, сои, залежи и пара (Хабаровский край, 2021)

Вид культуры	DG	DS	DF	DL	G	p
Гречиха	25,4± 5,8	35,4 ±10,5	21,5 ±5,0	36,0± 6,6	39,7± 11,3	p<0,05
Многолетние травы	20,0± 1,8	16,2 ±1,6	20,9 ±2,3	38,3± 1,4	23,5± 1,7	
Соя	18,8± 4,9	21,1 ±5,9	16,2 ±3,6	46,5± 8,9	33,3± 5,5	
Залежь	12,5± 4,7	11,7 ±3,5	10,7 ±3,2	37,7± 8,2	14,8± 5,0	
Пар	25,0± 7,4	20,7± 6,0	21,7± 4,0	48,8± 13,8	33,6± 5,7	
p	p<0,05					

Как следует из таблицы 9, была установлена значимость различий средней ошибки аппроксимации NDVI для пахотных земель Хабаровского края для разных типов функций. В частности, наилучший результат – с ошибкой в 8,5% и 8,9% продемонстрировали DF и DL. Достоверно наиболее низкая точность аппроксимации –

16,1% – была получена при использовании гауссианы. Приближение кривой NDVI функциями DG и DS в принципе показало достаточно хорошие результаты: ошибка, соответственно 10,6% и 9,8%.

Таблица 9. Апостериорный анализ (критерий Тьюки) по аппроксимирующим функциям для сезонного хода NDVI пахотных земель Хабаровского края

	DG	DS	DF	DL	G
DG		$p>0,05$	$p<0,05$	$p<0,05$	$p<0,05$
DS	$p>0,05$		$p<0,05$	$p>0,05$	$p<0,05$
DF	$p<0,05$	$p<0,05$		$p>0,05$	$p<0,05$
DL	$p<0,05$	$p>0,05$	$p>0,05$		$p<0,05$
G	$p<0,05$	$p<0,05$	$p<0,05$	$p<0,05$	
MAPE, %	10,6	9,8	8,5	8,9	16,1

В таблице 10 представлены результаты апостериорного анализа – попарного сравнения точности аппроксимации сезонного хода EVI пахотных земель Хабаровского края в зависимости от типа функции.

Таблица 10. Апостериорный анализ (критерий Тьюки) по аппроксимирующим функциям для сезонного хода EVI пахотных земель Хабаровского края

	DG	DS	DF	DL	G
DG		$p>0,05$	$p<0,05$	$p<0,05$	$p<0,05$
DS	$p>0,05$		$p<0,05$	$p<0,05$	$p<0,05$
DF	$p<0,05$	$p<0,05$		$p<0,05$	$p<0,05$
DL	$p<0,05$	$p<0,05$	$p<0,05$		$p<0,05$
G	$p<0,05$	$p<0,05$	$p<0,05$	$p<0,05$	
MAPE, %	18,5	18,3	16,0	42,8	26,4

Как видно, наиболее низкая ошибка также наблюдалась при использовании DF (16,0%), при этом она в два раза превышала соответствующее значение MAPE для кривой NDVI. Наименее удачной для аппроксимации кривой EVI представляется функция DL, со средней ошибкой 42,8%. MAPE при применении гауссианы составила 26,4%, а функции DG и DS продемонстрировали примерно одинаковые результаты – 18,5% и 18,3%.

Представляет интерес оценка потенциальной возможности применения полученных результатов для аппроксимации временных рядов VI сельскохозяйственных земель других территорий, в частности районов со схожими климатическими характеристиками. На рисунке 4 представлены аппроксимированные с помощью DF

усредненные значения сезонного хода NDVI для полей с соей Амурской области (Архаринский, Октябрьский, Михайловский, Тамбовский р-ны), Еврейской Автономной области (Ленинский, Октябрьский р-ны), Хабаровского края (Хабаровский, Вяземский, им. Лазо р-ны).

Как видно из рисунка, кривые сезонного хода NDVI сои схожи для трех регионов южной части Дальнего Востока, при этом как общая форма кривой, так и период наступления максимума в 2022 году (219-233 календарные дни) также соответствовали 2021 году. Для пахотных земель провинции Хэйлунцзян (северо-восточный Китай) основные фазы развития сои совпадали на временной шкале с фазами развития культуры в субъектах ДФО, а графики сезонного хода NDVI вполне схожи с представленными выше [31, 32]. Таким образом, предложенный подход к аппроксимации временных рядов может быть применен в регионах с климатическими условиями, соответствующими условиям южной части Хабаровского края.

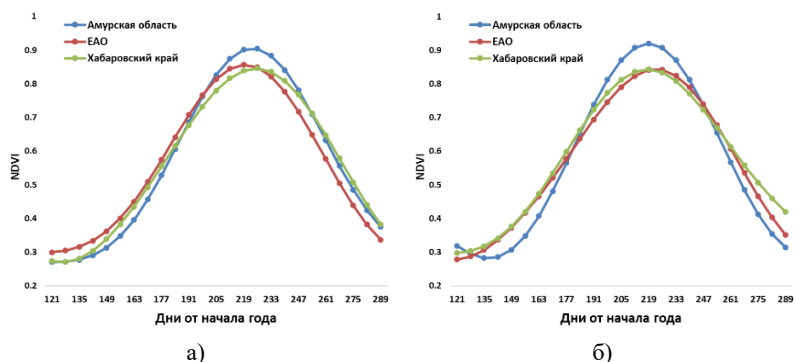


Рис. 4. Аппроксимация функцией DF временных рядов NDVI сои в Амурской области, ЕАО, Хабаровском крае: а) 2021 г., б) 2022 г.

Далее проводилась сравнительная оценка точности классификации на основе временных рядов NDVI с использованием рассмотренных аппроксимирующих функций, а также полиномиальной функции третьего порядка (куб) на примере пахотных земель Хабаровского района (2021-2022 гг.). Для классификации был использован метод RF (Random Forest), в обучающую выборку случайным образом включались 70% от общего числа пикселей, в тестовую 30%. В таблице 11 представлены значения общей точности (ОА, %), полученные по результатам трехкратной кросс-валидации.

Таблица 11. Точность классификации пахотных земель Хабаровского района на основе аппроксимированных рядов NDVI (RF, 2021-2022 гг.)

Общая точность	Полином 3-й степени	DG	DS	DF	DL	G
$OA_{cp} \pm \Delta OA_{cp}, \%$	$85,7 \pm 0,7$	$90,6 \pm 0,6$	$91,7 \pm 0,8$	$92,3 \pm 0,6$	$90,2 \pm 0,8$	$89,2 \pm 1,0$
p	$p < 0,05$; $p_{DS-DF} > 0,05$; $p_{DG-DF} > 0,05$; $p_{DL-DF} < 0,05$; $p_{G-DF} < 0,05$; $p_{куб-DF} < 0,05$;					

Методами однофакторного дисперсионного анализа с использованием апостериорного попарного сравнения было установлено, что ошибка классификации при использовании DF, DG и DS была существенно ниже, чем при применении для подготовки временных рядов DL, G, а также кубического полинома. OA классификации на основе DF превысила 92%.

4. Заключение. В ходе исследования по спутниковым снимкам Sentinel 2A за 2021 год были построены временные ряды NDVI и EVI для пахотных земель Хабаровского края. Было установлено, что аппроксимированные графики сезонного хода NDVI и EVI для основных сельскохозяйственных культур Хабаровского края имели характерный вид, позволяющий в дальнейшем использовать ряды сезонного хода в задачах идентификации культур и классификации пахотных земель, в том числе в регионах со схожими климатическими условиями. Методами двухфакторного дисперсионного анализа с апостериорным попарным сравнением на основе критерия Тьюки было установлено, что наиболее высокая точность аппроксимации временного ряда NDVI для всех культур достигалась при применении DF и DL (ошибка не превышала 9%), а временного ряда EVI – при применении DF (ошибка составила 16%). Таким образом, при необходимости подготовки временных рядов VI пахотных земель Хабаровского края для классификации (при отсутствии части данных, необходимости синхронизации данных нескольких лет наблюдения и т.д.) целесообразно использование DF. Вместе с тем, для большинства сельхозкультур функции DG и DS также показали хорошие результаты для приближения рядов сезонного хода NDVI и EVI. Использование разработанного подхода, в отличие от методов сглаживания временного ряда, позволит, во-первых, не только повысить точность автоматизированной классификации, но и использовать аппроксимированные ряды для выявления отклонений в развитии, а также прогнозирования урожайности. Изучение временных рядов VI, характерных для южной части Дальнего Востока, обеспечит возможность экстраполяции исследований на прочие субъекты ДФО, а также сопредельные провинции КНР.

Литература

1. Ferchichi A., Abbas A.B., Barra V., Farah I.R. Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review // *Ecological Informatics*. 2022. vol. 68. no. 101552.
2. Liu J., Song W. Mapping abandoned cropland using Within-Year Sentinel-2 time series // *Catena*. 2023. vol. 223. no. 106924.
3. Терехин Э.А. Оценка сезонных значений вегетационного индекса (NDVI) для детектирования и анализа состояния посевов сельскохозяйственных культур // *Исследование земли из космоса*. 2015. № 1. DOI: 10.7868/S0205961415010108.
4. Савин И.Ю., Лупян Е.А., Барталев С.А. Оперативный спутниковый мониторинг состояния посевов сельскохозяйственных культур в России // *Геоматика*. 2011. № 2. С. 69–76.
5. Черепанов А.С., Дружинин Е.Г. Спектральные свойства растительности и вегетационные индексы // *Геоматика*. 2009. № 3. С. 28–32.
6. Воронина П.В., Мамаш Е.А. Классификация тематических задач мониторинга сельского хозяйства с использованием данных дистанционного зондирования MODIS // *Вычислительные технологии*. 2014. Т. 19. № 3. С. 76–102.
7. Zhang C., Marzougui A., Sankaran S. High-resolution satellite imagery applications in crop phenotyping: An overview // *Computers and Electronics in Agriculture*. 2020. vol. 175. no. 105584.
8. Kong D., Zhang Y., Gu X., Wang D. A robust method for reconstructing global MODIS EVI time series on the Google Earth Engine // *Journal of Photogrammetry and Remote Sensing*. 2019. vol. 155. pp. 13–24.
9. Chen Y., Hu J., Cai Z., Yang J., Zhou W., Hu Q., Wang C., You L., Xu B. A phenology-based vegetation index for improving ratoon rice mapping using harmonized Landsat and Sentinel-2 data // *Journal of Integrative Agriculture*. 2023. 20 p. DOI: 10.1016/j.jia.2023.05.035.
10. Hao P., Tang H., Chen Z., Meng Q., Kang Y. Early-season crop type mapping using 30-m reference time series // *Journal of Integrative Agriculture*. 2020. vol. 19. no. 7. pp. 1897–1911.
11. Singh R.K., Rizvi J., Behera M.D., Biradar C. Automated crop type mapping using time-weighted dynamic time warping-A basis to derive inputs for enhanced food and Nutritional Security // *Current Research in Environmental Sustainability*. 2021. vol. 3. no. 100032.
12. Shammi S.A., Meng Q. Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling // *Ecological Indicators*. 2021. vol. 121. no. 107124.
13. Bandaru V., Yaramasu R., Koutilya P.N.V.R., He J., Fernando S., Sahajpal R., Wardlow B.D., Suyker A., Justice C. PhenoCrop: An integrated satellite-based framework to estimate physiological growth stages of corn and soybeans // *International Journal of Applied Earth Observation and Geoinformation*. 2020. vol. 92. no. 102188.
14. Tian F., Cai Z., Jin H., Hufkens K., Scheffinger H., Tagesson T., Smets B., Van Hoolst R., Bonte K., Ivits E., Tong X., Ardo J., Eklundh L. Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe // *Remote Sensing of Environment*. 2021. vol. 260. no. 112456.
15. Катаев М.Ю., Бекеров А.А., Медвецкий Д.В. Методика сглаживания временного ряда вегетационного индекса NDVI // *Доклады ТУСУРа*. 2017. Т. 20. № 1. С. 85–88.
16. Li S., Xu L., Jing Y., Yin H., Li X., Guan X. High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques // *International*

- Journal of Applied Earth Observation and Geoinformation. 2021. vol. 105. no. 102640.
17. Миклашевич Т.С., Барталев С.А. Метод определения фенологических характеристик растительного покрова на основе временных рядов спутниковых данных // Современные проблемы дистанционного зондирования Земли из космоса. 2016. Т. 13. № 1. С. 9–24.
18. Shao Y., Lunetta R.S., Wheeler B., Liame J.S., Campbella J.B. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data // Remote Sensing of Environment. 2016. vol. 174. pp. 258–265.
19. Плотников Д.Е., Миклашевич Т.С., Барталев С.А. Восстановление временных рядов данных дистанционных измерений методом полиномиальной аппроксимации в скользящем окне переменного размера // Современные проблемы дистанционного зондирования Земли из космоса. 2014. Т. 11. № 2. С. 103–110.
20. Atkinson P.M., Jeganathan C., Dash J., Atzberger C. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology // Remote Sensing of Environment. 2012. vol. 123. pp. 400–417.
21. Cao R., Chen J., Shen M., Tang Y. An improved logistic method for detecting spring vegetation phenology in grasslands from MODIS EVI time-series data // Agricultural and Forest Meteorology. 2015. vol. 200. pp. 9–20.
22. Seo B., Lee J., Lee K., Hong S., Kang S. Improving remotely-sensed crop monitoring by NDVI-based crop phenology estimators for corn and soybeans in Iowa and Illinois, USA // Field Crops Research. 2019. vol. 238. pp. 113–128.
23. Воробьёва Н.С., Чернов А.В. Аппроксимация временных рядов NDVI в задаче раннего распознавания видов сельскохозяйственных культур по космическим снимкам // Сборник трудов III международной конференции и молодежной школы «Информационные технологии и нанотехнологии» (ИТНТ-2017). Самара: Новая техника, 2017. С. 390–399.
24. Hird J.N., McDermid G.J. Noise reduction of NDVI time series: An empirical comparison of selected techniques // Remote Sensing of Environment. 2009. vol. 113. no. 1. pp. 248–258.
25. Julien Y., Sobrino J.A. Optimizing and comparing gap-filling techniques using simulated NDVI time series from remotely sensed global data // International Journal of Applied Earth Observation and Geoinformation. 2019. vol. 76. pp. 93–111.
26. Stepanov A., Dubrovin K., Sorokin A., Aseeva T. Predicting soybean yield at the regional scale using remote sensing and climatic data // Remote Sensing. 2020. vol. 12(12). no. 1936.
27. Stepanov A., Dubrovin K., Sorokin A. Function fitting for modeling seasonal normalized difference vegetation index time series and early forecasting of soybean yield // The Crop Journal. 2022. vol. 10. no. 5. pp. 1452–1459.
28. Lin S., Hao D., Zheng Y., Zhang H., Wang C., Yuan W. Multi-site assessment of the potential of fine resolution red-edge vegetation indices for estimating gross primary production // International Journal of Applied Earth Observation and Geoinformation. 2022. vol. 113. no. 102978.
29. Gavin H.P. The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems. Department of civil and environmental engineering, Duke University, 2019. vol. 19.
30. Zhao J., Zhang X. An Adaptive Noise Reduction Method for NDVI Time Series Data Based on S-G Filtering and Wavelet Analysis // Journal Indian Society Remote Sensing. 2018. vol. 46. no. 12. pp. 1975–1982.

31. Han Y., Meng J., Xu J. Soybean growth assessment method based on NDVI and phenological calibration // Transactions of the Chinese Society of Agricultural Engineering. 2017. vol. 33. no. 2. pp. 177–182.
32. Zhang H., Kang J., Xu X., Zhang L. Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi'an County, Heilongjiang province, China // Computers and Electronics in Agriculture. 2020. vol. 176. no. 105618.

