p-ISSN 1607-3274
e-ISSN 2313-688X

Радіоелектроніка
Інформатика
Управління

Radio Electronics
Computer Science
Control

Радиоэлектроника
Информатика
Управление

9 771607 327005 34

2023/4

Національний університет «Запорізька політехніка»

Радіоелектроніка, інформатика, управління

Науковий журнал
Виходить чотири рази на рік
№ 4(67) 2023
Заснований у 1998 році, видається з 1999 року.
Засновник і видавець – Національний університет «Запорізька політехніка».
ISSN 1607-3274 (друкований), ISSN 2313-688X (електронний).
Запоріжжя
НУ «Запорізька політехніка»
2023

National University «Zaporizhzhia Polytechnic»

Radio Electronics, Computer Science, Control

The scientific journal
Published four times per year
№ 4(67) 2023
Founded in 1998, published since 1999.
Founder and publisher – National University «Zaporizhzhia Polytechnic».
ISSN 1607-3274 (print), ISSN 2313-688X (on-line).
Zaporizhzhia
NU «Zaporizhzhia Polytechnic»
2023

Национальный университет «Запорожская политехника»

Радиоэлектроника, информатика, управление

Научный журнал
Выходит четыре раза в год
№ 4(67) 2023
Основан в 1998 году, издается с 1999 года.
Основатель и издатель – Национальный университет «Запорожская политехника».
ISSN 1607-3274 (печатный), ISSN 2313-688X (электронный).
Запорожье
НУ «Запорожская политехника»
2023

Науковий журнал «Радіоелектроніка, інформатика, управління» (скорочена назва – РІУ) видається Національним університетом «Запорізька політехніка» (НУ «Запорізька політехніка») з 1999 р. періодичністю чотири номери на рік.

Зареєстровано у Міністерстві юстиції України 19.11.2019 р. (Свідоцтво про державну реєстрацію друкованого засобу масової інформації серія КВ № 24220-14060 ПР.)

ISSN 1607-3274 (друкований), **ISSN** 2313-688X (електронний).

Наказом Міністерства освіти і науки України № 409 від 17.03.2020 р. «Про затвердження рішень Атестаційної колегії Міністерства щодо діяльності спеціалізованих вчених рад від 06 березня 2020 року» **журнал включений до переліку наукових фахових видань України в категої «А» (найвищій рівень),** в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук і доктора філософії (кандидата наук).

Журнал включений до польського Переліку наукових журналів та рецензованих матеріалів міжнародних конференцій з присвоєною кількістю балів (додаток до оголошення Міністра науки та вищої освіти Республіки Польща від 31 липня 2019 р.: № 16981).

В журналі безкоштовно публікуються наукові статті англійською, російською та українською мовами.

Правила оформлення статей подано на сайті: http://ric.zntu.edu.ua/information/authors.

Журнал забезпечує **безкоштовний відкритий он-лайн доступ** до повнотекстових публікацій.

Журнал дозволяє авторам мати авторські права і зберігати права на видання без обмежень. Журнал дозволяє користувачам читати, завантажувати, копіювати, поширювати, друкувати, шукати або посилатися на повні тексти своїх статей. Журнал дозволяє повторне використання його вмісту у відповідності Creative Commons ліцензією CC BY-SA..

Опублікованим статтям присвоюється унікальний ідентифікатор цифрового об'єкта DOI.

Журнал входить до наукометричної бази Web of Science.

Журнал реферується та індексується у провідних міжнародних та національних реферативних журналах і наукометричних базах даних, а також розміщується у цифрових архівах та бібліотеках з безкоштовним доступом у режимі on-line, повний перелік яких подано на сайті: http://ric.zntu.edu.ua/about/editorialPolicies#custom-0.

Журнал розповсюджується за Каталогом періодичних видань України (передплатний індекс – 22914).

Тематика журналу: телекомунікації та радіоелектроніка, програмна інженерія (включаючи теорію алгоритмів і програмування), комп'ютерні науки (математичне і комп'ютерне моделювання, оптимізація і дослідження операцій, управління в технічних системах, міжмашинна і людино-машинна взаємодія, штучний інтелект, включаючи системи, засновані на знаннях, і експертні системи, інтелектуальний аналіз даних, розпізнавання образів, штучні нейронні і нейро-нечіткі мережі, нечітку логіку, колективний інтелект і мультиагентні системи, гібридні системи), комп'ютерна інженерія (апаратне забезпечення обчислювальної техніки, комп'ютерні мережі), інформаційні системи та технології (структури та бази даних, системи, засновані на знаннях та експертні системи, обробка даних і сигналів).

Усі статті, пропоновані до публікації, одержують **об'єктивний розгляд**, що оцінюється за суттю без урахування раси, статі, віросповідання, етнічного походження, громадянства або політичної філософії автора(ів).

Усі статті проходять двоступінчасте закрите (анонімне для автора) **резензування** штатними редакторами і незалежними рецензентами – провідними вченими за профілем журналу.

РЕДАКЦІЙНА КОЛЕГІЯ

РЕДАКЦІЙНО-КОНСУЛЬТАТИВНА РАДА

Рекомендовано до видання Вченою радою НУ «Запорізька політехніка», протокол № 5 від 19.12.2023.

Журнал зверстаний редакційно-видавничим відділом НУ «Запорізька політехніка».

Веб-сайт журналу: http://ric.zntu.edu.ua.

Адреса редакції: Редакція журналу «РІУ», Національний університет «Запорізька політехніка», вул. Жуковського, 64, м. Запоріжжя, 69063, Україна.

Тел: (061) 769-82-96 – редакційно-видавничий відділ Факс: +38-061-764-46-62

E-mail: rvv@zntu.edu.ua

The scientific journal «Radio Electronics, Computer Science, Control» is published by the National University «Zaporizhzhia Polytechnic» NU «Zaporizhzhia Polytechnic» since 1999 with periodicity four numbers per year.

The journal is registered by the Ministry of Justice of Ukraine in 19.11.2019. (State Registration Certificate of printed mass media series КВ № 24220-14060 ПР).

ISSN 1607-3274 (print), **ISSN** 2313-688X (on-line).

By the Order of the Ministry of Education and Science of Ukraine from 17.03.2020 № 409 "On approval of the decision of the Certifying Collegium of the Ministry on the activities of the specialized scientific councils dated 06 March 2020" **journal is included in the list of scientific specialized periodicals of Ukraine in category "A" (highest level),** where the results of dissertations for Doctor of Science and Doctor of Philosophy may be published.

The journal is included to tne Polish List of scientific journals and peer-reviewed materials from international conferences with assigned number of points (Annex to the announcement of the Minister of Science and Higher Education of Poland from July 31, 2019: Lp. 16981).

The journal publishes scientific articles in English, Russian, and Ukrainian free of charge.

The **article formatting rules** are presented on the site: http://ric.zntu.edu.ua/information/authors.

The journal provides policy of **on-line open (free of charge) access** for full-text publications. The journal allow the authors to hold the copyright without restrictions and to retain publishing rights without restrictions. The journal allow readers to read, download, copy, distribute, print, search, or link to the full texts of its articles. The journal allow reuse and remixing of its content, in accordance with Creative Commons license CC BY-SA.

Published articles have a unique digital object identifier (DOI).

The journal is included into Web of Science.

The journal is abstracted and indexed in leading international and national abstractig journals and scientometric databases, and also placed to the digital archives and libraries with a free on-line access, full list of which is presented at the site: http://ric.zntu.edu.ua/about/editorialPolicies#custom-0.

The journal is distributed:by the Catalogue of Ukrainian periodicals (the catalog number is 22914).

The journal scope: telecommunications and radio electronics, software engineering (including algorithm and programming theory), computer science (mathematical modeling and computer simulation, optimization and operations research, control in technical systems, machine-machine and man-machine interfacing, artificial intelligence, including data mining, pattern recognition, artificial neural and neuro-fuzzy networks, fuzzy logic, swarm intelligence and multiagent systems, hybrid systems), computer engineering (computer hardware, computer networks), information systems and technologies (data structures and bases, knowledge-based and expert systems, data and signal processing methods).

All articles proposed for publication receive an **objective review** that evaluates substantially without regard to race, sex, religion, ethnic origin, nationality, or political philosophy of the author(s).

All articles undergo a two-stage **blind peer review** by the editorial staff and independent reviewers – the leading scientists on the profile of the journal.

EDITORIAL BOARD

EDITORIAL-ADVISORY COUNCIL

Recommended for publication by the Academic Council of NU «Zaporizhzhia Polytechnic», protocol № 5 dated 19.12.2023.

The journal is imposed by the editorial-publishing department of NU «Zaporizhzhia Polytechnic».

The journal web-site is http://ric.zntu.edu.ua.

The address of the editorial office: Editorial office of the journal «Radio Electronics, Computer Science, Control», National University «Zaporizhzhia Polytechnic», Zhukovskiy street, 64, Zaporizhzhia, 69063, Ukraine.

Tel.: +38-061-769-82-96 – the editorial-publishing department. Fax: +38-061-764-46-62

E-mail: rvv@zntu.edu.ua

ЗМІСТ

CONTENTS

ДВАДЦЯТЬ П’ЯТЬ РОКІВ НАУКОВОМУ ЖУРНАЛУ «РАДІОЕЛЕКТРОНІКА, ІНФОРМАТИКА, УПРАВЛІННЯ»

У 2023 р. виповнилося 25 років науковому журналу «Радіоелектроніка, інформатика, управління» (РІУ), засновником та видавцем якого є НУ «Запорізька політехніка». Журнал було зареєстровано як періодичне видання у 1998 році. Перший номер було видано у 1999 році. За час існування з 1998 р. до середини 2023 р. редколегією було отримано, а рецензентами – зрецензовано понад три з половиною тисячі рукописів статей. За цей період вийшло 65 номерів журналу, де опубліковано понад півтори тисячі наукових статей.

Опублікованим статтям журналу присвоюється унікальний ідентифікатор цифрового об’єкта DOI. Статті журналу індексуються провідними наукометричними та реферативними базами світу, зокрема Web of Science. Журнал РІУ включено Міністерством освіти і науки України до переліку наукових фахових видань України в категорії «А» (найвищий рівень), в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук і доктора філософії (кандидата наук), а також до Переліку наукових журналів та рецензованих матеріалів міжнародних конференцій з присвоєною кількістю балів Міністерства науки та вищої освіти Республіки Польща.

У різні роки журнал РІУ очолювали головні редактори: Потапенко Є. М. (Україна), Піза Д. М. (Україна), Погосов В. В. (Україна), Субботін С. О. (Україна), до складів редакційної колегії та редакційно-консультативної ради журналу входили: Андроулідакіс І. (Греція), Аррас П. (Бельгія), Ахметшин А. М. (Україна), Безрук В. М. (Україна), Бодянський Є. В. (Україна), Веннекенс Ю. (Бельгія), Волков О. В. (Україна), Вольф К. (Німеччина), Вуттке Г.-Д. (Німеччина), Гімпілевич Ю. Б. (Україна), Горбань О. М. (Велика Британія), Горбань О. М. (Україна), Городничий Д. О. (Канада), Горр Г. В. (Україна), Гостєв В. І. (Україна), Дробахін О. О. (Україна), Дубровін В. І. (Україна), Зайцева О. М. (Словаччина), Камеяма М. (Японія), Карпуков Л. М. (Україна), Карташов В. М. (Україна), Ковальов О. М. (Україна), Корніч Г. В. (Україна), Кулик А. С. (Україна), Лебедєв Д. В. (Україна), Левашенко В. Г. (Словаччина), Лісняньский А. (Ізраїль), Луенго Д. (Іспанія), Мадритщ Х. (Австрія), Марковска-Качмар У. (Польща), Маркосян М. В. (Вірменія), Матюшин В. М. (Україна), Олещук В. О. (Норвегія), Олійник А. О. (Україна), Онуфрієнко В. М. (Україна), Павліков В. В. (Україна), Павлов О. А. (Україна), Папшицький М. (Польща), Піза Д. М. (Україна), Рубель О. В. (Канада), Скрупський С. Ю. (Україна), Табунщик Г. В. (Україна), Тавхелідзе А. (Грузія), Толок В. О. (Україна), Тригано Т. (Ізраїль), Труфанов І. Д. (Україна), Урсутью Д. (Румунія), Хаханов В. І. (Україна), Хенке К. (Німеччина), Чумаченко В. П. (Україна), Шарпанських О. А. (Нідерланди), Шульц П. (Німеччина) та інші.

Організаційно-технічну підтримку діяльності редакції журналу забезпечували у різні роки начальники редакційно-видавничого відділу (Рибіна Ю. А., Дмитренко А. А., Савчук Н. О.) та його співробітники (Зуб С. В., Омесь О. В.), а також співробітники типографії НУ «Запорізька політехніка» (М’ясников М. Л., Винник А. П., Глінкін Є. О.).

Вітаємо редакцію журналу, спільноту його авторів, рецензентів та читачів, співробітників редакційно-видавничого відділу і типографії та бажаємо їм творчої наснаги, плідної роботи та подальших успіхів!

РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

RADIO ELECTRONICS AND TELECOMMUNICATIONS

UDC 621.391

TELETRAFFIC FORECASTING IN MEDIA SERVICE SYSTEMS

Gusiev O. Yu. – PhD, Associate Professor, Professor of the Department of Information Security and Telecommunications, Dnipro University of Technology, Dnipro, Ukraine.

Magro V. I. – PhD, Associate Professor, Professor of the Department of Information Security and Telecommunications, Dnipro University of Technology, Dnipro, Ukraine.

Nikolska O. I. – Senior Lecturer of the Department of Information Security and Telecommunications, Dnipro University of Technology, Dnipro, Ukraine.

ABSTRACT

Context. The development of information and communication technologies has led to an increase in the volume of information sent over the network. Media service platforms play an important role in the creation and processing of bitrate in the information network. Therefore, there is a need to develop a methodology for predicting bitrate in various media service platforms by creating an effective algorithm that minimizes the forecast error.

Objective. The aim of the work is to synthesize in analytical form the state transition matrix of the Kalman filter for non-stationary self-similar processes when predicting the bitrate in telecommunication networks.

Method. A methodology has been developed for predicting teletraffic in media service platforms, based on a modification of the Kalman filter for non-Gaussian processes. This methodology uses an original procedure for calculating statistics, which makes it possible to reduce the filtering and forecast error that arises due to the uncertainty of the analytical model of the process under study. The methodology does not require knowledge of the analytical model of the process, as well as strict restrictions on its stochastic characteristics.

Results. A methodology for estimating and forecasting bitrate in telecommunication systems is proposed. This methodology was used to study teletraffic processes in the media service platforms Google Meet, Zoom, Microsoft Teams. The passage of real bitrate through the specified media service platforms was studied. A comparison of real teletraffic with predicted teletraffic was carried out. The influence of the order of the state transition matrix of the Kalman filter on the error of estimation and prediction has been studied. It has been established that even a low (second) order of the state transition matrix allows one to obtain satisfactory forecast results. It is shown that the use of the proposed methodology makes it possible to predict traffic with a relative error of the order of 3–4%.

Conclusions. An original algorithm for assessing and forecasting the characteristics of media traffic has been developed. Recommendations for improving the technology for building media service platforms are formulated. It is shown that the bitrates generated by various media service platforms, in the case of applying the proposed estimation and forecasting methodology, are invariant with respect to the type of stochastic processes being processed.

KEYWORDS: Kalman filter, teletraffic, media platform, stochastic process, self-similar process.

ABBREVIATIONS

QoS – Quality of service;
LSTM – Long short-term memory;
ARIMA – Autoregressive integrated moving average;
ML – Machine learning;
CNN – Convolutional Neural Network.

NOMENCLATURE

F_n is a state transition matrix;
H_n are observation conditions matrix;
h is a disorder threshold value;
P_n is an error matrix;
S_n is a measured signal;
X_n is a useful signal;
Q_n is an additive noise;
R is a noise dispersion;
B_M is a discord statistics;
MO is an expectation operator.

INTRODUCTION

The beginning of the fourth industrial revolution (Industrie 4.0) has created new preconditions for the development of digital and intelligent production and devices that interact with each other and provide personalized product output. One of the most important components of Industrie 4.0 is not the product, but the data. Digitaliza-

DOI 10.15588/1607-3274-2023-4-1

tion of production is associated with large amounts of data (big data), which need to be read, analyzed, systematized, stored, transmitted, presented in the required form, etc. This requires appropriate infocommunication systems, software, wireless data transmission tools, cloud services exchange and storage of data. These circumstances increase the importance of network traffic forecasting as a function of data network management and operation, since traffic forecasting plays a crucial role in improving the quality of services (QoS) [1–4]. Thus, an important component of the process of studying network traffic in order to improve methods for predicting its characteristics is the development of new, more advanced methods that provide QoS.

One of the means of wireless data transmission within the framework of Industrie 4.0 is the new generation 5G, which provides very fast data transmission and which must be characterized by its QoS. This circumstance necessitates the study of various media traffic forecasting models [5–8]. In addition, predicting changes in the base station load can reduce energy consumption [7], and increasing the efficiency of big data processing by predicting traffic improves network performance [8].

Reliable and accurate prediction of QoS characteristics is essential for making effective web service recommendations to improve user experience and service management [9–12]. Network state forecasting is the process of in-depth study of network traffic in order to obtain the most complete information about the processes occurring in networks. In turn, the accuracy of analysis and estimation of network traffic characteristics is of great importance to achieve guaranteed QoS.

Computer network traffic control has been a topic of much research as it is addressed in various applications such as anomaly detection, congestion control, and bandwidth control [13–15]. On the other hand, network traffic forecasting aims to predict subsequent network traffic using previous network traffic data. This approach is used in network control and solving planning problems [16–27]. Network traffic forecasting is a critical tool for monitoring and estimation network security [19]. By predicting network traffic, you can effectively identify network failures, optimize performance, and ensure network security. Predicting telecommunications network data traffic with high accuracy is a challenging task for the network control function. It improves dynamic resource allocation and power control [22].

In this paper, the authors propose an original technique for predicting teletraffic in various media service platforms, which is based on approximating the state transition matrix of the Kalman filtering process in the case where the analytical description of the process under study is unknown.

The object of study is the process of forecasting teletraffic, created by various media service platforms in telecommunication networks.

The subject of the study is a methodology for synthesizing the Kalman filter state transition matrix for non-stationary stochastic processes in the absence of their analytical description.

The purpose of the work is to synthesize in analytical form the state transition matrix of the Kalman filter for non-stationary self-similar processes when predicting the bitrate in telecommunication networks.

1 PROBLEM STATEMENT

We consider a self-similar process that describes the behavior of traffic in media service platforms of telecommunication networks. This process is fractal in nature and, therefore, cannot be processed by classical methods, inherent in Gaussian processes.

To solve the problem, it is proposed to use a modification of the Kalman filtering method, which does not require knowledge of the process under study in analytical form. This approach is invariant to the type of stochastic process and removes the restrictions imposed on its characteristics.

The solution to the problem is sought by approximating the state transition matrix of the Kalman filter by an n-th order Taylor series at each reference point.

2 REVIEW OF THE LITERATURE

In recent years, well-known media platforms such as Microsoft Teams, Zoom, Google Meet, etc. have been widely used. When many users work simultaneously in some segments of the information and telecommunications network, overload or a significant delay may occur when processing information in the media service platform [28].

In this regard, there is a need to develop methods for optimal network traffic control in order to eliminate these negative phenomena, an integral attribute of which is the process of forecasting traffic flows.

Neural networks are currently most often used to predict network traffic, and the forecast is based on the long short-term memory (LSTM) of the neural network [1–10, 12–14, 16–23, 25–27]. In some cases, such a model is used independently [9,10, 14, 18, 21, 23, 25]. However, its composition with other models is more common: with the autoregressive integrated moving averages (ARIMA) model [4, 11, 13], with recurrent neural networks (RNN) [3, 6, 13, 16, 17, 19, 20, 22]. Less commonly used are approaches such as machine learning (ML) [5, 24], clustering model [12], and deep learning method of convolutional neural network (CNN) [27].

Forecasting methods using neural networks have a significantly complicated computational procedure; in addition, the choice of input layers and input values is, as a rule, subjective.

The use of the classical Kalman filtering method [29] also has its disadvantages, associated primarily with the requirement of the Gaussian nature of the process under study and knowledge of the analytical expression of the state transition matrix.

DOI 10.15588/1607-3274-2023-4-1

3 MATERIALS AND METHODS

The classical Kalman filtering method assumes knowledge of the initial values of the process under study and its description in analytical form, which allows one to determine the state transition matrix. In practice, as a rule, we are not able to obtain an analytical expression of the process being processed, since this process has not yet been studied.

The structure of the telecommunications system model includes traffic sources (voice traffic and video traffic) and the Internet network, which connects traffic receivers and transmitters with media service platforms. Using specialized programs, you can obtain data about teletraffic on the network. Estimation and forecast of teletraffic characteristics can be obtained using the Kalman filtering procedure. In this case, the obtained data can be used as input measurements for the Kalman filter. Kalman filters use system state information as well as measurements to estimate the current state of the system. It calculates a real-time forecast of the state of the system, taking into account new data and measurements. The resulting forecast allows the system to adapt to changes in traffic and promptly identify anomalies or problems in the telecommunications network.

In this paper, we propose a modification of the Kalman filter, which consists in approximating the state transition matrix at each reference point by a Taylor series. This approach has one significant drawback due to the divergence of the filtering procedure (due to the finite order of the polynomial) when processing non-stationary processes. It would seem that this drawback can be eliminated by increasing the order of the approximating polynomial. However, it has been established (easy to verify in practice) that as the order of the polynomial increases, the filter becomes more prone to self-excitation, since this increases the depth of feedback and, consequently, a large number of poles and zeros appears, leading to an increase in the probability of self-excitation. Therefore, a low order of the polynomial should be chosen, and to eliminate the divergence of the filter, an original procedure for detecting a disorder in the filtering process is proposed.

The procedure is as follows. Along the time series, a time window is allocated, with a dimension of, for example, ten samples. A certain threshold is assigned, with a dimension, for example, seven. In this case, each new sample is accompanied by a shift of the window one sample forward. Within the window between two adjacent samples, the difference between the true (measured) signal and its filtered (predicted) value is determined and the sign of this difference is determined. Next, statistics are calculated, which is the algebraic sum of positive and negative signs. The calculated statistics are compared with the threshold and, if its value is equal to or exceeds the threshold value, a decision is made to disturb the filter. In this case, the transition to the beginning of the time window occurs, the initial filter parameters are set, and filtering is repeated within the window boundaries, and the previously obtained values are replaced with the current ones. In this case, a slight increase in the variance of the estimation noise is possible; however, studies have shown that this increase is small and can be neglected. Next, filtering occurs as usual until a new disorder occurs. Thus, the considered procedure allows us to avoid the divergence of the filtering process, that is, the bias of the estimate, and to study non-stationary processes with a significant degree of correctness.

In analytical form, the modified Kalman filter can be represented as follows. For the case of discrete measurements of the signal S_n, which is an additive mixture,

$$S_n = X_n + Q_n,$$

where X_n is the useful signal, Q_n is additive noise with mathematical expectation MO[Q_n] = 0 and variance R, the Kalman filtering procedure can be represented in the following form

$$\overline{X}_n = F_n \overline{X}_{n-1} + P_n H_n^{'} R^{-1}(S_n - H_n F_n \overline{X}_{n-1});$$
$$P_n = (A_n^{-1} + H_n^{'} R^{-1} H_n)^{-1};$$
$$A_n = F_n P_{n-1} F_n^{'},$$

where index "′" means the transposition of the matrix.

The procedure for detecting a filter discord comes down to calculating statistics of the form

$$B_M = \sum_{l=1}^{M} b_l, \quad B_0 = 0, \quad l = 1,2,...M;$$
$$b_l = \operatorname{sgn}(S_l - \overline{X}) = +l, \quad S_l - \overline{X} \geq 0;$$
$$b_l = \operatorname{sgn}(S_l - \overline{X}) = -l, \quad S_l - \overline{X} < 0$$

on the interval [$n - M$, n]. Values determined on this interval ($B_m - \min B_m$) and ($\max B_m - B_m$) are compared with the threshold h. When the value h exceeds one of the quantities, a decision is made on discord, the filter parameters are assigned initial values, and filtering continues from moment $n - M$.

The above algorithm was once used by the authors to study processes in blast furnace production, but it has never been used to study temporary self-similar traffic.

4 EXPERIMENTS

During the experiment, teletraffic analysis was carried out in the following media service platforms: Microsoft Teams, Google Meet, Zoom. For this purpose, the cross-platform client-server program iperf3 was used, which allows testing network throughput. With its help, the maximum network bandwidth was measured and load testing of the communication channel was carried out. Statistical information about the state of the communication channel was collected using the Wireshark traffic analyzer program, which allows us to view all traffic passing through the network in real time. Various types of traffic were considered: audio traffic; audio traffic plus video traffic with minimal change over time; audio traffic

DOI 10.15588/1607-3274-2023-4-1

plus video traffic with a maximum change in video image in video time [30].

This information served as the basis for forecasting the bitrate in the media service platform under study. Based on the same information, the relative forecast error was calculated.

5 RESULTS

A study was conducted of the dynamic changes in voice and video traffic over time (Fig. 1–6). Here the real traffic statistics are shown in red; traffic estimate is shown in blue; the predicted bitrate is shown in green.

The dependence of traffic intensity (bit/s) on time at the input of the Google Meet media service platform was obtained for real statistics, estimation and forecast.

The studies were carried out for the case of voice traffic (Fig. 1).

The dependence of traffic intensity on time at the output of the Google Meet media service platform was studied for real statistics, eшtшated and forecast. The studies were carried out for the case of voice traffic (Fig. 2).

The dependence of traffic intensity on time at the input of the Zoom media service platform was obtained for real statistics, estimated and forecast. The studies were carried out for the case of voice traffic with minimal changes in the video image over time (Fig. 3).

The dependence of traffic intensity on time at the output of the Zoom media service platform was studied for real statistics, estimated and forecast. The studies were carried out for the case of voice traffic with minimal changes in the video image over time (Fig. 4).

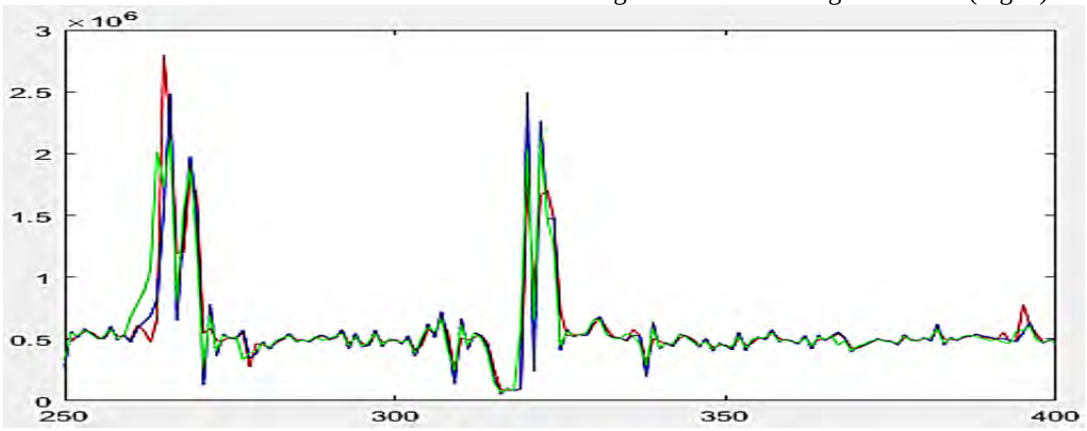


Figure 1 – Dependence of traffic intensity on time at the input of the Google Meet media service platform

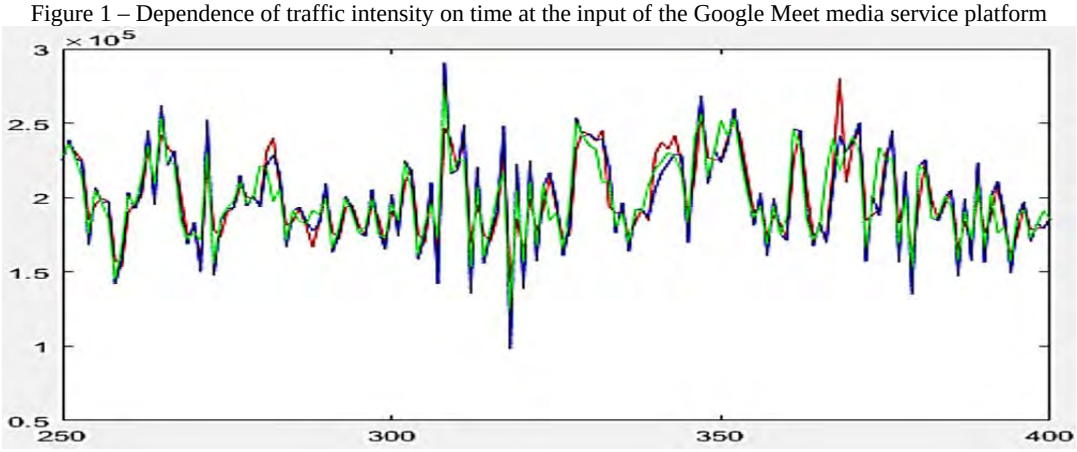


Figure 2 – Dependence of traffic intensity on time at the output of the Google Meet media service platform

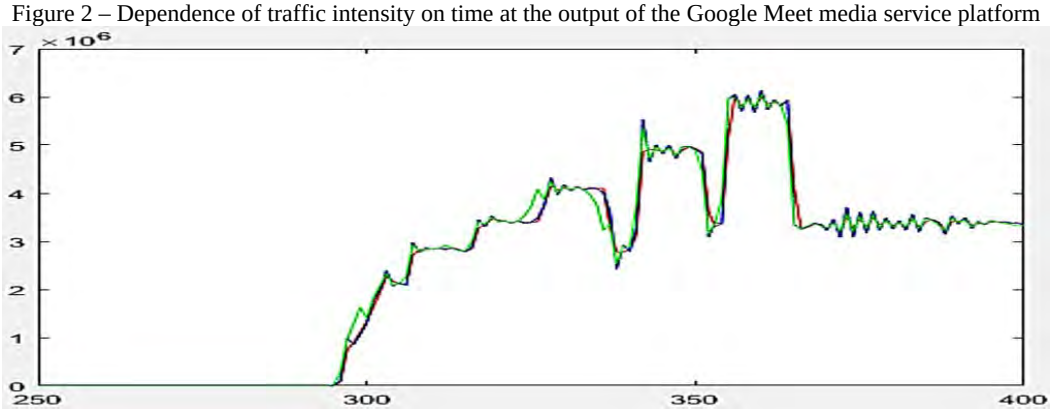


Figure 3 – Dependence of traffic intensity on time at the input of the Zoom media service platform

DOI 10.15588/1607-3274-2023-4-1

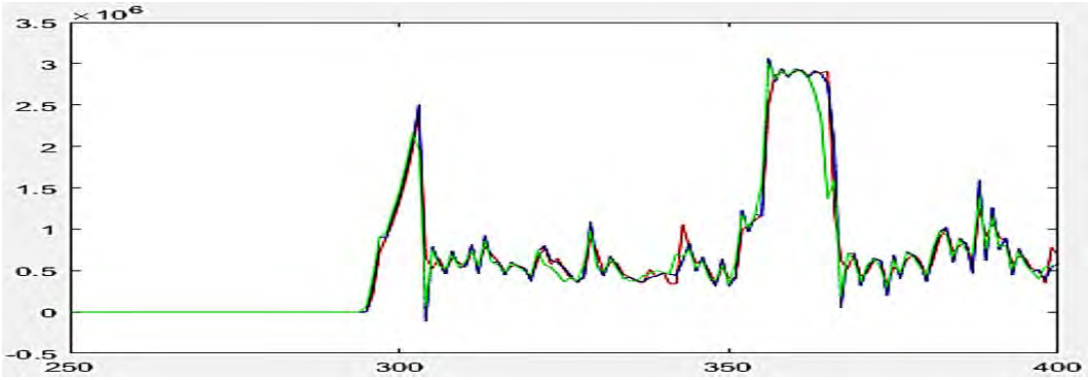


Figure 4 – Dependence of traffic intensity on time at the output of the media service Zoom platforms

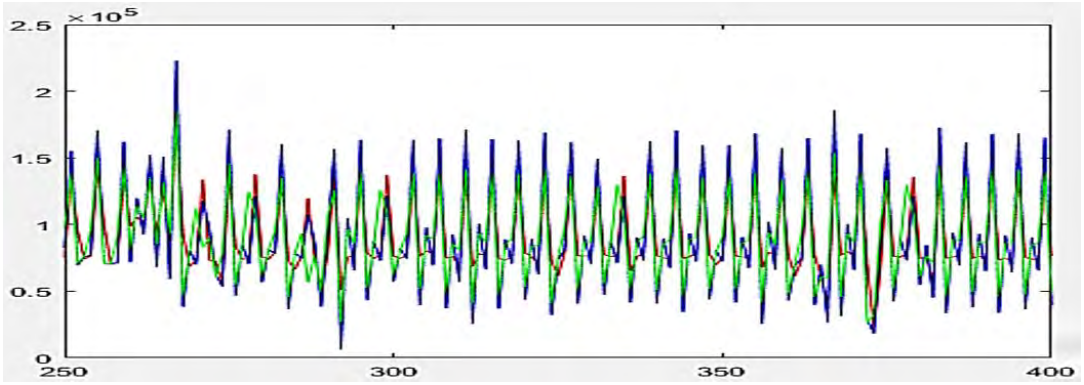


Figure 5 – Dependence of traffic intensity on time at the input of the Microsoft Teams media service platform for the case of voice traffic only

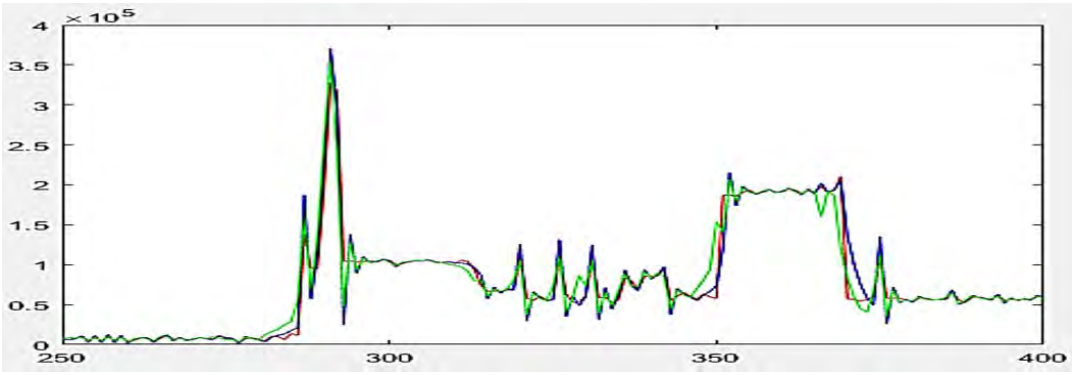


Figure 6 – Dependence of traffic intensity on time at the input of the Microsoft Teams media service platform for the case of video traffic with a maximum change in the video image over time

The dependence of traffic intensity on time at the input of the Microsoft Teams media service platform was obtained for real statistics, estimated and forecast. The studies were carried out for the case of voice traffic only (Fig. 5).

The dependence of traffic intensity on time at the input of the Microsoft Teams media service platform was studied for real statistics and forecast. The studies were carried out for the case of video traffic with a maximum change in the video image over time (Fig. 6).

6 DISCUSSION

In this work, we implemented a method for estimating and forecasting non-stationary time series based on a polynomial representation of the state transition matrix of the Kalman filtering procedure. The problem of filter discord due to the finite order of the approximating polynomial was solved by means of an original algorithm for detecting discord in tempo with the process. The algorithm is quite simple and does not require large time and software resources. It should also be noted that the proposed modification removes the requirements for stationarity and normality of the processed processes, as well as for knowledge of the initial conditions.

The analysis of the dependence of traffic intensity on time showed that the assessment of incoming traffic and traffic forecast in media service platforms (Fig. 1, Fig. 3) have relative errors of 4.44% and 0.64%, respectively.

At the same time, traffic leaving the media service platform has a significantly lower level of relative errors,

DOI 10.15588/1607-3274-2023-4-1

0.028% and 0.037%, respectively. This indicates fairly accurate forecasting.

The analysis of experimental data showed that the method can be effectively (with a small error) used to predict teletraffic in media service systems.

In addition, since the stochastic characteristics of the processes under study differ from classical Gaussian ones, it can be argued that the proposed estimation and forecast procedure is invariant to the stochastic properties of the processes and does not require strict restrictions.

CONCLUSIONS

The explosive and self-similar type of traffic in modern information and communication networks requires the development of forecasting methodology to prevent network "overload". The analysis of teletraffic in media service platforms requires special attention. The development of a universal methodology for analyzing traffic behavior in media service platforms will make it possible to "include" various methodology in order to prevent "overload" of the network.

Traffic forecasting allows network operators to monitor the health of the network and respond to the occurrence of anomalies or problems, including reporting a cyber-attack on the network. Traffic forecasting methodology allow you to plan customer service, including setting limits on the volume of traffic during peak load.

The scientific novelty of the results obtained lies in the fact that for the first time an original methodology for predicting bitrate in various media service platforms was proposed, and proposals were also developed for choosing the optimal characteristics of the signal model to achieve a minimum prediction error.

The practical significance of the results obtained is that an original algorithm for estimation and predicting the characteristics of media traffic has been developed. Recommendations for improving the technology for building media service platforms are formulated. It is shown that the bitrates generated by various media service platforms, in the case of applying the proposed estimation and forecast methodology, are invariant with respect to the type of stochastic processes undergoing assessment and forecast. The results obtained can be applied to the study of fractal random processes.

Prospects for further research is research in the direction of comparative analysis of forecasting results for various types of approximating polynomials.

ACKNOWLEDGEMENTS

The work was carried out with the assistance of the Department of Information Security and Telecommunications of the Dnipro University of Technology.

REFERENCES

1. Abdellah A. R., Mahmood O. A. K., Paramonov A., Koucheryavy A. IoT traffic prediction using multi-step ahead prediction with neural network, *IEEE 11th International congress on ultra modern telecommunications and control systems and workshops (ICUMT)*, 2019, IEEE, 2019, pp. 1–4. https://doi.org/ 10.1109/ICUMT48472.2019.8970675
2. Pan C., Wang Y., Shi H., Shi J., Cai R. Network traffic prediction incorporating prior knowledge for an intelligent network, *Sensors*, 2022, Vol. 22, No. 7, 2674. https://doi.org/ 10.3390/s22072674
3. Kumar B. P., Hariharan K. Multivariate time series traffic forecast with long short term memory based deep learning model, *IEEE International conference on power, instrumentation, control and computing (PICC)*, 2020, IEEE, 2020. pp. 1–5. https://doi.org/ 10.1109/PICC51425.2020.9362455
4. Liu B., Tang X., Cheng J., Shi P. Traffic flow combination forecasting method based on improved LSTM and ARIMA, *International Journal of Embedded Systems*, 2020, Vol. 12, No. 1, pp. 22–30. https://doi.org/ 10.1504/IJES.2020.10026902
5. Refaee A., Volkov A., Muthanna A., Gallyamov D., Koucheryavy A. Deep Learning for IoT traffic prediction based on edge computing, *Distributed computer and communication networks: control, computation, communications*, 2021, pp. 18–29. https://doi.org/ 10.1007/978-3-030-66242-4_2
6. Jaffry S., Hasan S. F. Cellular traffic prediction using recurrent neural networks, *IEEE 5th International Symposium on Telecommunication Technologies (ISTT)*, 2020, IEEE, 2020, pp. 94–98. https://doi.org/ 10.1109/ISTT50966.2020.9279373
7. Xu X., Gao S., Jiang Z. LSTCN: An attention-based deep neural network model combining LSTM and TCN for cellular network traffic prediction, *IEEE 5th International conference on communication and information systems (ICCIS)*, 2021, IEEE, 2021, pp. 34–38. https://doi.org/ 10.1109/ICCIS53528.2021.9645961
8. De Klerk M. L., Saha A. K. A review of the methods used to model traffic flow in a substation communication network, *IEEE Access*, 2020, Vol. 8, pp. 204545–204562. https://doi.org/ 10.1109/ACCESS.2020.3037143
9. Jirsik T., Trčka Š., Celeda P. Quality of service forecasting with LSTM neural network, *IFIP/IEEE Symposium on integrated network and service management (IM)*, 2019, IEEE, 2019, pp. 251–260
10. Zhang L., Zhang H., Tang Q., Dong P., Zhao Z., Wei Y., Mei J., Xue K. LNTP: An End-to-End Online Prediction Model for Network Traffic, *IEEE Network*, 2021, Vol. 35, pp. 226–233. https://doi.org/ 10.1109/MNET.011.1900647
11. Fanjiang Y.-Yi., Huang Y. S. W.-L. Time series QoS forecasting for Web services using multi-predictor-based genetic programming, *IEEE Transactions on Services Computing*, 2022, Vol. 15, pp. 1423–1435. https://doi.org/ 10.1109/TSC.2020.2994136
12. Aldhyani T. H. H., Alrasheedi M., Alqarni A. A., Alzahrani M.Y., M.Y. Bamhdi M.Y. Intelligent hybrid model to enhance time series models for predicting network traffic, *IEEE Access*, 2020, Vol. 8, pp. 130431–130451. https://doi.org/ 10.1109/ACCESS.2020.3009169
13. Madan R., Mangipudi P. S. Predicting computer network traffic: A time series forecasting approach using DWT, ARIMA and RNN, *IEEE Eleventh International conference on contemporary computing (IC3)*, 2018, IEEE, 2018, pp. 1–5. https://doi.org/ 10.1109/IC3.2018.8530608
14. Nihale S., Sharma S., Parashar L., Singh U. Network traffic prediction using long short-term memory, *IEEE International conference on electronics and sustainable communication systems (ICESC)*, 2020, IEEE, 2020, pp. 338–343. https://doi.org/ 10.1109/ICESC48915.2020.9156045

DOI 10.15588/1607-3274-2023-4-1

15. Drieieva H., Smirnov O., Drieiev O., Polishchuk Y., Brzhanov R., Aleksander M. Method of fractal traffic generation by a model of menerator on the graph, *2nd International Workshop on Control, Optimisation and Analytical Processing of Social Networks (COAPSN)*, 2020, pp. 366–379.
16. Alizadeh M., Beheshti M., Ramezani A., Saadatinezhad H. Network traffic forecasting based on fixed telecommunication data using deep learning, *IEEE 6th Iranian conference on signal processing and intelligent systems (ICSPIS)*, 2020, IEEE, 2020, pp. 1–7. https://doi.org/ 10.1109/ICSPIS51611.2020.9349573
17. Vinayakumar R., Soman K. P., Poornachandran P. Applying deep learning approaches for network traffic prediction, *IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, IEEE, 2017, pp. 2353–2358. https://doi.org/ 10.1109/ICACCI.2017.8126198
18. Aloraifan D., Ahmad I., Alrashed E. Deep learning based network traffic matrix prediction, *International Journal of Intelligent Networks*, 2021, Vol. 2, pp. 46–56. https://doi.org/10.1016/j.ijin.2021.06.002
19. 19. Fan J., Mu D., Liu Y. Research on network traffic prediction model based on neural network, *IEEE 2nd International conference on information systems and computer aided education (ICISCAE)*, 2019, IEEE, 2019, pp. 554–557. https://doi.org/ 10.1109/ICISCAE48440.2019.221694
20. Hua Y., Zhao Z., Liu Z., Chen X., Li R., Zhang H. Traffic prediction based on random connectivity in deep learning with long short-term memory, *IEEE 88th Vehicular technology conference (VTC-Fall)*, 2018, IEEE, 2018, pp. 1–6. https://doi.org/ 10.1109/VTCFall.2018.8690851
21. Lu H., Yang F. Research on network traffic prediction based on long short-term memory neural network, *IEEE 4th International conference on computer and communications (ICCC)*, 2018, IEEE, 2018, pp. 1109–1113. https://doi.org/ 10.1109/CompComm.2018.8781071
22. Do Q. H., Doan T. T. H., Nguyen T. V. A., Linh V. V. Prediction of data traffic in telecom networks based on deep neural networks, *Journal of computer science*, 2020, Vol. 16, No. 9, pp. 1268–1277. https://doi.org/ 10.3844/jcssp.2020.1268.1277
23. Guo D., Xia X., Zhu L., Zhang Y. Dynamic modification neural network model for short-term traffic prediction, *Procedia Computer Science*, 2021, Vol. 187, pp. 134–139. https://doi.org/10.1016/j.procs.2021.04.043
24. Jain G., Prasad R. R. Machine learning, prophet and XGBoost algorithm: analysis of traffic forecasting in telecom networks with time series data, *IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2020, IEEE, 2020, pp. 893–897. https://doi.org/ 10.1109/ICRITO48877.2020.9197864
25. Shihao W., Qinzheng Z., Han Y., Qianmu L., Yong Q. A network traffic prediction method based on LSTM, *ZTE Communications*, 2019, Vol. 17, No. 2. pp. 19–25. https://doi.org/ 10.12142/ZTECOM.201902004
26. Bi J., Zhang X., Yuan H., Zhang J., Zhou M.C. A hybrid method for realistic network traffic with temporal convolutional network and LSTM, *IEEE Transactions on Automation Science and Engineering*, 2022, Vol. 19, No. 3, pp. 1869–1879. https://doi.org/ 10.1109/TASE.2021.3077537
27. Ko T., Raza S. M., Binh D. T., Kim M., Choo H. Network prediction with traffic gradient classification using convolutional neural networks, *IEEE 14th International conference on ubiquitous information management and communication (IMCOM)*, 2020, IEEE, 2020, pp. 1–4. https://doi.org/ 10.1109/IMCOM48794.2020.9001712
28. Magro V., Svyatoshenko V., Tymofieiev D. Method for evaluating the delay time in a stream broadcast process, *Information Processing Systems*, 2019, Vol. 159, No. 4, pp. 28–35. https://doi.org/10.30748/soi.2019.159.03
29. Sage A.P., Melsa J.L. Estimation theory with applications to communications and control, New York, McGraw-Hill, 1971, 529 pp.
30. Magro V. I., Plaksin S. V., Syatoshenko V. O. Investigation of information network loading in the condition of remote education and remote monitoring, *Applied questions of mathematical modeling*, 2021, Vol. 4, No. 2.1, pp. 142–149. https://doi.org/10.32782/KNTU2618-0340/2021.4.2.1.15

Received 05.10.2023.
Accepted 28.11.2023.

УДК 621.391

ПРОГНОЗУВАННЯ ТЕЛЕТРАФІКА В МЕДІАСЕРВІСНИХ СИСТЕМАХ

Гусєв О. Ю. – канд. фіз.-мат. наук, доцент, професор кафедри безпеки інформації та телекомунікацій Національного технічного університету «Дніпровська політехніка», Дніпро, Україна.

Магро В. И. – канд. фіз.-мат. наук, доцент, професор кафедри безпеки інформації та телекомунікацій Національного технічного університету «Дніпровська політехніка», Дніпро, Україна.

Нікольська О. І. – старший викладач кафедри безпеки інформації та телекомунікацій Національного технічного університету «Дніпровська політехніка», Дніпро, Україна.

АНОТАЦІЯ

Актуальність. Розвиток інформаційно-комунікаційних технологій призвело до зростання обсягу інформації, що пересилається через мережу. Медіасервісні платформи відіграють важливу роль у створенні та обробці бітрейту в інформаційній мережі. Тому існує необхідність у розробці методики прогнозування бітрейту в різних медіасервісних платформах шляхом створення ефективного алгоритму, що мінімізує помилку прогнозу.

Мета. Метою роботи є синтез в аналітичній формі матриці переходу фільтра Калмана для нестаціонарних самоподібних процесів при прогнозуванні бітрейту в телекомунікаційних мережах.

Метод. Розроблено методику прогнозування телетрафіку в медіасервісних платформах, засновану на модифікації фільтра Калмана для негаусівських процесів. Ця методика використовує оригінальну процедуру підрахунку статистики, яка дозволяє знижувати помилку фільтрації та прогнозу, що виникає внаслідок невизначеності аналітичної моделі досліджуваного

DOI 10.15588/1607-3274-2023-4-1

процесу. Методика не вимагає знання аналітичної моделі процесу, а також жорстких обмежень на його стохастичні характеристики.

Результати. Запропоновано методику оцінки та прогнозу бітрейту в телекомунікаційних системах. Дана методика застосована для дослідження процесів телетрафіку в медсервісних платформах Google Meet, Zoom, Microsoft Teams. Досліджено проходження реального бітрейту через зазначені медіасервісні платформи. Проведено порівняння реального телетрафіку з прогнозованим телетрафіком. Досліджено вплив порядку матриці переходу фільтра Калмана на похибку оцінки та прогнозу. Встановлено, що навіть невисокий (другий) порядок матриці переходу дозволяє отримати задовільні результати прогнозу. Показано, що застосування запропонованої методики дозволяє прогнозувати трафік з відносною помилкою близько 3–4%.

Висновки. Розроблено оригінальний алгоритм оцінки та прогнозу характеристик медіатрафіку. Сформульовано рекомендації щодо удосконалення технології побудови медіасервісних платформ. Показано, що бітрейти, що породжуються різними медіасервісними платформами, у разі застосування запропонованої методики оцінки та прогнозу інваріантні щодо типу оброблюваних стохастичних процесів.

КЛЮЧОВІ СЛОВА: фільтр Калмана, телетрафік, медіасевісна платформа, стохастичний процес, самоподібний процес.

ЛІТЕРАТУРА

1. IoT traffic prediction using multi-step ahead prediction with neural network / [A. R. Abdellah, O. A. K. Mahmood, A. Paramonov, A. Koucheryavy] // Proceeding of the IEEE 11th International congress on ultra modern telecommunications and control systems and workshops (ICUMT). – 2019. – P. 1–4. https://doi.org/ 10.1109/ICUMT48472.2019.8970675
2. Network traffic prediction incorporating prior knowledge for an intelligent network / [C. Pan, Y. Wang, H. Shi et al.] // Sensors. – 2022. – Vol. 22, No. 7. – 2674. https://doi.org/ 10.3390/s22072674
3. Kumar B. P. Multivariate time series traffic forecast with long short term memory based deep learning model / B. P. Kumar, K. Hariharan // Proceeding of the IEEE International conference on power, instrumentation, control and computing (PICC). – 2020. – P. 1–5. https://doi.org/ 10.1109/PICC51425.2020.9362455
4. Traffic flow combination forecasting method based on improved LSTM and ARIMA / [B. Liu, X. Tang, J. Cheng, P. Shi] // International Journal of Embedded Systems. – 2020. – Vol. 12, No. 1. – P. 22–30. https://doi.org/ 10.1504/IJES.2020.10026902
5. Deep Learning for IoT traffic prediction based on edge computing / [A. Refaee, A. Volkov, A. Muthanna et al.] // Distributed computer and communication networks: control, computation, communications. – 2021. – P. 18–29. https://doi.org/ 10.1007/978-3-030-66242-4_2
6. Jaffry S. Cellular traffic prediction using recurrent neural networks / S. Jaffry, S. F. Hasan // Proceeding of the IEEE 5th International Symposium on Telecommunication Technologies (ISTT). – 2020. – P. 94–98. https://doi.org/ 10.1109/ISTT50966.2020.9279373
7. Xu X. LSTCN: An attention-based deep neural network model combining LSTM and TCN for cellular network traffic prediction / X. Xu, S. Gao, Z. Jiang // Proceeding of the IEEE 5th International conference on communication and information systems (ICCIS). – 2021. – P. 34–38. https://doi.org/ 10.1109/ICCIS53528.2021.9645961
8. De Klerk M. L. A review of the methods used to model traffic flow in a substation communication network / M. L. De Klerk, A. K. Saha // IEEE Access. – 2020. – Vol. 8. – P. 204545–204562. https://doi.org/ 10.1109/ACCESS.2020.3037143
9. Jirsik T. Quality of service forecasting with LSTM neural network / T. Jirsik, Š. Trčka, P. Celeda // Proceeding of the IFIP/IEEE Symposium on integrated network and service management (IM). – 2019. – P. 251–260.
10. LNTP: An End-to-End Online Prediction Model for Network Traffic / [L. Zhang, H. Zhang, Q. Tang et al.] // IEEE Network. – 2021. – Vol. 35. – P. 226–233. https://doi.org/ 10.1109/MNET.011.1900647
11. Fanjiang Y.-Yi. Time series QoS forecasting for Web services using multi-predictor-based genetic programming / Y.-Yi. Fanjiang, Y. S. W.-L. Huang // IEEE Transactions on Services Computing. – 2022. – Vol. 15. – P. 1423–1435. https://doi.org/ 10.1109/TSC.2020.2994136
12. Intelligent hybrid model to enhance time series models for predicting network traffic / [T. H. H. Aldhyani, M. Alrasheedi, A. A. Alqarni et al.] // IEEE Access. – 2020. – Vol. 8. – P. 130431–130451. https://doi.org/ 10.1109/ACCESS.2020.3009169
13. Madan R. Predicting computer network traffic: A time series forecasting approach using DWT, ARIMA and RNN / R. Madan, P. S. Mangipudi // Proceeding of the IEEE Eleventh international conference on contemporary computing (IC3). – 2018. – P. 1–5. https://doi.org/ 10.1109/IC3.2018.8530608
14. Network traffic prediction using long short-term memory / [S. Nihale, S. Sharma, L. Parashar, U. Singh] // Proceeding of the IEEE International conference on electronics and sustainable communication systems (ICESC). – 2020. – P. 338–343. https://doi.org/ 10.1109/ICESC48915.2020.9156045
15. Method of fractal traffic generation by a model of menerator on the graph / [H. Drieieva, O. Smirnov, O. Drieiev et al.] // Proceedings of the 2nd International Workshop on Control, Optimisation and Analytical Processing of Social Networks (COAPSN). – 2020. – P. 366–379.
16. Network traffic forecasting based on fixed telecommunication data using deep learning / [M. Alizadeh, M. Beheshti, A. Ramezani, H. Saadatinezhad] // Proceeding of the IEEE 6th Iranian conference on signal processing and intelligent systems (ICSPIS). – 2020. – P. 1–7. https://doi.org/ 10.1109/ICSPIS51611.2020.9349573
17. Vinayakumar R. Applying deep learning approaches for network traffic prediction / R. Vinayakumar, K. P. Soman, P. Poornachandran // Proceeding of the IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI). – 2017. – P. 2353–2358. https://doi.org/ 10.1109/ICACCI.2017.8126198
18. Aloraifan D. Deep learning based network traffic matrix prediction / D. Aloraifan, I. Ahmad, E. Alrashed // International Journal of Intelligent Networks. – 2021. – Vol. 2. – P. 46–56. https://doi.org/10.1016/j.ijin.2021.06.002
19. Fan J. Research on network traffic prediction model based on neural network / J. Fan, D. Mu, Y. Liu // Proceeding of the IEEE 2nd International conference on information systems and computer aided education (ICISCAE). – 2019. – P. 554–557. https://doi.org/ 10.1109/ICISCAE48440.2019.221694

DOI 10.15588/1607-3274-2023-4-1

20. Traffic prediction based on random connectivity in deep learning with long short-term memory / [Y. Hua, Z. Zhao, Z. Liu et al.] // Proceeding of the IEEE 88th Vehicular technology conference (VTC-Fall). – 2018. – P. 1–6. https://doi.org/ 10.1109/VTCFall.2018.8690851
21. Lu H. Research on network traffic prediction based on long short-term memory neural network / H. Lu, F. Yang // Proceeding of the IEEE 4th International conference on computer and communications (ICCC). – 2018. – P. 1109–1113. https://doi.org/ 10.1109/CompComm.2018.8781071
22. Prediction of data traffic in telecom networks based on deep neural networks / [Q. H. Do, T. T. H. Doan, T. V. A. Nguyen, V. V. Linh] // Journal of computer science. – 2020. – Vol. 16, No. 9. – P. 1268–1277. https://doi.org/ 10.3844/jcssp.2020.1268.1277
23. Dynamic modification neural network model for short-term traffic prediction / [D. Guo, X. Xia, L. Zhu, Y. Zhang] // Procedia Computer Science. – 2021. – Vol. 187. – P. 134–139. https://doi.org/10.1016/j.procs.2021.04.043
24. Jain G. Machine learning, prophet and XGBoost algorithm: analysis of traffic forecasting in telecom networks with time series data / G. Jain, R. R. Prasad // Proceeding of the IEEE 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). – 2020. – P. 893–897. https://doi.org/ 10.1109/ICRITO48877.2020.9197864
25. A network traffic prediction method based on LSTM / [W. Shihao, Z. Qinzheng, Y. Han et al.] // ZTE Communications. – 2019. – Vol. 17, No. 2. – P. 19–25. https://doi.org/ 10.12142/ZTECOM.201902004
26. A hybrid method for realistic network traffic with temporal convolutional network and LSTM / [J. Bi, X. Zhang, H. Yuan et al.] // IEEE Transactions on Automation Science and Engineering. – 2022. – Vol. 19, No. 3. – P. 1869–1879. https://doi.org/ 10.1109/TASE.2021.3077537
27. Network prediction with traffic gradient classification using convolutional neural networks / [T. Ko, S. M. Raza, D. T. Binh et al.] // Proceeding of the IEEE 14th International conference on ubiquitous information management and communication (IMCOM). – 2020. – P. 1–4. https://doi.org/ 10.1109/IMCOM48794.2020.9001712
28. Magro V. Method for evaluating the delay time in a stream broadcast process / V. Magro, V. Svyatoshenko, D. Tymofieiev // Information Processing Systems. – 2019. – Vol. 159, No. 4. – P. 28–35. https://doi.org/10.30748/soi.2019.159.03
29. Sage A. P. Estimation theory with applications to communications and control / A. P. Sage, J. L. Melsa. – New York : McGraw-Hill, 1971. – 529 p.
30. Magro V. I. Investigation of information network loading in the condition of remote education and remote monitoring / V. I. Magro, S. V. Plaksin, V. O. Syatoshenko // Applied questions of mathematical modeling. – 2021. – Vol. 4, No. 2.1. – P. 142–149. https://doi.org/10.32782/KNTU2618-0340/2021.4.2.1.15

DOI 10.15588/1607-3274-2023-4-1

UDC 621. 618:519

THE METHOD OF OPTIMIZING THE DISTRIBUTION OF RADIO SUPPRESSION MEANS AND DESTRUCTIVE SOFTWARE INFLUENCE ON COMPUTER NETWORKS

Sholokhov S. M. – PhD, Associate Professor, Associate Professor of the Department Special Telecommunication Systems, Institute of Special Communication and Information Protection of National Technical University of Ukraine "Kyiv Polytechnic Institute named Igor Sikorsky", Kyiv, Ukraine.

Pavlenko P. M. – Dr. Sc., Professor, Professor of Department Organization of Air Transportation of the National Aviation University, Kyiv, Ukraine.

Nikolaienko B. A. – PhD, Associate Professor, Associate Professor of the Department Special Telecommunication Systems, Institute of Special Communication and Information Protection of National Technical University of Ukraine "Kyiv Polytechnic Institute Named Igor Sikorsky", Kyiv, Ukraine.

Samborsky I. I. – PhD, Senior Research Officer, Associate Professor of the Department Special Telecommunication Systems, Institute of Special Communication and Information Protection of National Technical University of Ukraine "Kyiv Polytechnic Institute named Igor Sikorsky", Kyiv, Ukraine.

Samborsky E. I. – Post-graduate student of Department Organization of Air Transportation of the National Aviation University, Kyiv, Ukraine.

ABSTRACT

Context. Currently, generalized methodical approaches to the development of scenarios of complex radio suppression and electromagnetic influence of typical special telecommunication systems have been developed. However, during the development of possible cases for the complex application of radio suppression and destructive software influence,the problem of optimizing the resource of these means and its distribution according to the goals of radio suppression and objects of destructive computer influence arose, which has not yet been fully resolved.Especially in the literature known to the authors, there is no method for optimizing the resource distribution of radio and computer influence, used for the development and practical implementation of optimal scenarios of destructive influence on computer networks of enemy military groups in military operations.

Therefore, it is necessary to formulate a problem and develop a method of optimizing the distribution of the resource of radio suppression and destructive software influence for the development of possible scenarios of the enemy's violation of information exchange in a standart telecommunication network.

Objective. The purpose of the research is to develop a method for optimizing the distribution of the resource of radio suppression and destructive software influence for the development of scenarios of information exchange violations by the enemy in the telecommunications network.

Method. To achieve the purpose of the research, the methods of nonlinear optimization of heterogeneous resource distribution, mass service theory, and expert evaluation were comprehensively applied and developed in the field of modeling of information conflict.

To determine the coefficients of protection of objects from radio-electronic and destructive computer influence, expert evaluation methods are used, in particular, the method of frequencies of preferences of the decision-maker using the Thurstone method. This method requires only one expert (a decision-maker), minimal communication time with him, minimal expert information (full ordering of weighting factors) and can be applied with a small number of evaluated weighting factors.

To solve the problem of optimal distribution of a heterogeneous resource of means of destructive influence, to ensure the value of the multiplicative objective function of an arbitrary form is not less than the given one, the method of successive increments is applied.

To determine the efficiency indicator of information exchange violation, the methods of mass service theory are applied, which allows to formalize special telecommunication systems as a set of mass service systems – subsystems of digital communication and computer networks.