Степанов Алексей Сергеевич — д-р фармацевт. наук, ведущий научный сотрудник, лаборатория селекции зерновых и колосовых культур, Дальневосточный научно-исследовательский институт сельского хозяйства (ДВ НИИСХ). Область научных интересов: интересов сельскохозяйственная экономика прогнозирование урожайности сельскохозяйственных культур, математическое моделирование в сельском хозяйстве. Число научных публикаций — 80. stepanfx@mail.ru; улица Клубная, 13, 680521, Восточное, Россия; р.т.: +7(924)210-9102.

Фомина Елизавета Андреевна — инженер, лаборатория численных методов математической физики, Вычислительный центр Дальневосточного отделения Российской академии наук (ВЦ ДВО РАН). Область научных интересов: математическое моделирование, аппроксимация. Число научных публикаций — 3. eliz37@mail.ru; улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(914)316-0703.

Илларионова Любовь Викторовна — канд. физ.-мат. наук, старший научный сотрудник, лаборатория численных методов математической физики, Вычислительный центр Дальневосточного отделения Российской академии наук (ВЦ ДВО РАН). Область научных интересов: машинное обучение, программирование, численные методы. Число научных публикаций — 40. illarionova_l@list.ru; улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(924)200-4940.

Дубровин Константин Николаевич — младший научный сотрудник, лаборатория численных методов математической физики, Вычислительный центр Дальневосточного отделения Российской академии наук (ВЦ ДВО РАН). Область научных интересов: математическое моделирование, машинное обучение, применение методов дистанционного зондирования в сельском хозяйстве. Число научных публикаций — 14. pobforward@gmail.com; улица Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(909)859-0881.

Федосеев Денис Владимирович — аспирант, Вычислительный центр ДВО РАН. Область научных интересов: система управления базами данных, программирование, численные методы. d.fedoseev@mail.ru; Ким Ю Чена, 65, 680000, Хабаровск, Россия; р.т.: +7(962)151-34-35.

Поддержка исследований. Исследования выполнены при финансовой поддержке РФФ, проект № 23-76-00007 на тему «Разработка научных методов и подходов устойчивого управления почвенными ресурсами на основе технологий дистанционного зондирования Земли (на примере юга Дальнего Востока)».

A. STEPANOV, E. FOMINA, L. ILLARIONOVA, K. DUBROVIN, D. FEDOSEEV
**VEGETATION INDICES (NDVI AND EVI) TIME SERIES
 APPROXIMATION FOR MONITORING CROPS OF
 Khabarovsk Territory**

Stepanov A., Fomina E., Illarionova L., Dubrovin K., Fedoseev D. Vegetation Indices (NDVI and EVI) Time Series Approximation for Monitoring Crops of Khabarovsk Territory.

Abstract. Approximation of the series of the seasonal vegetation index time series is the basis for monitoring agricultural crops, their identification and cropland classification. For cropland of the Khabarovsk Territory in the period from May to October 2021, NDVI and EVI time series were constructed using Sentinel-2A (20 m) multispectral images using a cloud mask. Five functions were used to approximate time series: Gaussian function; double Gaussian; double sine wave; Fourier series; double logistic. Characteristics of extremums for approximated time series for different types of arable land were built and calculated: buckwheat, perennial grasses, soybeans, fallow and ley. It was shown that each type requires a characteristic species. It was found ($p < 0.05$) that Fourier approximation showed the highest accuracy for NDVI and EVI series (average error, respectively, 8.5% and 16.0%). Approximation of the NDVI series using a double sine, double Gaussian and double logistic function resulted in an error increase of 8.9-10.6%. Approximation of EVI series based on double Gaussian and double sine wave causes an increase in average errors up to 18.3-18.5%. The conducted a posteriori analysis using the Tukey criterion showed that for soybean, fallow and ley lands, it is better to use the Fourier series, double Gaussian or double sine wave to approximate vegetation indices, for buckwheat it is advisable to use the Fourier series or double Gaussian. In general, the average approximation error of the NDVI seasonal time series is 1.5-4 times less than the approximation error of the EVI series.

Keywords: vegetation index, Khabarovsk Territory, approximation, arable land, crop, time series.

References

1. Ferchichi A., Abbes A.B., Barra V., Farah I.R. Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review. *Ecological Informatics*. 2022. vol. 68. no. 101552.
2. Liu J., Song W. Mapping abandoned cropland using Within-Year Sentinel-2 time series. *Catena*. 2023. vol. 223. no. 106924.
3. Terekhin E.A. [Assess the seasonal NDVI values for detection and analysis of agricultural crops]. *Issledovanie Zemli iz kosmosa – Earth exploration from space*. 2015. no. 1. DOI: 10.7868/S0205961415010108. (In Russ.).
4. Savin I.Y., Bartalev S.A., Loupian E.A. [Operational satellite monitoring of the state of agricultural crops in Russia]. *Geomatika – Geomatics*. 2011. no. 2. pp. 69–76. (In Russ.).
5. Cherepanov A.S., Druzhinina E.G. [Spectral properties of vegetation and vegetation indices]. *Geomatika – Geomatics*. 2009. no. 3. pp. 28–32. (In Russ.).
6. Voronina P.V., Mamash E.A. [Classification of thematic monitoring for agriculture problems using remote sensing MODIS data]. *Vychislitel'nye tekhnologii – Computing technologies*. 2014. vol. 19. no. 3. pp. 76–102. (In Russ.).
7. Zhang C., Marzougui A., Sankaran S. High-resolution satellite imagery applications in crop phenotyping: An overview. *Computers and Electronics in Agriculture*. 2020. vol. 175. no. 105584.

8. Kong D., Zhang Y., Gu X., Wang D. A robust method for reconstructing global MODIS EVI time series on the Google Earth Engine. *Journal of Photogrammetry and Remote Sensing*. 2019. vol. 155. pp. 13–24.
9. Chen Y., Hu J., Cai Z., Yang J., Zhou W., Hu Q., Wang C., You L., Xu B. A phenology-based vegetation index for improving ratoon rice mapping using harmonized Landsat and Sentinel-2 data. *Journal of Integrative Agriculture*. 2023. 20 p. DOI: 10.1016/j.jia.2023.05.035.
10. Hao P., Tang H., Chen Z., Meng Q., Kang Y. Early-season crop type mapping using 30-m reference time series. *Journal of Integrative Agriculture*. 2020. vol. 19. no. 7. pp. 1897–1911.
11. Singh R.K., Rizvi J., Behera M.D., Biradar C. Automated crop type mapping using time-weighted dynamic time warping-A basis to derive inputs for enhanced food and Nutritional Security. *Current Research in Environmental Sustainability*. 2021. vol. 3. no. 100032.
12. Shammi S.A., Meng Q. Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. *Ecological Indicators*. 2021. vol. 121. no. 107124.
13. Bandaru V., Yaramasu R., Koutilya P.N.V.R., He J., Fernando S., Sahajpal R., Wardlow B.D., Suyker A., Justice C. PhenoCrop: An integrated satellite-based framework to estimate physiological growth stages of corn and soybeans. *International Journal of Applied Earth Observation and Geoinformation*. 2020. vol. 92. no. 102188.
14. Tian F., Cai Z., Jin H., Hufkens K., Scheffinger H., Tagesson T., Smets B., Van Hoolst R., Bonte K., Ivits E., Tong X., Ardo J., Eklundh L. Calibrating vegetation phenology from Sentinel-2 using eddy covariance, PhenoCam, and PEP725 networks across Europe. *Remote Sensing of Environment*. 2021. vol. 260. no. 112456.
15. Kataev M.Yu., Bekerov A.A., Medveckiy D.V. [Method to smooth time series of NDVI]. *Doklady TUSURa – TUSUR reports*. 2017. vol. 20. no. 1. pp. 85–88. (In Russ.).
16. Li S., Xu L., Jing Y., Yin H., Li X., Guan X. High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques. *International Journal of Applied Earth Observation and Geoinformation*. 2021. vol. 105. no. 102640.
17. Miklashevich T.S., Bartalyov S.A. [Method for estimating vegetation cover phenological characteristics from satellite data time series]. *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa – Modern problems of remote sensing of the Earth from space*. 2016. vol. 13. no. 1. pp. 9–24. (In Russ.).
18. Shao Y., Lunetta R.S., Wheeler B., Liame J.S., Campbella J.B. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. *Remote Sensing of Environment*. 2016. vol. 174. pp. 258–265.
19. Plotnikov D.E., Miklashevich T.S., Bartalev S.A. [Using local polynomial approximation within moving window for remote sensing data time-series smoothing and data gaps recovery]. *Sovremennyye problemy distantsionnogo zondirovaniya Zemli iz kosmosa – Modern problems of remote sensing of the Earth from space*. 2014. vol. 11. no. 2. pp. 103–110. (In Russ.).
20. Atkinson P.M., Jeganathan C., Dash J., Atzberger C. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*. 2012. vol. 123. pp. 400–417.
21. Cao R., Chen J., Shen M., Tang Y. An improved logistic method for detecting spring vegetation phenology in grasslands from MODIS EVI time-series data. *Agricultural and Forest Meteorology*. 2015. vol. 200. pp. 9–20.

22. Seo B., Lee J., Lee K., Hong S., Kang S. Improving remotely-sensed crop monitoring by NDVI-based crop phenology estimators for corn and soybeans in Iowa and Illinois, USA. *Field Crops Research*. 2019. vol. 238. pp. 113–128.
23. Vorobyova N.S., Chernov A.V. [Approximation of NDVI time series in the problem of early recognition of crop species from satellite images] *Sbornik trudov III mezhdunarodnoj konferencii i molodezhnoj shkoly «Informacionnye tekhnologii i nanotekhnologii»* [Proceedings of the III International Conference and Youth School «Information Technology and Nanotechnology»]. Samara: Novaya tekhnika, 2017. pp. 390–399. (In Russ.).
24. Hird J.N., McDermid G.J. Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*. 2009. vol. 113. no. 1. pp. 248–258.
25. Julien Y., Sobrino J.A. Optimizing and comparing gap-filling techniques using simulated NDVI time series from remotely sensed global data. *International Journal of Applied Earth Observation and Geoinformation*. 2019. vol. 76. pp. 93–111.
26. Stepanov A., Dubrovin K., Sorokin A., Aseeva T. Predicting soybean yield at the regional scale using remote sensing and climatic data. *Remote Sensing*. 2020. vol. 12(12). no. 1936.
27. Stepanov A., Dubrovin K., Sorokin A. Function fitting for modeling seasonal normalized difference vegetation index time series and early forecasting of soybean yield. *The Crop Journal*. 2022. vol. 10. no. 5. pp. 1452–1459.
28. Lin S., Hao D., Zheng Y., Zhang H., Wang C., Yuan W. Multi-site assessment of the potential of fine resolution red-edge vegetation indices for estimating gross primary production. *International Journal of Applied Earth Observation and Geoinformation*. 2022. vol. 113. no. 102978.
29. Gavin H.P. The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems. Department of civil and environmental engineering, Duke University, 2019. vol. 19.
30. Zhao J., Zhang X. An Adaptive Noise Reduction Method for NDVI Time Series Data Based on S–G Filtering and Wavelet Analysis. *Journal Indian Society Remote Sensing*. 2018. vol. 46. no. 12. pp. 1975–1982.
31. Han Y., Meng J., Xu J. Soybean growth assessment method based on NDVI and phenological calibration. *Transactions of the Chinese Society of Agricultural Engineering*. 2017. vol. 33. no. 2. pp. 177–182.
32. Zhang H., Kang J., Xu X., Zhang L. Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi'an County, Heilongjiang province, China. *Computers and Electronics in Agriculture*. 2020. vol. 176. no. 105618.

Stepanov Alexey — Ph.D., Dr.Sci., Leading researcher, Laboratory of breeding cereals and ear crops, Far Eastern Research Institute of Agriculture (FEARI). Research interests: agricultural economics, forecasting the yield of crops, mathematical modeling in agriculture. The number of publications — 80. stepanfx@mail.ru; 13, Clubnaya St., 680521, Vostochnoye, Russia; office phone: +7(924)210-9102.

Fomina Elizaveta — Engineer, Laboratory of numerical methods of mathematical physics, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences (CC FEB RAS). Research interests: mathematical modeling, approximation. The number of publications — 3. eliz37@mail.ru; 65, Kim Yu Chen St., 680000, Khabarovsk, Russia; office phone: +7(914)316-0703.

Illarionova Lyubov — Ph.D., Senior researcher, Laboratory of numerical methods of mathematical physics, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences (CC FEB RAS). Research interests: machine learning, numerical methods. The number of publications — 40. illarionova_l@list.ru; 65, Kim Yu Chen St., 680000, Khabarovsk, Russia; office phone: +7(924)200-4940.

Dubrovin Konstantin — Junior researcher, Laboratory of numerical methods of mathematical physics, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences (CC FEB RAS). Research interests: mathematical modeling, machine learning, application of remote sensing methods in agriculture. The number of publications — 14. nobforward@gmail.com; 65, Kim Yu Chen St., 680000, Khabarovsk, Russia; office phone: +7(909)859-0881.

Fedoseev Denis — Post-graduate student, Computing Center of the Far Eastern Branch of the Russian Academy of Sciences (CC FEB RAS). Research interests: database management system, programming, numerical methods. d.fedoseev@mail.ru; 65, Kim Yu Chen St., 680000, Khabarovsk, Russia; office phone: +7(962)151-34-35.

Acknowledgements. The studies were financially supported by the Russian Science Foundation project № 23-76-00007 «Development of scientific methods and approaches for sustainable management of soil resources based on remote sensing technologies (in the south of the Far East)».

Ю.В. БИБИК
**ПРОГНОЗИРОВАНИЕ НА ФОНДОВЫХ РЫНКАХ С
ИСПОЛЬЗОВАНИЕМ ФОРМАЛИЗМА СТАТИСТИЧЕСКОЙ
МЕХАНИКИ**

Бибик Ю.В. Прогнозирование на фондовых рынках с использованием формализма статистической механики.

Аннотация. Аналитически исследована возможность и целесообразность прогнозирования на фондовых рынках с помощью методов и подходов статистической механики. Аппарат статистической механики применен для анализа и прогноза одного из важнейших показателей рынка – распределения логарифмической доходности. В качестве исходной модели использована модель Лотки-Вольтерра, применяемая в экологии для описания систем типа «хищник-жертва». Она адекватно аппроксимирует динамику рынка. В статье использована ее гамильтоновость, позволяющая применить аппарат статистической механики. Аппарат статистической механики (с использованием принципа максимальной энтропии) позволяет реализовать вероятностный подход, который адаптирован к условиям неопределенности фондового рынка. Канонические переменные гамильтониана представлены в виде логарифмов цен акций и облигаций, совместная функция распределения вероятности цен акций и облигаций получена в виде распределения Гиббса. Больцмановский фактор, входящий в распределение Гиббса, позволяет оценить вероятность появления тех или иных цен на акции и облигации и получить аналитическое выражение для вычисления логарифмической доходности, дающее более точные результаты, чем широко используемое нормальное (Гауссово) распределение. По своим характеристикам полученное распределение напоминает распределение Лапласа. Вычислены основные характеристики полученного распределения – среднее значение, дисперсия, асимметрия, эксцесс. Математические результаты представлены графически. Дано объяснение причинно-следственного механизма, вызывающего изменение доходности рынка. Для этого развита идея Теодора Модиса о конкуренции между акциями и облигациями за внимание и деньги инвесторов (по аналогии с оборотом биомассы в моделях типа «хищник-жертва» в биологии). Результаты исследования представляют интерес для инвесторов, теоретиков и практиков фондового рынка. Они позволяют принимать продуманные и взвешенные решения по инвестированию за счет более реального представления об ожидаемой доходности и более адекватной оценки инвестиционного риска.

Ключевые слова: динамика фондового рынка, функция распределения доходности, принцип максимальной энтропии, распределение Гиббса, распределение Лапласа.

1. Введение

1.1. Общее состояние вопроса исследования. В работе предлагается метод прогнозирования фондового рынка с использованием формализма статистической механики. Учитывая, что реальность фондовых рынков всегда сложнее и непредсказуемее любых действующих и предлагаемых методов и подходов – предложение нового метода задача крайне непростая. Чтобы разобраться в перипетиях фондового рынка его участникам нужно

постоянно и напряженно следить как минимум за следующими показателями:

- насколько сильное влияние на современный фондовый рынок оказывает политическая нестабильность;
- каковы темпы роста или падения ВВП, в каком состоянии кредитно-денежная политика, инфляция, а также, существует ли опасность сползания в кризис, и есть ли системная программа по выходу из него;
- как поведут себя цены на акции в условиях экономического спада (при таких условиях нужно ждать роста или обвала цен на акции, к какому сроку и за счет чего цены на акции могут начать выравниваться в кризисных ситуациях, существуют ли финансовые и или сырьевые резервы, которые могут стабилизировать скачки цен на фондовых рынках);
- удастся ли удержать внимание инвесторов к наиболее ликвидным, устойчивым и перспективным акциям компаний, занимающим верхние строчки рейтингов фондового рынка или их потеснят акции новых компаний, предложивших новейшие инновационные продукты;
- имеет ли смысл вкладываться в стартапы и/или высокодоходные облигации, учитывая, что по статистике только одна-две из десяти таких компаний показывают устойчивый рост;
- какой в сложившихся условиях будет реальная конкуренция акций за внимание инвесторов (будет ли стоимость акций снижаться в случае избыточного предложения или расти в случае избыточного спроса).

Широкомасштабные и важные последствия воздействия на фондовый рынок каждого из вышеперечисленных пунктов из этого далеко не полного списка, а также их исключительная сложность, требует от аналитиков глубоких знаний, кропотливого труда и обязательного наличия и использования серьезного математического обеспечения. При этом математический аппарат должен постоянно совершенствоваться, учитывать статус рынка, его структуру, улавливать как основные тенденции его развития, так и новые особенности, повышать возможности качественного и количественного анализа на новый уровень. Таким образом, в настоящее время прогнозирование фондового рынка представляет собой постоянно развивающуюся область исследований. Конечной целью этих исследований является желание сделать прогнозирование более обоснованным, а значит более понятным, прозрачным и доступным.

Отличительной особенностью современных научных исследований фондового рынка является широта взглядов, мнений и суждений по всем вопросам его состояния и развития. Поражает количество предлагаемых и используемых теорий, методов и методик, имеющих целью улучшение результативности его деятельности [1, 2]. Результаты исследовательских работ помогают современным инвесторам самостоятельно или через различных посредников и консультантов более глубоко понять тонкости и адекватно оценить возможности безрисковой портфельной теории и практики [3, 4]. Они также используются аналитиками и практиками рынка, вооружая их более продвинутыми инструментами анализа. Уже сегодня обычным для современных аналитиков является использование для анализа состояния рынка серьезного математического аппарата [5], ведущих информационных технологий [6], применение искусственного интеллекта [7 – 9]. При этом широко используются как дающие хорошие результаты методики тридцатилетней давности [10], так и методики искусственного интеллекта, в надежде на то, что несмотря на трудности их настройки и подгонки, они помогут вскрыть до сих пор не разгаданные глубинные факторы воздействия на рынок и сделают его более эффективным. По эффективности старых и новых теорий и методов у участников рынка нет единого мнения.

Результаты современных исследований показывают, что в связи с существенными отличиями в работе развитых, развивающихся и пограничных рынков предложенные исследователями методы, хорошо зарекомендовавшие себя на развитых рынках, не работают или плохо работают на остальных [11 – 14].

Развитые страны отличает сильный и активный рынок, вносящий значительный вклад в развитие экономики этих стран. Экономика стран развивающихся и пограничных рынков очень нуждается в привлечении финансовых средств как собственных, так и иностранных инвесторов. Ситуация осложняется тем, что собственные инвесторы этих рынков малоопытны и малопрофессиональны, экономическая ситуация нестабильна. Не хватает средств для формирования инвестиционных портфелей, мешают элементы тенезации рынка, законодательная и нормативно-правовая база этих рынков нуждается в совершенствовании для повышения доверия инвесторов и эффективности рынка.

На первый взгляд кажется, что в дополнительных научных исследованиях нуждаются именно эти рынки, поскольку они находятся в начале пути и инвестиции в них рассматриваются инвесторами как шаги в неизвестность. Но это ошибочное

представление, поскольку многие масштабные потрясения и потери инвесторов отмечены именно на развитых рынках. Казалось бы, при успешной многолетней разработке и внедрении научных методов и подходов, ставших классическими, при наличии аналитиков высочайшего класса, включая нобелевских лауреатов, получивших премии за научный вклад в развитие экономики, какие могут быть потрясения на развитых рынках? Однако фондовые рынки развитых стран (с начала функционирования и до сегодняшнего времени) изобилуют историями о потерях инвесторов во время характерных для фондовых рынков неожиданных крахов.

Таким образом, на сегодняшнем этапе развития фондового рынка, ни использование самых современных теорий, методов и подходов, ни высокий статус рынка, ни обеспеченность высококлассным персоналом и аналитическими инструментами, не являются гарантией от потерь и не позволяют уверенно спрогнозировать последствия принятых инвестиционных решений. В современных условиях не представляется возможным предложить универсальную формулу, которая бы учитывала все тонкости прогнозирования фондового рынка. Результативными остаются исследования главных показателей рынка (доходности [15, 16], волатильности [17, 18], асимметрии [19 – 21], эксцесса [22], риска хвоста [23]). Именно они являются фундаментом осознанных инвестиций и ожидаемой доходности.

1.2. Цели и ключевые шаги исследования. Цель данного исследования:

- выяснить можно ли с помощью методов и подходов статистической механики получить адекватное аналитическое выражение для распределения доходности, которое можно будет использовать не только для математического прогнозирования изменений доходности рынка, но и для обоснования причинно-следственного механизма, приводящего к изменениям доходности;

- позволяет ли аппарат статистической механики в условиях неопределенности фондового рынка реализовать такой вероятностный подход, с помощью которого можно будет получить более точные результаты распределения доходности по сравнению с эталонным нормальным (Гауссовым) распределением?

- если такое распределение будет получено, то к характеристикам какого из известных распределений оно будет ближе, к нормальному (Гауссовому) или, например, к распределению Лапласа?

Ключевые шаги исследования – построение гамильтониана системы, получение распределения Гиббса для цен акций и облигаций, вывод аналитической формулы для функции распределения логарифмической доходности, визуальный анализ полученных результатов.

Как известно, в настоящее время для оценки динамики рынка широко используется нормальное (Гауссово) распределение. Также известно, что оно не в полной мере отражает реальную динамику фондового рынка. И хотя существует целый ряд альтернативных видов распределений доходности (распределение богатства Парето [24], распределение цен акций Леви [25], распределение Больцмана [26], распределение Лапласа [27],) фондовый рынок пока на них не переходит и ни одно из них не является общепринятым. В тоже время целый ряд исследований подтверждает, что например, распределение Лапласа показывает большую точность при оценке ежедневной доходности рынка. Вопрос перехода на другое распределение остается открытым.

Для реализации поставленных в работе целей в качестве исходной модели использована модель Лотки-Вольтерра, применяемая в экологии для описания систем типа «хищник-жертва». Предлагаемый в работе метод исследования развивает идею, согласно которой в основе динамики фондового рынка лежит конкуренция. Математически эта идея наилучшим образом описывается с помощью модели Лотки-Вольтерра [28 – 31], которая сравнительно недавно стала использоваться для моделирования оборота акций и облигаций на фондовом рынке. Целесообразность использования этих уравнений для оценки оборота акций и облигаций (по аналогии с оборотом биомассы в моделях типа «хищник-жертва» в биологии) подробно обоснована в работе Теодора Модиса (T. Modis) [32]. Как отмечено в этой работе, подход с рассмотрением фондового рынка как экосистемы и использование для анализа и прогноза системы уравнений Лотки-Вольтерра *«обеспечивает новый, уникальный и научно обоснованный способ прогнозирования тенденций и получение понимания на фондовом рынке»*. Эта исследовательская работа была очень успешной и привела к публикации на ее основе широко известной книги Т. Модиса [33]. (С результатами успешного применения в экономике принципов статистической механики можно ознакомиться в работах [34 – 36], с результатами использования системы уравнений Лотки-Вольтерра – в работах [37 – 44]).

Следует отметить определенное отличие модели Лотки-Вольтерра, используемой в настоящей работе от модели Теодора

Модиса [32]. Основным отличием является то, что в данной работе не учитывается влияние на динамику системы конечности размера экологической ниши. В работе используется предположение, что ресурсы экологической ниши не ограничены. Однако, несмотря на это отличие, исследуемая модель также хорошо аппроксимирует динамику фондового рынка. (Подробности и объяснения выбора модели приведены в разделе 2).

Оригинальность и новизна настоящей работы заключается в том, что связь динамического и статистического описания исходной системы осуществляется через ее гамильтониан, использование которого дает возможность описывать не только динамику финансовых показателей, но и их статистические характеристики. Учитывая, что на динамику фондового рынка влияет большое количество факторов, многие из которых неизвестны, для исследования распределения доходности акций использованы такие подходы и инструменты статистической механики как вероятностный подход, реализованный в рамках статистической механики с помощью принципа максимальной энтропии, Больцмановский фактор, входящий в распределение Гиббса, производящий функционал, позволяющий определить первые моменты и его предельный случай – статистическую сумму, температура – как макроскопический параметр, релаксация, использованная для определения совместной функции распределения цен акций.

Для исследования функции распределения логарифмической доходности применен метод преобразования Фурье, который позволил получить результаты в компактном и наглядном виде. На заключительных этапах исследования (при вычислении функции распределения логарифмической доходности) в комбинации с преобразованием Фурье применен асимптотический метод перевала, позволивший получить аналитические результаты в явном виде в элементарных функциях.

Настоящая статья организована следующим образом:

- **в разделе 2** представлена исходная модель и построен ее гамильтониан;
- **в разделе 3** построен вероятностный подход и получено аналитическое выражение совместной функции распределения вероятностей цен акций и облигаций с помощью распределения Гиббса с использованием принципа максимальной энтропии;
- **в разделе 4** вычислена функция распределения логарифмической доходности акций на основе полученного в разделе 3 распределения Гиббса;

- **в разделе 5** приведены результаты аналитического вычисления эксцесса и других основных характеристик функции распределения логарифмической доходности акций;
- **в разделе 6** представлено описание графиков и выполнен визуальный анализ полученных результатов;
- **в разделе 7** приведены основные выводы по итогам исследования;
- **в приложении А** доказана факторизуемость полученного в разделе 3 распределения Гиббса;
- **в приложении В** дано пояснение и вычислена комбинация параметров $\alpha_1\beta$, входящих в формулы (4.17)–(4.20) для функции распределения логарифмической доходности;
- **в приложении С** дано вычисление в явном виде функции распределения логарифмической доходности для конкретного случая, когда интервал $t = \tau$ (τ – время релаксации);
- **в приложении Д** методом перевала вычислена аналитическая формула (4.18).

2. Исходная модель. Описание исходной модели и построение ее гамильтониана. Подбор модели для обеспечения качественного анализа доходности фондового рынка с помощью формализма статистической механики осуществлялся исходя из следующих критериев:

- исследуемая динамическая модель должна быть гамильтоновой;
- гамильтониан системы должен быть аддитивен;
- используемые канонические переменные должны быть приспособлены для описания логарифмической доходности;
- алгебраическая структура гамильтониана должна быть такой, чтобы полученное вероятностное распределение допускало интегрирование в явном виде;
- исходная динамическая система должна хорошо аппроксимировать динамику фондового рынка.

Кроме того, модель должна раскрывать причинно-следственный механизм, вызывающий изменение направления динамики рынка акций и облигаций. Для этого в работе развита идея Т. Модиса [32]. В ее основу положено предположение о главенствующей роли конкуренции на фондовом рынке между акциями и облигациями за внимание и деньги инвесторов (по аналогии с оборотом биомассы в моделях типа «хищник-жертва»).

Для решения этих вопросов в настоящей работе также использована модель Лотки-Вольтерра типа «хищник-жертва». Основным ее отличием от модели Т. Модиса является то, что в данной работе не учитывается влияние на динамику системы конечности размера экологической ниши. В этой работе используется предположение, что ресурсы экологической ниши не ограничены. Такой выбор продиктован изложенными выше требованиями к модели. Несмотря на определенное упрощение, как будет показано ниже, исследуемая модель хорошо аппроксимирует динамику фондового рынка.

В оригинальной работе Теодора Модиса [32] для описания взаимодействия между акциями и облигациями использована разновидность системы Лотки-Вольтерра с квадратичными членами, учитывающими конечный размер экологической ниши. Она имеет вид:

$$\frac{dX}{dt} = \alpha_x X - b_x X^2 + c_{xy} XY, \quad (2.1)$$

$$\frac{dY}{dt} = \alpha_y Y - b_y Y^2 + c_{yx} YX. \quad (2.2)$$

Здесь параметры α , b и c влияют на скорость роста. Параметры α определяют способность каждого вида размножаться, параметры b связаны с ограничением емкости ниши, параметры c описывают взаимодействие с другими видами.

X – численность акций, Y – численность облигаций.

Данная разновидность системы Лотки-Вольтерра в общем случае не гамильтонова. Она содержит шесть произвольных параметров и требование ее гамильтоновости налагает на них три условия. При выполнении этих условий она гамильтонова. В противном случае – нет. Ее гамильтониан имеет вид:

$$H = aXY - bXY^2 - cX^2Y. \quad (2.3)$$

Уравнение Гамильтона для данного гамильтониана совпадает с уравнениями (2.1), (2.2) при условии:

$$\alpha_x = a; \quad \alpha_y = -a, \quad (2.4)$$

$$b_x = 2c; b_y = -2b, \quad (2.5)$$

$$c_{xy} = -2b; c_{yx} = 2c. \quad (2.6)$$

Каноническими координатами и каноническими импульсами данной системы являются переменные X и Y соответственно. Для приведения системы к гамильтоновому виду не требуется каких-либо их преобразований. Однако применительно к условиям реального фондового рынка эти довольно жесткие условия скорее всего не будут выполняться и в большинстве случаев система не будет гамильтоновой.

Для реализации поставленных в работе целей (прогнозирование фондового рынка с использованием формализма статистической механики) гамильтоновость системы является одним из важнейших основных условий. Поэтому в данной работе используется другая разновидность системы Лотки-Вольтерра, которая позволяет построить удовлетворяющий всем вышеперечисленным требованиям аддитивный гамильтониан. Она также отражает идею Т. Модиса о конкуренции акций и облигаций за внимание инвесторов и имеет следующий вид:

$$\frac{dx}{dt} = \alpha_1 x - \gamma_1 xy, \quad (2.7)$$

$$\frac{dy}{dt} = -\alpha_2 y + \gamma_2 xy, \quad (2.8)$$

где x – средняя цена облигаций, y – средняя цена акций, α_1 – коэффициент, которому в биологии соответствует коэффициент рождаемости жертв, α_2 – коэффициент, которому соответствует коэффициент смертности хищников, γ_1 и γ_2 – коэффициенты, отражающие взаимодействия акций и облигаций, t – время.

Эта система является частным случаем системы (2.1), (2.2) при условии равенства нулю коэффициентов b_x и b_y .