Results. The formulated problem and the entered indicators made it possible to solve the problem of determining the minimum resource of means of destructive influence and their optimal distribution according to the purposes of radio suppression on the objects of destructive program influence in order to achieve the required level of disruption of the efficiency of information exchange in special telecommunication systems.

Conclusions. According to the results of the article, a method for optimizing the distribution of the resource of radio suppression and destructive software influence has been developed for the development of possible scenarios of information exchange violations by the enemy in a typical telecommunications network.The verification of the proposed method was carried out by comparing the theoretical results with the results of simulated modeling of scenarios of violation of the information exchange in the telecommunications network by the enemy.

KEYWORDS: information exchange, computer radio network, computer attack, protection of information, radio suppression, optimization of resource allocation, destructive influence.

ABBREVIATIONS

TN is a telecommunication network;
TSM is a tactical section of management;
LCRN is a local computer radio network;

DOI 10.15588/1607-3274-2023-4-2

DPI is a destructive program influence;
LCN is a local computing network;
TPN is a tactical packet network;
CCCN is a combat control computer network;
RES is a radio-electronic struggle;
DST is a destructive software tool;
UAV is an unmanned aircraft vehicle;
WS is a workstation;
UAVTM is an unmanned aerial vehicle – trouble maker;
MSSis a mass service system;
COCC is acombat operations control center;
COT is a compact obstacle transmitter;
DCI is a destructive computer influence;
GCN is a global computing network;
OTS is an operational tactical situation;
SRG is a subversive reconnaissance group;
CIT is a compact interference transmitter;
TS is a telecommunication system;
EI is an electromagnetic influence.

NOMENCLATURE

β is a intensity of information aging;

H is a number of control units in which the enemy solves the task of disorganization;

h is a control link number;

k^{pws} is a security weight factor of workstations;

k^{pp} is a security weight factor of DCI targets;

k^{pnc} is a security weight factor of typical TPN nodal centers;

k_h^{pws} is a security weighting coefficient of the PC in the h-th link of management when the DST is introduced;

k_v^{pp} is a security weighting coefficient of the DCItarget of the v-th type;

k_h^{pnc} is a security weight factor nodal center TNin the$h$-th management chain;

l is a number of ratings on the x_i and x_j indicator scales;

m is a number of compared pairs in the plane Pk_iOPk_j;

N_h^{nc} is a number of workstations in LCNTN;

N_h^{nc} is a number of nodal centerTNin the h-th chain of management;

p_{ij} is a percentage ratio of the number of pairs of objects for which the arrow is “directed from criterion$i$ to criterion$j$”;

P_h^{pui}, P_{nec}^{pui} is a probability of untimely receipt of information about the crisis situation in the h-th link of management and its necessary importance for the disorganization of state administration, respectively;

$Q_{hv}(r_{hv})$ is a hv-th component of the transformed objective function;

$R\left(w^{mcsd}, w^{dis}\right)$ is a resource spent by the enemy on the implementation of the PIІ scenario of the nodal centers of TN and DCI on computer networks;

$R_h\left(w_{hv}^{mcsd}, w_h^{dis}\right)$ is a function of the resource used to implement the method of radio suppression and destructive software influence on computer networks in the h-th link of control;

$r_{hv}\left(w_{hv}^{mcsd}, w_h^{dis}\right)$ is a hv-th resource function component;

T^{ri} is a time of receiving information about changes in the situation in the event of a crisis situation;

T^{ct} is a critical delay time for information about a crisis situation;

$T_h^{ri}\left(w_{hv}^{mcsd}, w_h^{dis}\right)$ is a time of receiving information about the operational and tactical situation in the h-th chain of command;

T_{h-1}^{ri} is a time of receiving information about the crisis situation in the relevant management link, at which it achieves the fulfillment of the suppression efficiency criterion;

$\overline{T}_{hv}\left(w_{hv}^{mcsd}, w_h^{dis}\right)$ is an average time of transaction data processing in the target of the v-th type of DPI in the h-th management link;

T_h^{ct} is an aging time of information about the crisis situation in the h-th management link;

V is a number of types of DST targets in computer networks of TN;

v is a number of the DST target type;

w^{mcsd} is a number of DST used to suppress computer networks;

w^{dis} is a number of UAV-TM (TPN) used to suppress TN nodal centers;

w_{hv}^{mcsd} is a number of DST, which are implemented in PCs of local computing networks in each h control link to suppress each v type of DCI targets;

w_h^{dis} is a number of UAV-TM s used to suppress TPN nodal centers in each h control link;

z_{ij} is an unknown normalized variable;

$Q_h(R_h)$ is a transformed objective function.

INTRODUCTION

The development of organizational and technical methods of protection of TN, containing radio networks, mobile and computer means of TSM, which are consolidated by means of protected radio lines in LCRN and integrated into global (regional) LCRN (hereinafter, typical TN) is extremely relevant in the conditions of conducting hybrid military operations [1–9]. At the same time, there is an urgent need to develop scenarios of possible actions of the enemy during the conduct of radio suppression and computer attacks on TN [3–11].

DOI 10.15588/1607-3274-2023-4-2

The objectof study – is the processes of distribution of a heterogeneous resource of means of destructive influence on information exchange.

The subjectof study – is a method of optimizing the distribution of a heterogeneous resource of means of destructive influence on information exchange.

The purpose of the work is to develop a method for optimizing the distribution of resources of radio suppression and DPI for the development of scenarios of violation of information exchange in TN by the enemy.

1 PROBLEM STATEMENT

A military operation is being considered, during which the task of disorganization by the enemy of the command of the opposing army corps by means of radio suppression and destructive influence on the computer network is solved [1–3, 4, 6, 9].

Analysis of military control systems shows that computer networks of combat control, the structural diagram of which is shown in Fig. 1, consist of LCN, LCRN, TPN, which are hierarchically combined in the CC CN [1, 3, 6, 7, 9–14]:

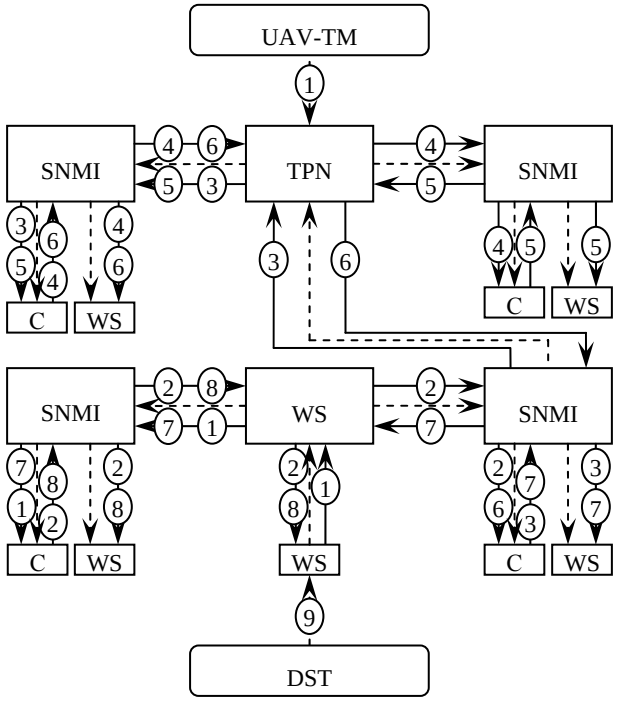


Figure 1 – Structural diagram of a model of the construction of computer networks of a typical army corps of the armed forces, as an object of RESand DCI in modern military operations

– workstations of platoon, company, and battalion commanders united by the ADDSI protocol (analogous to X.25) by means of the combat radio communication system (EPLRS, NTDR, SINCGARS SIP);

– workstations and servers of automated information systems of combat control centers of battalions, brigades, divisions and army corps, connected by cable lines using the Ethernet protocol.

The GCN is formed by local computer networks of various control units, united by means of a tactical packet network according to the X.25 protocol. TPN consists of routers, packet switches of node centers and separate radio relay communication channels of the MSE type system [1, 2]. The structural diagram of the mathematical calculation model of radio and software-computer suppression of army corps computer networks is shown in Fig. 1. Accordingly, data is exchanged between all vertical and horizontal links of the typical army corps of the armed forces of the leading countries of the world.

The adversary, in the conditions of conducting hybrid military operations, affects the CC CN in a complex way, combining the radio-electronic influence of RES devices on lines, node centers and communication lines of a typical TN with means of radio suppression and DST on elements of wireless computer networks [1, 2, 11–17].

The grouping of the enemys RES forces and means includes software-computer and radio suppression subsystems.

As part of the software-computer suppression subsystem, there is a set of DST.The radio suppression subsystem includes jamming transmitters on the UAV, Fig. 1.

DST using special sabotage actions are introduced in the LCN working stations of the army corps; UAV are launched in the areas of the TPN nodal center and create radio interference.

Introduced into the DST workstation, fig. 1, carries out a “denial of service” attack, causing a decrease in the speed of data processing by servers, overloading the LCN bus and duplex TPN communication channels with data packets; UAV-TM suppress the TPN nodal center, causing redirection of data flows through unsuppressed nodal centers and overloading of the corresponding TPN duplex communication channels, fig. 1.This leads to an increase in the time of transmission of information about the operational and tactical situation and, ultimately, its aging at the time of receipt by the concerned officials.

Computer networks of the army corps, fig. 1, are presented as a set of hierarchically united MSS, and DCI, UAV-TM affect their functioning by overloading [12, 18]. Here, the MSS are marked with rectangles: C is a two-node, two-phase MSS with an intermediate storage device of finite capacity, which simulates a server; WS – a two-node two-phase MSS with an intermediate storage device of finite capacity, which simulates a workstation; SNMI – single-node multi-input with random selection of a service request, which simulates an Ethernet bus; PM – single-node multi-input round-robin polling in the case of time distribution, which simulates the EPLRS radio network; TPN is a multi-node single-phase with waiting, which simulates the TPN of the MSE system. Circles indicate the process of transferring data about OTS: 1 – from the WS of reconnaissance means, platoon, company, battalion commanders to the battalion’s COCC server; 2 – from the battalion’s COCC server to the brigade’s COCC server and the battalion’s WS; 3 – from the brigade’s COCC server to the division’s

DOI 10.15588/1607-3274-2023-4-2

COCC server and brigade's COCC server; 4 – from the division's COCC server to the army corps'COCCserver and the division's WSCOCC; 5 – from the army corps COCC server to the division's COCC server and the army corps'WSCOCC; 6 – from the division's COCC server to the brigades COCC server and the divisions WS COCC;7 – from the brigade COCC server to the battalion COCC server and the brigade WSCOCC; 8 – from the battalion's COCC server to the battalion's WSCOCC and the WS of platoon, company, and battalion commanders. Dashed arrows indicate the impact on the elements of the MSS;9 – introduction of DCI into the computer network by the software-computer suppression subsystem; 10 – radio suppression of the TPN nodal center.

It is necessary to optimize (solve the task of minimizing the resource objective function under a set of constraints) the resource of means of radio suppression and destructive software influence of the enemy on a typical computer network of the AK (according to Fig. 1 for use in a modern military operation.

With,as a resource target function for the implementation of the appropriate scenario of radio suppression and DCI on computer networks, it is advisable to consider the weighted additive number of DST intended for suppression of various targets of DCI and jamming transmitters on UAV of CIT, entered by SRG, intended for radio suppression of nodal centers typical TN:

$$R\left(w^{mcsd}, w^{dis}\right) = k^{pws} k^{pp} w^{mcsd} + k^{pnc} w^{dis}. \qquad (1)$$

The task of optimal distribution of the resource used to implement scenarios of radio suppression of the nodal center, communication lines and DCI on computer networks of a typical TN is specified as follows: form the following matrix $W^{mcsdopt} = \left\| w_{hv}^{mcsdopt} \right\|$, $h = 1...H$, $v = 1...V$ of minimum quantities of DPI means, which are introduced in LCNworkstations in each h chain of command to suppress each v type of DCItargets, and the following matrix $W^{disopt} = \left\| w_h^{disopt} \right\|$, $h = 1...H$, minimum quantitiesof minimum number of UAV-TM, COT, used to suppress TNnodal centers in each h control chain, which ensure the fulfillment of the criterion of effectiveness of radio suppression and DSI $P_h^{pui} \ge P_{pec}^{pui}$, taking into account the weighting coefficients of security of WS k_h^{pws}, DPI goals k_v^{pp} and nodal centers, TN communicationlines k_h^{pws}:

$$S^{opt} = \arg\min\left\{R_h\left(w_{hv}^{mcsd}, w_h^{dis}\right)\right\}, \qquad (2)$$

$$R_h\left(w_{hv}^{mcsd}, w_h^{dis}\right) = \sum_{v=1}^{V} r_{hv}\left(w_{hv}^{mcsd}, w_h^{dis}\right) =$$

$$= \sum_{v=1}^{V} k_h^{pws} k_v^{pp} w_{hv}^{mcsd} + k_h^{pnc} w_h^{dis}, \qquad (3)$$

$$h = 1...H,\ v = 1...V,$$

subject to restrictions:

$$p_h^{pui}\left(w_{hv}^{mcsd}, w_h^{dis}\right) \ge p_{nec}^{pui}, \qquad (4)$$

$$w_{hv}^{mcsd} \in \{0,1,2,..., N_{ws}\},\ w_h^{dis} \in \left\{0,1,2,..., N_h^{nc}\right\}, \qquad (5)$$

$h = 1...H,\ v = 1...V.$

2 REVIEW OF THE LITERATURE

General issues of conducting radio suppression and DCI on computer networks are considered in [1–10].

Approaches to computer networks modeling are also described in [12–18], criteria for assessing their security are presented in [15–18].

However, the known results are not specified for solving the tasks of applying the scenario approach in the development of methods of protecting TN from radio suppression and DST of the enemy, taking into account indicators of violation of the efficiency of information exchange [4–18].The specifics of the targeted practical use of the object of research considered in this article are relatively little presented in the publications open to print.According to the authors, a certain generalization and the possibility of civilian use of the research results allow the results to be put to the public's discretion.

The simulation modeling carried out by the authors showed that in practice, under the conditions of modern measures for the radio-electronic protection of TN of the army corps, the enemy needs to increase the number of radio suppression means of UAV-TM, COT and DCI (hereinafter – the resource) in order to achieve the required level of disruption of the efficiency of information exchange, but the procedure for optimizing means has not been developed [4–9, 15–18].

Methodical approaches to the development of complex radio suppression and EI scenarios of typical TN are developed in [4–9, 12, 15–20]. It was concluded that in the practice of conducting hybrid military operations, the adversary for radio suppression of the hub center and communication lines of typical TN can comprehensively use radio suppression means on UAV-TM, COT, which carry SRG, and for the impact on wireless computer networks – DST [4–7].

At the same time, during the development of possible scenarios for the complex use of radio-suppression means and DST, the task of optimizing the resource of these means and its distribution for the purposes of radio-suppression and the DCI object arose.

The procedures for optimizing the resources of the means of destructive influence on TN and optimal

DOI 10.15588/1607-3274-2023-4-2

distribution for radio suppression targets and DST objects in order to achieve the desired level of violation of the efficiency of information exchange in TN by the enemy in a special period have not been developed in the foreign and domestic literature known to the authors.

Therefore, the task of developing a method of optimizing the distribution of the resource of radio suppression means and destructive software influence for the development of scenarios of violation by the enemy of information exchange in the telecommunications network is relevant and is considered in this article.

3 MATERIALS AND METHODS

To solve the optimization problem in the statement (1)–(6) and achieve the goal of the article, it is necessary to solve a number of partial scientific problems.

First of all, developing the procedure for determining the weighting coefficients of the objective function (1) requires a further scientific solution.

The issue of determining the weighting coefficients of the objective function (1) requires further scientific resolution.

The conducted analysis showed that in order to determine the weighting coefficients when searching for the extremum (1), it is potentially possible to consider methods of approximation of the utility function in generalized convolutions [4–6]. It was concluded that known methods of pairwise comparisons, point estimates on a frequency scale, and individual preferences determine the weighting factors that are difficult to use in generalized convolutions. At the same time, the methods of approximation of the utility function are used only when the utility function can be represented in an additive form, while the weighting coefficients are determined according to the contribution of the components to the total utility [18–20].

However, the weighting coefficients calculated by such methods differ in essence from the weighting coefficients k^{pws}, k^{pp}, k^{pnc} of the security of radio suppression targets, because the latter are not part of the efficiency function and, on the contrary, express the resistance of radio suppression targets to its increase. Then, to estimate k^{pws}, k^{pp} the weighting coefficients k^{pws}, k^{pp}, k^{pnc} of the security of radio suppression targets, methods of pairwise comparisons, precise evaluations on a scale, frequencies of person preferences can be applied, but their application requires, as input data, expert information. In such a situation, the solution to the problem lies in the field of integration of methods for determining weighting factors and expert evaluation.

The analysis of the methods showed that a compromise option is the choice of the frequency method of the decision-maker's preferences [15–20] using the Thurstone method [14–20]. This method requires only one expert (a decision-maker), minimal communication time with him, minimal expert information (full ordering of weighting factors) and can be applied with a small number of evaluated weighting factors.

The essence of the frequency method of the decision-maker's preferences for solving problem (1) is as follows.It is necessary to determine the weighting coefficients k^{pws}, k^{pp}, k^{pnc} of the qualitative indicators "security of the WS" x^{pws}, "security of the DPI target" x^{pp}, "security of the nodal center of the TN" x^{pws} respectively. For each of the specified groups of indicators, a procedure is carried out, which consists of the following stages:

1) a single ordinal scale is developed for all indicators so that the minimum quality for each indicator corresponds to the origin of the coordinates of the space of indicators $X = x_1 \times x_2 \times ... \times x_n$;

2) the person making the decision compares all the objects located in each coordinate plane $x_i O x_j$,connecting them with arrows, and the arrows areplaced from the best object to the worst;

3) the number of arrows directed from indicator*i*to indicator *j* is counted, which characterizes the importance of indicator*i*in relation to indicator*j*. The total number of arrows a_{ij} will be the number of cases in which indicator*i*is more important than indicator*j*. According to the results of the calculation, the matrix $A = \|a_{ij}\|$ of the Thurstone method is formed;

4) the *P* matrix is constructed. *P*– is the proportion of cases when indicator *i* was more important than indicator *j*: $P = \|p_{ij}\|, 1 \le i, j \le n$, where p_{ij} – percentage ratio of the number of pairs of objects for which the arrow is "directed from criterion *i* to criterion *j*", moreover $p_{ij} = \frac{a_{ij}}{m}$, where $m = \frac{l^2\left(l^2 - 2l + 1\right)}{4}$. The matrix element *P* satisfies the condition $p_{ij} + p_{ji} = 1$;

5) p_{ij} expressed in standard deviations are determined by the iterative method:

$$p_{ij} = \int_{-\infty}^{z_{ij}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz; \qquad (6)$$

6) the "importance" of indicator *i*, expressed in standard deviations, is calculated $\bar{z}_i = \frac{z_i}{n}$ де $z_i = \sum_{j=1}^{n} z_{ij}$;

7) formula (6) determines the probability p_i, corresponding to $\bar{z}_i$;

8) normalization of p_i is carried out, and the weighting factor is found k_i.

So, the proposed method of frequency preferences of the decision-maker allows solving the problem of determining the security coefficients of radio suppression

DOI 10.15588/1607-3274-2023-4-2

and DCI targets, which can be used in optimizing the allocation of the resource used for the implementation of radio suppression and DPI scenarios on TN elements. The method requires minimal communication time with only one expert and minimal expert information.

Secondly, when solving problems (1)–(6) the next important stage of optimization is the determination of constraints on the parameter of the objective function (1).

Based on the results of the imitative simulation, it was determined that in practice, radio suppression and DCI on TS elements leads to an increase in the probability of untimely "delivery" of information about the crisis situation by officials (authorities) of various branches of state administration. In Fig. 1 presents the graphs of the dependence of the probability $p_{rs}^{ld}(t)$ in different control links on the intensity of interfering packets for different values of the number of WS s in which DST are introduced. An increase in the intensity of packets leads to an increase in the probability of untimely receipt of information about a crisis situation, while suppressing server disruptions is more effective at lower levels of management.

Graphs of the dependence of the probability of untimely "delivery" of information about a crisis situation on the number of WS s, in which DST are implemented to suppress the server, Ethernet LCN buses, packet networks have a non-linear nature, therefore, when solving the problem of optimizing the resource allocation of radio suppression means and DPI of TN, it is necessary to apply non-linear methods.

Taking into account the limitations on the scope of the publication, we specify the least covered in the literature and researched in Ukraine issues of probability determination [4].

$$p_{rs}^{ld}(t) = P\left(T^{ri} \le T^{ct}\right) = 1 - e^{-\beta}, \text{ where } \beta = \frac{T^{ri}}{T^{ct}}. \quad (7)$$

The methods of determining the indicator T^{ri} in each specific case are determined by the features of the TNconstruction. At the same time, the TScan be formalized as a set of MSS`s – subsystems of digital communication and computer networks [4–9, 12, 18].

When determining the parameter T^{ri} in the conditions of the DPI for a typical TN, it is advisable to model computer networks as a set of hierarchically combined MSS`s. To implement the proposed approach, in [4] a methodological basis for mathematical modeling of DCI on TN computer networks was developed, in which computer networks are represented as a set of hierarchically combined MSS`s, and the introduced DST and radio suppression means affect their functioning by overloading.

Thirdly, it is necessary to choose a method of solving problems (1)–(6). Problem (1)–(6) is the problem of determining the minimum necessary resource to ensure the value of the multiplicative objective function of an arbitrary form is not less than the specified one.The optimal distribution of the resource can be solved by the method of successive increments [19–20].

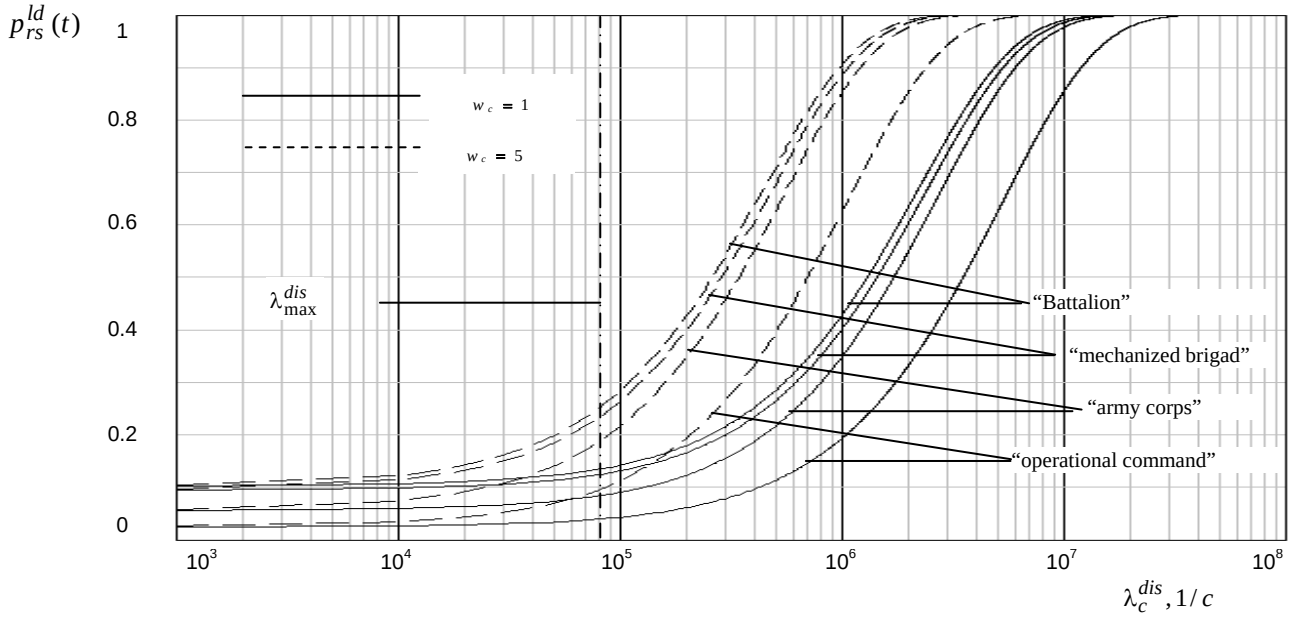


Figure 2 – Dependence of the probability of untimely "delivery" of information about a crisis situation in the control units "tactical unit", "connection", "unification" on the intensity of interfering packets for different values of the number of WS s of the TN computer network

DOI 10.15588/1607-3274-2023-4-2

The method of successive increments allows you to effectively solve problems such as: finding a series $X^o = \{x_i^o\}_n = \min\limits_{x_i}\{x_i\}_n = \min\limits_{x_i}\sum\limits_{i=1}^{n} x_i$, which ensures the value of the additive function is not less than the specified:

$$F(X) = \sum_{i=1}^{n} F_i(x_i) \geq F^{giv}(X) \tag{8}$$

with a linear restriction on the variables:

$$\left(X = \sum_{i=1}^{n} x_i \right) \leq b,\ b > 0, \tag{9}$$

under the conditiontheir integers and non-negativity:

$$x_i \in \{0,1,2,...,b\}, i = 1,2,...,n. \tag{10}$$

To reduce the problem (1)–(6) to the conditions (7)–(9), it is necessary to ensure the additive nature of the target function, the independence of its components, as well as the transition from variables w_{hv}^{mcsd}, w_h^{dis} to the function $R_h\left(w_{hv}^{mcsd}, w_h^{dis}\right)$.

Fulfillment of the condition of additivity of the target function will be ensured by the transition from the function of the probability of untimely receipt of information about the crisis situation to the function of the time of receipt of information about the operational-tactical situation, which is additive. Then, we get the objective function:

$$T_h^{ri}\left(w_{hv}^{mcsd}, w_h^{dis}\right) = T_{h-1}^{ri} + \sum_{v-1}^{V} \overline{T}_{hv}\left(w_{hv}^{mcsd}, w_h^{dis}\right) \geq$$

$$\geq -T_h^{ct} \ln\left(1 - p_{nec}^{pui}\right),\ h = 1...H. \tag{11}$$

In cases where some components $\overline{T}_{hv}\left(w_{hv}^{mcsd}, w_h^{dis}\right)$ of the objective function turn out to be dependent, it is advisable to switch to their combinations, which are formed by sequentially sorting the values of one parameter while the second parameter is fixed, while the formed components will be independent.

Sinceforeachvalue w_{hv}^{mcsd}, w_h^{dis} thevaluesareknown $T_h^{ri}\left(w_{hv}^{mcsd}, w_h^{dis}\right)$ and $R_h\left(w_{hv}^{mcsd}, w_h^{dis}\right)$, then, putting the value R_h as independent, it is possible to assume that the value of the function is also known $Q_h(R_h)$, because it is equal to the value of $T_h^{ri}\left(w_{hv}^{mcsd}, w_h^{dis}\right)$. So, the set of function values $T_h^{ri}\left(w_{hv}^{mcsd}, w_h^{dis}\right)$ is uniquely mapped to the set of function values $Q_h(R_h)$, which ensures the transition from variables w_{hv}^{mcsd}, w_h^{dis} to the function $R_h\left(w_{hv}^{mcsd}, w_h^{dis}\right)$:

$$T_h^{ri}\left(w_{hv}^{mcsd}, w_h^{nc}\right) = T_{h-1}^{ri} + \sum_{v=1}^{V} \overline{T}_{hv}\left(w_{hv}^{mcsd}, w_h^{nc}\right) =$$

$$= T_{h-1}^{ri} + \sum_{v=1}^{V} F_{hv}\left(w_{hv}^{mcsd}(R_h), w_h^{nc}(R_h)\right) =$$

$$= T_{h-1}^{ri} + \sum_{v=1}^{V} Q_{hv}(r_{hv}) = Q(R_h). \tag{12}$$

Then, taking into account (8)–(12), problem (4)–(7) will take the form:

$$S^{opt} = \arg\min\left\{R_h\left(w_{hv}^{mcsd}, w_h^{dis}\right)\right\}, \tag{13}$$

$$R_h\left(w_{hv}^{mcsd}, w_h^{dis}\right) = \sum_{v=1}^{V} r_{hv}\left(w_{hv}^{mcsd}, w_h^{dis}\right) =$$

$$= \sum_{v=1}^{V} k_h^{pws} k_v^{pp} w_{hv}^{mcsd} + k_h^{pnc} w_h^{dis}, \tag{14}$$

$$h = 1\ldots H, \quad v = 1\ldots V,$$

subject to restrictions:

$$\Theta_h\left(R_h\left(w_{hv}^{mcsd}, w_h^{nc}\right)\right) \geq -T_h^{ct} \ln\left(1 - P_{nec}^{pui}\right), \tag{15}$$

$$w_{hv}^{mcsd} \in \{0,1,2,\ldots,N_{ws}\}, \quad w_h^{dis} \in \{0,1,2,\ldots,N_h^{nc}\}, \tag{16}$$

$$h = 1...H,\ v = 1...V.$$

For optimization, we will use the process of sequential distribution of the resource w_{hv}^{mcsd}, w_h^{dis} in some portions $\Delta w_{hv}^{mcsd(t)}, \Delta w_h^{dis(t)}$ respectively, at the *t*-th step of the process. If, at the *t*-thstep of the process, the argument of the *hv*-thcomponents of the resource function receives an increment $\Delta w_{hv}^{mcsd(t)}, \Delta w_h^{dis(t)}$, then will receive the increment $\Delta r_{hv}^{(t)}$ of *hv*-thcomponent of the resource functionand, accordingly, the transformed target function:

$$\Delta Q_h\left(R_h^{(t)}\right) = \Delta Q_{hv}\left(r_{hv}^{(t)}\right) =$$

$$= \Delta Q_{hv}\left(r_{hv}^{(t-1)} + \Delta r_{hv}^{(t)}\right) - Q_{hv}\left(r_{hv}^{(t-1)}\right) = \Delta Q_{hv}. \tag{17}$$

DOI 10.15588/1607-3274-2023-4-2

The average efficiency of use of each of the Δr_{hv} resource units at the t-th step of the process is determined by the ratio:

$$e_{hv}^{(t)} = \frac{\Delta Q_{hv}\left(r_{hv}^{(t)}\right)}{\Delta r_{hv}^{(t)}},\ h = 1...H,\ v = 1...V. \qquad (18)$$

The optimization method will consist in the sequential distribution of the resource in portions $\Delta r_{hv_t}^{(t)}$, the value of which and the index hv_t are determined in accordance with the maximum efficiency of the use of each unit of the resource at each step of the process.

The optimal step value $\Delta r_{hv}^{(m)}$ is determined from the condition:

$$\left(e_{hv}^{(m)} = \frac{\Delta Q_{hv}^{(m)}}{\Delta r_{hv}^{(m)}}\right) = \max_{\Delta r_{hv}}\left(e_{hv} = \frac{\Delta Q_{hv}}{\Delta r_{hv_t}}\right), \qquad (19)$$

$$h = 1...H,\ v = 1...V.$$

The index hv_t of the component of the resource function, at which the largest value of the quantity $e_{hv}^{(m)}$, is reached at the t-th step of the process, is determined in accordance with the condition:

$$e_{hv_t}^{(m)} = \max e_{hv_t}^{(m)},\ h = 1...H,\ v = 1...V. \qquad (20)$$

As a result, the matrices are obtained:

$$W^{mcsd\ opt} = \left\| w_{hv}^{mcsd\ opt} \right\|,\quad W^{dis\ opt} = \left\| w_{h}^{dis\ opt} \right\|,$$

$$h = 1...H,\ v = 1...V,$$

which form the optimal distribution of the resource used for the implementation of scenarios of radio suppression and DPI of the enemy on the TS.

So, the method makes it possible to optimize the allocation of the resource used to implement possible scenarios of radio suppression and computer impact of TS.

4 EXPERIMENTS

Based on the likely nature of the application, the initial data regarding the composition of the connections, parts and subdivisions of a typical army corps is shown in Table 1.

Control of connections and parts of army corps is carried out using an automated control system of the ATCCS type, the technical basis of which is computer networks.

Table 1 – Initial data for the experimental application of the resource optimization method

	Output data for scenario optimization	A typical stage of a military operation: conducting hostilities to maintain the defense line
"battalion"	the number of PC in the network of commanders	30
	the number of LCNCOCCat the most important radio-electronic objects	130
	the number of battalions in the brigade	14
"brigad	number of PC in LCNCOCC	34
	the number of LCNCOCCat the most important radio-electronic objects	11
	the number of brigades in the division	3
"division"	number of PC in LCNCOCC	30
	the number of LCNCOCCat the most important radio-electronic objects	8
	number of divisions in army corps	3
	the number of nodal centerTPN	15
	the deployment distance of the nodal center TPN from the battle line	60
"army corps"	number of PC in LCNCOCC	30
	the number of LCNCOCCat the most important radio-electronic objects	12
	the number of PC	60
	the number of nodal center TPN	14
	the deployment distance of the nodal center TPN from the battle line	60–80

The main components of computer networks are:

– computer tools:

– Appliqué-type workstations of company (battery) and platoon commanders under the control of Microsoft Windows operating systems;

– Appliqué-type workstations of officials of combat control centers of battalions, brigades, divisions, army corps under the control of Microsoft Windows operating systems;

– servers of automated information systems of combat control centers of battalions, brigades, divisions and army corps under the control of Microsoft Windows operating systems;

– routers of node centers of a tactical packet network under the control of the Sun Solaris operating system;

– means of communication:

– Means of the combat radio communication system of the "brigade and below" control units such as EPLRS, NTDR, SINCGARS SIP;

– radio relay means of the AN/TTC-46, AN/TTC-48, AN/TRC-190 type of expansion nodes of the automated communication system of the MSE type;

– radio relay means of the AN/TTC-47, AN/TRC-113, AN/TRC-143, AN/TRC-190 type of nodal centers of the automated communication system of the MSE type.

Officials' workstations and COCC servers operate under the control of Microsoft Windows and Sun Solaris family operating systems[1].

In order to disorganize the management of connections and parts of army corps, the enemy needs to achieve the effectiveness of methods of radio and software-computer suppression of computer networks not lower than 0.8 in the control links: "battalion" – $h = 1$;

DOI 10.15588/1607-3274-2023-4-2

"brigade" – $h=2$; "division" – $h=3$; "army corps"– $h=4$. For this purpose: 1) introduce software and computer suppression tools of the TCP, UDP flood type into the workstations of local computer networks – to suppress servers ($v=1$), type ICMP flood – to suppress the Ethernet bus ($v=2$), type ICMP flood – to suppress TPN lines ($v=3$), which carry out a "distributed denial of service" attack;2) suppress the TPN nodal centers in the areas of responsibility of divisions ($h=3$) and army corps ($h=4$) with jamming transmitters on UAV.

Then the tasks (1)–(4) of the optimal distribution of the means of radio and DST computer networks of the army corps in operations are specified as follows: to form the following matrix $W^{mcsdopt}=\left\|w_{hv}^{mcsdopt}\right\|$, $h=1,2...4$, $v=1,2,3$ of the minimum quantities of DST, which are implemented in WS s of local computer networks in each h chain of command to suppress each v type of DCI targets, and the following matrix $W^{disopt}=\left\|w_{hv}^{disopt}\right\|, h=3,4$ of the minimum number of UAV-TM s used to suppress TPNnodal centers in each h chain of control, which ensure the fulfillment of the intelligence electronic device efficiency criterion $p_h^{pui}\ge p_{nec}^{pui}=0.9$ taking into account the weighting coefficients of WS protection k_h^{pws}, the objectives of DCI k_v^{pp} and TPN nodal centers k_h^{pnc}:

$$S^{opt}=\arg\min\left\{R_h\left(w_{hv}^{mcsd},w_h^{dis}\right)\right\},$$

$$R_h\left(w_{hv}^{mcsd},w_h^{dis}\right)==\sum_{v=1}^{3}k_h^{pws}k_v^{pp}w_{hv}^{mcsd}+k_h^{pnc}w_h^{dis},$$

$$h=1,2,3,4;\ v=1,2,3.$$

under constraints of type (6)–(7).

In order to solve the task of optimal distribution of means of radio suppression of computer networks and DST, it is necessary to determine the aging time of information about the operational and tactical situation and the weighting factors of the security of LCNworkstations k_h^{pws}, DCI targets k_v^{pp}, TPN nodal centers k_h^{pnc}, Table 2.

Table 2 – The results of determining the aging time of information for various control units of a typical army corpsin a military operation

Probability of finding a typical object of shooting damage in the area	Information aging time in management units T_h^{ct}, hours.			
	"battalion"	"brigade"	"battalion"	"army corps"
$P^{id}=0.9$	0.3	1.4	1.6	4.2

To determine the weighting coefficients k^{pws},k^{pp},k^{pnc} $h=1,2,3,4;$ $v=1,2,3$ of the quality indicators "security of the WS" x^{pws}, "security of the target of the DCI" x^{pp} ,"security of the nodal centerTPN" x^{pnc} respectively, the following procedure was carried out:

– A single ordinal scale was developed for all indicators so that the minimum quality for each indicator corresponds to the origin of the coordinates of the indicator space $X^{pws}=x_1^{pws}\times x_2^{pws}\times x_3^{pws}\times x_4^{pws}$; $X^{pp}=x_1^{pp}\times x_2^{pp}\times x_3^{pp}$; $X^{pnc}=x_3^{pnc}\times x_4^{pnc}$; all objects lying in each coordinate plane are compared $X_i^{pws}OX_j^{pws}, X_i^{pp}Ox_j^{pp}, X_i^{pnc}Ox_j^{pnc}$, connecting them with arrows, and the arrows are placed from the best object to the worst;

– The calculated number of arrows directed from indicator i to indicator j, which characterizes the importance of indicator i in relation to indicator j. The total number of arrows $a_{ij}^{pws},a_{ij}^{pp},a_{ij}^{pnc}$; is the number of cases in which indicator$i$is more important than indicator j. According to the calculation results, the matrices of the Thurstone method are composed:

$$A^{pws}=\begin{Vmatrix}0&1&2&1\\8&0&1&0\\7&8&0&2\\8&9&7&0\end{Vmatrix},\quad A^{pp}=\begin{Vmatrix}0&8&5\\1&0&2\\4&7&0\end{Vmatrix},\quad A^{pnc}=\begin{Vmatrix}0&2\\7&0\end{Vmatrix};$$

– Constructed matrices P^{pws},P^{pp},P^{pnc} – are the fractions of cases when indicator$i$was more important than indicator j:

$$P^{pws}=\begin{Vmatrix}0.00&0.11&0.22&0.11\\0.79&0.00&0.11&0.00\\0.78&0.89&0.00&0.22\\0.89&1.50&0.82&0.00\end{Vmatrix},$$

$$P^{pp}=\begin{Vmatrix}0.00&0.93&0.59\\0.11&0.00&0.22\\0.42&0.78&0.00\end{Vmatrix},\quad P^{pnc}=\begin{Vmatrix}0.00&0.26\\0.88&0.00\end{Vmatrix}.$$

– Defined matrix elements P^{pws},P^{pp},P^{pnc}, expressed in standard deviations:

$$Z^{pws}=\begin{Vmatrix}0.00&-1.22&-0.76&-1.22\\1.23&0.00&-1.22&-2.00\\0.77&1.23&0.00&-0.76\\1.23&6.35&0.77&0.00\end{Vmatrix},$$

DOI 10.15588/1607-3274-2023-4-2

$$Z^{pp} = \begin{Vmatrix} 0.00 & 1.23 & 0.14 \\ -1.22 & 0.00 & -0.76 \\ -0.13 & 0.77 & 0.00 \end{Vmatrix}, \quad Z^{pnc} = \begin{Vmatrix} 0.00 & -0.76 \\ 0.77 & 0.00 \end{Vmatrix}.$$

– Calculated "importance" of indicator*i*, expressed in standard deviations:

$$\overline{Z}^{pws} = \begin{Vmatrix} -0.80 \\ -0.50 \\ 0.31 \\ 2.09 \end{Vmatrix}, \quad \overline{Z}^{pp} = \begin{Vmatrix} 0.46 \\ -0.66 \\ 0.21 \end{Vmatrix}, \quad \overline{Z}^{pnc} = \begin{Vmatrix} -0.38 \\ 0.39 \end{Vmatrix};$$

– Defined probabilities P^{pws}, P^{pp}, P^{pnc} that correspond to $\overline{Z}^{pws}, \overline{Z}^{pp}, \overline{Z}^{pnc}$:

$$P^{pws} = \begin{Vmatrix} 0.21 \\ 0.31 \\ 0.62 \\ 0.98 \end{Vmatrix}, \quad P^{pp} = \begin{Vmatrix} 0.68 \\ 0.25 \\ 0.58 \end{Vmatrix}, \quad P^{pnc} = \begin{Vmatrix} 0.35 \\ 0.65 \end{Vmatrix};$$

– normalization was carried out P^{pws}, P^{pp}, P^{pnc}, and the weighting coefficients of the security of workstations, targets of software and computer suppression, nodal centers of TPNarmy corps were found, which are presented in Table 3.

The determined optimal composition of radio suppression and DST means allow to disrupt army corps management at the appropriate stage of the operation, provided that the place and time of application are agreed upon.

Table3 – Results of optimizing the distribution of means of radio and software-computer suppression of computer networks of a typical army corps at the stage of maintaining the first line of defense

Control link of atypicalarmy corps	Number of DST introduced in PC LCNCOCC of control units to suppress targets			The number of interference transmitters on UAVs to suppress TPN nodal centers	Time of receiving information about OTS, which is achieved as a result of suppression, h.	The limit time of receiving information about OTS, at which the required suppression efficiency is achieved, h.
		Ethernet bust LCN	TPN			
"battalion"	8	0	2	8	2.8	2.4
"brigade"	0	0	4	2	4.2	4.1
"division"	5	0	4	0	7.31	6.1
"army corps"	0	0	4	2	12.41	12.2

Based on the simulation results, the dependences of the probability of untimely "delivery" of information about the situation (see Fig. 2) on the intensity of "interfering"packets were obtained. At the same time, the values of the number of PC s of the TN computer network, in which the DST were introduced, were chosen equal to 1 and 5, respectively. the intensity of interfering packets varied between $10^3$and 10^8.

It was concluded that an increase in intensity leads to an increase in the probability of untimely receipt of information about a crisis situation, while suppression of server disruptions is more effective at lower levels of management.

Based on the experimental application of the developed method, it was concluded that, in practice, the complex influence of radio suppression and DST on the elements of the TS of military systems leads to an increase in the probability of untimely "delivery" of information about the situation on the battlefield, which can lead to a breakdown in management in the relevant sections.

5 RESULTS

The task was formulated and a method was elaborated to optimize the resource of radio suppressors and DCI in the development of scenarios of the enemy's influence on the elements of a typical network of combat control of a military unification of army in military operations.

To determine the coefficients of protection of objects from radio-electronic influence and destructive computer influence of the target additive function, the method of the frequency of individual preferences is used, which finally combined the solution with the Terstoune`s method. The research showed that in order to solve the problem of optimal distribution of a heterogeneous resource of means of destructive influence, provided that the value of the multiplicative objective function of an arbitrary form is not less than the given one, it is advisable to apply the method of successive increments. At the same time, in order to determine the indicator of the efficiency of the contravention of information exchange, the methods of the mass service theory are used, which allows to formalize the combat control networks as a set of mass service systems – subsystems of digital communication and computer networks, and to propose a general method of optimizing the distribution of the resource of radio suppression and DPI, and development of scenarios of interruption by the enemy of information exchange in combat control networks in modern military operations. The essence of the proposed method is the sequential distribution of the resource in portions $\Delta r_{hv_t}^{(t)}$, the value of which and the index $_{hv_t}$ are determined in accordance with

DOI 10.15588/1607-3274-2023-4-2

the maximum efficiency of use of each resource unit at each step of the process. At the same time, the weighting coefficients of the security of radio targets and software-computer suppression of the target function (1) are determined based on the frequency method of the decision-maker's preferences.

Based on the results of the imitative simulation, it was concluded that in practice, radio suppression and DCI on TS elements leads to an increase in the probability of untimely "delivery" of information about the crisis situation by officials (bodies) of various branches of state administration.

6 DISCUSSION

The results obtained in the article are the development of a scenario approach for predicting possible actions of the enemy to disrupt the efficiency of information exchange in the networks of combat control of units (combinations) of army in military operations.

The obtained results are different from the known results in the field of optimization of resource distribution of computing networks. In the article, for the first time, it became possible to comprehensively apply the mathematical apparatus of the theory of optimizing the distribution of heterogeneous resources, mass service and expert evaluation, based on the development in the field of substantiation of scenarios of disruption of information exchange in combat control networks within the framework of information confrontation in modern military operations. This made it possible for the first time to carry out a task statement and develop a method of optimizing the resource of radio suppression and DCI in modern military operations.

The method uses the well-known mathematical apparatus of the theory of optimal resource distribution in conditions of non-linearity of the additive resource function. However, for the first time, additional restrictions on the resource target function specific to military operations, which are due to the achievement of the desired infringement of the operational efficiency of information exchange in combat control networks that are in the information conflict on the battlefield, were taken into account.

In contrast to the known results, which are close in terms of the research direction, the peculiarities of the additive nature of the target function are taken into account, when developing scenarios of radio suppression and DPI of elements of combat control networks, an approach is developed to take into account the security coefficients of the corresponding elements of combat control networks, which are objects of radio suppression and DPI in the information struggle. The solution to this problem can be obtained by the frequency method of the decision-maker's preferences [14–20] using the Terstoune`s method [14–20]. It is substantiated that this method requires only one expert (a decision-maker), minimal communication time with him, minimal expert information (full arrangement of weighting factors) and can be applied with a small number of evaluated weighting factors.

To fulfill the condition of additivity of the target function, a transition from the function of the probability of untimely receipt of information about the situation on the battlefield to the function of the time of receipt of information about the operational-tactical situation, which is additive, was ensured. At the same time, in order to determine the minimum necessary resource to ensure the value of the multiplicative target function of an arbitrary form not less than the specified one, the possibility of using the known method of successive increments is substantiated.At the same time, in order to determine the minimum necessary resource to ensure the value of the multiplicative objective function of an arbitrary form not less than the specified one, the possibility of using the known method of successive increments is substantiated.

The application of a set of methods is specified to the level of a generalized method that can be used in calculations of the efficiency of information exchange in TS.

The correctness of the application of the developed mathematical apparatus is shown by its practical application for a specific simplified operational-tactical situation that may arise on the battlefield. The obtained results do not contradict the known theoretical propositions and logical understanding regarding the organization of information confrontation in the defense operation of a typical army corps.

CONCLUSIONS

The research proposes a method for optimizing the resource allocation of radio suppression and DPI during information exchange. The advantage of the method is that it allows for the first time to formalize and solve the problem of optimal resource allocation used to implement methods of radio and software-computer suppression of computer networks in an information conflict, as a problem of determining the minimum necessary resource to ensure the value of the multiplicative target functions of an arbitrary form not less than a given one, using the method of successive increments.

The method is quite simple from the point of view of numerical implementation and is not critical to the choice of the optimization procedure. The results of the experiments confirm the effectiveness of the proposed method in the tasks of protecting TN from radio suppression and DPI of the enemy, taking into account indicators of infringement of the efficiency of information exchange.

The scientific novelty for the first time, the method of optimizing the distribution of the radio suppression resource and DPIwas proposed for the development of possible scenarios of infringement of the information exchange by the enemy in a typical TN when organizing an information confrontation.

The developed method differs from the known ones in that, for the first time, it takes into account when developing scenarios of destructive impact on TS, the enemy's complex use of radio suppression and software influence.

DOI 10.15588/1607-3274-2023-4-2

The practical significance of the research results is the developed method that allows determining in practice the minimum set of radio suppression and DCI means, which ensures the required level of disruption of the efficiency of information exchange in various branches of state administration.

The generalized method is concretized to the level of algorithms, which simplifies its further implementation in the form of applied software products that can be used in the development of scenarios of an attacker's actions to disorganize goverment administration.

The results obtained in the article are normalized and can be used in engineering calculations of the efficiency of information exchange in TS.

Prospects for further research in this direction consist of the development of theoretical bases for determining the scenarios of possible enemy actions in the conditions of the enemy's complex use of radio suppression, electromagnetic and computer influence on TN elements.

REFERENCES

1. Bihun N., Shyshatskyi A., Bondar O., Bogrieiev S., Nalapko O., Sova O., Trotsko O. Analysis of the peculiarities of the communication organization in NATO countries, *Advanced Information Systems*, 2019, Vol. 3 (4), pp. 39–44. DOI: https://doi.org/10.20998/2522-9052.2019.4.05.
2. Shyshatskiy A. V., Bashkirov O. M., Kostina O. M. Development of integrated systems and data for Armed Forces, *Arms and military equipment*, 2015. No. 1 (5), pp. 35–40. DOI: http://journals.uran.ua/ index.php/2414-0651/issue/view/1%285%29%202015.
3. Zhuk A. V., Shyshatskiy P. V., Zhuk O. G., Zhyvotovskyi R. M. Methodological substances of management of the radio-resource managing systems of military radio communication, *Information Processing Systems*, 2017. Vol. 5 (151), pp. 16–25. DOI: https://doi.org/10.30748/soi.2017.151.02.
4. Sholokhov S., Samborsky I., Nikolayenko B., Vasylenko S., Hordiienko Y. Optimization of resources distribution of radio suppression means and destructive program impact on electronics networks, *Information Technology and Security*, 2022, Vol. 10, Issue 2 (19), pp. 230–240. DOI:https://doi.org/10.20535/24111031.2022.10.2.270464
5. Romanenko I., Shyshatskyi A. Analysis of modern condition of military radiocommunication system, *Advanced Information Systems*, 2017, Vol. 1, No. 1, pp. 28–33. DOI:https://doi.org/10.20998/2522-9052.2017.1.05.
6. Tomisla K. Improving the integrity of military-defence communication systems using network access points with a focus on terrestrial radio-relay links. Strategos. 2022. Vol. 6 (2), pp. 177–206. DOI: https://hrcak.srce.hr/file/421158.
7. NATO Glossary of Terms and Definitions: AAP-6 (2018), NATO Standardization Agency, 2018, 2019 p.
8. Luchuk E. V. Model' radio- ta programno-komp'yuternogo podavlennya komp'yuternix merezh protivnika v operaciyax, *Vijs'kovo-texnichnijzbirnik*, 2011. Vol. 5. pp. 104–109. DOI:https://doi.org/10.33577/2312-4458.5.2011.104-109
9. Roma O. M., Vasylenko S. V., Peleshok Ye. V., Honenko S. V., Nikolaienko B. A. Analysis of the synchronism entering process robustness in uav's radio control line with FHSS, *Radio Electronics, Computer Science, Control*, 2020, No. 2, pp. 15–24. DOI 10.15588/1607-3274-2020-2-2
10. Fernandez de Gorostiza E., Berzosa J., Mabe J., Cortiñas R. Method for Dynamically Selecting the Best Frequency Hopping Technique in Industrial Wireless Sensor Network Applications, *Sensors*, 2018, Vol. 18, Issue 2, P. 657. DOI: https://doi.org/10.3390/s18020657
11. Evaluation Criteria for IT Security. Part 1 : *Introduction and general model*. ISO/IEC 15408_1: 2005.
12. Evaluation Criteria for IT Security. Part 2: *Security functional requirements.* ISO/IEC 15408_2: 2005.
13. Evaluation Criteria for IT Security. Part 3: *Security assurance requirements.* ISO/IEC 15408_3: 2005.
14. Bhat, Srinidi and KVSSSS Sairam. Optimization of resource allocation in optical networks. *IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT).* 2022. Bangalore, India, pp. 1–9.
15. Srinivasa K. G., Anil Kumar Muppalla. Guide to High Performance Distributed Computing.Case Studies with Hadoop, Scalding and Spark, *Computer Communication and Networks*, 2015, 321 p. DOI: https://doi.org/10.1007/978-3-319-13497-0
16. Borovkov A. A. Stochastic Processes in Queueing Theory, *Stochastic Modelling and Applied Probability* (SMAP). 1976, Vol. 4, 280 p. DOI: https://doi.org/10.1007/978-1-4612-9866-3
17. Chen L., Cheng J., and Tseng Y., Optimal path planning with spatial-temporal mobility modeling for individualbased emergency guiding, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2015, Vol. 45, No. 12, pp. 1491–1501. DOI: 10.1109/TSMC.2015.2445875
18 Srinivasan V. R, Chiasserini C. F., Nuggehalli P. S., and Rao R. R. Optimal rate allocation for energy-efficient multipath routing in wireless ad hoc networks, *IEEE Transactions on Wireless Communications*, 2004, Vol. 3, No. 3, pp. 891–899. DOI: https://doi.org/10.1109/TWC.2004.826343
19. Raskin L. G., Kyrychenko I. O. Multi-index tasks of linear programming. Radio and communication, 1982, 240 p.
20. Bj¨ornson E., Jorswieck E. Optimal Resource Allocation in Coordinated Multi-Cell Systems, *Foundations and Trends in Communications and Information Theory*, 2013, Vol. 9, No. 2. pp. 113–381. DOI: 10.1561/0100000069

Received 25.08.2023.
Accepted 19.10.2023.

DOI 10.15588/1607-3274-2023-4-2

УДК 621. 618:519

МЕТОД ОПТИМІЗАЦІЇ РОЗПОДІЛУ ЗАСОБІВ РАДІОПОДАВЛЕННЯ ТА ДЕСТРУКТИВНОГО ПРОГРАМНОГО ВПЛИВУ НА КОМПЮТЕРНІ МЕРЕЖІ

Шолохов С. М. – канд. техн. наук, доцент, доцент спеціальної кафедри № 3 Інституту спеціального звязку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Павленко П. М. – д-р техн. наук, професор, професор кафедри організації авіаційнихперевезень Національного авіаційного університету, Київ, Україна.

Ніколаєнко Б. А. – канд. техн. наук, доцент спеціальної кафедри № 3 Інституту спеціального звязку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Самборський І. І. – канд. техн. наук, с.н.с., доцент спеціальної кафедри № 3 Інституту спеціального звязку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Самборський Є. І. – аспірант кафедри організації авіаційних перевезень Національного авіаційного університету, Київ, Україна.

АНОТАЦІЯ

Актуальність. На даний час узагальнені методичні підходи до розробки сценаріїв комплексного радіоподавлення та електромагнітного впливу типових спеціальних телекомунікаційних систем розроблені. Однак, при розробці можливих сценаріїв комплексного застосування засобів радіоподавлення та деструктивного програмного впливу виникла задача оптимізації ресурсу цих засобів та його розподілу по цілях радіоподавлення та обєктах деструктивного компютерного впливу яка досі не вирішена у повному обсязі. Зокрема, у відомій авторам літературі не розроблений метод оптимізації розподілу ресурсу засобів радіо та компютерного впливу, використовуваного для розробки та практичної реалізації оптимальних сценаріїв деструктивного впливу на компютерні мережі військових угруповань противника у військових операціях.

Тому, необхідно провести постановку задачі та розробити метод оптимізації розподілу ресурсу засобів радіоподавлення та деструктивного програмного впливу для розробки можливих сценаріїв порушення противником інформаційного обміну у типовій телекомунікаційній мережі.

Мета. Метою дослідження є розробка метода оптимізації розподілу ресурсу засобів радіоподавлення та деструктивного програмного впливу для розробки сценаріїв порушення противником інформаційного обміну у телекомунікаційній мережі.

Метод.Для досягнення мети дослідження комплексно застосовані та розвинені у галузь моделювання інформаційного протиборства методи нелінійної оптимізації розподілу різнорідного ресурсу, теорії масового обслуговування та експертного оцінювання.

Для визначення коефіцієнтів захищеності обєктів радіоелектронного та деструктивного компютерного впливу застосовані методи експертного оцінювання, зокрема, метод частот переваг особи, що приймає рішення із застосуванням методу Терстоуна. Цей метод вимагає лише одного експерта (особу, що приймає рішення), мінімального часу спілкування з ним, мінімальної експертної інформації (повного упорядкування вагових коефіцієнтів) та може застосовуватись при невеликій кількості оцінюваних вагових коефіцієнтів.

Для вирішення задачі оптимального розподілу різнорідного ресурсу засобів деструктивного впливу, для забезпечення значення мультиплікативної цільової функції довільного вигляду не менше заданого, застосований метод послідовних приростень.

Для визначення показника оперативності порушення інформаційного обміну застосовані методи теорії масового обслуговування, що дозволяє формалізувати спеціальні телекомунікаційні системи, як сукупність системи масового обслуговування – підсистем цифрового звязку та компютерних мереж.

Результати. Здійснена постановка задачі та введені показники дозволили вирішити задачу визначення мінімального ресурсу засобів деструктивного впливу та оптимального розподілу їх по цілях радіоподавлення на обєктах деструктивного програмного впливу для досягнення потрібного рівня порушення оперативності інформаційного обміну у спеціальних телекомунікаційних системах.

Висновки. За результатами статтірозроблено метод оптимізації розподілу ресурсу засобів радіоподавлення та деструктивного програмного впливу для розробки можливих сценаріїв порушення противником інформаційного обміну у типовій телекомунікаційній мережі. Перевірку запропонованого методу здійснено шляхом порівняння теоретичних результатів з результатами імітаційного моделювання сценаріїв порушення противником інформаційного обміну у телекомунікаційній мережі.

КЛЮЧОВІ СЛОВА: інформаційний обмін, компютерна радіомережа, компютерна атака, захист інформації, радіоподавлення, оптимізація розподілу ресурсу, деструктивний вплив.

ЛІТЕРАТУРА

1. Analysis of the peculiarities of the communication organization in NATO countries / [N. Bihun, A. Shyshatskyi, O. Bondar et al.] // Advanced Information Systems. – 2019. – Vol. 3(4). – P. 39–44. DOI: https://doi.org/10.20998/2522-9052.2019.4.05.
2. Shyshatskiy A. V. Development of integrated systems and data for Armed Forces / A. V. Shyshatskiy, O. M. Bashkirov, O. M. Kostina // Arms and military equipment. – 2015. – No 1(5). – P. 35–40. DOI: http://journals.uran.ua/index.php/2414-0651/issue/view/1%285%29%202015.

DOI 10.15588/1607-3274-2023-4-2

3. Methodological substances of management of the radio-resource managing systems of military radio communication / [A. V. Zhuk, P. V. Shyshatskiy, O. G. Zhuk et al.] // Information Processing Systems. – 2017. – Vol. 5(151). – P. 16–25. DOI: https://doi.org/ 10.30748/soi.2017.151.02.
4. Optimization of resources distribution of radio suppression means and destructive program impact on electronics networks / [S. Sholokhov, I. Samborsky, B. Nikolayenko et al.] // Information Technology and Security. – 2022. – Vol. 10, Issue 2 (19). – P. 230–240. DOI: https://doi.org/10.20535/2411-1031.2022.10.2.270464.
5. Romanenko I. Analysis of modern condition of military radiocommunication system / I. Romanenko, A. Shyshatskyi // AdvancedInformationSystems. – 2017. – Vol. 1, No. 1. – P. 28–33. DOI:https://doi.org/10.20998/2522-9052.2017.1.05.
6. Tomisla K. Improving the integrity of military-defence communication systems using network access points with a focus on terrestrial radio-relay links / K. Tomisla // Strategos. – 2022. – Vol. 6(2). – P. 177–206. DOI: https://hrcak.srce.hr/file/421158.
7. NATO Glossary of Terms and Definitions: AAP-6 (2018), NATO Standardization Agency, 2018, 2019 p.
8. Лучук Е. В. Модель радіо- та програмно-комп'ютерного подавлення комп'ютерних мереж противника в операціях / Е. В. Лучук // Військово-технічний збірник. – 2011. – Vol. 5. – P. 104–109. DOI:https://doi.org/10.33577/2312-4458.5.2011.104-109
9. Analysis of the synchronism entering process robustness in uav's radio control line with FHSS / [O. M. Roma, S. V. Vasylenko, B. A. Nikolaienko et al.] //Radio Electronics, Computer Science,Control. – 2020. – No. 2. – P. 15–24. DOI 10.15588/1607-3274-2020-2-2
10. Method for Dynamically Selecting the Best Frequency Hopping Technique in Industrial Wireless Sensor Network Applications / E. Fernandez de Gorostiza, J. Berzosa, J. Mabe, Cortiñas R. // Sensors. – 2018. – Vol. 18, Issue 2. – P. 657. DOI: https://doi.org/10.3390/s18020657
11. ISO/IEC 15408_1. Evaluation Criteria for IT Security. Part 1. Introduction and general model. – 2005.
12. ISO/IEC 15408_2. Evaluation Criteria for IT Security. Part 2. Security functional requirements. – 2005.
13. ISO/IEC 15408_3. Evaluation Criteria for IT Security. Part 3: Security assurance requirements. – 2005.
14. Optimization of resource allocation in optical networks / Bhat, Srinidi and KVSSSS Sairam // IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT). – Bangalore, India. – 2022. – P. 1–9.
15. Srinivasa K. G. Guide to High Performance Distributed Computing.Case Studies with Hadoop, Scalding and Spark / K. G. Srinivasa, Anil Kumar Muppalla // Computer Communication and Networks. – 2015. – 321 p. DOI: https://doi.org/10.1007/978-3-319-13497-0.
16. Borovkov A. A. Stochastic Processes in Queueing Theory / A. A. Borovkov // Stochastic Modelling and Applied Probability (SMAP). – 1976. – Vol. 4. – 280 p. DOI: https://doi.org/10.1007/978-1-4612-9866-3.
17. Chen L. Optimal path planning with spatial-temporal mobility modeling for individualbased emergency guiding / L. Chen, J. Cheng, Y. Tseng // IEEE Transactions on Systems, Man, and Cybernetics*:* Systems. – 2015. – Vol. 45, No. 12. – P. 1491–1501. DOI: 10.1109/TSMC.2015.2445875
18 Srinivasan V. R. Optimal rate allocation for energy-efficient multipath routing in wireless ad hoc networks / [V. R. Srinivasan, C. F.Chiasserini, P. S. Nuggehalli, and R. R. Rao] // IEEE Transactions on Wireless Communications. – 2004. – Vol. 3, No. 3. – P. 891–899. DOI: https://doi.org/10.1109/TWC.2004.826343
19. Raskin L. G. Multi-index tasks of linear programming / L. G. Raskin, I. O. Kyrychenko // Radio and communication. – 1982. – 240 p.
20. Bj¨ornson E. Optimal Resource Allocation in Coordinated Multi-Cell Systems / E. Bj¨ornson, E. Jorswieck // Foundations and Trends in Communications and Information Theory. – 2013. – Vol. 9, No. 2. – P. 113–381. DOI: 10.1561/0100000069

DOI 10.15588/1607-3274-2023-4-2

МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

MATHEMATICAL AND COMPUTER MODELING

UDC 004.827:519.816

DEVELOPMENT OF TECHNIQUE FOR STRUCTURING OF GROUP EXPERT ASSESSMENTS UNDER UNCERTAINTY AND INCONCISTANCY

Davydenko Ye. O. – PhD, Associate professor, Head of Department of Software Engineering, Petro Mohyla Black Sea National University, Mykolayiv, Ukraine.