Перейдем далее к построению гамильтониана исходной системы (2.7), (2.8). Необходимо отметить, что не существует единого рецепта приведения системы к гамильтоновому виду. Все зависит от конкретной ситуации. В рассматриваемом в работе случае гамильтоновость обусловлена факторизацией правых частей системы

(2.7), (2.8). Поясним сказанное. Система (2.7), (2.8) может быть представлена в виде с факторизованными правыми частями:

$$\frac{dx}{dt} = x(\alpha_1 - \gamma_1 y), \quad (2.9)$$

$$\frac{dy}{dt} = y(-\alpha_2 + \gamma_2 x). \quad (2.10)$$

Разделим левую и правую части уравнения (2.9) на x и левую и правую части уравнения (2.10) на y . Получим следующую систему:

$$\frac{1}{x} \frac{dx}{dt} = \frac{d}{dt} \int_1^x \frac{dx}{x} = \alpha_1 - \gamma_1 y, \quad (2.11)$$

$$\frac{1}{y} \frac{dy}{dt} = \frac{d}{dt} \int_1^y \frac{dy}{y} = -\alpha_2 + \gamma_2 x. \quad (2.12)$$

В итоге получена формула, левая часть которой зависит только от x , а правая только от y , и, наоборот, для второго уравнения системы. Такая алгебраическая структура характерна для гамильтоновых систем с аддитивным гамильтонианом. Для приведения системы к гамильтоновому виду теперь достаточно сделать замену переменных.

Вид замены переменных подсказан структурой интегралов в левых частях формул (2.11), (2.12):

$$q(y) = \int_1^y \frac{dy}{y} = \ln y, \quad (2.13)$$

$$p(x) = \int_1^x \frac{dx}{x} = \ln x. \quad (2.14)$$

Благодаря удачно выполненной замене переменных полученные канонические переменные p и q (представляющие собой логарифм цены облигаций и логарифм цены акций соответственно) намного

упрощают дальнейшее исследование. В этих переменных система (2.11), (2.12) приобретает гамильтонов вид:

$$\frac{dq}{dt} = \gamma_2 e^p - \alpha_2 = \frac{\partial H}{\partial p}, \quad (2.15)$$

$$\frac{dp}{dt} = \alpha_1 - \gamma_1 e^q = -\frac{\partial H}{\partial q}. \quad (2.16)$$

При этом гамильтониан принимает следующую форму:

$$H = \gamma_2 e^p - \alpha_2 p + \gamma_1 e^q - \alpha_1 q = H_1(p) + H_2(q). \quad (2.17)$$

В разделе представлена исходная система (2.7), (2.8) и построено ее гамильтониан (2.17), удовлетворяющие перечисленным в начале раздела требованиям. Гамильтониан аддитивен, канонические переменные p и q представляют собой логарифмы цен акций и облигаций, что значительно упрощает дальнейшее исследование распределения доходности акций. В следующем разделе с помощью заданного гамильтониана перейдем к построению вероятностного подхода, позволяющего в рамках формализма статистической механики получить аналитическое выражение совместной функции распределения вероятностей цен акций и облигаций.

3. Построение вероятностного подхода. Получение аналитического выражения совместной функции распределения вероятностей цен акций и облигаций с помощью распределения Гиббса с использованием принципа максимальной энтропии. Учитывая, что на динамику фондового рынка влияет большое количество факторов, многие из которых неизвестны, совершенно очевидно, что исследование распределения доходности акций требует использования вероятностных подходов. В данной работе вероятностное описание реализуется с помощью инструментов статистической механики. Наличие закона сохранения – гамильтониана исследуемой системы позволяет ввести каноническое распределение Гиббса. Для его построения в данном разделе будут выполнены следующие шаги:

1. Будет использована энтропия – один из основных инструментов статистической механики.

2. Для определения функционала энтропии будет использовано классическое представление энтропии Больцмана или Шеннона.

3. Для получения распределения Гиббса к функционалу энтропии будут добавлены множители Лагранжа, учитывающие сохранение энергии и нормировку функции вероятности.

4. Будет найден максимум построенного функционала, представляющий собой распределение Гиббса.

Скажем несколько слов о природе энтропии. Очень грубо можно сказать, что энтропия является «вероятностью вероятностей». Энтропия является счетчиком числа вариантов, которыми может реализоваться данное распределение вероятностей. Точнее – она является логарифмом этого числа. Поэтому на практике проявляется то распределение, для которого энтропия максимальна. В свете сказанного становится ясным содержание принципа максимальной энтропии – выживает распределение вероятностей, реализующееся максимальным числом способов. На практике часто максимизируется энтропия с учетом ограничений, присущих данной системе. Ниже будет использоваться два таких ограничения – нормировка интеграла от функции распределения вероятностей и сохранение полной энергии системы. Последнее условие непосредственно использует в явном виде полученный в предыдущем разделе гамильтониан.

Перейдем к реализации принципа максимальной энтропии. Формула классической энтропии Больцмана или Шеннона имеет следующий вид:

$$S = - \int \int_{-\infty}^{\infty} P(p, q) \ln P(p, q) dp dq . \quad (3.1)$$

Здесь $P(p, q)$ – функция распределения вероятности.

Согласно принципу максимальной энтропии распределение вероятностей большого ансамбля переменных (в нашем случае большого количества финансовых инструментов, которые могут принимать разные значения) должно быть таким, чтобы энтропия Больцмана или Шеннона максимизировалась при соответствующих ограничениях, учитывающих необходимость сохранения основных параметров системы. Эти ограничения аналогичны ограничениям, применяемым в статистической физике к идеальному газу.

Первое ограничение состоит в том, что интеграл по всей области интегрирования от входящей в энтропию функции

распределения вероятностей равен единице. Это ограничение соответствует сохранению числа частиц в аналогичной модели для газа:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(p, q) dp dq = 1. \quad (3.2)$$

Второе ограничение состоит в необходимости сохранения полученной с помощью имеющегося гамильтониана средней энергии исследуемой системы, что соответствует сохранению энергии в аналогичной модели газа:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(p, q) P(p, q) dp dq = E. \quad (3.3)$$

Для максимизации энтропии и реализации этих ограничений используем далее множители Лагранжа. Они используются для того, чтобы в процессе поиска максимума энтропии эти ограничения автоматически выполнялись. При поиске максимума энтропии строится функционал, содержащий функционал энтропии и множители Лагранжа:

$$F = S - \beta \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(p, q) P(p, q) dp dq - \mu \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(p, q) dp dq. \quad (3.4)$$

Здесь S – функционал энтропии, а два идущих за ней выражения – множители Лагранжа, обеспечивающие автоматическое выполнение условий (3.2), (3.3).

Как известно, в точке максимума энтропии вариационная производная этого функционала должна обращаться в ноль. Вначале вычислим вариационную производную от этого функционала:

$$\frac{\delta F}{\delta P(p, q)} = -\ln P(p, q) - 1 - \beta H(p, q) - \mu. \quad (3.5)$$

Далее приравняем левую часть формулы (3.5) к нулю в силу обращения в ноль вариационной производной, получим следующую формулу:

$$\ln P(p, q) = -(1 + \mu) - \beta H(p, q). \quad (3.6)$$

Возьмем в уравнении (3.6) экспоненты от его левой и правой частей, получим совместную функцию распределения вероятностей цен акций и облигаций $P(p, q)$, которая может быть представлена в виде следующего распределения Гиббса:

$$P(p, q) = e^{-(1+\mu)} e^{-\beta H} = C e^{-\beta H(p, q)}. \quad (3.7)$$

Здесь $C = e^{-(1+\mu)}$. Параметр C представляет собой нормировочный множитель, который подбирается из условия, указанного в формуле (3.2). Он нормирует функцию распределения вероятности таким образом, чтобы интеграл от нее был равен единице.

Обычно он записывается в виде $C = \frac{1}{Z}$, где Z называется статистической суммой, $e^{-\beta H(p, q)}$ – представляет собой больцмановский фактор, который будет использован далее для определения вероятности конфигурации исследуемой системы.

Полученное распределение Гиббса будет использовано в следующем разделе для статистического описания цен акций и построения распределения логарифмической доходности акций на фондовом рынке.

4. Вычисление функции распределения логарифмической доходности акций на основе полученного в разделе 3 распределения Гиббса. Основной целью данной работы является получение функции распределения логарифмической доходности акций $W(r, t)$. В основе вычисления этой функции лежат две основные идеи. Первая идея состоит в вычислении функции распределения логарифмической доходности для произвольного интервала времени по функции распределения логарифмической доходности для интервала времени, равного времени релаксации с использованием мультипликативности ее Фурье-образа. Вторая идея состоит в использовании факторизации совместной функции распределения цен акций и облигаций, через которые выражается функция распределения логарифмической доходности для интервала времени, равного времени релаксации. При этом совместная функция распределения цен акций и облигаций выражается через произведение распределений Гиббса, найденных в предыдущем разделе.

Приступим к осуществлению этой программы.

Функция распределения логарифмическая доходности акций зависит от значения интервала t , для которого вычисляется доходность. Функции распределения доходности для разных интервалов связаны следующим соотношением:

$$W(r, t_1 + t_2) = \int_{-\infty}^{\infty} W(r - r_0, t_1) W(r_0, t_2) dr_0. \quad (4.1)$$

Это соотношение означает, что функции распределения доходности для интервала времени $t = t_1 + t_2$ равна свертке функции распределения доходности для интервалов времени t_1 и t_2 . Соотношение (4.1) особенно просто выглядит для Фурье-образов функций распределения логарифмической доходности:

$$\tilde{W}(k, t_1 + t_2) = \tilde{W}(k, t_1) \tilde{W}(k, t_2). \quad (4.2)$$

Свойство, выражаемое формулой (4.2), можно назвать свойством мультипликативности Фурье-образов функции распределения логарифмической доходности. Здесь Фурье-образ функций распределения логарифмической доходности имеет вид:

$$\tilde{W}(k, t) = \int_{-\infty}^{\infty} e^{ikr} W(r, t) dr. \quad (4.3)$$

Из формулы (4.2) следует, что если известна функция распределения доходности и ее Фурье-образ для интервала t_0 , то Фурье-образ функции распределения доходности для произвольного интервала t определяется формулой:

$$\tilde{W}(k, t) = \tilde{W}(k, t_0)^{\frac{t}{t_0}}. \quad (4.4)$$

Из формулы (4.4) следует, что для получения функции распределения логарифмической доходности для произвольного интервала времени t достаточно знать функцию распределения доходности для одного определенного интервала t_0 . Для вычисления функцию распределения доходности для интервала времени t_0 необходимо знать совместную плотность распределения цен акций

$P(q, q_0, t_0)$. (Здесь q_0 – цена акций в момент времени t_0 , q – цена акций в момент времени $t_0 + t$). Функция распределения логарифмической доходности определяется по следующей формуле:

$$W(r, t_0) = \int_{-\infty}^{\infty} P(q_0 + r, q_0, t_0) dq_0. \quad (4.5)$$

В свою очередь совместная плотность распределения цен акций выражается через совместную плотность распределения цен акций и облигаций по формуле:

$$P(q, q_0, t_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(p, q, p_0, q_0, t_0) dp dp_0. \quad (4.6)$$

Из формул (4.5), (4.6) вытекает следующая формула для функции распределения логарифмической доходности:

$$W(r, t_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(p, q_0 + r, p_0, q_0, t_0) dp dp_0 dq_0. \quad (4.7)$$

В качестве интервала времени t_0 выбран интервал, равный времени релаксации τ . (В связи с использованием термина «время релаксации» здесь и далее подразумевается аналогия между случайными блужданиями цены и случайными движениями молекул в газе). Для этого интервала совместная плотность распределения цен акций и облигаций факторизуется и выражается через пару распределений Гиббса:

$$P(p, q, p_0, q_0, \tau) = P(p, q) P(p_0, q_0). \quad (4.8)$$

Поведение совместной функции распределения цен акций и облигаций взято в разные моменты времени для некоторого интервала t . Зависимость поведения функции от интервала t понять достаточно легко. Если интервал времени t мал, то трудно рассчитывать на то, что зависимость от переменных p и q будет сильно отличаться от зависимости от переменных p_0 и q_0 . В пределе $t \rightarrow 0$ должно получиться что-то, похожее на выражение

$\delta(p - p_0)\delta(q - q_0)P(p_0, q_0)$. Нетрудно проверить, что интеграл от этого выражения по переменным p_0 и q_0 даст функцию $P(p, q)$, а интеграл по переменным p и q даст функцию $P(p_0, q_0)$, как это и должно быть по определению совместной функции распределения цен акций и облигаций. С увеличением интервала t начинает проявляться большая независимость в поведении функции для переменных p , q и p_0 , q_0 . Для некоторого, достаточно большого интервала t , поведение функции в зависимости от переменных p, q должно забыть о влиянии зависимости от переменных p_0 и q_0 . Для этого интервала времени совместная функция распределения цен акций и облигаций приближенно факторизуется. Она будет иметь вид, представленный формулой (4.8). Этот интервал времени и есть время релаксации. Это именно тот интервал, для которого совместная функция распределения цен акций и облигаций просто выражается через распределение Гиббса. Взять меньшее время нельзя, потому, что нарушится факторизация, а для большего времени выражение (4.8) перестает согласовываться с мультипликативностью Фурье-образов функции доходности (формула (4.2)). Поэтому, для вычисления функции доходности для производного интервала времени (формула (4.4)), необходимо выбрать в качестве времени t_0 время релаксации, а в качестве Фурье-образа функции доходности $\tilde{W}(k, \tau)$ – Фурье-образ функции, определенный ниже формулой (4.9).

Формулы (4.6), (4.8) определяют функцию распределения логарифмической доходности для интервала времени $t_0 = \tau$:

$$W(r, \tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(p, q_0 + r) P(p_0, q_0) dp dq_0 . \quad (4.9)$$

Формула (4.9) может быть значительно упрощена, если воспользоваться формулой (А.7) факторизации распределения Гиббса, подробно описанной в **приложении А**:

$$P(p, q) = P(p)P(q) . \quad (4.10)$$

В итоге функция распределения доходности для интервала τ примет следующий вид:

$$W(r, \tau) = \int_{-\infty}^{\infty} P(q_0 + r)P(q_0)dq_0. \quad (4.11)$$

Для того чтобы перейти к другим временным интервалам необходимо вычислить Фурье-образ функции распределение доходности $W(r, \tau)$ и воспользоваться формулой (4.4). Фурье-образ функции распределение логарифмической доходности $W(r, \tau)$ определяется следующей формулой:

$$\tilde{W}(k, \tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{ikr} P(q_0 + r)P(q_0)dq_0 dr = \tilde{P}(k)\tilde{P}(-k). \quad (4.12)$$

Здесь $\tilde{P}(k)$ – Фурье-образ входящего в распределение Гиббса сомножителя $P(q)$, определенного формулой (A.7). Этот Фурье-образ имеет вид:

$$\tilde{P}(k) = \int_{-\infty}^{\infty} e^{ikq} P(q)dq = \frac{1}{Z_1} \int_{-\infty}^{\infty} e^{ikq} e^{-\beta\gamma_1 e^q + \beta\alpha_1 q} dq. \quad (4.13)$$

В интеграле (4.13) нетрудно узнать интеграл Эйлера, если сделать следующую замену переменных:

$$\theta = \beta\gamma_1 e^q. \quad (4.14)$$

После этой замены получим выражение для функции $\tilde{P}(k)$ в терминах гамма-функций:

$$\tilde{P}(k) = \frac{1}{Z_1} \int_{-\infty}^{\infty} \frac{e^{-\theta}}{\theta} \frac{\theta^{(\beta\alpha_1 + ik)}}{(\beta\gamma_1)^{(\beta\alpha_1 + ik)}} d\theta = \frac{1}{(\beta\gamma_1)^{ik}} \frac{\Gamma(\beta\alpha_1 + ik)}{\Gamma(\beta\alpha_1)}. \quad (4.15)$$

При выводе этой формулы использована формула (A.5) для статистической суммы Z_1 . Таким образом, получено окончательное аналитическое выражение для Фурье-образа $\tilde{P}(k)$ в явном виде. Для получения окончательного выражения для функции распределения

логарифмической доходности (с использованием формул (4.12) и (4.15)) определим далее выражение $\tilde{P}(k)\tilde{P}(-k)$:

$$\tilde{P}(k)\tilde{P}(-k) = \frac{\Gamma(\beta\alpha_1 - ik)}{\Gamma(\beta\alpha_1)} \frac{\Gamma(\beta\alpha_1 + ik)}{\Gamma(\beta\alpha_1)}. \quad (4.16)$$

Подставив полученное выражение в формулу (4.12) и воспользовавшись формулой (4.4), получим окончательное аналитическое выражение для функции распределения логарифмической доходности:

$$W(r, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikr} \left[\frac{\Gamma(\beta\alpha_1 - ik)}{\Gamma(\beta\alpha_1)} \frac{\Gamma(\beta\alpha_1 + ik)}{\Gamma(\beta\alpha_1)} \right]^{\frac{t}{\tau}} dk. \quad (4.17)$$

Входящая в формулу (4.17) комбинация параметров $\beta\alpha_1$ вычислена в **приложении В**.