Shved A. V. – Dr. Sc., Professor, Professor of Department of Software Engineering, Petro Mohyla Black Sea National University, Mykolayiv, Ukraine.

Honcharova N. V. – Post-graduate student of Department of Software Engineering, Petro Mohyla Black Sea National University, Mykolayiv, Ukraine.

ABSTRACT

Context. The issues of structuring group expert assessments are considered in order to determine a generalized assessment under inconsistency between expert assessments. The object of the study is the process of synthesis of mathematical models of structuring (clustering, partitioning) of expert assessments that are formed within the framework of Shafer model under uncertainty, inconsistency (conflict).

Objective. The purpose of the article is to develop an approach based on the metrics of theory of evidence, which allows to identify a number of homogeneous subgroups from the initial heterogeneous set of expert judgments formed within the framework of the Shafer model, or to identify experts whose judgments differ significantly from the judgments of the rest of the group.

Method. The research methodology is based on the mathematical apparatus of theory of evidence and cluster analysis. The proposed approach uses the principles of hierarchical clustering to form a partition of a heterogeneous (inconsistent) set of expert evidence into a number of subgroups (clusters), within which expert assessments are close to each other. Metrics of the theory of evidence are considered as a criterion for determining the similarity and dissimilarity of clusters. Experts' evidence are considered consistent in the formed cluster if the average or maximum (depending on certain initial conditions) level of conflict between them does not exceed a given threshold level.

Results. The proposed approach for structuring expert information makes it possible to assess the degree of consistency of expert assessments within an expert group based on an analysis of the distance between expert evidence bodies. In case of a lack of consistency within the expert group, it is proposed to select from a heterogeneous set of assessments subgroups of experts whose assessments are close to each other for further aggregation in order to obtain a generalized assessment.

Conclusions. Models and methods for analyzing and structuring group expert assessments formed within the notation of the theory of evidence under uncertainty, inconsistency, and conflict were further developed. An approach to clustering group expert assessments formed under uncertainty and inconsistency (conflict) within the framework of the Shafer model is proposed in order to identify subgroups within which expert assessments are considered consistent. In contrast to existing clustering methods, the proposed approach allows processing expert evidence of a various structure and taking into account possible ways of their interaction (combination, intersection).

KEYWORDS: theory of evidence, distance metric, dissimilarity measure, clustering, expert evidence, uncertainty, inconsistency.

ABBREVIATIONS

bpa is a basic probability assignment;
CCT is a cophenetic correlation test;
DI is a Dunn index;
DST is a Dempster-Shafer theory;
SSE is a sum of the squared error.

NOMENCLATURE

A is a set of alternatives;
avg(·) is an arithmetic average of its argument;
B is a set of expert preference profiles;
B_j reflects the preferences (choice) of expert E_j;
b_k^j is a k-th evidence formed, within the given scale of preferences, by the expert E_j;
Conf(E_k, G_q) is a measure reflecting the degree of conflict between E_k and group G_q;
ConfLev is a given limit level of conflict;
$d(m_i, m_j)$ is a distance metric value;
$d_J(m_i, m_j)$ is a value of the *Jousselme's* distance measure between two groups of evidence;
Dst is a matrix of pairwise distances;
E is a group of experts;
E^* is a set of experts candidates for the subgroup with consistent estimates G_q;
E_o is an expert whose preferences are selected as a reference element;
E^{conf} is a group of experts whose assessments differ significantly from the assessments of the rest of the group;

DOI 10.15588/1607-3274-2023-4-3

f is a single group decision;

f_t is an individual expert preference;

G_q is a group of experts with consistent assessments;

G_i^j is a subgroup of expert evidence for which the level of conflict l_i is acceptable;

P is a preference relation of the type P={$\succ$} (strict ordering), or P={$\succ, \sim$} (non-strict ordering);

R_{rez} is a result ranking;

l_q is a predetermined threshold level of conflict responsible for expert E_j belonging to the subgroup G_q;

l_0 is considered equal to 0;

m_i is a 2^{A}-dimensional vector-column, the elements of which are the *bpa's* of focal elements formed over the *i*-th group of evidence;

$(\mathrm{m}_i)^{\mathrm{T}}$ is a transposed vector m_i (string vector);

m_j^q is a vector of *bpa's* formed by the expert E_j in group G_q;

(m1–m2) is a difference of the corresponding vectors;

n is a number of examination objects (alternatives);

p is a number of formed groups of experts G_q with consistent assessments;

r is a number of experts in G_q;

$S(B_i, B_j)$ is a *Jaccard* coefficient;

t is a number of experts in expert group E;

2^{A} is a set of all possible subsets formed on the set A;

[π] is an operator for processing individual expert assessments (methods, rules, algorithms);

|·| is a cardinality of its argument.

INTRODUCTION

Group choice usually means the development of an agreed group decision on the order of preference of analyzed objects based on the individual judgments of experts. In other words, the problem of group choice is the problem of structuring individual preferences $f_1, f_2, \ldots, f_t$ into a single group decision f [1]:

$$\begin{matrix}(f_1, f_2, \ldots, f_n) \Rightarrow f. \\ \Uparrow \\ [\pi]\end{matrix} \qquad (1)$$

To select a method for obtaining a generalized assessment based on a set of group expert assessments, first need to test them for homogeneity (consistency). The results of such testing can lead to one of two possible cases:

1) the set of expert assessments is characterized by a high degree of consistency (which indicates their homogeneity);

2) the group of experts contains those whose assessments may differ in value from the assessments of the majority, the presence of such assessments in the total set of group expert assessments violates its homogeneity (consistency).

If the analysis reveals a high degree of consistency, a procedure of expert evidence aggregation is performed in order to obtain a final (group) ordering (ranking) of the analyzed objects (alternatives) in form of:

$$R_{rez} : A_j \mathrm{P} A_k \mathrm{P} \ldots \mathrm{P} A_z, \quad \forall (A_j, A_k, A_z) \in \mathrm{A}. \qquad (2)$$

The lack of consistency (homogeneity) indicates the presence in the commission of such experts who have different (but similar (homogeneous, agreed upon) within the same subgroup) points of view on solving the problem under consideration. Such situation arises, for example, due to the presence among the group of experts of representatives of different scientific schools or even teams. In the worst case, as a result of an expert survey, a significant number of small subgroups of experts are formed, with consistent judgments.

As a result, two tasks arise:

1) identifying and excluding outlier observations;

2) division (clustering) of the initial set of experts' judgments into several subgroups (clusters) of experts with similar (agreed, homogeneous) assessments, for their further analysis and determination of the aggregated assessment.

The object of study is the process of synthesis of mathematical models of structuring (clustering, partitioning) of expert assessments that are formed within the framework of Shafer model under uncertainty, inconsistency (conflict).

The subject of study is the models and methods of the group expert assessment analysis and structuring in the context of multi-alternative, inconsistency, conflict, uncertainty and their combinations.

The purpose of the work is a development of an approach based on the metrics of theory of evidence, which allows to identify a number of homogeneous subgroups from the initial heterogeneous set of expert judgments formed within the framework of the Shafer model, or to identify experts whose judgments differ significantly from the judgments of the rest of the group.

1 PROBLEM STATEMENT

Let a group of experts $\mathrm{E} = \{E_j \mid j = \overline{1,t}\}$, evaluating some initial set of objects of expertise (alternatives) $\mathrm{A} = \{A_i \mid i = \overline{1,n}\}$, forms profiles of expert preferences $\mathrm{B} = \{B_j \mid j = \overline{1,t}\}$, where B_j is a 2^{A}-dimensional vector. Profile $B_j = \{b_k^j \mid k = \overline{1,s}\}$, $s = 2^{|\mathrm{A}|}$, reflects the preferences (choice) of expert E_j, each element of which is built on the basis of a system of rules:

$$\begin{aligned} &1.\ b_k^j = \{\varnothing\}; \\ &2.\ b_k^j = \{A_i\}; \\ &3.\ b_k^j = \{A_i \mid i = \overline{1,v}\},\ v < n; \\ &4.\ b_k^j = \mathrm{A} = \{A_i \mid i = \overline{1,n}\}. \end{aligned} \qquad (3)$$

The task consists (in case of absence of agreement between the opinions of the members of the expert commission) to identify from the total set of expert judgments, subgroups of experts $\mathrm{E} \Rightarrow \{G_1\}, \{G_2\}, \ldots, \{G_q\}, \ldots, \{G_p\}$ ($G_q \subseteq \mathrm{E}$, $\{G_q\} = \{E_1, \ldots, E_r\}$, $t \geq r \geq 1$, $t \geq p \geq 1$), who have

DOI 10.15588/1607-3274-2023-4-3

a similar opinion and identify such experts E_l, who do not belong to any of these subgroups, that is, $E_l \subseteq G_q$, provided, that $|G_q| = 1$ (if any).

We will assume that:

1) judgments of $E_j \subseteq G_q$, $l \geq 2$ are considered consistent;

2) judgments of $E_j \subseteq G_q$, $|G_q| = 1$ are considered atypical, that is, significantly different (conflict) from other expert judgments.

Provided that $p = 1$ (and therefore $t = r$) the evidence of the entire group E are considered consistent.

If there is a trend $p \to t$ and $r \to 1$ (formation of a significant number of small groups G_q) the further analysis is inappropriate.

An example of the worst situation is the formation of the maximum possible number of subgroups, such that $\forall\, G_q : |G_q| = 1$ ($q = \overline{1, p}$, $p = t$); moreover, the best situation is considered to be in which $|G_q| = t$, q=1.

Thus, it is necessary to construct a decision rule that allows one to unambiguously determine whether the expert E_l belongs to the group Gr_q.

Further, additional procedures can be applied to bring together the opinions of different subgroups. Or, provided that the expert evidence are stable and final (formed taking into account the positions of all survey participants), the procedure for aggregating expert evidence is carried out for each of the resulting subgroups of experts Gr_q separately.

2 REVIEW OF THE LITERATURE

An analysis of methods that can be used to solve the problem of dividing group expert assessments into homogeneous, in a certain sense, subgroups has shown that their effective implementation is not always possible. For example, when analyzing expert assessments formed within the framework of numerical scales (absolute), the following methods have become widely used: cluster analysis methods based on the determination of distance functions, for example, Euclidean distance, Manhattan distance, Chebyshev distance, etc. [2–4]; clustering based on mathematical programming methods (dynamic programming, integer programming) [5, 6]; clustering based on estimation of probability density functions [7], etc.

To analyze expert judgments formed in ratio or order scales, non-numeric data clustering methods, for example, the Kemeny median method [8], can be used.

A justified choice and use of the considered methods for solving the problem of dividing group expert assessments in order to search for homogeneous subgroups can be carried out provided that various types of ignorance that arise in the process of obtaining and processing expert information are correctly taken into account. It is also necessary to take into account the possible structure of expert evidence (consonant, consistent, arbitrary, etc.), and take into account possible ways of their interaction (intersection, union, absorption) [9].

An effective mathematical apparatus that allows to correctly operate with such types of structures of expert evidence is the theory of evidence (Dempster-Shafer theory, DST) [10–12]. To solve the problem of assessing the distance between different types of structures of expert evidence in order to determine the degree of similarity of expert evidence, distance measures of evidence in Dempster-Shafer theory [13–16] can be applied.

3 MATERIALS AND METHODS

Let $\mathrm{A} = \{A_i \mid i = \overline{1,n}\}$ be a set of alternatives and a group of experts $\mathrm{E} = \{E_j \mid j = \overline{1,t}\}$ carrying out the examination. Within the notation of the DST, the set of initial data (alternatives, objects of examination) called the frame of discernment is a set of exhaustible and mutually exclusive elements [10–12]. Based on the analysis of A, according to the results of the expert survey, a subset system $\mathrm{B} = \{B_j \mid j = \overline{1,t}\}$ can be formed, where B_j is a 2^{A}-dimensional vector reflecting the preferences (choice) of the expert E_j, each element of which is built according to a system of rules (3).

So, for example, by assessing the initial set of alternatives A={a, b, c} by a group of experts E={E_1, E_2} the following profiles of expert preferences can be formed:

$$B_1=\{\{a\},\{b, c\}\}; \qquad B_2=\{\{a\},\{b\},\{c\}\}.$$

If the condition $\forall b_k^j \in B_j : (|\, b_k^j \,| = 1) \wedge (|\, B_j \,| = n)$ satisfied for $\forall\, B_j \subset \mathrm{B}$, then the results of expert survey in form of a set of group expert judgments (evidence) can be presented in the form of $n \times t$ dimension matrix:

$$\mathrm{B} = \begin{pmatrix} B_1 \\ B_2 \\ \cdots \\ B_j \\ \cdots \\ B_t \end{pmatrix} = \begin{pmatrix} b_1^1 & b_2^1 & \ldots & b_n^1 \\ b_1^2 & b_2^2 & \ldots & b_n^2 \\ \ldots & \ldots & \ldots & \ldots \\ b_1^j & b_2^j & \ldots & b_n^j \\ \ldots & \ldots & \ldots & \ldots \\ b_1^t & b_2^t & \ldots & b_n^t \end{pmatrix}. \qquad (4)$$

In matrix (4), each row includes the judgments of an expert E_j, for all objects, and the column includes judgments of the entire group of experts for a given object A_i.

For each subset B_j, $j = \overline{1,t}$, a vector of *bpa's* $\mathrm{m}_j = \{m_i \mid i = \overline{1,s}\}$, $s = 2^{|\mathrm{A}|}$, will be constructed whose elements satisfy the condition [11, 12], m: $2^{\mathrm{A}} \to [0,1]$:

$$0 \leq m(b_k^j) \leq 1, \quad m(\varnothing) = 0, \quad \sum_{b_k^j \in B_j} m(b_k^j) = 1, \qquad (5)$$

One of the metrics of the distance between expert evidence is taken as a measure of conflict [13–16]. Since expert evidence cannot be expressed in numerical terms, it is possible to establish that the original objects (experts) belong to any groups (classes) only on the basis of their similarity to each other.

The choice of metric is one of the main factors influencing the results of partitioning the initial set of expert

DOI 10.15588/1607-3274-2023-4-3

evidence and forming subgroups of experts with fairly close estimates. As a rule, the choice of metric is quite subjective and is determined by an analyst independently based on his / her own experience.

Let us consider the procedure for generation of G_q ($\forall\, G_q^{conf} \subset \mathrm{E}$) provided that the evidence of $E_j \subseteq G_q$ do not exceed the specified threshold *ConfLev* of the conflict coefficient.

1. Assessing the degree of similarity of expert evidence. For each pair $<\mathrm{m}_i, \mathrm{m}_j>$, $\forall (i, j) = \overline{1,t}$, $i \neq j$, estimates of the distance measure are determined, for example, the *Jousselme* distance [15]:

$$d_J(\mathrm{m}_1, \mathrm{m}_2) = \sqrt{\frac{1}{2}(\mathrm{m}_1 - \mathrm{m}_2)^{\mathrm{T}} \underline{\underline{D}}(\mathrm{m}_1 - \mathrm{m}_2)} \qquad (6)$$

where $\underline{\underline{D}}$ is a matrix of $2^{\mathrm{A}} \times 2^{\mathrm{A}}$ dimension, the elements of which are defined as

$$D(B, B) = \begin{cases} 1, & \text{if } B = B; \\ S(B, B), & \forall B, B \in \mathrm{A}. \end{cases} \qquad (7)$$

The $S(B_i, B_j)$ function corresponds to the *Jaccard* coefficient $S(B_i, B_j) = |\, B_i \cap B_j \,| / |\, B_i \cup B_j \,|$.

The results are stored in the form of a matrix of pairwise distances, which is symmetrical about the main diagonal in form of:

$$Dst = \begin{pmatrix} - & d(\mathrm{m}_1, \mathrm{m}_2) & \dots & d(\mathrm{m}_1, \mathrm{m}_t) \\ d(\mathrm{m}_2, \mathrm{m}_1) & - & \dots & d(\mathrm{m}_2, \mathrm{m}_t) \\ \dots\dots & \dots\dots & - & \dots\dots \\ d(\mathrm{m}_t, \mathrm{m}_1) & d(\mathrm{m}_t, \mathrm{m}_2) & \dots & - \end{pmatrix} \qquad (8)$$

where $d(\mathrm{m}_i, \mathrm{m}_j) = d(\mathrm{m}_j, \mathrm{m}_i)$, $\forall (i, j) = \overline{1,t}$, $i \neq j$.

2. Formation of a set of candidates $\mathrm{E}^* = \mathrm{E} = \{E_j \mid j = \overline{1,t}\}$ in group G_q.

3. Determination of the acceptable conflict level *ConfLev*.

4. Formation of a subgroup of experts $G_q \subseteq \mathrm{E}$, $q = \overline{1, p}$.

4.1 In matrix (8), the minimum value of measure $d(\mathrm{m}_i, \mathrm{m}_j)$ is sought, which corresponds to the distance between the two closest evidence E_i and E_j.

If $d(m_i, m_j)$ does not exceed a given *ConfLev* level, then evidence E_i and E_j ($E_i, E_j \in \mathrm{E}^*$) are added to the cluster G_q and removed from the set $\mathrm{E}^* = \mathrm{E}^* \setminus (E_i \cup E_j)$.

If such a pair is not found, the algorithm stops. It is assumed that $|\mathrm{E}^*|$ single-element $E_k \in \mathrm{E}^*$ subgroups are formed from the elements of the set E^*.

4.2 For $\forall\, E_k \in \mathrm{E}^*$ in (8) the minimum value of the measure is sought, which reflects the degree of conflict between E_k and the group G_q [17]:

$$Conf(E_k, G_q) = \frac{1}{r} \sum_{j=1}^{r} d(\mathrm{m}_k, \mathrm{m}_j^q), \; r = |\, G_q \,|. \qquad (9)$$

If the value (9) does not exceed the specified *ConfLev* level (if necessary, an additional condition is imposed: $\forall\, E_j \in G_q$: $d(\mathrm{m}_k, \mathrm{m}_j) \leq ConfLev$, $k \neq j$), then E_k evidence is added to the cluster G_q and removed from the set $\mathrm{E}^* = \mathrm{E}^* \setminus E_k$.

If all elements of E^* have been sorted out, then proceed to step 5.

5. Correction of the matrix *Dst* by removing elements belonging to a set $\mathrm{E} \setminus \mathrm{E}^*$.

6. Repeat steps 4–5 until $\mathrm{E}^* \neq \varnothing$.

Let us consider the procedure for generation of G_q provided that the estimates of $E_j \subseteq G_q$ do not exceed the specified threshold conflict level l_q.

1. Assessing the degree of similarity of expert evidence. Formation of matrix (8) elements.

2. Formation of a set of candidates $\mathrm{E}^* = \mathrm{E} = \{E_j \mid j = \overline{1,t}\}$.

3. Establishment of threshold values l_q, $q = \overline{1, p}$, responsible for certain levels of conflict (for example, low, medium, high conflict).

4. Selection a reference element $E_0 \in \mathrm{E}^*$.

*Algorithm*_1:

4.1*a* $\forall\, E_k \in \mathrm{E}^*$ it is defined estimates characterizing the degree of conflict between E_k and $\mathrm{E} \setminus E_k$ [17]:

$$Conf(E_k, \mathrm{E}^*) = \frac{1}{t^* - 1} \sum_{j=1,\, j \neq k}^{t^*} d(\mathrm{m}_k, \mathrm{m}_j), \; t^* = |\, \mathrm{E}^* \,|. \qquad (10)$$

4.2*a* The reference element $E_0 \in \mathrm{E}^*$ is selected, which is provide $\min(Conf(E_0, \mathrm{E}^*))$. Element E_0 is a least conflicting in relation to the entire group of experts.

*Algorithm*_2**:**

4.1*b* $\forall\, E_k \in \mathrm{E}^*$ are determined estimates in accordance with (10).

4.2*b* A subgroup of elements $\mathrm{E}^{conf} \subset \mathrm{E}^*$ is formed such that for $E_j \in \mathrm{E}^{conf}$ the value of measure (10) is significantly different (sharply different) from the value of measure (10) for the rest of the group $\mathrm{E}^* \setminus \mathrm{E}^{conf}$.

4.3*b* The reference element $E_0 \in \mathrm{E}^*$ is selected, which is provide $\min(Conf(E_0, \mathrm{E}^* \setminus \mathrm{E}^{conf}))$. Element E_0 is a least conflicting in relation to a group of experts from which experts with conflicting evidence are excluded.

*Algorithm*_3:

4.1*c* Based on the values of matrix (8), a set G_1^j, $j = \overline{1,t}$ is formed. The subgroup G_1^j includes estimates of $E_k \in \mathrm{E}^*$ for which the following condition is satisfied:

$$\forall (E_j, E_k) \in G_1^j : d(\mathrm{m}_j, \mathrm{m}_k) \leq l_1, \; k = \overline{1,t}, \; j \neq k. \qquad (11)$$

Thus, for a subgroup G_1^j, the $E_j \in \mathrm{E}^*$ is a reference element.

DOI 10.15588/1607-3274-2023-4-3

4.2*c* The reference element $E_o = E_j$, $E_j \in \mathrm{E}^*$, is selected such that $\max_j(|G_1^j|)$. That is, a reference element E_0 ensures the formation of the largest group of consistent (with the lowest specified level of conflict) evidence.

5. Based on the values of matrix (8), $\forall\, l_q$, $q = \overline{1, p}$ according to the reference element E_o, the resulting subgroups are formed on the set of E^* in such a way that $\forall E_j \in G_q$, $q = \overline{1, p}$, $r \geq 1$ the following condition is satisfied:

$$j = \overline{1, r}\,,\ l_{q-1} < d(\mathrm{m}_o, \mathrm{m}_j) \leq l_q. \qquad (12)$$

5.1 When forming a cluster G_q, $q = \overline{1, p}$, all elements of the set E^* are searched for compliance with condition (12).

The element E_j does not fall into the class G_q if the condition $\forall\, E_j, E_s \in G_q$: $d(\mathrm{m}_j, \mathrm{m}_s) \leq l_q$, $j \neq s$, is not satisfied.

If E_j is added to the cluster G_q, then it is removed from the set $\mathrm{E}^* = \mathrm{E}^* \setminus E_j$.

5.2 The procedure provided for in clause 5.1 is repeated p–1 times or terminated early if $\mathrm{E}^* = \varnothing$.

4 EXPERIMENTS

A comparative analysis of the proposed approaches to identification of homogeneous subgroups of expert assessments among an inconsistent initial set of expert evidence and agglomerative clustering methods have been carried out. The following classical methods have been considered: Ward's method (*Ward*), single-linkage (*Single*), complete-linkage (*Complete*), centroid (*Centroid*).

The class of agglomerative clustering methods was chosen due to the fact that, firstly, the proposed approach is based on the principles underlying agglomerative algorithms. Secondly, the goal of the proposed approach is to obtain such coverage (partitioning) of the initial set of expert evidence that ensures the formation of subgroups of experts with consistent assessments (consistent in the sense that the level of conflict between expert evidence belonging to the same group does not exceed a given threshold level of conflict) rather than determining the optimal number of classes. Accordingly, it is the principles and mechanisms underlying agglomerative algorithms that make it possible to terminate the agglomeration process at an iteration ahead of schedule, when the merging of clusters occurs at an unacceptable level of conflict. Thereby reducing the running time of the algorithm.

Case 1. For studying the effectiveness of the proposed approach (*Method*_1), which makes it possible to form a partition of a set of assessments into consistent (homogeneous) subgroups, provided that a certain threshold (acceptable) level of conflict *ConfLev* is specified, five test samples were formed, Table. 1.

The task was to form consistent subgroups of expert evidence with a $ConfLev \leq 0.3$. Testing was carried out for samples of ten, 20 and 30 elements. The maximum sample size did not exceed 30 values, since usually a group of experts does not exceed 25–30 people.

Table 1 – Principles for test samples formation

Sample	Sample formation method
A	consistent estimates (max distance between evidence is equal to 0.2)
B	moderately conflicting expert evidence (30% of the sample is a group of expert evidence with an average distance equal to 0.3 in relation to the expert evidence of the main group)
C	conflicting expert evidence (30% of the sample is a group of expert evidence with an average distance equal to 0.3 in relation to the expert evidence of the main group; 17% of the sample is a group of expert evidence with an average distance equal to 0.4 in relation to the expert evidence of the main group)
D	highly conflicting expert evidence (17% of the sample is a group of expert evidence with an average distance equal to 0.5 in relation to the expert evidence of the main group)
E	highly conflicting expert evidence (17% of the sample is a single expert evidence with an average distance equal to 0.6 in relation to the expert evidence of the main group)

Case 2. To study the effectiveness of the approach (*Method*_2), which makes it possible to form a partition of a set of estimates into consistent (homogeneous) subgroups, provided that several different threshold levels of conflict l_q are specified, a method was chosen based on the search for a reference element using *Algorithm*_3.

Testing was carried out for samples of ten, 20 and 30 values.

Rule for generating a test sample:

– 50% of the sample is a group of expert evidence with max distance between evidence equal to: 0.170 (n = 10); 0.234 (n = 20); 0.220 (n = 30);

– 13% of the sample is a group of expert evidence with an average distance of 0.1 in relation to the expert evidence of the main group;

– 13% of the sample is a group of expert evidence with an average distance of 0.2 in relation to the expert evidence of the main group;

– 13% of the sample is a group of expert evidence with an average distance of 0.3 in relation to the expert evidence of the main group;

– 11% of the sample is a group of expert evidence with an average distance of 0.4 in relation to the expert evidence of the main group.

5 RESULTS

Let's analyze the results obtained.

Case 1. Table 2 shows the values of the obtained cophenetic correlation coefficient using the Mantel test for clustering results.

As can be seen from Table 2, in most cases the proposed method gives the maximum value of cophenetic correlation coefficient (p-value = 0.001).

For samples *C* and *E*, testing was carried out only for the samples of 20 and 30 elements.

Table 3 shows the results of a comparative analysis of the considered clustering methods (F_0 is an average dis-

DOI 10.15588/1607-3274-2023-4-3

tance in a cluster; F_1 is an average distance between clusters).

For samples *B* and *D*, all considered methods gave the same result. Both samples were formed according to the rule: one group of evidence with a moderate (sample *B*) and significant (sample *D*) level of conflict was added to the main consistent population.

For samples *C* and *D*, the proposed method provides the highest value of the silhouette index; for sample *D*, the proposed method provides the maximum silhouette index and modified Dunn index (*DI*); the lowest value of the ratio of the average intra-cluster distance (F_0) to the average inter-cluster distance (F_1), which indicates better separation of clusters and greater compactness of elements in the cluster compared to other methods.

Case 2. The results of the analysis are given in the Table 4.

For samples of sizes ten and 20, the proposed method provides the highest value of the cophenetic correlation coefficient according to the Mantel test (p-value = 0.001), and the formation of a cluster with the largest number of consistent expert evidence.