Формула (4.17) является точной. Поскольку при численных расчетах работать с гамма-функциями неудобно, специально для этих целей упростим аналитическую формулу (4.17) с использованием большой величины параметра $\beta\alpha_1$. В отличие от формулы (4.17) упрощенная аналитическая формула будет выражена в терминах элементарных функций. Это делает очевидным ее поведение для аналитиков, позволяет более глубоко понять и оценить основные особенности исследуемой системы, а также значительно упрощает представленную в разделе 6 процедуру визуального анализа результатов исследования. Упрощенная формула имеет следующий вид:

$$W(r, t) = \frac{A}{\left(2 \cosh\left(\frac{r}{2(t/\tau)}\right)\right)^{\left((M-1)\frac{t}{\tau}+1\right)}}. \quad (4.18)$$

Отправной точкой получения формул (4.17), (4.18) является функция $W(r, \tau)$, определенная формулой (4.11). Ее Фурье-образ находится в квадратных скобках в подынтегральном выражении в формуле (4.17). При вычислении функции $W(r, \tau)$ с помощью формулы (4.17) можно стартовать не с Фурье-образа функции $W(r, \tau)$,

как это предлагается в формуле (4.17), а непосредственно с функции $W(r, \tau)$, как это сделано ниже, в **приложении D**. При этом будет дважды применяться прямое и обратное преобразование Фурье. Результаты обоих преобразований Фурье будут вычислены методом перевала. Он также будет применен дважды. Ввиду важности функции $W(r, \tau)$ для процесса вычисления функции $W(r, t)$, представленной формулой (4.17), ее вывод приведен в **приложении C**, где она будет получена в терминах элементарных функций. Вывод формулы (4.18) выполнен с использованием метода перевала. Подробности вывода формулы приведены в **приложении D**.

В формулы (4.17) и (4.18) входит комбинация параметров $\beta\alpha_1$. В формулу (4.18) она входит в виде параметра $M = 2\beta\alpha_1$. Эта комбинация представляет собой единственный параметр, полученный путем обработки исторических данных (ежедневные цены закрытия для индекса S&P 500 SPY – источник Yahoo Finance). Остальные данные вычисляются аналитически. В **приложении B** этот параметр выражен через отношение квадрата среднего значения логарифмов цен акций и дисперсии логарифмов цен акций (формула (B.9)).

Таким образом, формулы (4.17) и (4.18) позволяют получить функцию распределения логарифмической доходности акций, используя всего два первых момента для массива логарифмов цен акций вместо использования всего массива, как это обычно делается при построении графика функции распределения доходности с использованием исторических данных.

Численная проверка аналитических формул (4.17), (4.18) показала, что они адекватно описывают только однодневные доходности. Поэтому для улучшения прогностических характеристик формул (4.17), (4.18) выполняется их перенормировка. Она сводится только к мультипликативной перенормировке большого параметра $M = 2\beta\alpha_1$, входящего в эти формулы. При этом $M \rightarrow \lambda M$, где λ – параметр перенормировки. После такой перенормировки и численной проверки ее результатов формулы (4.17), (4.18) описывают весь диапазон доходностей. Это позволит аналитикам адекватно оценивать не только ежедневную доходность, но и доходность за произвольный промежуток времени (неделю, месяц и т.д.). Для случая использования перенормировки в перенормированных формулах (4.17), (4.18) следует заменить параметр $M = 2\beta\alpha_1$ на $\lambda M = 2\lambda\beta\alpha_1$. Универсальные перенормированные формулы (4.17), (4.18) будут иметь следующий вид:

$$W(r, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikr} \left[\frac{\Gamma(\lambda\beta\alpha_1 - ik)}{\Gamma(\lambda\beta\alpha_1)} \frac{\Gamma(\lambda\beta\alpha_1 + ik)}{\Gamma(\lambda\beta\alpha_1)} \right]^{\frac{t}{\tau}} dk. \quad (4.19)$$

$$W(r, t) = \frac{A}{\left(2 \cosh\left(\frac{r}{2(t/\tau)}\right)\right)^{((\lambda M - 1)\frac{t}{\tau} + 1)}}. \quad (4.20)$$

Здесь перенормированная нормировочная константа A имеет следующий вид:

$$A = \frac{\Gamma((\lambda M - 1)(t/\tau) + 1)}{\Gamma^2(0.5(\lambda M - 1)(t/\tau) + 1)} \frac{1}{(t/\tau)}. \quad (4.21)$$

После получения аналитического выражения для функции распределения логарифмической доходности акций (4.20) перейдем к определению первых моментов функции – среднего значения доходности, дисперсии, асимметрии и эксцесса.

5. Аналитическое вычисление эксцесса и других основных характеристик для функции распределения логарифмической доходности акций. Функция распределения логарифмической доходности позволяет получить аналитические выражения для четырех первых моментов функции. Этими моментами являются – среднее значение μ , дисперсия σ^2 , асимметрия S и эксцесс K . Асимметрия и эксцесс выражаются через моменты с точностью до множителей $\frac{1}{\sigma^3}$ и $\frac{1}{\sigma^4}$. Хорошо известно, что моменты выражаются через производные в нуле от Фурье-образа функции распределения логарифмической доходности. Поэтому существенным фактором является то, что получено не только аналитическое выражение для функции распределения логарифмической доходности, но также аналитическое выражение для ее Фурье-образа. Это позволяет автоматически получить четыре первых момента и выражающиеся через них характеристики функции распределения логарифмической доходности μ , σ^2 , S и K .

Перейдем к их получению.

Первый и третий моменты не требуют дополнительных вычислений, поскольку в силу симметрии функции распределения доходности относительно замены $r \rightarrow -r$ они равны нулю:

$$\mu(\beta) = \langle r \rangle = \int_{-\infty}^{\infty} r W(r, t) dr = 0, \quad (5.1)$$

$$S(\beta) = \frac{\langle (r - \mu)^3 \rangle}{\langle (r - \mu)^2 \rangle^{3/2}} = \frac{1}{\sigma^3(\beta)} \int_{-\infty}^{\infty} (r - \mu)^3 W(r, t) dr = 0. \quad (5.2)$$

Таким образом, величины полученного среднего значения и асимметрии получены автоматически. Поэтому фактически необходимо вычислить только дисперсию и эксцесс. Выражение для дисперсии имеет следующий вид:

$$\sigma^2 = \langle (r - \mu)^2 \rangle = \int_{-\infty}^{\infty} (r - \mu)^2 W(r, t) dr. \quad (5.3)$$

Интеграл (5.3) может быть вычислен аналитически:

$$\begin{aligned} \int_{-\infty}^{\infty} r^2 W(r, t) dr &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r^2 \frac{e^{-ikr}}{2\pi} \tilde{W}(k, t) dk dr = \\ &= - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{-ikr}}{2\pi} \frac{\partial^2}{\partial k^2} \tilde{W}(k, t) dk dr = - \frac{d^2}{dk^2} [\tilde{P}(k) \tilde{P}(-k)]^t \Big|_{k=0}. \end{aligned} \quad (5.4)$$

В формуле (5.4) подразумевалось интегрирование по частям по переменной k для учета множителя r^2 . В итоге из формулы (5.4) следует окончательное выражение для дисперсии:

$$\sigma^2 = -\frac{2t}{\tau} \left[\left(\frac{\Gamma^I(\lambda\beta\alpha_1)}{\Gamma(\lambda\beta\alpha_1)} \right)^2 - \left(\frac{\Gamma^{II}(\lambda\beta\alpha_1)}{\Gamma(\lambda\beta\alpha_1)} \right) \right]. \quad (5.5)$$

Эксцесс вычисляется по формуле:

$$K(\beta) = \frac{\langle (r - \mu)^4 \rangle}{\langle (r - \mu)^2 \rangle^2} = \frac{1}{\sigma^4(\beta)} \int_{-\infty}^{\infty} (r - \mu)^4 W(r, t) dr. \quad (5.6)$$

Входящий в выражение (5.6) для эксцесса интеграл может быть преобразован к следующему виду:

$$\int_{-\infty}^{\infty} r^4 W(r, t) dr = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{-ikr}}{2\pi} \frac{\partial^4}{\partial k^4} \tilde{W}(k, t) dk dr =$$

$$= \int_{-\infty}^{\infty} \delta(k) \frac{\partial^4}{\partial k^4} \tilde{W}(k, t) dk = \frac{d^4}{dk^4} [\tilde{P}(k) \tilde{P}(-k)]^{\frac{t}{\tau}} \Big|_{k=0} \quad (5.7)$$

В формуле (5.7) подразумевается использование интегрирования по частям по переменной k для учета множителя r^4 . В итоге формула эксцесса принимает следующий вид:

$$K(\beta) = \frac{1}{\sigma^4} \left(\frac{2t}{\tau} \left[\frac{\Gamma^{IV}(\lambda\beta\alpha_1)}{\Gamma(\lambda\beta\alpha_1)} \right] - \frac{4\Gamma^{III}(\lambda\beta\alpha_1)}{\Gamma(\lambda\beta\alpha_1)} \frac{\Gamma^I(\lambda\beta\alpha_1)}{\Gamma(\lambda\beta\alpha_1)} + \right.$$

$$\left. + 3 \left(\frac{\Gamma^{II}(\lambda\beta\alpha_1)}{\Gamma(\lambda\beta\alpha_1)} \right)^2 \right] + 12 \frac{t}{\tau} \left(\frac{t}{\tau} - 1 \right) \left[\left(\frac{\Gamma^I(\lambda\beta\alpha_1)}{\Gamma(\lambda\beta\alpha_1)} \right)^2 - \frac{\Gamma^{II}(\lambda\beta\alpha_1)}{\Gamma(\lambda\beta\alpha_1)} \right]^2 \right) \quad (5.8)$$

Таким образом, выражения для среднего значения, дисперсии, асимметрии и эксцесса получены в явном виде.

6. Описание графиков и визуальный анализ полученных результатов. Результаты визуального анализа полученных данных представлены на рисунках 1 и 2. Обработка всех данных (включая исходные и полученные в работе) и построение графиков выполнено с использованием программы Matlab. Для построения графиков использованы исторические данные ежедневной доходности за двадцать лет (за период с 05.06.2003 по 05.06.2023) для индекса S&P 500 SPY (источник данных – Yahoo Finance, цены закрытия). В работе используется массив логарифмов цен акций, полученный логарифмированием массива цен акций.

На рисунке 1 совмещены три следующих графика:

- график функции распределения логарифмической доходности, построенный с использованием формулы (4.20), представлен на рисунке сплошной линией. Единственным параметром (входящим в формулы (4.19) и (4.20) и использующим исторические данные, является комбинация параметров $\beta\alpha_1$. Для ее вычисления используются только два первых момента для массива логарифмов цен акций (формула (B.9)) вместо всего массива логарифмов цен акций (как это обычно делается при построении графиков распределения доходности с использованием исторических данных);

- график функции распределения логарифмической доходности S&P 500 SPY, построенный с использованием

исторических данных ежедневной доходности за двадцать лет для цен закрытия, представлен штриховой линией;

– график нормального распределения выполнен с

использованием стандартного распределения Гаусса $P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$

и представлен на рисунке штрихпунктирной линией.

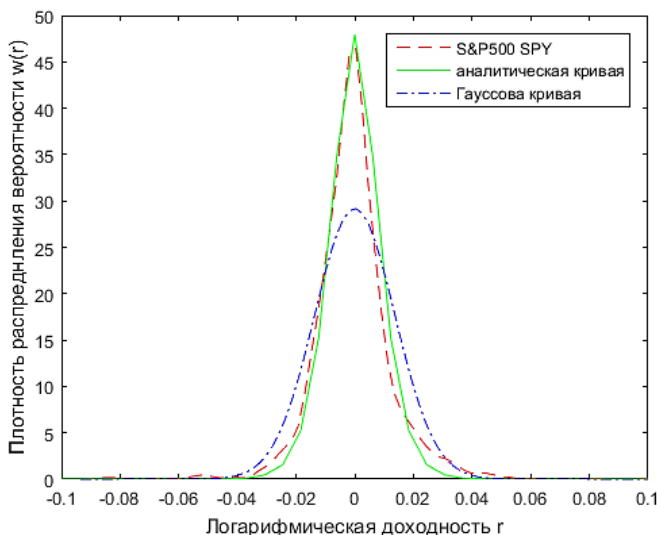


Рис. 1. Графики S&P 500 SPY, аналитической и Гауссовой кривой

Из рисунка 1 видно, что полученная в работе аналитическая кривая с использованием формулы (4.20) намного точнее, чем Гауссова аппроксимирует кривую, описывающую исторические данные.

На рисунке 2 также совмещены три графика.

Первые два графика те же, что и на рисунке 1. Третий график представляет собой график функции распределения логарифмической доходности Лапласа. Он выполнен с использованием формулы

$$P(x) = \frac{\alpha}{2} e^{-\alpha|x|}.$$

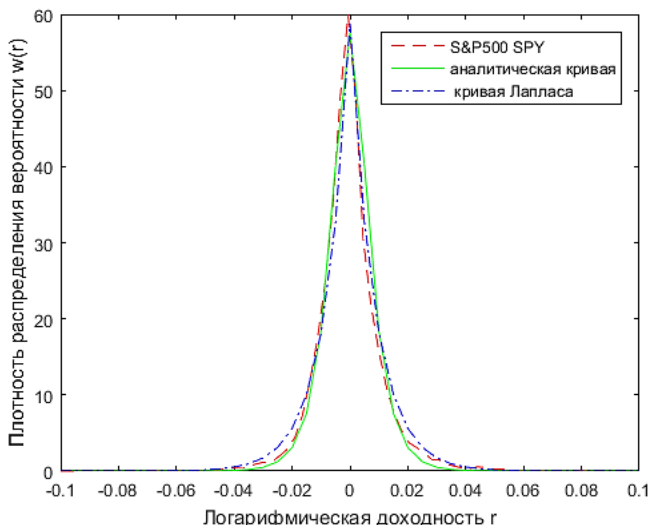


Рис. 2. График S&P 500 SPY, аналитической и кривой Лапласа

Из рисунка 2 видно, что все три графика хорошо согласуются, при этом аналитическая кривая точнее, чем кривая Лапласа аппроксимирует исторические данные. Визуальный анализ обоих рисунков позволяет сделать вывод о том, что по сравнению с нормальным распределением Гаусса, как распределение Лапласа, так и распределение, полученное в данной работе, дают более точную и реалистичную картину ежедневной доходности акций фондового рынка.

В настоящей работе формализм статистической механики применен для прогнозирования распределения логарифмической доходности акций с использованием исторических данных по котировкам одного из основных индикаторов и барометров мирового фондового рынка – индекса S&P 500 SPY, который торгуется как акции.

Однако использованный в работе подход может быть применен для анализа и прогнозирования распределения логарифмической доходности акций конкретных компаний.

В качестве примера выполнен визуальный анализ распределения логарифмической доходности обыкновенных акций ПАО «Сбербанк». График функции распределения логарифмической доходности представлен на рисунке 3.

Для построения графика использованы фактические исторические данные по ежедневным котировкам акций ПАО

«Сбербанк» за десять лет (за период с 05.06.2013 по 05.06.2023, источник данных Investing.com).

На этом же рисунке показан график функции распределения логарифмической доходности, построенный с использованием полученной в работе аналитической формулы (4.20) и график нормального распределения Гаусса.

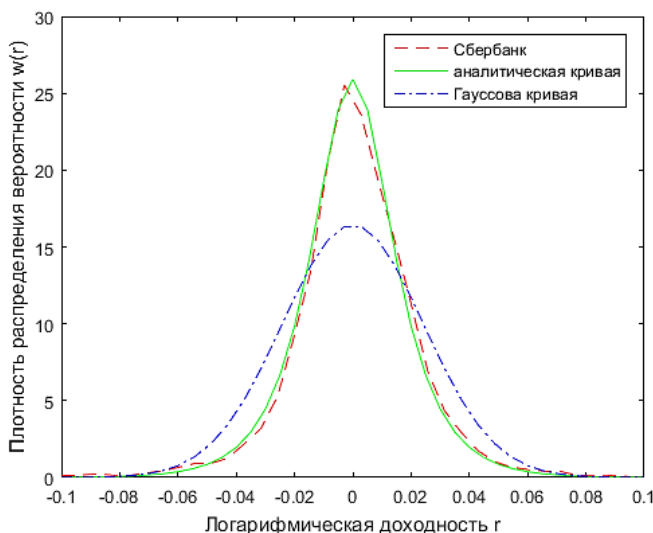


Рис. 3. График кривой Сбербанка, аналитической и Гауссовой кривой

Из рисунка 3 явно видно, что полученная в работе аналитическая кривая с использованием формулы (4.20) намного точнее, чем Гауссова аппроксимирует кривую, описывающую фактические исторические данные, что подтверждает эффективность предложенного метода прогнозирования.

Как известно, акции ПАО «Сбербанк» относятся к «голубым фишкам», что повышает их надежность и привлекательность в глазах инвесторов. Однако, для предложения инвесторам и аналитикам аргументированной дальновидной стратегии, необходим углубленный статистический анализ возможной доходности и риска инвестиций в акции этой компании, который выходит за рамки данного исследования.

Оценка ключевых факторов инвестирования (доходности и риска) только с помощью метода, предложенного в данной работе, равно как и с использованием распределения Гаусса и/или Лапласа, не

будет полностью объективной. Для принятия окончательного решения об инвестировании необходимо учитывать факторы, которые невозможно спрогнозировать с помощью этих инструментов (санкции, уровень инфляции, рост ВВП, стабильность рубля и т.д.).

Для определения средней процентной доходности за n дней необходимо перейти от логарифмической доходности $r(n)$ к однодневной процентной доходности $R(n)$ по формуле $R(n) = e^{r(n)} - 1$. Ожидаемая средняя процентная доходность $\bar{R}(n)$ определяется по формуле:

$$\bar{R}(n) = \int_{-\infty}^{\infty} W(r, n)(e^r - 1)dr = \int_{-\infty}^{\infty} \frac{A(e^r - 1)}{(2 \cosh(\frac{r}{n/\tau}))^{((\lambda M - 1)\frac{n}{\tau} + 1)}} dr. \quad (6.1)$$

Обезопасить инвестицию можно, определив по формуле (5.5) степень риска путем вычисления стандартного отклонения от средней доходности акций.

Но любом случае (даже если вероятность результатов инвестирования, определенная с помощью формул и графиков оказывается многообещающей, и/или инвестор, исходя из собственного опыта считает, что будущая доходность акций ПАО «Сбербанк» превысит исторические данные), в обязательном порядке необходимо учитывать степень неопределенности существующей на момент инвестиций экономической ситуации.

7. Заключение. Основные выводы по результатам исследования:

1. Результаты работы подтверждают возможность и целесообразность прогнозирования распределения доходности акций фондового рынка с помощью методов и подходов статистической механики. Полученное аналитическое выражение для функции распределения логарифмической доходности акций дает возможность участникам рынка использовать его для математического прогнозирования изменений доходности рынка. Полученные результаты дают участникам рынка возможность выбора распределения, более глубокого понимания причинно-следственного механизма фондового рынка (если рассматривать его как экосистему), а также позволяют использовать для финансового моделирования хорошо изученный математический аппарат статистической механики.

2. Несмотря на устоявшиеся традиции использования нормального распределения для оценки динамики фондового рынка

(включая удобство его использования, наличие соответствующего программного обеспечения и результативность на больших временных масштабах), для оценки ежедневной доходности акций и своевременного улавливания риска, лучшим выбором являются более точные распределения, включая распределение, полученное в данной работе. Как показывает практика, обвалы рынка происходят гораздо чаще, чем они улавливаются распределениями Гаусса.

3. Краткий сравнительный анализ данной работы с аналогичными работами по этой же тематике дал следующие результаты:

- в работе (27) выполнено моделирование доходности с использованием распределения Лапласа и его сравнение с нормальным распределением Гаусса. Подтверждена большая эффективность распределения Лапласа. В данной работе аналогичные преимущества обнаружены у предлагаемого распределения, близкого к распределению Лапласа;

- в работе (26) функция распределения логарифмической доходности рассматривается как распределение Больцмана и вычислена в терминах специальных функций. В настоящей работе она получена в терминах элементарных функций, что значительно упрощает анализ ее поведения;

- в работе (16) аналитическое выражение для функции распределения логарифмической доходности получено с использованием исчисления Ито. В настоящей работе для этого использован альтернативный подход, основанный на применении формализма статистической механики. Он позволяет рассмотреть проблему с новой стороны, и направлен на поиск не рассмотренных ранее глубинных факторов.

4. Перспективными и важными являются следующие направления дальнейших исследований в области анализа и прогнозирования поведения фондового рынка:

- исследования по совершенствованию прогностических возможностей, обеспечивающие максимальную прозрачность и лучшее понимание всех процессов фондового рынка;

- изучение особенностей хаотического поведения системы, включая выявление и использование в прогностических целях индикаторов раннего предупреждения обвалов рынка;

- работы по тщательному и всестороннему анализу накопленного практического опыта по выходу из кризисных ситуаций и обвалов рынка, результаты которых можно будет использовать для

выработки предложений по своевременной компенсации возможных травмирующих воздействий на рынок.

Приложение А.

Факторизация распределения Гиббса. В приложении А будет доказана факторизуемость полученного в разделе 3 распределения Гиббса. Совместная функция распределения вероятностей цен акций и облигаций (3.7) может быть представлена следующим образом:

$$P(p, q) = \frac{1}{Z} e^{-\beta(\gamma_2 e^p - \alpha_2 p + \gamma_1 e^q - \alpha_1 q)} = \frac{e^{-\beta H}}{Z}. \quad (\text{A.1})$$

Здесь β – обратная температура, Z – нормирующий множитель, называемый статистической суммой.

Покажем далее, что функция $P(p, q)$ факторизуется. Очевидно, что больцмановский фактор в формуле (A.1) факторизуется. Осталось показать, что статистическая сумма Z тоже факторизуется. Поскольку интеграл от совместной функции распределения вероятностей цен акций и облигаций равен единице, то из формулы (4.1) следует формула:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(p, q) dp dq = \frac{1}{Z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\beta H} dp dq = 1. \quad (\text{A.2})$$

С использованием формулы (A.2) статистическая сумма определяется следующим образом:

$$Z = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\beta(\gamma_2 e^p - \alpha_2 p + \gamma_1 e^q - \alpha_1 q)} dp dq. \quad (\text{A.3})$$

В силу аддитивности гамильтониана (2.17) статистическая сумма Z мультипликативна и распадается на произведение множителей Z_1 и Z_2 :

$$Z = Z_1 Z_2, \quad (\text{A.4})$$

$$Z_1 = \int_{-\infty}^{\infty} e^{-\beta \gamma_1 e^q + \beta \alpha_1 q} dq = \frac{\Gamma(\alpha_1 \beta)}{(\gamma_1 \beta)^{\alpha_1 \beta}}, \quad (\text{A.5})$$

$$Z_2 = \int_{-\infty}^{\infty} e^{-\beta \gamma_2 e^p + \beta \alpha_2 p} dp = \frac{\Gamma(\alpha_2 \beta)}{(\gamma_2 \beta)^{\alpha_2 \beta}}. \quad (\text{A.6})$$

Из формул (A.1), (A.4) следует факторизация функции $P(p, q)$:

$$P(p, q) = \left(\frac{e^{-\beta H_1(q)}}{Z_1} \right) \left(\frac{e^{-\beta H_2(p)}}{Z_2} \right) = P(q)P(p). \quad (\text{A.7})$$

Здесь $P(q)$ – вероятность распределения логарифма цены акций, $P(p)$ – вероятность распределения логарифма цены облигаций.

Приложение В.

Вычисление комбинации параметров $\alpha_1 \beta$, входящих в формулы (4.17)-(4.20) для функции распределения логарифмической доходности. В формулы для функции распределения логарифмической доходности (4.17)-(4.20) в виде комбинации $\alpha_1 \beta$ входит параметр β – обратная температура и параметр α_1 , (формулы (2.7), (2.8)). Для вычисления величины $\alpha_1 \beta$ будут вычислены первый и второй моменты. Перейдем к их вычислению. Вначале, выполним переход от статистической суммы $Z_1(\beta)$ (формула (A.4) к производящей функции $Z_1(\beta, J)$. Производящая функция зависит от параметра J , называемого в физике источником. Статистическая сумма $Z_1(\beta)$ представляет собой предельный случай производящей функции в пределе $J = 0$. Формула для производящей функции может быть представлена в общем виде:

$$Z_1(\beta, J) = \int e^{-\beta(\gamma_1 e^q - \alpha_1 q) + Jq} dq. \quad (\text{B.1})$$

Легко видеть, что при таком определении производящей функции моменты переменной y (2.7) будут иметь следующий вид:

$$\langle (e^q)^J \rangle = \langle y^J \rangle = \frac{Z_1(\beta, J)}{Z_1(\beta, 0)}. \quad (\text{B.2})$$

Для вычисления производящей функции в явном виде выполним следующую замену переменных:

$$t = \gamma_1 \beta e^q. \quad (\text{B.3})$$

После этой замены переменных формула (B.1) преобразуется в интеграл Эйлера:

$$Z_1(\beta, J) = \int_0^\infty \frac{t^{\alpha_1 \beta + J - 1} e^{-t}}{(\gamma_1 \beta)^{\alpha_1 \beta + J}} dt = \frac{\Gamma(\alpha_1 \beta + J)}{(\gamma_1 \beta)^{\alpha_1 \beta + J}}. \quad (\text{B.4})$$

Из формулы (B.4) очевидно, что удачная структура гамильтониана модели позволяет вычислить производящую функцию в явном виде в терминах гамма-функций. Рекуррентное соотношение $\Gamma(x+1) = x\Gamma(x)$ для гамма-функции позволяет легко находить моменты, используя формулу (B.4). Первый и второй моменты для переменной x будут иметь вид:

$$\langle y \rangle = \langle e^q \rangle = \frac{Z_1(\beta, 1)}{Z_1(\beta, 0)} = \frac{\alpha_1}{\gamma_1}, \quad (\text{B.5})$$

$$\langle y^2 \rangle = \langle e^{2q} \rangle = \frac{Z_1(\beta, 2)}{Z_1(\beta, 0)} = \frac{\alpha_1^2}{\gamma_1^2} + \frac{\alpha_1}{\gamma_1^2 \beta}. \quad (\text{B.6})$$

Комбинируя первый и второй моменты, можно вычислить среднее квадратичное отклонение для функции распределения цен акций:

$$\langle (y - \langle y \rangle)^2 \rangle = \langle y^2 \rangle - \langle y \rangle^2 = \frac{\alpha_1}{\gamma_1^2 \beta} = \left(\frac{\alpha_1}{\gamma_1} \right)^2 \frac{1}{(\alpha_1 \beta)}. \quad (\text{B.7})$$

Используя формулу (B.5) заменим выражение $\left(\frac{\alpha_1}{\gamma_1} \right)^2$ в формуле (B.7) на выражение $\langle y \rangle^2$. После произведенной замены формула (B.7) примет вид:

$$\langle (y - \langle y \rangle)^2 \rangle = \frac{\langle y \rangle^2}{\alpha_1 \beta}. \quad (\text{B.8})$$

С использованием формулы (В.8) получим окончательное выражение для комбинации параметров $\alpha_1\beta$:

$$\alpha_1\beta = \frac{\langle y \rangle^2}{\langle (y - \langle y \rangle)^2 \rangle}. \quad (\text{В.9})$$

На этом вычисление входящей в формулы (4.17)–(4.20) комбинации параметров $\alpha_1\beta$ завершено.

Приложение С.

Вычисление функции распределения логарифмической доходности для интервала $t = \tau$ в явном виде. В приложении вычисляется функция распределения логарифмической доходности для конкретного случая, когда интервал времени равен времени релаксации.

Эта функция играет важную роль при получении формулы (4.18). Функция распределения логарифмической доходности для интервала $t = \tau$ может быть вычислена в явном виде. В качестве отправной точки используем формулу (4.11):

$$W(r, \tau) = \int_{-\infty}^{\infty} P(q)P(q+r)dq. \quad (\text{С.1})$$

Здесь $W(r, \tau)$ – функция распределения логарифмической доходности для интервала $t = \tau$.

Для получения аналитического выражения этой функции подставим в формулу (С.1) распределение вероятностей логарифма цен $P(q)$ и $P(q+r)$ из формулы (А.7):

$$W(r, \tau) = \frac{1}{Z_1^2} \int_{-\infty}^{\infty} e^{-\beta\gamma_1 e^q (1+e^r)} e^{2\beta\alpha_1 q} e^{\beta\alpha_1 r} dq. \quad (\text{С.2})$$

Для осуществления интегрирования проведем перегруппировку слагаемых в показателях экспонент и замену дифференциала dq на $d(e^q)$. Получим следующий интеграл:

$$W(r, \tau) = \frac{e^{-\beta\alpha_1 r}}{Z_1^2} \int_0^{\infty} (e^q)^{2\beta\alpha_1 - 1} e^{-2\beta\gamma_1 e^q (1+e^r)} d(e^q). \quad (\text{С.3})$$

Удачный вид интеграла (С.3) позволяет преобразовать его к виду интеграла Эйлера. Как известно, это один из немногих интегралов, допускающих вычисление в аналитическом виде. Выполним далее следующую замену переменных:

$$t = \beta\gamma_1 e^{q(1+e^r)}. \quad (\text{C.4})$$

После проведения этой замены получим интеграл Эйлера:

$$W(r, \tau) = \frac{e^{-\beta\alpha_1 r}}{Z_1^2} \int_0^\infty \frac{t^{2\beta\alpha_1-1} e^{-t}}{(\beta\gamma_1(1+e^r))^{2\beta\alpha_1}} dt. \quad (\text{C.5})$$

Вычислим интеграл Эйлера в формуле (С.5) и получим окончательный результат для функции распределения логарифмической доходности акций:

$$W(r, \tau) = \frac{\Gamma(2\beta\alpha_1)}{\Gamma^2(\beta\alpha_1)} \frac{1}{(e^{-\frac{r}{2}} + e^{\frac{r}{2}})^{2\beta\alpha_1}}. \quad (\text{C.6})$$

Формула (С.6) представляет собой частный случай аналитического выражения функции распределения логарифмической доходности акций фондового рынка. Она получена в явном виде, выражена в элементарных функциях. Полученная достаточно простая структура формулы (С.6) используется при выводе формул (4.18)–(4.20).