Table 2 – Analysis of the quality of clustering results

Sample	Method	$n = 10$		$n = 20$		$n = 30$		Sample	$n = 20$		$n = 30$	
		max $d(m_i,m_j)$	***CCT***	**max** $d(m_i,m_j)$	***CCT***	**max** $d(m_i,m_j)$	***CCT***		**max** $d(m_i,m_j)$	***CCT***	**max** $d(m_i,m_j)$	***CCT***
A	Method_1	0.092	**0.700**	0.169	**0.784**	0.169	**0.696**	*C*	0.411	**0.930**	0.441	**0.926**
	Ward		0.684		0.622		0.612			0.923		0.920
	Single		0.669		0.760		0.600			0.912		0.878
	Complete		0.698		0.769		0.613			0.906		0.915
	Centroid		0.626		0.719		0.641			0.929		0.920
B	Method_1	0.334	**0.976**	0.341	**0.960**	0.354	**0.932**	*E*	0.657	**0.973**	0.654	**0.956**
	Ward		0.975		0.957		0.926			0.947		0.937
	Single		0.974		0.957		0.914			0.957		0.783
	Complete		0.975		0.925		0.907			0.965		0.201
	Centroid		0.975		**0.960**		0.923			0.965		0.951
D	Method_1	0.576	**0.981**	0.530	**0.939**	0.646	**0.978**					
	Ward		0.979		0.932		0.976					
	Single		**0.981**		0.935		0.975					
	Complete		**0.981**		0.841		0.952					
	Centroid		0.980		0.938		**0.978**					

Table 3 – Comparative analysis of clustering methods ($n = 30$)

Sample	Distance		Method	Clusters						Silhouette score		*SSE*	*DI*	F_0 / F_1
				Generated		Detected								
	max	**min**		**№**	**Size**	**№**	**Size**	**Diameter**	**avg** $d(m_i,m_j)$	S_i	**avg(S)**			
B	0.354	0.004	All methods	1	20	1	20	0.188	0.082	0.695	0.728	0.115	3.425	0.263
				2	10	2	10	0.111	0.057	0.796				
C	0.441	0.004	Method_1, Ward, Single, Centroid	1	16	1	16	0.188	0.078	0.712	**0.674**	0.098	1.997	0.240
				2	9	2	9	0.111	0.060	0.603				
				3	5	3	5	0.080	0.050	0.681				
			Complete	1	16	1	14	0.117	0.060	0.464	0.562	**0.050**	**2.32**	**0.207**
				2	9	2	9	0.111	0.059	0.603				
				3	5	3	5	0.080	0.050	0.681				
						4	2	0.023	0.023	0.771				
D	0.646	0.004	All methods	1	25	1	25	0.200	0.071	0.863	0.850	0.116	4.65	0.150
				2	5	2	5	0.180	0.112	0.783				
E	0.654	0.004	Method_1, Centroid	1	25	1	25	0.192	0.074	0.777	**0.738**	0.106	**2.590**	**0.158**
						2	2	0.100	0.100	0.612				
				2	5	3	1	–	–	–				
						4	2	0.087	0.087	0.745				
			Ward	1	25	1	25	0.192	0.074	0.752	0.734	0.149	2.127	0.180
				2	5	2	3	0.278	0.206	0.536				
						3	2	0.087	0.087	0.800				
			Single	1	25	1	22	0.228	0.126	0.408	0.375	0.284	0.960	0.300
				2	5	2	4	0.078	0.045	0.569				
						3–6	1	–	–	–				
			Complete	1	25	1	21	0.169	0.068	0.189	0.274	**0.067**	0.926	0.201
						2	4	0.048	0.038	0.389				
				2	5	3	2	0.100	0.100	0.612				
						4	1	–	–	–				
						5	2	0.087	0.087	0.745				

DOI 10.15588/1607-3274-2023-4-3

Table 4 – Comparative analysis of clustering methods when forming the largest group of consistent evidence ($l_1 = 0.2$)

Sample	Distance		Method	*CCT*	Clusters			Elements of the largest cluster with $l_1 = 0.2$	
	max	min			Count	max size	Diameter	Detected	Generated
10	0.421	0.030	Method_2	**0.934**	**3**	**8**	**0.183**	$\{E_1,\boldsymbol{E_2},E_3,\boldsymbol{E_4},\boldsymbol{E_6},E_7,\boldsymbol{E_8},\boldsymbol{E_{10}}\}$	$\{E_2,E_4,E_6,E_8,E_{10}\}$
			Single	0.925					
			Centroid	0.929					
			Ward	0.828	2	8	0.183	$\{E_1,\boldsymbol{E_2},E_3,\boldsymbol{E_4},\boldsymbol{E_6},E_7,\boldsymbol{E_8},\boldsymbol{E_{10}}\}$	
			Complete	0.836					
20	0.399	0.021	Method_2	0.869	4	**10**	**0.134**	$\{E_1,\boldsymbol{E_3},E_4,\boldsymbol{E_5},E_8,\boldsymbol{E_{10}},\boldsymbol{E_{11}},E_{12},\boldsymbol{E_{18}},\boldsymbol{E_{20}}\}$	$\{\boldsymbol{E_2},E_3,E_5,E_7,E_{10},E_{11},\boldsymbol{E_{14}},\boldsymbol{E_{16}},E_{18},E_{20}\}$
			Centroid	0.861					
			Single	0.870	4	8	0.124	$\{\boldsymbol{E_3},\boldsymbol{E_5},\boldsymbol{E_7},\boldsymbol{E_{10}},\boldsymbol{E_{11}},E_{15},\boldsymbol{E_{18}},\boldsymbol{E_{20}}\}$	
			Ward	**0.872**					
			Complete	0.766	10	4	0.145	$\{E_3,\boldsymbol{E_8},\boldsymbol{E_{18}},\boldsymbol{E_{20}}\}$	
30	0.394	0.030	Method_2	**0.881**	5	**16**	**0.189**	$\{E_2,E_3,\boldsymbol{E_4},E_5,E_{10},\boldsymbol{E_{12}},E_{15},\boldsymbol{E_{18}},E_{19},E_{20},\boldsymbol{E_{21}},\boldsymbol{E_{22}},E_{25},\boldsymbol{E_{26}},E_{29},\boldsymbol{E_{30}}\}$	$\{E_1,E_4,\boldsymbol{E_7},E_8,E_{12},E_{14},E_{17},E_{18},E_{21},E_{22},E_{24},E_{26},E_{28},E_{30}\}$
			Ward	0.827	4	12	0.144	$\{E_3,\boldsymbol{E_4},E_{10},\boldsymbol{E_{12}},\boldsymbol{E_{18}},E_{19},E_{20},\boldsymbol{E_{21}},\boldsymbol{E_{22}},E_{25},\boldsymbol{E_{26}},\boldsymbol{E_{30}}\}$	
			Single	0.849	11	9	0.132	$\{E_3,E_{10},\boldsymbol{E_{18}},E_{19},E_{20},\boldsymbol{E_{21}},\boldsymbol{E_{22}},E_{25},\boldsymbol{E_{30}}\}$	
			Complete	0.752	5	12	0.144	$\{E_3,\boldsymbol{E_4},E_{10},\boldsymbol{E_{12}},\boldsymbol{E_{18}},E_{19},E_{20},\boldsymbol{E_{21}},\boldsymbol{E_{22}},E_{25},\boldsymbol{E_{26}},\boldsymbol{E_{30}}\}$	
			Centroid	0.864	8	14	0.155	$\{\boldsymbol{E_1},E_3,\boldsymbol{E_4},\boldsymbol{E_8},E_{10},\boldsymbol{E_{12}},\boldsymbol{E_{14}},\boldsymbol{E_{17}},E_{19},\boldsymbol{E_{21}},\boldsymbol{E_{24}},\boldsymbol{E_{26}},\boldsymbol{E_{28}},\boldsymbol{E_{30}}\}$	

As can be seen from Table 4, for a sample size of 20, none of the methods under consideration identified the evidence of experts E_2, E_{14} and E_{16} (for a sample size of 30, this is the evidence of expert E_7) as belonging to the initially formed consensus subgroup. But this is explained by the fact that when forming the initial group of consistent expert evidence (which is 50% of the test sample) for $n = 20$, the maximum distance between expert evidence was 0.234 (with $n = 30$, the maximum distance between expert evidence was 0.220), and the splitting of the totality of expert evidence into clusters occurred at the level of conflict (distance) $l_1 = 0.200$.

6 DISCUSSION

The analysis of tasks and methods for processing group expert assessments allows to conclude that solving the problem of finding generalized (aggregated) assessments, on the basis of which recommendations are formed for the decision maker, largely depends on the effective solution of clustering and ranking problems.

The problem of clustering (partitioning) expert assessments arises in situations where the results of the examination are characterized by a lack of consistency, which creates certain difficulties in determination of generalized assessments.

To solve the problem, two approaches are proposed. The first is to form subgroups of experts that have agreed upon assessments, provided that a certain threshold (acceptable) level of conflict *ConfLev* is specified. The evidence of experts included in subgroup G_q does not exceed a certain conflict level *ConfLev*. In this case, p subgroups of experts can be formed, within which the expert opinions can be considered consistent, but formed subgroups can be in conflict with each other.

The second approach allows to identify subgroups of experts within which expert opinions can be considered consistent, but with different threshold levels of conflict l_q. Thus, for example, a group of experts G_1 will be obtained with a low level of conflict between the expert evidence belonging to it; group of experts G_2 – with a moderate level of conflict; a group of experts G_3 – with a significant level of conflict, etc.

CONCLUSIONS

The paper proposes a technique for structuring group expert assessments, which is based on the mathematical apparatus of the theory of evidence. The proposed approach allows, in the absence of an acceptable level of consistency (consensus, homogeneous) between expert evidence, to identify from the original set of experts subgroups with similar (in a certain sense) assessments (preferences). Various distance measures of evidence (in the framework of DST) were used as a degree of similarity.

The scientific novelty of obtained results is that the models and methods of group expert assessment analysis and structuring under inconsistency, conflict, uncertainty and their combinations are received the further development.

Unlike existing methods for clustering expert assessments, the proposed approach allows processing expert evidence of a various structure: consonant, consistent, arbitrary, etc.; take into account possible combinations and overlaps of expert evidence.

The proposed approach is based on the mathematical apparatus of distances in evidence theory, which allows to assess the degree of dissimilarity (conflict) between selected groups of expert evidence, taking into account their structure. Expert evidence is considered consistent (homogeneous) if the value of the selected metric for all evidence of the selected subgroup does not exceed a specified threshold level.

The practical significance of the obtained results is that the proposed approach can be used as an additional tool for identifying experts (one or more) whose assessments are based on the results of several examinations, largely from the assessments of the main group. Next, it

DOI 10.15588/1607-3274-2023-4-3

can be studied the reason for such behavior of the expert: is it his / her creative opinion, reflecting a non-standard approach to solving the current problem; an attempt to manipulate the results of an expert survey or lack of sufficient knowledge of the subject area.

The prospects for further research are to study of the influence of the choice of distance measure on the results of partitioning under different structures of expert judgments (consonant, consistent, arbitrary).

ACKNOWLEDGEMENTS

The work is partially supported by the state research project of Petro Mohyla Black Sea National University "Development of modern information and communication technologies for the management of intellectual resources to decision-making support of operational management" (research project no. 0121U107831, financed by the Government of Ukraine).

REFERENCES

1. Blokdyk G. Group Decision Making A Complete Guide. Toronto, 5STARCooks, 2021, 208 p.
2. Cichosz P. Data Mining Algorithms: Explained Using R. Chichester, Wiley, 2015, 720 p.
3. Hair J. F. et al. Multivariate Data Analysis (6th ed.). New Jersy, Pearson Prentice Hall, 2006, 899 p.
4. Scitovski R., Sabo K., Martínez-Álvarez F., Ungar S. Cluster Analysis and Applications. Switzerland. Springer Cham, 2021, 271 p. DOI: 10.1007/978-3-030-74552-3
5. Alhameli F., Elkamel A., Alkatheri M., Betancourt-Torcat A., Almansoori A. New Class of Simple and Efficient Clustering Algorithms for Multiscale Mathematical Programming with Demand Data Applications, *Industrial Engineering and Operations Management*, Fourth North American International Conference, Toronto, 23–25 October 2019: proceedings. Canada, IEOM Society International, 2019, pp. 497–505.
6. Hubert L. J., Arabie P., Meulman J. Dynamic Programming in Clustering. In: Floudas C., Pardalos P. (eds) Encyclopedia of Optimization. Boston, Springer, 2008, pp. 837–844. DOI: 10.1007/978-0-387-74759-0_145
7. Nguyentrang T., Vovan T. Fuzzy clustering of probability density functions, *Journal of Applied Statistics*, 2017, Vol. 44(4), pp. 583–601. DOI: 10.1080/02664763.2016.1177502
8. Kemeny J. G., Snell J. L. Mathematical models in the social sciences. Introduction to higher mathematics. New York, Toronto, London, Blaisdell Publishing Company, A Division of Ginn and Company, 1963, 145 p.
9. Sentz K., Ferson S. Combination of evidence in Dempster-Shafer theory. Technical report SAND 2002-0835. Albuquerque, Sandia National Laboratories, 2002, 94 p.
10. Dempster A. P. Upper and lower probabilities induced by a multi-valued mapping, *Annals of Mathematical Statistics*, 1967, Vol. 38(2), pp. 325–339. DOI: 10.1214/aoms/1177698950
11. Shafer G. A mathematical theory of evidence. Princeton, Princeton University Press, 1976, 297 p.
12. Smarandache F., Dezert J. Advances and applications of DSmT for information fusion. Rehoboth, American Research Press, 2004, Vol. 1, 760 p.
13. Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distribution, *Bulletin of the Calcutta Mathematical Society*, 1943, Vol. 35, pp. 99–110.
14. Cuzzolin F. A geometric approach to the theory of evidence, *Transactions on Systems, Man, and Cybernetics*, Part C: Applications and Reviews, 2007, Vol. 38(4), pp. 522–534. DOI: 10.1109/TSMCC.2008.919174
15. Jousselme A. L., Grenier D., Boss´e E. A new distance between two bodies of evidence, *Information Fusion*, 2001, Vol. 2, pp. 91–101. DOI: 10.1016/S1566-2535(01)00026-4
16. Tessem B. Approximations for efficient computation in the theory of evidence, *Artificial Intelligence*, 1993, Vol. 61, pp. 315–329. DOI: 10.1016/0004-3702(93)90072-J
17. Martin A., Jousselme A. L., Osswald C. Conflict measure for the discounting operation on belief functions, *Information Fusion* (FUSION 2008): the 11th International Conference, Cologne. Germany, 30 June–3 July 2008: proceedings. Cologne, IEEE, 2008, pp. 1–8.

Received 28.09.2023.
Accepted 20.11.2023.

УДК 004.827:519.816

РОЗРОБКА МЕТОДИКИ СТРУКТУРИЗАЦІЇ ГРУПОВИХ ЕКСПЕРТНИХ ОЦІНОК В УМОВАХ НЕВИЗНАЧЕНОСТІ ТА НЕУЗГОДЖЕНОСТІ

Давиденко Є. О. – канд. техн. наук, доцент, завідувач кафедри інженерії програмного забезпечення Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

Швед А. В. – д-р наук, професор, професор кафедри інженерії програмного забезпечення Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

Гончарова Н. В. – аспірантка кафедри інженерії програмного забезпечення Чорноморського національного університету імені Петра Могили, Миколаїв, Україна.

АНОТАЦІЯ

Актуальність. Розглянуті питання структуризації групових експертних оцінок з метою визначення узагальненої оцінки у випадку відсутності узгодженості експертних оцінок. Об'єктом дослідження є процеси синтезу математичних моделей структуризації (кластеризації, розбиття) експертних оцінок, що формуються в рамках моделі Шейфера в умовах невизначеності, неузгодженості (конфлікту). Мета роботи – розробка підходу на основі метрик теорії свідоцтв, що дозволяє із вихідної неоднорідної сукупності експертних оцінок, сформованих в рамках моделі Шейфера, виділяти ряд однорідних підгруп, або ідентифікувати експертів чиї оцінки в значній мірі відрізняються від оцінок решти групи.

Метод. Методика дослідження ґрунтується на математичному апараті теорії свідоцтв, кластерному аналізі. Запропонований підхід використовує принципи ієрархічної кластеризації при формуванні розбиття неоднорідної (неузгодженої) суку-

DOI 10.15588/1607-3274-2023-4-3

пності експертних свідоцтв на ряд підгруп (кластерів), всередині яких оцінки експертів близькі між собою. В якості критерію визначення схожості та відмінності кластерів розглянуті метрики теорії свідоцтв. Оцінки експертів вважаються узгодженими у сформованому кластері, якщо середній або максимальний (в залежності від визначених початкових умов) рівень конфлікту між ними не перевищує заданий пороговий рівень.

Результати. Запропонована методика структуризації експертної інформації дозволяє оцінювати рівень узгодженості експертних оцінок усередині експертної групи на основі аналізу відстані між експертними свідоцтвами. У разі відсутності узгодженості всередині експертної групи запропоновано виділяти з неоднорідної сукупності оцінок підгрупи експертів, оцінки яких близькі для подальшого їх агрегування з метою отримання узагальненої оцінки. Наявність у комісії небагатьох груп експертів із узгодженими оцінками може свідчити про наявність експертів, що мають різний погляд на аналізовану проблему.

Висновки. Дістали подальшого розвитку моделі та методи аналізу та структуризації групових експертних оцінок, сформованих в рамках нотації теорії свідоцтв в умовах невизначеності, неузгодженості, конфлікту. Запропоновано метод кластеризації групових експертних оцінок, що формуються в умовах невизначеності та неузгодженості (конфлікту) в рамках моделі Шейфера, з метою виділення підгруп, всередині яких оцінки експертів вважаються узгодженими. На відміну від існуючих методів кластеризації, запропонований підхід дозволяє обробляти експертні свідоцтва довільної структури, враховувати можливі способи їх взаємодії (об'єднання, перетин).

КЛЮЧОВІ СЛОВА: теорія свідоцтв, метрики теорії свідоцтв, кластеризація, міри відстані, експертні свідоцтва, невизначеність, неузгодженість.

ЛІТЕРАТУРА

1. Blokdyk G. Group Decision Making A Complete Guide / G. Blokdyk. – Toronto: 5STARCooks, 2021. – 208 p.
2. Cichosz P. Data Mining Algorithms: Explained Using R / P. Cichosz – Chichester : Wiley, 2015. – 720 p.
3. Multivariate Data Analysis (6th ed.) / J. F. Hair et al. – New Jersy: Pearson Prentice Hall, 2006. – 899 p.
4. Cluster Analysis and Applications / [R. Scitovski, K. Sabo, F. Martínez-Álvarez, S. Ungar]. – Switzerland : Springer Cham, 2021. – 271 p. DOI: 10.1007/978-3-030-74552-3
5. New Class of Simple and Efficient Clustering Algorithms for Multiscale Mathematical Programming with Demand Data Applications / [F. Alhameli, A. Elkamel, M. Alkatheri et al.] // Industrial Engineering and Operations Management: Fourth North American International Conference, Toronto, 23–25 October 2019: proceedings. – Canada: IEOM Society International, 2019. – P. 497 – 505.
6. Hubert L. J. Dynamic Programming in Clustering / L. J. Hubert, P. Arabie, J. Meulman. In: Floudas C., Pardalos P. (eds) // Encyclopedia of Optimization. – Boston: Springer, 2008. – P. 837–844. DOI: 10.1007/978-0-387-74759-0_145
7. Nguyentrang T. Fuzzy clustering of probability density functions / T. Nguyentrang, T. Vovan // Journal of Applied Statistics. – 2017. – Vol. 44(4). – P. 583–601. DOI: 10.1080/02664763.2016.1177502
8. Kemeny J. G. Mathematical models in the social sciences. Introduction to higher mathematics / J. G. Kemeny, J. L. Snell. – New York, Toronto, London: Blaisdell Publishing Company, A Division of Ginn and Company, 1963. – 145 p.
9. Sentz K. Combination of evidence in Dempster-Shafer theory. Technical report SAND 2002-0835 / K. Sentz, S. Ferson. – Albuquerque: Sandia National Laboratories, 2002. – 94 p.
10. Dempster A. P. Upper and lower probabilities induced by a multi-valued mapping / A. P. Dempster // Annals of Mathematical Statistics. – 1967. – Vol. 38(2). – P. 325–339. DOI: 10.1214/aoms/1177698950
11. Shafer G. A mathematical theory of evidence / G. Shafer. – Princeton: Princeton University Press, 1976. – 297 p.
12. Smarandache F. Advances and applications of DSmT for information fusion / F. Smarandache, J. Dezert. – Vol. 1. – Rehoboth: American Research Press, 2004. – 760 p.
13. Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distribution / A. Bhattacharyya // Bulletin of the Calcutta Mathematical Society. – 1943. – Vol. 35. – P. 99–110.
14. Cuzzolin F. A geometric approach to the theory of evidence / F. Cuzzolin // Transactions on Systems, Man, and Cybernetics (Part C: Applications and Reviews). – 2007. – Vol. 38(4). – P. 522–534. DOI: 10.1109/TSMCC.2008.919174
15. Jousselme A. L. A new distance between two bodies of evidence / A. L. Jousselme, D. Grenier, E. Boss´e // Information Fusion. – 2001. – Vol. 2. – P. 91–101. DOI: 10.1016/S1566-2535(01)00026-4
16. Tessem B. Approximations for efficient computation in the theory of evidence / B. Tessem // Artificial Intelligence. – 1993. – Vol. 61. – P. 315–329. DOI: 10.1016/0004-3702(93)90072-J
17. Martin A. Conflict measure for the discounting operation on belief functions / A. Martin, A. L. Jousselme, C. Osswald // Information Fusion (FUSION 2008): the 11th International Conference, Cologne, Germany, 30 June–3 July 2008: proceedings. – Cologne: IEEE, 2008. – P. 1–8.

DOI 10.15588/1607-3274-2023-4-3

UDC 621.396.962

METHOD OF MINIMIZATION SIDELOBES LEVEL AUTOCORRELATION FUNCTIONS OF SIGNALS WITH NON-LINEAR FREQUENCY MODULATION

Kostyria O. O. – Dr. Sc., Senior Research, Leading Research Scientist, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Hryzo A. A. – PhD, Associate Professor, Head of the Research Laboratory, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Dodukh O. M. – PhD, Leading Research Scientist, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Lisohorskyi B. A. – PhD, Senior Research Scientist, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

Lukianchykov A. A. – Senior Research Scientist, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

ABSTRACT

Context. At present, when creating new and upgrading existing radar systems, solid-state generator devices are widely used, which imposes certain restrictions on the peak power of probing signals. To overcome this limitation, longer duration signals with internal pulse modulation are used. The main efforts of the researchers are focused on reducing the maximum level of the side lobes of the autocorrelation function of such signals, which, without taking additional measures, has a significant level, which complicates the work of systems for detecting and stabilizing the level of false alarms. Attention is paid to signals with non-linear frequency modulation, which consist of two and three linearly frequency-modulated fragments. The maximum level of the side lobes of such signals depends significantly on the frequency-time parameters of the fragments, and therefore it is very difficult to obtain its stable value. Searching for signals with minimal side lobe level values by optimizing their time-frequency parameters is a difficult task, because changing the parameters of previous signal fragments leads to changes in the parameters of subsequent fragments

Objective. The aim of the work is to develop a method for simplifying the search for local minima of the level of side lobes of two- and three-fragment signals with nonlinear frequency modulation by using a modified mathematical model with a whole number of periods of radio oscillations of linear-frequency modulated fragments.

Method. The developed method is based on the proposed modification of the mathematical model, which corrects the frequency-time parameters of two- and three-fragment signals with non-linear frequency modulation by modifying the values of the frequency modulation speed while providing an integer number of complete periods of radio frequency oscillations for each of the fragments, which simplifies the process of finding local minima of the level of side lobes.

Results. Modification of the initial mathematical model leads to the expansion of the possible range of values of frequency-time parameters, ratios of durations and frequency deviations of linearly-frequency modulated fragments and ensures stability of the mathematical model with a decrease in the maximum level of side lobes of the autocorrelation function.

Conclusions. It has been experimentally confirmed that the use of the proposed method of modifying the input frequency-time parameters of signals with non-linear frequency modulation in the vast majority of cases reduces the maximum level of side lobes and simpli-fies the process of finding its local minima. The optimal ratios of durations and deviations of the frequency of the signal frag-ments are determined, subject to these, stable operation of the models is ensured and, in most cases, - less than the value of the maximum level of the side lobes.

KEYWORDS: mathematical model; a non-linear frequency modulation signal; autocorrelation function; maximum level of side lobes.

ABBREVIATIONS

ACF is a autocorrelation function;
SL is a side lobe;
Wt is a weighting;
LFM is a linear frequency modulation;
MM is a mathematical model;
MPSLL is a maximum side lobes level;
NLFM is a non-linear frequency modulation;
SLL is a side lobe level;
FM is a frequency modulation;
FMR is a frequency modulating rate.

NOMENCLATURE

n=1, 2, 3 is a fragment sequence number LFM signal;

f_0 is a initial signal frequency, Hz;

t is a time, sec;

T_n is a duration n-th fragment LFM signal, sec;

T_S is a duration NLFM signal, sec;

$\varphi_n(t)$ is a instantaneous phase n-th fragment LFM signal, rad;

$\beta_n = \Delta f_n / T_n$ is a FMR n-th fragment LFM signal, Hz/sec;

$\Delta\beta_{21}$ is a difference of FMR 2-nd and 1-st LFM fragments, Hz/sec;

$\Delta\beta_{31}$ is a difference of FMR 3-rd and 1-st LFM fragments, Hz/sec;

DOI 10.15588/1607-3274-2023-4-4

$\Delta\beta_{32}$ is a difference of FMR 3-rd and 2-nd LFM fragments, Hz/sec;

$\tilde{\beta}_n$ is a modified FMR *n*-th fragment LFM signal, Hz/sec;

Δf_n is a frequency deviation *n*-th fragment LFM signal, Hz;

$\Delta\tilde{f}_n$ is a modified deviation *n*-th fragment LFM signal, Hz;

$\delta\varphi_{mn}$ is a phase jump during transition from $m=n-1$-st fragment LFM signal on *n*-th, rad;

INTRODUCTION

The modern stage of development of radar facilities is characterized by wide introduction of modular construction of receiving-transmitting devices using the technology of phased antenna arrays on solid-state generator devices and application of internal pulse modulation of frequency (phase) of probing signals [1–9]. These technologies and technical solutions are interrelated, since they aim to achieve the required radiated power under conditions of limiting the peak power of an individual transmitting module. Signals from LFM [1–10] are used as wide sounding signals, the methods of forming and processing of which are constantly being improved.

The main efforts of the researchers are focused on reducing the MPSLL of the autocorrelation function of LFM signals, which, without additional measures, is approximately –13 dB. As a rule, for this purpose, Wt is used in the time (spectral) region of the received radar signal [10–14] and signals with a rounded spectrum [10, 15–20] are used, which is equivalent to Wt in time.

Rounding the spectrum LFM signal brings its shape closer to the bell-shaped one and leads to a decrease in the effective spectrum width and, as a result, to the expansion of the main lobe of the ACF signal. That is, the deterioration of the range discriminating ability in the case of the use of Wt or the rounding of the signal spectrum is a fee for reducing the MPSLL. This is acceptable because from the point of view of detecting weak signals, minimizing MPSLL is a more important task. In [12, 19, 20] it is noted that the application of Wt to signals with a rounded spectrum gives a better final result than the separate use of these methods.

A common method of obtaining signals with a rounded spectrum is the use of intra-pulse NLFM, which, in comparison with LFM signals, provides significantly lower MPSLL [1, 2, 8–10, 12, 15–31]. NLFM signals are widely used, consisting of two or three fragments combined in time with linear or nonlinear FM, or their combinations [10, 12, 15, 17, 19–22, 27–31]. Unlike LFM signals, which have practically fixed MPSLL regardless of the values of the input variables of their MM [1–14], MPSLL of multifragment NLFM signals significantly depends on the ratio of the duration of fragments and deviations of their frequency and varies widely depending on their magnitude.

The studies carried out by the authors [30, 31] showed that the known MM NLFM signals, consisting of two and three LFM fragments [15, 25, 27–29, 32–34], have a significant drawback – they do not take into account frequency and phase jumps (or only phases, depending on the temporal representation of MM). These jumps are caused by an instantaneous change in the FM value, that is, the ratio of the frequency deviation of the FM fragment to its duration, at the moment of transition from one fragment to another. Several methods of their compensation are offered, which allow to take full advantage of such signals. It was also found that their MPSLL depends on what value the initial and final phase of each of the LMF fragments has.

Searching for signals with minimum MPSLL values by optimizing their frequency-time parameters is a complex problem that belongs to the class of dynamic programming problems, because changing the parameters of previous fragments of the NLFM signal leads to changes in the parameters of subsequent fragments. Therefore, the paper proposes a method that simplifies the search for local minima of the MPSLL by adjusting the input MM variables under the conditions of providing an integer number of radio oscillation periods in each of the fragments of the NLFM signal.

The object of study is the process of formation and processing of two- and three-fragment NLFM signals.

The subject of study are mathematical models of NLFM signals, which consist of two and three LFM fragments.

The purpose of the work is to develop a method for simplifying the search for local minima of two- and three-fragment NLFM signals by using a modified MM with a whole number of periods of radio oscillations of LFM fragments.

1 PROBLEM STATEMENT

The authors propose [30, 31] MM for calculating the instantaneous phase of two- and three-fragment NLFM signals in the current time, in which frequency and phase jumps are compensated during the transition from the previous to the next LFM fragment. Let's use MM [31] for the three-fragment NLFM signal:

$$\varphi(t)=2\pi\begin{cases} f_0 t+\dfrac{\beta_1 t^2}{2}, 0\le t\le T_1; \\ \left(f_0-\Delta\beta_{21}T_1\right)t+\dfrac{\beta_2 t^2}{2}+\delta\varphi_{12},\ T_1\le t\le T_{12}; \\ \left(f_0-\Delta\beta_{31}T_1-\Delta\beta_{32}T_2\right)t+\dfrac{\beta_3 t^2}{2}-\delta\varphi_{23}, \\ T_{12}\le t\le T_{\Sigma}, \end{cases} \quad (1)$$

in which, to ensure compactness of mathematical records, the designation of the difference FMR between the second and first fragments is introduced $\Delta\beta_{21}=\beta_2-\beta_1$, the difference FMR between third and first fragments

DOI 10.15588/1607-3274-2023-4-4

$\Delta\beta_{31} = \beta_3 - \beta_1$, the difference FMR between third and second fragments: $\Delta\beta_{32} = \beta_3 - \beta_2$.

Similarly, the total duration of the first and second LFM fragments $T_{12} = T_1 + T_2$ and the total duration of the NLFM signal are introduced $T_S = T_1 + T_2 + T_3$. Additional MM variables (1), which are determined using already specified parameters, are the phase jump during the transition from the first to the second LFM fragment:

$$\delta\varphi_{12} = \frac{1}{2}\Delta\beta_{21}T_1^2 \qquad (2)$$

and phase jumps between the second and third LFM fragments:

$$\delta\varphi_{23} = \frac{1}{2}\left(\Delta\beta_{31}T_1^2 + \Delta\beta_{32}T_2^2\right) \qquad (3)$$

Thus, in addition to the current time *t* and the initial frequency f_0, the input parameters of the model (1) are Δf_n and T_n that can be corrected.

In the presentation of the material, the designation of the operation of finding the nearest larger integer ceil was used. The two-step model is a separate case (1) when using the first two components. Model (1) can be adapted for the descending law of FM, if you use the negative sign of FMR in the calculated ratios.

The result of the research is to modify the model (1) so that the radio oscillations of each LFM fragment have an integer number of periods. The results of the assessment of MPSLL and the rate of decline of SLL are compared with those obtained earlier [15, 17, 19–22, 27–34].

2 REVIEW OF THE LITERATURE

The first publications on the use of NLFM signals, which have fragments with LFM and NLFM, appeared in the 60s of the last centuries [16, 17]. Subsequently, these developments received a more detailed theoretical basis, which emphasized the need to ensure the continuity of the instantaneous phase of such signals. For this, NLFM signals with symmetrical rounding of the spectrum were proposed, which in theory should have provided an MPSLL of –42.8 dB [10]. However, even today such an MPSLL is unattainable for known multi-fragment NLFM signals. Interest in using NLFM signals, which consist of LFM fragments, is associated with the results of developments in the field of their formation and processing [1–11, 13, 14].

A number of researchers [15–29, 32–34], along with tri-fragmental signals that can reduce MPSLL by an average of 6 dB, consider two-fragment NLFM signals, which, in comparison with LFM, provide a decrease in MPSLL by about 3dB. The peculiarity of the use of NLFM signals, which consist of two and three LFM fragments, is that the MPSLL of the resulting signal depends significantly on the frequency-time parameters of the fragments, and therefore it is very difficult to obtain a stable value of the MPSLL even with slight changes in the parameters of the signals. Therefore, the choice of parameters must be approached carefully, for example, in [27] an algorithm for selecting parameters is proposed in order to minimize the MPSLL of a two-fragment NLFM signal.

Frequency and phase jumps at the junctions of previous LFM fragments are included as components in the frequency and phase expressions of subsequent fragments (see the second and third fragments in expression (1)), which complicates the optimization of parameters. In the previously developed MM of both current and shifted time [10, 12, 15, 16, 17, 19–22, 25, 27–29, 32–34], such distortions were not taken into account. In works [30, 31] on the example of MM of the current time of two- and three-fragment NLFM signals it is shown that such frequency-phase distortions arise as a result of instantaneous change in the value of FMR and a method of their compensation is proposed. The range of change of initial parameters of NLFM signals (input variables of their MM) after compensation of frequency and phase jumps at the joints of LFM fragments is somewhat expanded, however, as practice shows, the possibilities of such expansion have not yet been exhausted. At the same time, the actual task remains to minimize the SLL of the correlation functions of NLFM signals.

3 MATERIALS AND METHODS

We apply the concept of the full phase Ψ_n *-nth* LFM fragment, which is equal to the phase of the radio frequency oscillation during its duration. For the first LFM fragment, based on (1), its total phase is:

$$\Psi_1 = 2\pi\left(f_0T_1 + \frac{\beta_1}{2}T_1^2\right). \qquad (4)$$

We impose the condition that the complete phase of the LFM fragment (4) has an integer number of complete periods of radio frequency oscillations $2\pi N_1$. In this case:

$$N_1 = ceil\left(f_0T_1 + \frac{\beta_1}{2}T_1^2\right). \qquad (5)$$

In order to satisfy (5) without applying the *ceil* operation, it is necessary to solve this equation with respect to β_1 and obtain a new modified FMR $\tilde{\beta}_1$ value, which for the first LFM fragment is:

$$\tilde{\beta}_1 = \frac{2\left(N_1 - f_0T_1\right)}{T_1^2}. \qquad (6)$$

Since the frequency jump, and therefore the instantaneous phase at the junction of the fragments caused by the change in the FMR, remains, these jumps still need to be

DOI 10.15588/1607-3274-2023-4-4

compensated. Taking into account the above and on the basis of (1) for the second LFM fragment, by analogy we have:

$$\Psi_2 = 2\pi\left([f_0 - (\beta_2 - \tilde{\beta}_1)T_1]T_2 + \frac{\beta_2}{2}T_2^2 \right) = 2\pi N_2; \quad (7)$$

$$\tilde{\beta}_2 = \frac{2\left(N_2 - [f_0 - (\beta_2 - \tilde{\beta}_1)T_1]T_2\right)}{T_2^2}. \quad (8)$$

As you can see, in (7), (8), a modified value is already involved, that is, the calculation of the modified parameters $\tilde{\beta}_1$ of the next fragment is performed taking into account the modification of the parameters of the previous fragments of the NLFM signal. Similarly, for the third LFM fragment, we write:

$$\Psi_3 = 2\pi\left([f_0 - (\beta_3 - \tilde{\beta}_1)T_1 - (\beta_3 - \tilde{\beta}_2)T_2]T_3 + \frac{\beta_3}{2}T_3^2 \right) = 2\pi N_3,$$

where:

$$\tilde{\beta}_3 = \frac{2\left\{N_3 - [f_0 - (\beta_3 - \tilde{\beta}_1)T_1 - (\beta_3 - \tilde{\beta}_2)T_2]T_3\right\}}{T_3^2}. \quad (9)$$

Thus, the essence of the proposed method for minimizing ACF SLL signals from NLFM is as follows. Due to modification of FMR values, an integer number of periods of radio frequency oscillations is formed at each of the signal sections, which eliminates the cause of phase jumps due to an arbitrary value of the final phase of the LFM fragment. However, the frequency jumps due to the change of the FMR with the transition to the next fragment are preserved and must be compensated. Taking into account the above, we have a modified MM in the current time for the instantaneous phase of the NLFM signal, which consists of three LFM fragments, with an integer number of radio oscillation periods and compensation for frequency and phase jumps at the joints of the fragments:

$$\varphi(t) = 2\pi \times \begin{cases} f_0 t + \frac{\tilde{\beta}_1 t^2}{2}, \ 0 \le t \le T_1; \\ [f_0 - (\tilde{\beta}_2 - \tilde{\beta}_1)T_1]t + \frac{\tilde{\beta}_2 t^2}{2} + \delta\tilde{\varphi}_{12}, \ T_1 \le t \le T_{12}; \\ [f_0 - (\tilde{\beta}_3 - \tilde{\beta}_1)T_1 - (\tilde{\beta}_3 - \tilde{\beta}_2)T_2]t + \frac{\tilde{\beta}_3 t^2}{2} - \delta\tilde{\varphi}_{23}, \\ T_{12} \le t \le T_S. \end{cases} \quad (10)$$

The compensating phase components (2), (3) in (10) should be calculated using expressions (11) and (12) already taking into account the modification of the FMR:

$$\delta\tilde{\varphi}_{12} = \frac{1}{2}(\tilde{\beta}_2 - \tilde{\beta}_1)T_1^2; \quad (11)$$

$$\delta\tilde{\varphi}_{23} = \frac{1}{2}\left((-\tilde{\beta}_3 + \tilde{\beta}_1)T_1^2 + (\tilde{\beta}_3 - \tilde{\beta}_2)T_2^2\right). \quad (12)$$

The modified FMR values for (10)–(12) are in accordance with (6), (8), (9). It should be noted that with the transition from the first fragment to the second, the FMR decreases, that is, it has a negative increase, and at the junction of the second and third fragments it increases – the increase is positive. This is taken into account by the opposite signs of the compensating phase components $\delta\tilde{\varphi}_{12}$, $\delta\tilde{\varphi}_{23}$, in (10) and by changing the signs of the FMR, $\tilde{\beta}_3$, $\tilde{\beta}_1$ in the first term (12).

The FMR rate is a time-frequency parameter because it is determined by frequency deviation Δf_n and duration T_n of the corresponding fragment. In case of modification of the FMR $\tilde{\beta}_n$ with constant duration of the LFM fragment, its frequency deviation changes:

$$\Delta\tilde{f}_n = \tilde{\beta}_n T_n, \quad (13)$$

Thus, MM (10) is obtained, which changes the frequency-time parameters of two- and three-fragment NLFM signals by modifying the values of the FMR fragments while providing an integer number of complete periods of radio frequency oscillations for each of the FMR fragments. A feature of the proposed MM is the correction of the initial values of the frequency deviations of the LFM fragments, which simplifies the process of finding local minima of the MPSLL. It should be noted that modification of the initial MM (1) leads to expansion of the possible range of values of frequency-time parameters, ratios of durations and deviations of frequency of LFM fragments, which ensure stability of MM operation.

4 EXPERIMENTS

Operability of the developed MM with adjustment of frequency-time parameters of two- and three-fragment NLFM signals was checked in the MATLAB application package. For verification, an experimental scheme was used, identical for both two- and three-fragment signals – groups of ten signals with the same input variables for each type of signals were studied. The obtained MPSLL for two-fragment NLFM signals is not higher than –18.0 dB, and for three-fragment signals – not higher than –22.0 dB. The ranges of possible changes in the ratios of the duration of LFM fragments, as well as their frequency deviations, in which the stable operation of the MM is observed, were also determined.

5 RESULTS

MPSLL was first evaluated for MM (1) with compensation for frequency and phase jumps at the joints of LFM fragments, and then for MM (10) with frequency parameters correction by modifying FMR values.

Table 1 shows the values of frequency-time parameters of two-fragment NFLM signals and the obtained MLSS values for the case of MM (1) with compensation of frequency and phase jumps. Corrected values of time-frequency parameters and corresponding MPSLL are

DOI 10.15588/1607-3274-2023-4-4

given in parentheses using MM (10). The parameters in the table are arranged in the order of increasing the duration of NLFM signals, and in the case of the same duration – by increasing the total frequency deviation.

Table 1 – Frequency-time parameters and MPSLL of two-segment NLFM signals

No.	T_1, μs	T_2, μs	Δf_1, kHz	Δf_2, kHz	MPSLL, dB
1.	12	50	75 (100)	150 (280)	–18.25 (–20.85)
2.	25	140	100 (80)	180.0 (271.43)	–16.96 (–18.83)
3.	30	180	100.0 (66.66)	150.0 (233.33)	–15.61 (–18.91)
4.	30	180	150.0 (133.33)	300.0 (445.45)	–16.77 (–17.89)
5.	35	180	100.0 (114.29)	260.0 (388.89)	–17.20 (–18.40)
6.	40	180	100 (100)	220.0 (322.22)	–18.10 (–18.10)
7.	45	220	100.0 (88.88)	190 (280)	–16.84 (–18.12)
8.	45	220	150.0 (133.33)	300 (445.45)	–16.95 (–18.06)

Analysis of Table 1 indicates that the use of adjusted time-frequency parameters in 80% of cases for this data set provides a decrease in MPSLL. It should be noted that the use of the proposed method for selecting frequency-time parameters simplifies the process of finding local minima of the MPSLL. It should be noted that in some cases the modified values of frequency deviation of the LFM fragments differ quite significantly from the initial ones.

Frequency-time parameters of three-fragment NLFM signals and corresponding MPSLL values are given in Table 2.

Table 2 – Frequency-time parameters and MPSLL of three-fragment NLFM signals.

No.	T_1, μs	T_2, μs	T_3, μs	Δf_1, kHz	Δf_2, kHz	Δf_3, kHz	*MPSLL*, dB
1.	15	75	15	30 (133.33)	60 (77.3)	30 (72)	–13.80 (–24.62)
2.	25	100	25	80 (80)	200 (200)	100 (120)	–21.08 (–22.50)
3.	25	100	25	95 (80)	200 (200)	100 (120)	–22.68 (–22.50)
4.	30	150	30	95 (133.33)	200 (200)	90 (80)	–15.95 (–21.74)
5.	30	150	30	150 (200)	400 (400)	150 (133.33)	–18.96 (–22.61)
6.	30	150	30	260 (266.67)	590 (596)	260. (261.33)	–18.93 (–21.25)
7.	30	150	30	300 (333.33)	790 (796)	310 (328)	–21.13 (–21.58)
8.	40	200	40	100 (100)	200 (200)	90 (80)	21.13 (–21.96)

Comparison of the results of Table 2 indicates that the use of adjusted frequency-time parameters for three-fragment NLFM signals in the vast majority of cases (for the given data set – 90%) provides a decrease in MPSLL, which indicates the feasibility of using (10) in practical activities.

In the case of using modified frequency parameters of two- and three-fragment signals for MM (1) and (10), the obtained results completely coincide, which indicates the adequacy of the proposed method. So it should be, since model (10) is a separate case of model (1) under the condition of an integer number of periods of radio oscillations.

In case of FMR modification, the new values of fragment frequency deviations may differ from the given initial ones, which ensures their automatic selection and variability. Subsequently modified by (10) deviation frequency is advisable to use as input variables for MM (1).

The possibilities for obtaining potentially achievable MPSLL values and the rate of decline of the SLL in the case of adjusting the frequency-time parameters of three-fragment NLFM signals are shown in Fig. 1 – Fig. 4. Fig. 1 shows the results of application (10) for signal № 1 from Table 1, Fig. 1a shows the signal spectrum, and Fig. 1b shows its ACF on a logarithmic scale, the MPSLL is –20.85 dB. Note that such a MPSLL value is achieved with a relatively small value of the total frequency deviation when the spectrum type approaches the bell-shaped one more.

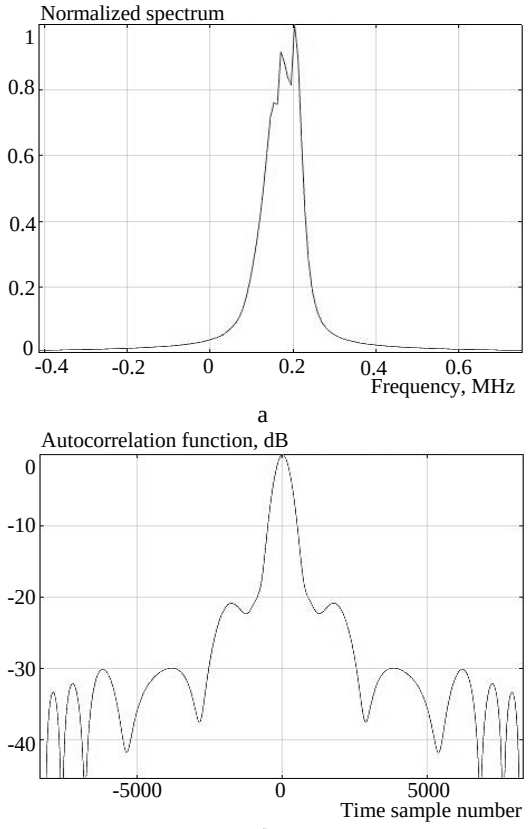
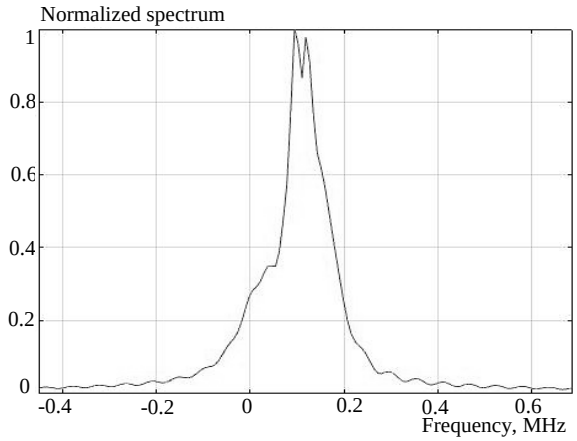


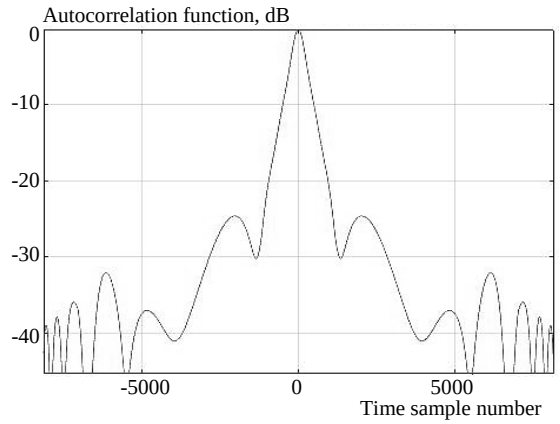
Figure 1– Spectrum (a), ACF (b) NLFM signal by model (10), parameter № 1 tab. 1

The simulation results shown in Fig. 2 are also obtained using (10), the set of parameters corresponds to signal № 1 of Table 2. The spectrum of Fig. 2a has a noticeable rounding, which led to a decrease in MPSLL to –24.62 dB (Fig. 2b).

DOI 10.15588/1607-3274-2023-4-4



a

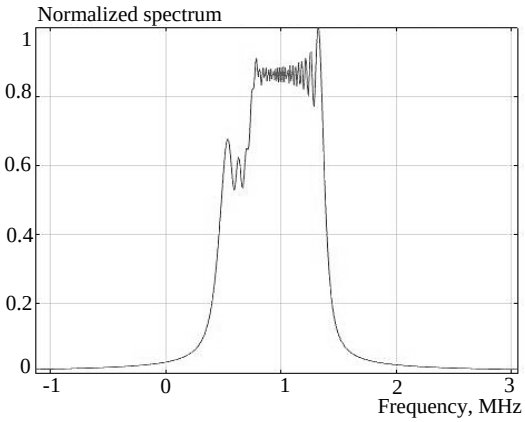


b

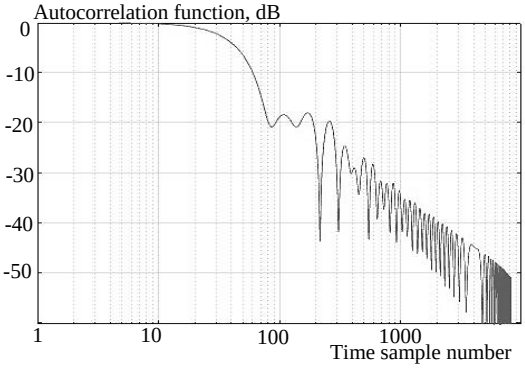
Figure 2– Spectrum (a), ACF (b) NLFM signal by model (10), parameters № 1 Tab. 2

To assess the rate of decline of the ACF SL, two- and three-fragment NLFM signals with larger values of the resulting duration and frequency deviation were used – parameter set № 8 from Table 1 and parameter set № 5 from Table 2. The results are shown in Fig. 3 and Fig. 4, respectively.

There is a clear decrease in the intensity of the signal spectrum in the low frequency region of Fig. 3a. Logarithmic scale along both coordinate axes is used for convenience of ACF signal analysis. For a two-segment NLFM signal, the SLL decay rate is estimated at about 22 dB/dec. (Fig. 3b). Analysis of Fig. 4a indicates a two-sided decrease in the intensity of the spectrum, as a result of which the ACF MPSLL decreased to the level of –22.61 dB. The rate of decline of the SLL is approximately 21.5 dB/dec.

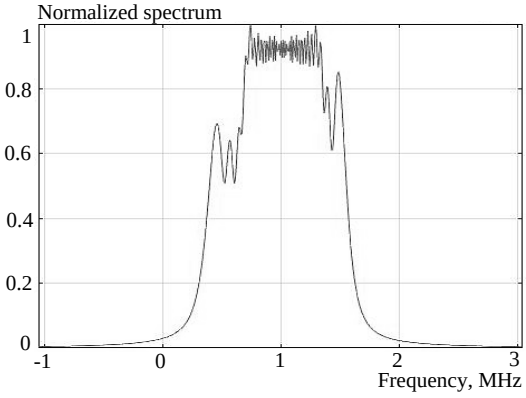


a

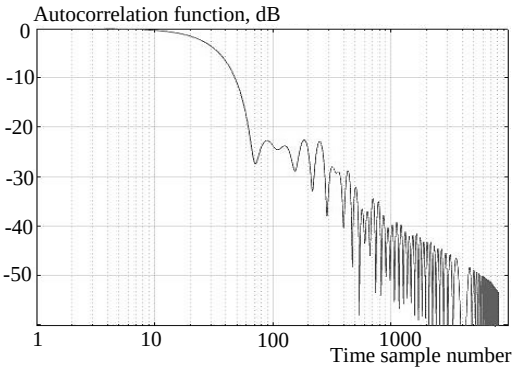


b

Figure 3 – Spectrum (a), ACF (b) NLFM signal by model (10), parameters № 10 Tab. 1



a



b

Figure 4 – Spectrum (a), ACF (b) NLFM signal by model (10), parameters № 7 tab. 2

DOI 10.15588/1607-3274-2023-4-4

Analysis of Fig. 1 – Fig. 4 indicates that the spectra of Fig. 1a, 2a, 3a, 4a do not have significant additional distortions in the form of breaks and dips, and the corresponding ACF of Fig. 1b, 2b, 3b, 4b do not have a sharp change in the frequency of side lobe pulsations and SLL drops. This confirms the absence of frequency and phase jumps in the resulting NLFM signals.

For the studied groups of signals, the ratio of durations and deviations of the frequency of the LFM fragments is determined, which ensures the stability of the MM operation and minimization of the SLL in the case when only the frequency-phase jumps are compensated, as well as when the FMR is further modified to provide a whole number of radio oscillation periods. The obtained results for the ratios of LFM fragment durations are grouped in Table 3, and for frequency deviations – in Table 4. The ratio is relative to the value corresponding to the parameter of the first LFM fragment.

Table 3 – Ratio of LFM fragment durations

Compensation of frequency and phase jumps			
2 fragments		3 fragments	
from 1:4.1	to 1:5	from 1:5:0.5	to 1:6.7:1
Modifications FMR			
from 1:4.5	to 1:6	from 1:4:1	to 1:5:1

Table 4 – Frequency deviation ratio LFM fragments

Compensation of frequency and phase jumps			
2 fragments		3 fragments	
from 1:2	to 1:2.3	from 1:2:0.9	to 1:2.5:1
Modifications FMR			
from 1:2.8	to 1:3.5	from 1:2:0.67	to 1:2.5:1.5

A comparative analysis of the results of Table 3 and Table 4 indicates that the ranges of change in the input parameters for (1) and (10) differ, that is, as a result of using (10), the total variability of the input variables for determining the parameters of NLFM signals increased.

6 DISCUSSIONS

Comparison of results obtained using (1) and (10) suggests that the proposed method of finding new values of input variables for (1) simplifies their selection and increases variability. When changing the input data, models (1) and (10) can give excellent results both in favor of one and in favor of another model. It should be noted that they mutually complement each other. That is, in studies of NLFM signals, both MM should be used and already on the basis of the obtained results, it should be concluded that it is advisable to use unmodified input variables or the results of their modification.

In the case of obtaining a lower MPSLL by adjusting the initial time-frequency parameters, it is advisable to use them as input for (1). This approach does not always work in the opposite direction, since as a result of modification in (10), the input variables from (1) may receive new values, and the achieved result may even deteriorate. In the course of experimental studies, it was found that the proposed method allows for two-fragment NLFM signals to obtain MPSLL values below –20 dB, and for three-fragment signals – below –24 dB with a relatively small total frequency deviation, which can be useful in a number of practical applications, for example, sonar, ultrasound diagnostics, etc. At the same time, signal spectra do not have significant distortions, which provides a potential possibility of achieving a high degree of their coordination with the frequency characteristics of the receiving-transmitting paths.

CONCLUSIONS

The scientific novelty. The paper proposes a new method of simplifying the search for local minima of MPSLL by adjusting the input variables of MM under the conditions of providing a whole number of periods of radio oscillations in each of the fragments of the NLFM signal. Thus, each LFM fragment begins with a zero phase, and ends with a value of 2 π radians, which ensures the stability of the model and in the overwhelming number of cases – an additional decrease in MPSLL. Use of the proposed method of modification of input frequency-time parameters for two-fragment NLFM signals for the investigated group of signals ensured reduction of MPSLL by 3 dB, and for three-fragment signals – up to 3.5 dB. It should be noted that the use of the proposed method for selecting frequency-time parameters simplifies the process of finding local minima of the MPSLL.

It has been determined that two- and three-fragment NLFM signals generated with the help of the proposed MM have a higher rate of decline of MPSLL, which is estimated at 21–22 dB/dec, compared to the LFM signals. The optimal ratios of durations and deviations of the frequency of LFM fragments have been determined experimentally, subject to these, stable operation of models is ensured and, in most cases, – less than the MPSLL value.

The practical value of the obtained results lies in the possibility of using the proposed method for selecting parameters of NLFM signals, which consist of two and three LFM fragments. This can be used to develop devices for generating and processing radio signals of various applications, for example, radar of air targets, aviation and space systems for inspecting the earth's surface, weather location, sonar, ultrasonic diagnostics, etc., in which NLFM signals can be used to reduce MPSLL independently or in combination with Wt in the receiving device.

Prospects for further research. In the future, it is planned to improve MM [15, 25, 27–29, 32–34] in terms of compensating for phase jumps at the joints of fragments of NLFM signals and to explore the possibilities of optimizing input variables for such MM.

ACKNOWLEDGEMENTS

We thank the management of Ivan Kozhedub Kharkiv National Air Force University for the opportunity to conduct scientific research.

DOI 10.15588/1607-3274-2023-4-4

REFERENCES

1. Skolnik M. I. Radar Handbook. Editor in Chief. Boston, McGraw-Hill Professional, 2-nd Edition, 1990, 846 p.
2. Meikle H. Modern radar systems. Boston, London, Artech House, 2-nd Edition, 2008, 701 p.
3. Van Trees H. L. Detection, Estimation, and Modulation Theory. Edition, reprint, John Wiley & Sons, 2004, 716 p.
4. Curry G. R. Radar System Performance Modeling. London, Artech House, 2-nd Edition, 2004, 410 p.
5. Levanon N. Mozeson E. Radar Signals. Hoboken, John Wiley & Sons, 2004, 403 p.
6. Barton D. K. Radar System Analysis and Modeling. London, Artech House, 2005, 545 p.
7. Richards M. A. Scheer J. A., Holm W. A. Principles of modern radar. SciTech Publishing, 2010, 924 p.
8. Mervin C., Budge J. , Shawn R. G. Basic Radar Analysis. London, Artech House, 2015, 727 p.
9. Melvin W. L., Scheer J. A. Principles of modern radar. New York, SciTech Publishing, 2013, 846 p.
10. Cook C. E. Bernfeld M. Radar Signals: An Introduction to Theory and Application. Boston, Ar-tech House, 1993, 552 p.
11. Heinzel G., Rüdiger A., Schilling R. Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flattop windows [Electronic resource]. Access mode: https://pure.mpg.de/rest/items/item_52164_1/component/file_152163/content.
12. Valli N. A., Rani D. E., Kavitha C. Windows For Reduction of ACF Side Lobes of Pseudo-NLFM Signal, *International Journal of Scientific & Technology Research*, 2019,Vol. 8, Issue 10, pp. 2155–2161.
13. Doerry A. W. Generating nonlinear FM chirp waveforms for radar. Sandia Report SAND2006-5856, 2006, 34 p. DOI:10.2172/894743.
14. Doerry A. W. Catalog of Window Taper Functions for Side lobe Control [Electronic resource]. Access mode: https://www.researchgate.net/publication/316281181_Catalog_of_Window_Taper_Functions_for_Sidelobe_Control.
15. Fan Z., Meng H.-Y. Coded excitation with Nonlinear Frequency Modulation Carrier in Ultrasound Imaging System, *IEEE Far East NDT New Technology & Application Forum (FENDT)*. Kunming, Yunnan province, China, 20–22 Nov. 2020, P. 31–35. DOI:10.1109/FENDT50467.2020.9337517.
16. Cook C. E., Paolillo J. A pulse compression predistortion function for efficient side lobe reduction in a high-power radar, *Proceedings of the IEEE*, 1964, Vol. 52, Issue 4, pp. 377–389. DOI:10.1109/proc.1964.2927.
17. Cook C.E. A class of nonlinear FM pulse compression signals, *Proceedings of the IEEE*, 1964, Vol. 52(11), pp. 1369–1371. DOI: 10.1109/PROC.1964.3393.
18. Xu Z., Wang X., Wang Y. Nonlinear Frequency-Modulated Waveforms Modeling and Optimization for Radar Applications, *Mathematics*, 2022, Vol. 10, Article № 3939. DOI: 10.3390/math10213939.
19. Ghavamirad R., Sebt M. A. Side Lobe Level Reduction in ACF of NLFM Waveform, *IET Radar, Sonar & Navigation*, 2019, Vol. 13, Issue 1, pp. 74–80. DOI: 10.1049/iet-rsn.2018.5095.
20. Alphonse S., Williamson G. A. Evaluation of a class of NLFM radar signals, *EURASIP Journal on Advances in Signal Processing*, 2019, Article № 62, 12 p.
21. Saleh M., Omar S.-M., Grivel E. et al. A Variable Chirp Rate Stepped Frequency Linear Frequency Modulation Waveform Designed to Approximate Wideband Non-Linear Radar Waveforms, *Digital Signal Processing*, 2021, Vol. 109, Article № 102884. DOI:10.1016/j.dsp.2020.102884.
22. Chukka A., Krishna B. Peak Side Lobe Reduction analysis of NLFM and Improved NLFM Radar signal, *AIUB Journal of Science and Engineering (AJSE)*, 2022, Vol. 21, Issue 2, pp. 125–131. DOI: https://doi.org/10.53799/ajse.v21i2.440.
23. Zhao Y., Ritchie M., Lu X. et al. Non-continuous piecewise nonlinear frequency modulation pulse with variable subpulse duration in a MIMO SAR Radar System, *Remote Sensing Letters*, 2020, Vol. 11, Issue 3, pp. 283–292. DOI:10.1080/2150704X.2019.1711237.
24. Kurdzo J. M. Cho John Y. N., Cheong B. L. et al. A Neural Network Approach for Waveform Generation and Selection with Multi-Mission Radar, *IEEE Radar Conference*. Boston, 22–26 April 2019, Article № 19043446. DOI: 10.1109/RADAR.2019.8835803.
25. Nettem A. V., Rani E. Doppler Effect Analysis of NLFM Signals, *International Journal of Scientific & Technology Research*, 2019, Vol. 8, Issue 11, pp. 1817–1821.
26. Jin G., Deng Y.-K., Wang R. et al. An Advanced Nonlinear Frequency Modulation Waveform for Radar Imaging With Low Side Lobe, *IEEE Transactions on Geoscience and Remote Sensing*, 2019, Vol. 57, Issue 8, pp. 6155–6168. DOI: 10.1109/TGRS.2019.2904627.
27. Adithyavalli N. An Algorithm for Computing Side Lobe Values of a Designed NLFM function, *International Journal of Advanced Trends in Computer Science and Engineering*, 2019, Vol. 8, Issue 4, pp. 1026–1031. DOI:10.30534/ijatcse/2019/07842019.
28. Valli N. A., Rani D. E., Kavitha C. Performance Analysis of NLFM Signals with Doppler Effect and Back-ground Noise, *International Journal of Engineering and Advanced Technology (IJEAT)*, 2020, Vol. 9, Issue 3, pp. 737–742. DOI: 10.35940/ijeat.B3835.029320.
29. Valli N. A., Rani D. E, Kavitha C. Modified Radar Signal Model using NLFM, *International Journal of Recent Technology and Engineering (IJRTE)*, 2019, Vol. 8, Issue 2S3, pp. 513–516. DOI: 10.35940/ijrte.B1091.0782S319.
30. Kostyria O. O., Hryzo A. A., Dodukh O. M. et al. Mathematical model of a two-fragment signal with a non-linear frequency modulation in the current period of time, *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia*, 2023, Vol. 92, pp. 60–67. DOI: 10.20535/RADAP.2023.92.60-67.
31. Kostyria O. O., Hryzo A. A., Dodukh O. M. et al. Improvement of mathematical models with time-shift of two- and trifragment signals with non-linear frequency modulation, *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia*, 2023, Vol. 93, pp. 22–30. DOI: 10.20535/RADAP.2023.93.22-30.
32. Widyantara M. R., Suratman S.-F. Y., Widodo S. et al.Analysis of Non-Linear Frequency Modulation (NLFM) Waveforms for Pulse Compression Radar, *Jurnal Elektronika dan Telekomunikasi*, 2018, Vol. 18, No. 1, pp. 27–34. DOI: 10.14203/jet.v18.27-34.
33. Kavitha C., Valli N. A., Dasari M. Optimization of two-stage NLFM signal using Heuristic approach, *Indian Journal of Science and Technology*, 2020, Vol. 13(44), pp. 4465–4473. DOI:10.17485/IJST/v13i44.1841.
34. Anoosha C., Krishna B. T. Peak Side Lobe Reduction analysis of NLFM and Improved NLFM Radar signal with Non-Uniform PRI, *Aiub Journal of Science and Engineering (AJSE)*, 2022, Vol. 21, Issue 2, pp. 125–131.

Received 02.10.2023.
Accepted 30.10.2023.