Приложение D.

Вычисление аналитической формулы (4.18) методом перевала. Для целей визуального анализа результатов исследования преобразуем содержащую гамма-функции точную аналитическую формулу (4.17). Конечная цель преобразования – представить формулу в аналитическом виде в терминах элементарных функций, что даст возможность аналитикам использовать ее для визуального и наглядного анализа поведения системы. В качестве инструмента преобразования используется метод перевала.

Перейдем к преобразованию формулы (4.17). Используем для ее преобразования заданную формулой (С.6) функцию $W(r, \tau)$. На первом этапе преобразуем ее с использованием преобразования Фурье. Она будет следующий вид:

$$\begin{aligned}\tilde{W}(k, \tau) &= \frac{\Gamma(\beta\alpha_1 - ik)}{\Gamma(\beta\alpha_1)} \frac{\Gamma(\beta\alpha_1 + ik)}{\Gamma(\beta\alpha_1)} = \int_{-\infty}^{\infty} e^{ikx} \frac{C}{(e^{\frac{x}{2}} + e^{-\frac{x}{2}})^M} dx = \\ &= C \int_{-\infty}^{\infty} e^{-M\varphi(x, M, k)} dx\end{aligned}\quad (D.1)$$

Здесь:

C – нормировочная константа,

M – большой параметр, равный $2\beta\alpha_1$.

Далее вычислим определенную формулой (D.1) функцию $\tilde{W}(k, \tau)$ с использованием метода перевала. Входящая в формулу (D.1) функция φ , умноженная на большой параметр M в показателе экспоненты, а также ее первая, вторая производные и точка экстремума $x^*(k)$ имеют следующий вид:

$$\varphi(x, M, k) = \ln(2 \cosh(\frac{x}{2})) - i \frac{k}{M} x, \quad (D.2)$$

$$\varphi'(x^*(k)) = \frac{1}{2} \operatorname{th}(\frac{x^*}{2}) - \frac{ik}{M} = 0, \quad (D.3)$$

$$\frac{x^*(k)}{2} = \operatorname{Arth}(\frac{i2k}{M}), \quad (D.4)$$

$$\varphi'' = \frac{1}{2} (1 - \operatorname{th}^2(\frac{x}{2})) = \frac{1}{2} (1 + \frac{4k^2}{M^2}). \quad (D.5)$$

С учетом формул (D.2)-(D.5) получим окончательное выражение для функции $\tilde{W}(k, \tau)$:

$$\tilde{W}(k, \tau) \sim \sqrt{\frac{\pi}{\varphi''(x^*)}} e^{-M\varphi(x^*)} = (\sqrt{1 + \frac{4k^2}{M^2}})^{M-1} e^{ik2\operatorname{Arth}(\frac{i2k}{M})}. \quad (D.6)$$

Возведем полученное окончательное выражение (D.6) в степень $\frac{t}{\tau}$ и возьмем обратное преобразование Фурье от полученного выражения:

$$\begin{aligned} W(r, t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikr} \tilde{W}(k)^{t/\tau} dk \sim \\ &\sim \int_{-\infty}^{\infty} e^{-ikr + i(t/\tau)k 2 \operatorname{Arth}(\frac{i2k}{M}) - (M-1)(t/\tau) \ln(2/\sqrt{1+4k^2/M^2})} dk = \\ &= \int_{-\infty}^{\infty} e^{-(M-1)\psi(k, M, r)} dk \end{aligned} \quad (\text{D.7})$$

Преобразуем формулу (D.7) методом перевала. Входящая в формулу (D.7) функция ψ , а также ее первая, вторая производные и точка экстремума $k^*(r)$ имеют следующий вид:

$$\begin{aligned} \psi(k, M, r) &= \left(\frac{t}{\tau}\right) \ln\left(\frac{2}{\sqrt{1+4k^2/M^2}}\right) - \\ &- i\left(\frac{t}{\tau}\right) \frac{k}{M-1} 2 \operatorname{Arth}\left(\frac{i2k}{M}\right) + \frac{ik}{M-1} r \end{aligned} \quad (\text{D.8})$$

$$\psi'(k^*(r)) = \frac{i(t/\tau)}{M-1} 2 \operatorname{Arth}\left(\frac{i2k}{M}\right) + \frac{ir}{M-1} = 0, \quad (\text{D.9})$$

$$\frac{i2k^*}{M} = th\left(\frac{r}{2t/\tau}\right), \quad (\text{D.10})$$

$$\psi''(k^*) = \frac{4(t/\tau)}{M^2} \cosh^2\left(\frac{r}{2(t/\tau)}\right). \quad (\text{D.11})$$

С учетом формул (D.8)-(D.11) получим окончательное выражение для функции распределения логарифмической доходности, представленной в разделе 4 как формула (4.18):

$$W(r, t) \sim \frac{\sqrt{\pi}}{\sqrt{\psi''(k^*)}} e^{-(M-1)\psi(k^*)} = \frac{A}{(2 \cosh(\frac{r}{2(t/\tau)})^{((M-1)(t/\tau)+1)}}. \quad (\text{D.12})$$

Здесь A – нормировочный множитель. Вид нормировочного множителя можно получить, используя формулу (С.6). Поскольку

$$\int_{-\infty}^{\infty} W(r, \tau) = 1, \text{ верна формула:}$$

$$\int_{-\infty}^{\infty} \frac{1}{(2 \cosh(\frac{x}{2}))^M} dx = \frac{\Gamma^2(0.5M)}{\Gamma(M)}. \quad (\text{D.13})$$

Используя формулу (D.13) получим вид нормировочного множителя A :

$$A = \frac{\Gamma((M-1)(t/\tau)+1)}{\Gamma^2(0.5((M-1)(t/\tau)+1))} \frac{1}{(t/\tau)}. \quad (\text{D.14})$$

На этом вывод упрощенной формулы (4.18) завершен.

Литература

1. Bhowmik R., Wang S. Stock market volatility and return analysis: A systematic literature review // Entropy. 2020. vol. 22(5). no. 522. DOI: 10.3390/e22050522.
2. Shah D., Isah H., Zulkernine F. Stock market analysis: A review and taxonomy of prediction techniques // International Journal of Financial Studies. 2019. vol. 7(2). no. 26. DOI: 10.3390/ijfs7020026.
3. Mavruk T. Analysis of herding behavior in individual investor portfolios using machine learning algorithms // Research in International Business and Finance. 2022. vol. 62. no. 101740. DOI: 10.1016/j.ribaf.2022.101740.
4. Lee S.C., Eid Junio W. Portfolio construction and risk management: theory versus practice // RAUSP Management Journal. 2018. vol. 53. pp. 345–365. DOI: 10.1108/RAUSP-04-2018-009.
5. Ingber L. Hybrid classical-quantum computing: Applications to statistical mechanics of financial markets // E3S Web of Conferences. 2021. vol. 307. no. 04001. DOI: 10.1051/e3sconf/202130704001.
6. Anyiam K., Oluigbo I., Eze U., Ezech G. An empirical investigation of the impact of information technology on global capital markets operation // International Journal of Research and Development Organisation. 2015. vol. 1. no. 1.
7. Song Y.-G., Zhou Y.-L., Han R.-J. Neural networks for stock price prediction // arXiv preprint arXiv:1805.11317v1. 2018. DOI: 10.48550/arXiv.1805.11317.

8. Lin Y., Liu S., Yang H., Wu H., Jiang B. Improving stock trading decisions based on pattern recognition using machine learning technology // PLOS ONE. 2021. vol. 16. no. 8. DOI: 10.1371/journal.pone.0255558.
9. Nou A., Lapitskaya D., Eratalay M.H., Sharma R. Predicting stock return and volatility with machine learning and econometric models: A comparative case study of the Baltic stock market // Tartu: The University of Tartu FEBA, 2021. 52 p. DOI: 10.2139/ssrn.3974770.
10. Maqsood A., Safdar S., Shafi R., Lelit N.J. Modeling stock market volatility using GARCH models: a case study of Nairobi securities exchange (NSE) // Open Journal of Statistics. 2017. vol. 7. no. 2. pp. 369–381. DOI: 10.4236/ojs.2017.72026.
11. Nazlioglu S., Kucukkaptan I., Kilic E., Altuntas M. Financial market integration of emerging markets: Heavy tails, structural shifts, nonlinearity, and asymmetric persistence // Research in International Business and Finance. 2022. vol. 62. no. 101742. DOI: 10.1016/j.ribaf.2022.101742.
12. Camilleri S.J., Vassallo S., Bai Y. Predictability in securities price formation: differences between developed and emerging markets // Journal of Capital Markets Studies. 2020. vol. 4. no. 2. pp. 145–166. DOI: 10.1108/JCMS-07-2020-0025.
13. Mallikarjuna M., Rao R.P. Evaluation of forecasting methods from selected stock market returns // Financial Innovation. 2019. vol. 5(1). no. 40. DOI: 10.1186/s40854-019-0157-x.
14. Al-Thaqeb S.A. Do international markets overreact? Event study: international market reaction to US local news events // Research in International Business and Finance. 2018. vol. 44. pp. 369–385. DOI: 10.1016/j.ribaf.2017.07.106.
15. Shen X., Wang G., Wang Y. The influence of research reports on stock returns: The mediating effect of machine-learning-based investor sentiment // Discrete Dynamics in Nature and Society. 2021. vol. 2021. no. 5049179. DOI: 10.1155/2021/5049179.
16. Liu Z., Dashti Moghaddam M., Serota R.A. Distributions of historic market data – stock returns // The European Physical Journal B. 2019. vol. 92. no. 60. DOI: 10.1140/epjb/e2019-90218-8.
17. Ozdemir L. Volatility spillover stock prices and trading volume: Evidence from the pre-, in-, and post global financial crisis period // Frontiers in Applied Mathematics and Statistics. 2020. vol. 5. no. 65. DOI: 10.3389/fams.2019.00065.
18. Cheteni P. Stock market volatility using GARCH models: Evidence from South Africa and China stock markets // Journal of Economics and Behavioral Studies. 2016. vol. 8. no. 6. pp. 237–245. DOI: 10.22610/jebis.v8i6(J).1497.
19. Pham D. P. T., Huynh N. Q. A., Duong D. The impact of US presidents on market returns: Evidence from Trump's tweets // Research in International Business and Finance. 2022. vol. 62. no. 101681. DOI: 10.1016/j.ribaf.2022.101681.
20. Olbrys J., Majewska E. Asymmetry effects in volatility on major European stock markets: the EGARCH based approach // Quantitative Finance and Economics. 2017. vol. 1. no. 4. pp. 411–427. DOI: 10.3934/QFE.2017.4.411.
21. Kouser R., Saba I., Anjum F. Impact of asymmetric information on the investment sensitivity to stock price and the stock price sensitivity to investment // Journal of Accounting and Finance in Emerging Economies. 2016. vol. 2. no. 1. pp. 1–16. DOI: 10.26710/jafee.v2i1.101.
22. Dhesi G., Shakeel B., Ausloos M. Modeling and forecasting the kurtosis and returns distributions of financial markets: irrational fractional Brownian motion model approach // Annals of Operations Research. 2021. vol. 299. pp. 1397–1410. DOI: 10.1007/s10479-019-03305-z.
23. Chevapatrakul T., Xu Z., Yao K. The impact of tail risk on stock market returns: The role of market sentiment // International Review of Economics and Finance. 2019. vol. 59. pp. 289–301. DOI: 10.1016/j.iref.2018.09.005.

24. Pareto V. Manual of political economy. Translated by: Schiavier A.S., Page A.N. // New-York: A.M. Kelley, 1971. 504 p.
25. Levy M. Market efficiency, the Pareto wealth distribution, and the Levy distribution of stock returns // The economy as an evolving complex system, III: Current perspectives and future directions. 2005. pp. 101–132. DOI: 10.1093/acprof:oso/9780195162592.003.0006.
26. Kleinert H., Chen X.J. Boltzmann Distribution and Market Temperature. *Physica A: Statistical Mechanics and its Applications*. 2007. vol. 383. no. 2. pp. 513–518
27. Toth D., Jones B. Against the norm: modeling daily stock returns with the Laplace distribution. arXiv preprint arXiv:1906.10325. 2019. Available at: https://www.researchgate.net/publication/334027873_Against_the_Norm_Modeling_Daily_Stock_Returns_with_the_Laplace_Distribution.
28. Volterra V. Fluctuations in the abundance of a species considered mathematically // *Nature*. 1926. vol. 118. pp. 558–560. DOI: 10.1038/118558a0.
29. Volterra V. Variazioni e fluttuazioni dei numero d'individui in specie animali conviventi // Società anonima tipografica "Leonardo da Vinci", 1927. 142 p.
30. Volterra V. Lessons on the mathematical theory of struggle for life. Paris, Gauthier-Villars, 1931. 214 p.
31. Lotka A.J. Elements of physical biology // Baltimore: Williams and Wilkins, 1925. 495 p.
32. Modis T. Technological forecasting at the Stock Market // *Technological Forecasting and Social Change*. 1999. vol. 62. no. 3. pp. 173–202. DOI: 10.1016/S0040-1625(99)00046-3.
33. Modis T. An S-shaped trail to Wall Street: survival of the fittest reigns at the Stock Market // Geneva, Growth Dynamics. 1999. 201 p.
34. Clement E.P., Jim U.S. Statistical mechanics in economics: An application of Brownian motion in modeling prices of assets // *International Journal of Statistics and Applied Mathematics*. 2021. vol. 6. no. 1. pp. 29–34.
35. Yakovenko Victor M. and Rosser J. Barkley Jr. Colloquium: Statistical mechanics of money, wealth, and income // *Reviews of Modern Physics*. 2009. vol. 81. no. 1703. DOI: 10.1103/RevModPhys.81.1703.
36. Chakraborti A., Chakrabarti B.K. Statistical mechanics of money: How saving propensity affects its distribution // *The European Physical Journal B-Condensed Matter and Complex Systems*. 2000. vol. 17. pp. 167–170. DOI: 10.1007/s100510070173.
37. Gatabazi P., Mba J.C., Pindza E. Fractional Gray Lotka-Volterra models with application to cryptocurrencies adoption // *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2019. vol. 29(7). no. 073116. DOI: 10.1063/1.5096836.
38. Zhang G., McAdams D.A., Shankar V., Mohammadi Darani M. Technology evolution prediction using Lotka-Volterra equations // *Journal of Mechanical Design*. 2018. vol. 140(6). no. 061101. DOI: 10.1115/1.4039448.
39. Титов В.А., Вейнберг Р.Р. Анализ существующих динамических моделей на базе системы уравнений Лотки-Вольтерры «хищник-жертва» // *Фундаментальные исследования*. 2016. № 8–2. С. 409–413.
40. Романов В.П., Ахмадеев Б.А. Моделирование инновационной экосистемы на основе модели «хищник-жертва» // *Бизнес-информатика*. 2015. № 1(31). С. 7–17.
41. Hung H.C., Tsai Y.S., Wu M.C. A modified Lotka-Volterra model for competition forecasting in Taiwan's retail industry // *Computers and Industrial Engineering*. 2014. vol. 77. pp. 70–79. DOI: 10.1016/j.cie.2014.09.010.
42. Sterpu M., Rocsoreanu C., Soava G., and Mehedintu A. A generalization of the Grey Lotka–Volterra model and application to GDP, export, import and investment for the

- European Union // Mathematics. 2023. vol. 11(15). no. 3351. DOI: 10.3390/math11153351.
43. Pan M.X., Wang S.Y., Wu X.L., Zhang M.W. Study on the growth driving model of the enterprise innovation community based on the Lotka–Volterra model: a case study of the Chinese Automobile Manufacturing Enterprise Community // Mathematical Problems in Engineering. 2023. vol. 2023. no. 8743167. DOI: 10.1155/2023/8743167.
 44. Bauer R., Schwarzmayr F., Brunner N., Kühleitner M. Dynamics of the Austrian food market: application of Lotka–Volterra differential equations // Open Journal of Modelling and Simulation. 2022. vol. 10. no. 2. pp. 152–164. DOI: 10.4236/ojmsi.2022.102009.

Бибик Юрий Викторович — канд. физ.-мат. наук, старший научный сотрудник, отдел №23 (механика сплошных сред), Федеральный исследовательский центр «Информатика и управление» Российской академии наук. Область научных интересов: статистическая механика, нелинейная динамика, Гамильтоновы системы, стохастические процессы, фазовые переходы, квантовая теория поля, применение математических моделей и методов для изучения и выявления кризисных ситуаций в природе, экологии, биологии, экономике. Число научных публикаций — 22. yvbibik@ccas.ru; улица Вавилова, 40, 119333, Москва, Россия; р.т.: +7(499)135-6498.

YU. BIBIK

FORECASTING IN STOCK MARKETS USING THE FORMALISM OF STATISTICAL MECHANICS*Bibik Yu. Forecasting in Stock Markets Using the Formalism of Statistical Mechanics.*

Abstract. The possibility and expediency of forecasting in the stock markets are analyzed analytically using the methods and approaches of statistical mechanics. The apparatus of statistical mechanics is used to analyze and forecast one of the most important indicators of the market – the distribution of its logarithmic profitability. The Lotka-Volterra model used in ecology to describe systems of the "predator-prey" type was used as the initial model. It approximates market dynamics adequately. In the article, its Hamiltonian property is used, which makes it possible to apply the apparatus of statistical mechanics. The apparatus of statistical mechanics (using the principle of maximum entropy) makes it possible to implement a probabilistic approach that is adapted to the conditions of stock market uncertainty. The canonical variables of the Hamiltonian are presented as logarithms of stock and bond prices, the joint probability distribution function of stock and bond prices is obtained as a Gibbs distribution. The Boltzmann factor, included in the Gibbs distribution, allows us to estimate the probability of the occurrence of certain stock and bond prices and obtain an analytical expression for calculating the logarithmic return, which gives more accurate results than the widely used normal (Gaussian) distribution. According to its characteristics, the resulting distribution resembles the Laplace distribution. The main characteristics of the resulting distribution are calculated – the mean value, variance, asymmetry, and kurtosis. Mathematical results are presented graphically. An explanation is given of the cause-and-effect mechanism that causes a change in the profitability of the market. For this, the idea of Theodore Modis about the competition between stocks and bonds for the attention and money of investors is developed (by analogy with the turnover of biomass in models of the "predator-prey" type in biology). The results of the study are of interest to investors, theorists, and practitioners of the stock market. They allow us to make thoughtful and balanced investment decisions due to a more realistic idea of the expected return and a more adequate assessment of investment risk.

Keywords: stock market dynamics, return distribution function, maximum entropy principle, Gibbs distribution, Laplace distribution.

References

1. Bhowmik R., Wang S. Stock market volatility and return analysis: A systematic literature review. *Entropy*. 2020. vol. 22(5). no. 522. DOI: 10.3390/e22050522.
2. Shah D., Isah H., Zulkernine F. Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies*. 2019. vol. 7(2). no. 26. DOI: 10.3390/ijfs7020026.
3. Mavruk T. Analysis of herding behavior in individual investor portfolios using machine learning algorithms. *Research in International Business and Finance*. 2022. vol. 62. no. 101740. DOI: 10.1016/j.ribaf.2022.101740.
4. Lee S.C., Eid Junio W. Portfolio construction and risk management: theory versus practice. *RAUSP Management Journal*. 2018. vol. 53. pp. 345–365. DOI: 10.1108/RAUSP-04-2018-009.
5. Ingber L. Hybrid classical-quantum computing: Applications to statistical mechanics of financial markets. *E3S Web of Conferences*. 2021. vol. 307. no. 04001. DOI: 10.1051/e3sconf/202130704001.

6. Anyiam K., Oluigbo I., Eze U., Ezech G. An empirical investigation of the impact of information technology on global capital markets operation. *International Journal of Research and Development Organisation*. 2015. vol. 1. no. 1.
7. Song Y.-G., Zhou Y.-L., Han R.-J. Neural networks for stock price prediction. *arXiv preprint arXiv:1805.11317v1*. 2018. DOI: 10.48550/arXiv.1805.11317.
8. Lin Y., Liu S., Yang H., Wu H., Jiang B. Improving stock trading decisions based on pattern recognition using machine learning technology. *PLOS ONE*. 2021. vol. 16. no. 8. DOI: 10.1371/journal.pone.0255558.
9. Nou A., Lapitskaya D., Eratalay M.H., Sharma R. Predicting stock return and volatility with machine learning and econometric models: A comparative case study of the Baltic stock market. *Tartu: The University of Tartu FEBA*, 2021. 52 p. DOI: 10.2139/ssrn.3974770.
10. Maqsood A., Safdar S., Shafi R., Lelit N.J. Modeling stock market volatility using GARCH models: a case study of Nairobi securities exchange (NSE). *Open Journal of Statistics*. 2017. vol. 7. no. 2. pp. 369–381. DOI: 10.4236/ojs.2017.72026.
11. Nazlioglu S., Kucukkaptan I., Kilic E., Altuntas M. Financial market integration of emerging markets: Heavy tails, structural shifts, nonlinearity, and asymmetric persistence. *Research in International Business and Finance*. 2022. vol. 62. no. 101742. DOI: 10.1016/j.ribaf.2022.101742.
12. Camilleri S.J., Vassallo S., Bai Y. Predictability in securities price formation: differences between developed and emerging markets. *Journal of Capital Markets Studies*. 2020. vol. 4. no. 2. pp. 145–166. DOI: 10.1108/JCMS-07-2020-0025.
13. Mallikarjuna M., Rao R.P. Evaluation of forecasting methods from selected stock market returns. *Financial Innovation*. 2019. vol. 5(1). no. 40. DOI: 10.1186/s40854-019-0157-x.
14. Al-Thaqeb S.A. Do international markets overreact? Event study: international market reaction to US local news events. *Research in International Business and Finance*. 2018. vol. 44. pp. 369–385. DOI: 10.1016/j.ribaf.2017.07.106.
15. Shen X., Wang G., Wang Y. The influence of research reports on stock returns: The mediating effect of machine-learning-based investor sentiment. *Discrete Dynamics in Nature and Society*. 2021. vol. 2021. no. 5049179. DOI: 10.1155/2021/5049179.
16. Liu Z., Dashti Moghaddam M., Serota R.A. Distributions of historic market data – stock returns. *The European Physical Journal B*. 2019. vol. 92. no. 60. DOI: 10.1140/epjb/e2019-90218-8.
17. Ozdemir L. Volatility spillover stock prices and trading volume: Evidence from the pre-, in-, and post global financial crisis period. *Frontiers in Applied Mathematics and Statistics*. 2020. vol. 5. no. 65. DOI: 10.3389/fams.2019.00065.
18. Cheteni P. Stock market volatility using GARCH models: Evidence from South Africa and China stock markets. *Journal of Economics and Behavioral Studies*. 2016. vol. 8. no. 6. pp. 237–245. DOI: 10.22610/jebbs.v8i6(J).1497.
19. Pham D. P. T., Huynh N. Q. A., Duong D. The impact of US presidents on market returns: Evidence from Trump's tweets. *Research in International Business and Finance*. 2022. vol. 62. no. 101681. DOI: 10.1016/j.ribaf.2022.101681.
20. Olbrys J., Majewska E. Asymmetry effects in volatility on major European stock markets: the EGARCH based approach. *Quantitative Finance and Economics*. 2017. vol. 1. no. 4. pp. 411–427. DOI: 10.3934/QFE.2017.4.411.
21. Kouser R., Saba I., Anjum F. Impact of asymmetric information on the investment sensitivity to stock price and the stock price sensitivity to investment. *Journal of Accounting and Finance in Emerging Economies*. 2016. vol. 2. no. 1. pp. 1–16. DOI: 10.26710/jafec.v2i1.101.
22. Dhesi G., Shakeel B., Ausloos M. Modeling and forecasting the kurtosis and returns distributions of financial markets: irrational fractional Brownian motion model

- approach. *Annals of Operations Research*. 2021. vol. 299. pp. 1397–1410. DOI: 10.1007/s10479-019-03305-z.
23. Chevapatrakul T., Xu Z., Yao K. The impact of tail risk on stock market returns: The role of market sentiment. *International Review of Economics and Finance*. 2019. vol. 59. pp. 289–301. DOI: 10.1016/j.iref.2018.09.005.
 24. Pareto V. *Manual of political economy*. Translated by: Schiwer A.S., Page A.N. New-York: A.M. Kelley, 1971. 504 p.
 25. Levy M. Market efficiency, the Pareto wealth distribution, and the Levy distribution of stock returns. *The economy as an evolving complex system, III: Current perspectives and future directions*. 2005. pp. 101–132. DOI: 10.1093/acprof:oso/9780195162592.003.0006.
 26. Kleinert H., Chen X.J. Boltzmann Distribution and Market Temperature. *Physica A: Statistical Mechanics and its Applications*. 2007. vol. 383. no. 2. pp. 513–518
 27. Toth D., Jones B. Against the norm: modeling daily stock returns with the Laplace distribution. *arXiv preprint arXiv:1906.10325*. 2019. Available at: https://www.researchgate.net/publication/334027873_Against_the_Norm_Modeling_Daily_Stock_Returns_with_the_Laplace_Distribution.
 28. Volterra V. Fluctuations in the abundance of a species considered mathematically. *Nature*. 1926. vol. 118. pp. 558–560. DOI: 10.1038/118558a0.
 29. Volterra V. Variazioni e fluttuazioni dei numero d'individui in specie animali conviventi // Società anonima tipografica "Leonardo da Vinci", 1927. 142 p.
 30. Volterra V. *Lessons on the mathematical theory of struggle for life*. Paris, Gauthier-Villars, 1931. 214 p.
 31. Lotka A.J. *Elements of physical biology*. Baltimore: Williams and Wilkins, 1925. 495 p.
 32. Modis T. Technological forecasting at the Stock Market. *Technological Forecasting and Social Change*. 1999. vol. 62. no. 3. pp. 173–202. DOI: 10.1016/S0040-1625(99)00046-3.
 33. Modis T. An S-shaped trail to Wall Street: survival of the fittest reigns at the Stock Market. Geneva, Growth Dynamics. 1999. 201 p.
 34. Clement E.P., Jim U.S. Statistical mechanics in economics: An application of Brownian motion in modeling prices of assets. *International Journal of Statistics and Applied Mathematics*. 2021. vol. 6. no. 1. pp. 29–34.
 35. Yakovenko Victor M. and Rosser J. Barkley Jr. Colloquium: Statistical mechanics of money, wealth, and income. *Reviews of Modern Physics*. 2009. vol. 81. no. 1703. DOI: 10.1103/RevModPhys.81.1703.
 36. Chakraborti A., Chakrabarti B.K. Statistical mechanics of money: How saving propensity affects its distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*. 2000. vol. 17. pp. 167–170. DOI: 10.1007/s100510070173.
 37. Gatabazi P., Mba J.C., Pindza E. Fractional Gray Lotka-Volterra models with application to cryptocurrencies adoption. *Chaos: An Interdisciplinary Journal of Nonlinear Science*. 2019. vol. 29(7). no. 073116. DOI: 10.1063/1.5096836.
 38. Zhang G., McAdams D.A., Shankar V., Mohammadi Darani M. Technology evolution prediction using Lotka-Volterra equations. *Journal of Mechanical Design*. 2018. vol. 140(6). no. 061101. DOI: 10.1115/1.4039448.
 39. Titov V.A., Vejnborg R.R. [Analysis of existing dynamic models based on the system of Lotka-Volterra equations “predator-prey”]. *Fundamental'nye issledovaniya – Basic Research*. 2016. no. 8–2. pp. 409–413. (In Russ.).
 40. Romanov V.P., Ahmadeev B.A. [Innovation ecosystem modeling based on ‘predator-prey’ model]. *Biznes-informatika – Business Informatics*. 2015. no. 1(31). pp. 7–17. (In Russ.).

41. Hung H.C., Tsai Y.S., Wu M.C. A modified Lotka-Volterra model for competition forecasting in Taiwan's retail industry. *Computers and Industrial Engineering*. 2014. vol. 77. pp. 70–79. DOI: 10.1016/j.cie.2014.09.010.
42. Sterpu M., Rocsoreanu C., Soava G., and Mehedintu A. A generalization of the Grey Lotka–Volterra model and application to GDP, export, import and investment for the European Union. *Mathematics*. 2023. vol. 11(15). no. 3351. DOI: 10.3390/math11153351.
43. Pan M.X., Wang S.Y., Wu X.L., Zhang M.W. Study on the growth driving model of the enterprise innovation community based on the Lotka–Volterra model: a case study of the Chinese Automobile Manufacturing Enterprise Community. *Mathematical Problems in Engineering*. 2023. vol. 2023. no. 8743167. DOI: 10.1155/2023/8743167.
44. Bauer R., Schwarzmayr F., Brunner N., Kühleitner M. Dynamics of the Austrian food market: application of Lotka-Volterra differential equations. *Open Journal of Modelling and Simulation*. 2022. vol. 10. no. 2. pp. 152–164. DOI: 10.4236/ojmsi.2022.102009.

Bibik Yuriy — Ph.D., Senior researcher, Department no. 23 (continuum mechanics), Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences. Research interests: statistical mechanics, nonlinear dynamics, Hamiltonian systems, stochastic processes, phase transitions, quantum field theory, application of mathematical models and methods for studying and identifying crisis situations in nature, ecology, biology, economics. The number of publications — 22. yvbibik@ccas.ru; 40, Vavilov St., 119333, Moscow, Russia; office phone: +7(499)135-6498.

Руководство для авторов

Взаимодействие автора с редакцией осуществляется через личный кабинет на сайте журнала «Информатика и автоматизация» <http://ia.spcras.ru/>. При регистрации авторам рекомендуется заполнить все предложенные поля данных. Подготовка статьи ведется с помощью текстовых редакторов MS Word 2007 и выше или LaTeX. Объем основного текста (до раздела Литература) - от 20 до 30 страниц включительно. Переносы запрещены. Номера страниц не проставляются. Основная часть текста статьи разбивается на разделы, среди которых являются обязательными: введение, хотя бы один «содержательный» раздел и заключение. Допускается также мотивированное содержанием и структурой материала выделение подразделов. В основную часть опускается помещать рисунки, таблицы, листинги и формулы. Правила их оформления подробно рассмотрены на нашем сайте в разделе «Руководство для авторов».

Author guidelines

Interaction between each potential author and the Editorial board is realized through the personal account on the website of the journal "Informatics and Automation" <http://ia.spcras.ru/>. At the registration the authors are requested to fill out all data fields in the proposed form. The submissions should be prepared using MS Word 2007, LaTeX. The text of the paper in the main part should not exceed 30 pages. Pages are not numbered; hyphenations are not allowed. Certain figures, tables, listings and formulas are allowed in the main section, and their typography is considered in more detail at the journal web.

Signed to print 07.11.2023. Passed for print 01.12.2023.

Printed in Publishing center GUAP.

Address: 67 litera A, B. Morskaya, St. Petersburg, 190000, Russia

Founder and Publisher: SPC RAS.

Address: 39 litera A, 14th Line V.O., St. Petersburg, 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 07.11.2023. Дата выхода в свет 01.12.2023.

Формат 60×90 1/16. Усл. печ. л. 17,7. Заказ № 461. 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»