DOI 10.15588/1607-3274-2023-4-4

УДК 621.396.962

СПОСІБ МІНІМІЗАЦІЇ РІВНЯ БІЧНИХ ПЕЛЮСТОК АВТОКОРЕЛЯЦІЙНИХ ФУНКЦІЙ СИГНАЛІВ З НЕЛІНІЙНОЮ ЧАСТОТНОЮ МОДУЛЯЦІЄЮ

Костиря О. О. – д-р техн. наук, с.н.с., провід. наук. співр. Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Гризо А. А. – канд. техн. наук, доцент, начальник НДЛ Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Додух О. М. – канд. техн. наук, пров. наук. співр. Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Лісогорський Б. А. – канд. техн. наук, старш. наук. співр. Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

Лук'янчиков А. А. – старш. наук. співр. Харківського національного університету Повітряних Сил імені Івана Кожедуба, Харків, Україна.

АНОТАЦІЯ

Актуальність. У теперішній час при створенні нових та модернізації існуючих радіолокаційних систем широко використовуються твердотільні генераторні прилади, що накладає певні обмеження на пікову потужність зондувальних сигналів. Для подолання цього обмеження застосовуються сигнали більшої тривалості з внутрішньо імпульсною модуляцією. Основні зусилля дослідників зосереджуються на зниженні максимального рівня бічних пелюсток автокореляційної функції таких сигналів, який без прийняття додаткових мір має суттєвий рівень, що утруднює роботу систем виявлення та стабілізації рівня хибних тривог. Увагою користуються сигнали з нелінійною частотною модуляцією, які складаються з двох та трьох лінійно-частотномодульованих фрагментів. Максимальний рівень бічних пелюсток таких сигналів суттєво залежить від частотно-часових параметрів фрагментів, а тому дуже складно отримати його стабільне значення. Пошук сигналів з мінімальними значеннями рівня бічних пелюсток шляхом оптимізації їх частотно-часових параметрів є складною задачею, бо зміна параметрів попередніх фрагментів сигналу призводить до змін параметрів наступних фрагментів.

Метою роботи є розробка способу для спрощення пошуку локальних мінімумів рівня бічних пелюсток дво- та трифрагментних сигналів з нелінійною частотною модуляцією за рахунок використання модифікованої математичної моделі з цілим числом періодів радіоколивань лінійно-частотномодульованих фрагментів.

Метод. Розроблений спосіб спирається на запропоновану модифікацію математичної моделі, яка здійснює коригування частотно-часових параметрів дво- та трифрагментних сигналів з нелінійною частотною модуляцією за рахунок модифікації значень швидкості частотної модуляції при забезпеченні цілого числа повних періодів радіочастотних коливань для кожного з фрагментів, що спрощує процес знаходження локальних мінімумів рівня бічних пелюсток.

Результати. Модифікація початкової математичної моделі призводить до розширення можливого діапазону значень частотно-часових параметрів, співвідношень тривалостей та девіацій частоти лінійно-частотномодульованих фрагментів та забезпечує стійкість роботи математичної моделі при зниженні значення максимального рівня бічних пелюсток автокреляційної функції.

Висновки. Експериментально підтверджено, що використання запропонованого способу модифікаціїї вхідних частотно-часових параметрів сигналів з нелінійною частотною модуляцією у переважній більшості випадків забезпечує зниження максимального рівня бічних пелюсток та спрощує процес знаходження його локальних мінімумів. Визначено оптимальні співвідношення тривалостей та девіацій частоти фрагментів сигналу, при дотриманні таких забезпечується стійка робота моделей та у більшості випадків – менше значення максимального рівня бічних пелюсток.

КЛЮЧОВІ СЛОВА: математична модель; сигнал з нелінійною частотною модуляцією; автокореляційна функція; максимальний рівень бічних пелюсток.

ЛІТЕРАТУРА

1. Skolnik M. I. Radar Handbook. Editor in Chief / M. I. Skolnik. – Boston: McGraw-Hill Professional, 2-nd Edition, 1990. – 846 p.
2. Meikle H. Modern radar systems / H. Meikle – Boston, London: Artech House, 2-nd Edition, 2008. – 701 p.
3. Van Trees H. L. Detection, Estimation, and Modulation Theory / H. L. Van Trees. – Edition, reprint: John Wiley & Sons, 2004. – 716 p.
4. Curry G. R. Radar System Performance Modeling / G. R. Curry. – London : Artech House, 2-nd Edition, 2004. – 410 p.
5. Levanon N. Radar Signals / N. Levanon, E. Mozeson. – Hoboken: John Wiley & Sons, 2004. – 403 p.
6. Barton D. K. Radar System Analysis and Modeling / D. K. Barton. – London : Artech House, 2005. – 545 p.
7. Richards M. A. Principles of modern radar / M. A. Richards, J. A. Scheer, W. A. Holm. – SciTech Publishing, 2010. – 924 p.
8. Mervin C. Basic Radar Analysis / C. Mervin, J. Budge, R. G. Shawn. – London : Artech House, 2015. – 727 p.
9. Melvin W. L. Principles of modern radar / W. L. Melvin, J. A. Scheer. – New York : SciTech Publishing, 2013. – 846 p.
10. Cook C. E. Radar Signals: An Introduction to Theory and Application / C. E. Cook, M. Bernfeld. – Boston : Ar-tech House, 1993. – 552 p.
11. Heinzel G. Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flattop windows [Electronic resource] / G. Heinzel, A. Rüdiger, R. Schilling. – Access mode: https://pure.mpg.de/rest/items/item_52164_1/component/file_152163/content.
12. Valli N. A. Windows For Reduction of ACF Side Lobes of Pseudo-NLFM Signal / N. A. Valli, D. E. Rani, C. Kavitha // International Journal of Scientific & Technology Research. – 2019. – Vol. 8, Issue 10. – P. 2155–2161.

DOI 10.15588/1607-3274-2023-4-4

13. Doerry A. W. Generating nonlinear FM chirp waveforms for radar. – Sandia Report SAND2006-5856, 2006. – 34 p. DOI:10.2172/894743.
14. Doerry A. W. Catalog of Window Taper Functions for Side lobe Control [Electronic resource] / A. W. Doerry. – Access mode: https://www.researchgate.net/publication/316281181_Catalog_of_Window_Taper_Functions_for_Sidelobe_Control.
15. Fan Z. Coded excitation with Nonlinear Frequency Modulation Carrier in Ultrasound Imaging System / Z. Fan, H.-Y. Meng // IEEE Far East NDT New Technology & Application Forum (FENDT): Kunming, Yunnan province, China, – 20–22 Nov. 2020. – P. 31–35. DOI:10.1109/FENDT50467.2020.9337517.
16. Cook C. E. A pulse compression predistortion function for efficient side lobe reduction in a high-power radar. / C. E. Cook, J. Paolillo // Proceedings of the IEEE. – 1964. – Vol. 52, Iss. 4. – P. 377–389. DOI:10.1109/proc.1964.2927.
17. Cook C.E. A class of nonlinear FM pulse compression signals / C. E. Cook // Proceedings of the IEEE. – 1964. – Vol. 52(11). – P. 1369–1371. DOI: 10.1109/PROC.1964.3393.
18. Xu Z. Nonlinear Frequency-Modulated Waveforms Modeling and Optimization for Radar Applications / Z. Xu, X. Wang, Y. Wang // Mathematics. – 2022. – Vol. 10. – Article № 3939. DOI: 10.3390/math10213939.
19. Ghavamirad R. Side Lobe Level Reduction in ACF of NLFM Waveform / R. Ghavamirad, M. A. Sebt // IET Radar, Sonar & Navigation. – 2019. – Vol. 13, Issue 1. – P. 74–80. DOI: 10.1049/iet-rsn.2018.5095.
20. Alphonse S. Evaluation of a class of NLFM radar signals. /S. Alphonse, G. A. Williamson // EURASIP Journal on Advances in Signal Processing. – 2019. – Article № 62. – 12 p.
21. A Variable Chirp Rate Stepped Frequency Linear Frequency Modulation Waveform Designed to Approximate Wideband Non-Linear Radar Waveforms / [M. Saleh, S.-M. Omar, E. Grivel et al.] // Digital Signal Processing. – 2021. – Vol. 109. – Article № 102884. DOI:10.1016/j.dsp.2020.102884.
22. Chukka A. Peak Side Lobe Reduction analysis of NLFM and Improved NLFM Radar signal. / A. Chukka, B. Krishna // AIUB Journal of Science and Engineering (AJSE). – 2022. – Vol. 21, Issue 2. – P. 125–131. DOI: https://doi.org/10.53799/ajse.v21i2.440.
23. Non-continuous piecewise nonlinear frequency modulation pulse with variable sub-pulse duration in a MIMO SAR Radar System / [Y. Zhao, M. Ritchie, X. Lu et al.] // Remote Sensing Letters. – 2020. – Vol. 11, Issue 3. – P. 283–292. DOI:10.1080/2150704X.2019.1711237.
24. A Neural Network Approach for Waveform Generation and Selection with Multi-Mission Radar / [J. M. Kurdzo, Y. N. Cho John, B. L. Cheong et al.] // IEEE Radar Conference: Boston – 22–26 April 2019. – Article № 19043446. DOI: 10.1109/RADAR.2019.8835803.
25. Nettem A. V. Doppler Effect Analysis of NLFM Signals / A. V. Nettem, E. Rani // International Journal of Scientific & Technology Research. – 2019. – Vol. 8, Issue 11. – P. 1817–1821.
26. An Advanced Nonlinear Frequency Modulation Waveform for Radar Imaging With Low Side Lobe / [G. Jin, Y.-K. Deng, R. Wang et al.] // IEEE Transactions on Geoscience and Remote Sensing. – 2019. – Vol. 57, Issue 8. – P. 6155–6168. DOI: 10.1109/TGRS.2019.2904627.
27. Adithyavalli N. An Algorithm for Computing Side Lobe Values of a Designed NLFM function / N. Adithyavalli // International Journal of Advanced Trends in Computer Science and Engineering. – 2019. – Vol. 8, Issue 4. – P. 1026–1031. DOI:10.30534/ijatcse/2019/07842019.
28. Valli N. A. Performance Analysis of NLFM Signals with Doppler Effect and Back-ground Noise / N. Valli, D. E. Rani, C. Kavitha // International Journal of Engineering and Advanced Technology (IJEAT). – 2020. – Vol. 9, Issue 3. – P. 737–742. DOI: 10.35940/ijeat.B3835.029320.
29. Valli N. A. Modified Radar Signal Model using NLFM / N. A. Valli, D. E Rani, C. Kavitha // International Journal of Recent Technology and Engineering (IJRTE). – 2019. – Vol. 8, Issue 2S3. – P. 513–516. DOI: 10.35940/ijrte.B1091.0782S319.
30. Mathematical model of a two-fragment signal with a non-linear frequency modulation in the current period of time / [O. O. Kostyria, A. A. Hryzo, O. M. Dodukh et al.] // Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia. – 2023. – Vol. 92. – P. 60–67. DOI: 10.20535/RADAP.2023.92.60-67.
31. Improvement of mathematical models with time-shift of two- and tri-fragment signals with non-linear frequency modulation / [O. O. Kostyria, A. A. Hryzo, O. M. Dodukh et al.] // Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia. – 2023. – Vol. 93. – P. 22–30. DOI: 10.20535/RADAP.2023.93.22-30.
32. Analysis of Non-Linear Frequency Modulation (NLFM) Waveforms for Pulse Compression Radar / [M. R. Widyantara, S.-F. Y. Suratman, S. Widodo et al.] // Jurnal Elektronika dan Telekomunikasi. – 2018. – Vol. 18, No. 1. – P. 27–34. DOI: 10.14203/jet.v18.27-34.
33. Kavitha C. Optimization of two-stage NLFM signal using Heuristic approach / C. Kavitha, N. A. Valli, M. Dasari // Indian Journal of Science and Technology. – 2020. – Vol. 13(44). – P. 4465–4473. DOI:10.17485/IJST/v13i44.1841.
34. Anoosha C. Peak Side Lobe Reduction analysis of NLFM and Improved NLFM Radar signal with Non-Uniform PRI / C. Anoosha, B.T.Krishna // Aiub Journal of Science and Engineering (AJSE). – 2022 – Vol. 21, Issue 2. – P. 125–131.

DOI 10.15588/1607-3274-2023-4-4

UDC 004.93

TEMPORAL EVENTS PROCESSING MODELS IN FINITE STATE MACHINES

Miroshnyk M. A. – Doctor of Science, Professor, Professor of the Department of Theoretical and Applied Systems Engineering, V. N. Karazin, Kharkiv National University, Kharkiv, Ukraine.

Shmatkov S. I. – Doctor of Science, Professor, Head of the Department of Theoretical and Applied Systems Engineering, V. N. Karazin, Kharkiv National University, Kharkiv, Ukraine.

Shkil O. S. – PhD, Associate Professor of the Department of Computer Engineering Design Automation, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine.

Miroshnyk A. M. – Assistant of the Department of Computer Engineering Design Automation, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine.

Pshenychnyi K. Y. – Postgraduate student of the Department of Computer Engineering Design Automation, Kharkiv National University of Radioelectronics, Kharkiv, Ukraine.

ABSTRACT

Context. The issue of a synthesizable finite state machine with temporal events processing using hardware description language pattern. The object of this study is external event processing in real-time systems.

Objective. The goal of this work is to introduce methods to express external temporal events on finite state machine state diagrams and corresponding HDL patterns of such events processing in control systems.

Method. The classification of external events in real-time systems is analyzed. A device class that changes its internal state depending on the temporal external events is introduced. A method to express these events on the temporal state diagram is introduced. Possible model behavior scenarios based on the external event duration are analyzed. A Verilog HDL external event processing pattern is introduced. The efficiency of the proposed model is proved by developing, verifying, and synthesis of a power-saving module in Xilinx ISE. The results and testing showed the model's correctness.

Results. External temporal events processing methods in real-time device models are proposed. The corresponding HDL pattern for the proposed model implementation is presented.

Conclusions. The real-time systems with external temporal events automated synthesis problem has been solved. To solve this problem, a finite state machine model-based device using the Verilog language was developed and tested. The scientific novelty lies in the introduction a method to express temporal events on the state diagram of the finite state machine as well as in a HDL when implementing the proposed model on CPLD and FPGA.

KEYWORDS: FSM pattern, HDL, real time devices, temporal events, electronic design automation.

ABBREVIATIONS

FSM is a finite state machine;
CAD is a computer aided design;
RTL is a register transfer level;
HDL is a hardware description language.

NOMENCLATURE

$X = \{X_C, X_E\}$ is a set of input variables;

X_C is a set of input signals from the control object;

X_E is a set of external events;

$Y = \{Y_C, Y_F\}$ is a set of output variables;

Y_c is a set of control functions;

Y_f is a set of initial functions;

Z is a set of internal variables that determine the encoding of FSM states;

f is a transition function;

g is an output function;

Z_0 is a initial FSM state;

$T_c = \{t_{c1}, t_{c2....}t_{cp}\}$ is a set of timed variables for timing constraints on each arc of the state diagram;

p is a number of arcs in the state diagram;

$t_{ci} = \{1, k\}$ is a maximum number of constraints on transitions to the i-th node of the state diagram in the event processing mode;

$T_{to} = \{t_{to1}, t_{to2}...t_{ton}\}$ is a set of timed variables for timing constraints for timeout of each FSM state;

$t_{toi} = \{1, n\}$ is a timeout for each state;

n is a number of FSM states;

$T_d = \{t_{d1}, t_{d2}...t_{dm}\}$ is a set delays for implementing output signal;

m is a number of output variables;

$t_{dm} = \{1, l\}$ is a range of possible delay values for each output signal;

l is a maximum number of clock cycles for implementing output signals in the indicated state of the FSM.

INTRODUCTION

Logic control systems represent a significant node any digital system. Such systems use the binary alphabet to determine the behavior of a control unit. The FSM pattern is known to be a widespread model for such systems. FSM itself is mathematical abstraction that can be represented in various ways such as state-transition table, state or flow-chart diagram, etc. When using the FSM pattern, it is essential to understand that the target logic unit behavior depends on the events happening in the external environment.

The clock cycles determine the machine time in which FSMs operate. However, the real-time devices work in the

DOI 10.15588/1607-3274-2023-4-5

metric time. In other words, such devices state depends both on the input signals and the time during which these signals are processed [2]. Thereby there are arises a requirement to express the metric time in terms of clock cycles because transitions between the FSM states directly depend on the timing aspect. It's also required to express the timing constraints on the state diagram as a starting modeling point.

HDLs has proven their indispensable role in both modern hardware development and verification processes. The high level of abstraction provided by language construct allows to implement algorithms of any difficulty. Modern CAD systems provide a range of tools to examine design before actual circuit implementation allowing to check the model correctness.

The automata-based description style is a way to structure HDL description when implementing a control logic device. It's the automata pattern that defines the behavior of the target device including timing parameters such as delays and duration requirements.

Thus, there exists a task of developing a unified HDL pattern for real time devices with external temporal events that can be used for the device implementation on PLD hardware platform (FPGA, CPLD).

The object of study is automatized real-time digital control logic devices development process.

The subject of study is events model and their processing in FSM-based real-time devices.

1 PROBLEM STATEMENT

Suppose given general temporal control unit model: $Y(t) = g(X(t), Z(t), T)$, $Z(t+1) = f(X(t), Z(t), T)$. Where $T = \{t_c, t_o, t_d\}$ represents automata time parameters: t_c – time constraints, t_o – output timeouts, t_d – input delays.

For given input constraints $t_c(X(i)) = [c1]$ that represents external event the problem of processing these events in real time control logic device models can be presented.

The main goal of this research is to study the usage of external temporal events in real time device. Such events have a requirement of a minimal duration required for a device response. Also, this paper addresses the following problems:

– define a class of devices that depend on the external temporal events;

– define a way to represent this event class on the FSM state diagram;

– simulate possible scenarios of such event processing;

– introduce HDL patterns that would express temporal events processing.

The object of study is real-time devices control algorithms.

The subject of study is real-time control device automata patterns.

2 REVIEW OF THE LITERATURE

In [1] any computing device is represented as a combination of two parts –operational unit and computing unit. This form of representation is based on the following principles:

– any operation can be represented as a sequence of basic logical and arithmetic operation upon input data;

– logical condition define the sequence of the operations;

– microprogram is representation of an algorithms in terms of microoperations and logical conditions;

– microprogram defines the structure of a device and the way it operates in time.

Operational unit stores words of information, performs operations upon them and calculates logical conditions required for algorithm. Operation unit us defined by a range of finite sets:

$Y = \{y_1, y_m\}$ defines microoperations;

$X = \{x_1, x_l\}$ defines logical conditions that are required during the algorithm flow;

$D = \{d_1, d_h\}$ represents operands;

$R = \{r_1, r_q\}$ represents operation results;

$S = \{s_1, s_n\}$ – internal words that represent information during computations.

The control unit generates a sequence of control signals based on logical conditions. This way the control unit defines the order of the operations.

The timed FSM as way to describe real-time devices was introduced in [2, 3]. In these works, the state diagram is extended with a finite state of real value timers. Each timer is reset to zero during a transition and increased with each FSM cycle. Each transition has an associated clock constraint. This means that transition can only be executed when all timer values satisfy this restriction. These papers introduce a generalized model of temporal control unit: $Y(t) = g(X(t), Z(t), T)$, $Z(t+1) = f(X(t), Z(t), T)$. Here X is the set of input signals, Z – the set of internal automata states, Y – the set of output signals, t – the machine time defined by clock cycles, d – the output function, t – the transition function. $T = \{t_c, t_o, t_d\}$ is the set of automata time parameters: t_c – time constraints, t_o – output timeouts, t_d – input delays.

In [2] all state machines are classified into three categories: regular, timed and recursive. Here the timed FSM are the FSM which have at least one time-dependent transition. For each machine category a general HDL pattern is proposed. As in [3] state loops and additional variable is used to implement the delays between state transitions.

In [4, 5] timed FSM that consider state timeouts and output signals delays. In the proposed model FSM makes a transition into a determined state in case no input signals were triggered during a timeout. Testing methods for TFSM were considered as well.

The method for implementing models of FSM in the VHDL language was proposed in [7, 8]. To implement delays in the states of the Moore FSM, it is proposed to use loops in states and a special variable count, which

DOI 10.15588/1607-3274-2023-4-5

decreases by 1 in each FSM cycle, which corresponds to a sync pulse.

Event based timed automata description and minimization method are described in [9, 10].

The method for constructing a HDL description for real-time systems, which can be synthesized into a programmable logic device, was developed in the [13, 14]. Models of real-time systems, such as a state diagram, a timed FSM with many timers, a timed FSM with one timer, an extended timed FSM, a timed FSM with timeouts and timing constraints were considered. Based on these models, a complete structural model of timed Moore FSM was proposed.

The issue of finding homing sequences for finite state machines and their length is addressed in [11]. Lower/Upper-bound parametric timed automata without invariants is described in [12].

In [16] timed systems in which some timing features are unknown parameters are addressed. The Upper-bound Parametric Timed Automata (U-PTAs), one of the simplest extensions of timed automata with parameters, in which parameters are only used as clock upper bounds are addressed.

In [17] logical discrete event systems modeling is addressed. It also discusses the diagnosability and opacity in the context of partially observed systems.

The issue of modeling and verification of digital discrete event systems is discussed in [18]. This work introduces the connection between Event-B methodology and automata modeling. This paper is important for verification of the automata based models.

In [19] a new classification of system events is during the proposed. This classification is used during the software and hardware specification development process. There are three classes of external events:

– business event – an action by a human user that stimulates a dialog with the software or hardware, as when the user presses a button;

– signal event – such event is registered when the system receives a control signal, data reading, or interrupt from an external hardware device or another software system;

– temporal event – a time-triggered event, as when the computer's clock reaches a specified time or when a preset duration has passed since a previous event (as in a system that logs a sensor's temperature reading every 10 seconds).

None of the works describes automata events with a duration longer than 1 automata cycle. In this paper, the FSM models with external temporal events are considered.

3 MATERIALS AND METHODS

The temporal state diagram is a visual representation of the real-time automata. This state diagram is extended with a timer used for state delays. The timer is used for keeping the automat in a certain state during a fixed amount of clock cycles. An example of timed FSM is shown on Fig. 1.

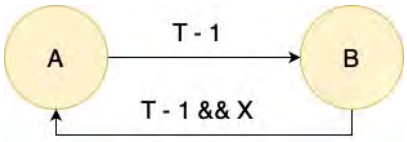


Figure 1 – Timed FSM example

Here the *AB* transition is a regular timed transition during which the automata remains in the state *A* during *T* clock cycles. The condition is written in the form $T - 1$ because the internal timer will run from $t = 0$ to $t = T - 1$. Here the *BA* transition depends on the auxiliary timer state and the input signal *X* (conditionally timed transition). Thereby the FSM will remain in state *B* at least during $T - 1$ clock cycles. From the HDL point of view, the time-related transitions are implemented via staying in a particular state, i.e., state loop. This transition type is called conditionally-timed.

Fig. 2 elaborates on both timed and conditionally timed transitions for the FSM from the Fig. 1 using an explicit timer value in the condition description.

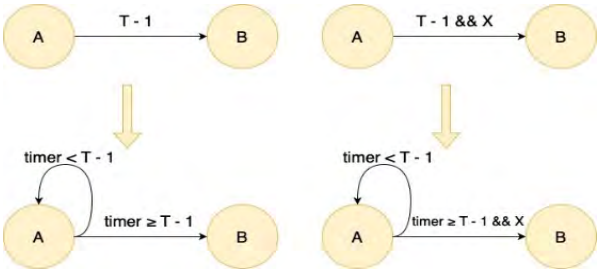


Figure 2 – Timed transitions

As it was already said there exists a class of devices that are sensible to external events with a certain duration. Event time parameters are considered when modeling such devices. For example, pushing the button for certain time turn-on a device; cooling must be enabled in a case a certain temperature level is maintained in a room; a system must transfer to the emergency mode in case the operator ignores error messages for 5 minutes. The mentioned device class must be described using the existing automata-based methods.

The event time constraints is a value $t_c(X(i)) = [c1]$ that represents the minimal duration of a signal. In the button example the $t = 3$ seconds. This constraint must be expressed on the temporal state diagram as transition. Such transition must explicitly compare the timer value with the $t_c(X(i))$ value.

Such transition is shown on Fig. 3. Like with regular timed transition the condition is based upon auxiliary timer value, but in this case the timer must have a minimal value for a state transition.

DOI 10.15588/1607-3274-2023-4-5

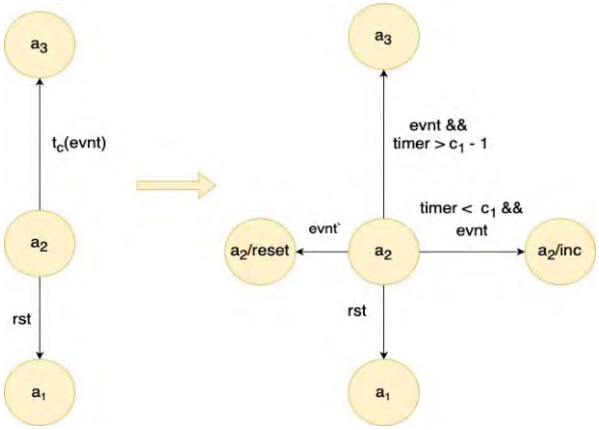


Figure 3 – Event with time constraint transition

For example, for the successful state transition between a_1/a_2 signal *evnt* must be active high and the timer value is greater than $t_c(evnt)$.

Internally the a_2 state consists of 2 states: *a*2/*reset* and *a*2/*inc*. In case the signal *evnt* is high, but the timer value is less than $t_c(evnt)$ the automata transfers into the and *a*2/*inc* substate where the timer is incremented. In case the signal *evnt* is low and the timer value is less that $t_c(evnt)$ – the automate goes into the substate *a*2/*reset* where the timer is reset to 0. It's essential that automata has a timer-independent transition to avoid livelocks. In this example this is the a_2/a_1 transition activated by signal *rst*.

There are four possible scenarios of automata behavior depending on the nature of external events. To illustrate these examples let's consider an automata transition that occur after signal *evnt* maintains high logical level during at least 4 clock cycles. The first scenario describes a situation when the external event lasts exactly 4 clock cycles. In this case the automata transfers into the next state and the internal counter is reset. Figure 4 shows a waveform of this scenario – *evnt* maintains high logical level during 4 clock cycles (the minimal amount of clock cycles required for the transition) and the automat transfers to state *a*3. The calculation points are marked with a red dotted line.

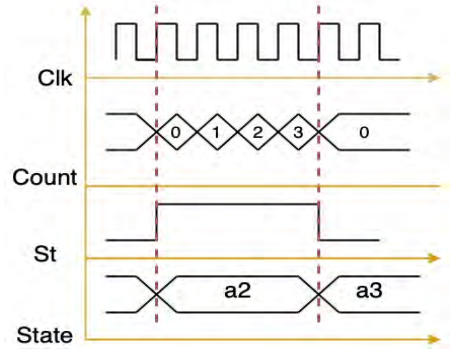


Figure 3 – Exact match event waveform

The next situation describes a case when the external event last longer than required. It worth noting that for the observed devices a longer event is not critical, because the minimal event duration condition is satisfied. Like in the previous case the automata will transfer to the next state and the internal timer will be reset. Corresponding waveform is shown on Figure 4. Signal *evnt* remains high during 5 clock cycles that is 1 clock cycle longer than required. After 4 cycles the automata transfers to the next state and the timer is reset. The event calculation points are marked with green dotted line, the automata transfer is marked with red dotted line.

The last scenario is when the event does not satisfy the timing requirements, i.e., lasts less than required. In this case the automata preserves its state with the timer being reset. Resetting the timer is an important operation, because the timer must be ready for the next event calculation. The waveform for this scenario is shown on Figure 5.

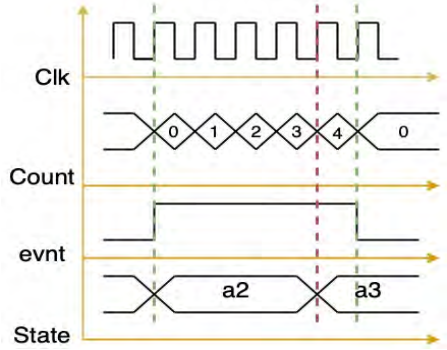


Figure 4 – Longer event waveform

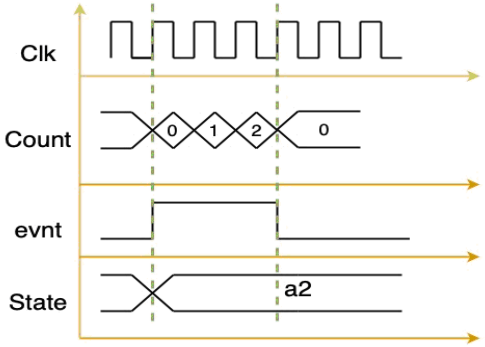


Figure 5 – Shorter event waveform

When describing temporal automata using state loops the deadlock situation must be addressed. The deadlock may occur in case the desired event does not occur at all. Thus, there must be a way to reset the automata into a predefined state. Typically a digital device has a reset signal with a higher priority than other signals. The priority is important when describing automata using hardware description languages like Verilog, VHDL. Figure 6 shows a waveform for the situation when both *evnt* and *rst* signals remain high, but the machine goes to the state a1, because *rst* has a higher priority.

DOI 10.15588/1607-3274-2023-4-5

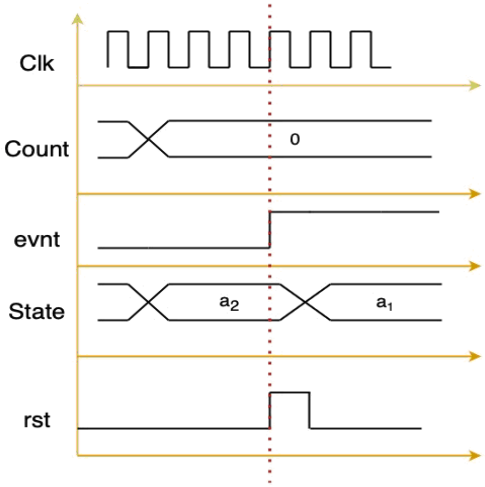


Figure 6 – A sync reset waveform

4 EXPERIMENTS

To illustrate the proposed event-based timed FSM model the power save module will be used. This module is usually used in power critical systems to decrease the energy usage whenever the system does not operate for a certain period.

The module operates in two modes: bypass and power saving. The input alphabet consists of the following binary signals X = {*onn*, *evnt*}, where *onn* – signal to enable power saving algorithm, *evnt* – signal notifying that an external event occurred, and the system must leave the power saving mode. The output binary signals set is Y = {*save*}, where *save* represent the power saving mode enter. The module interface and its connection with other system nodes are shown on Fig. 7.

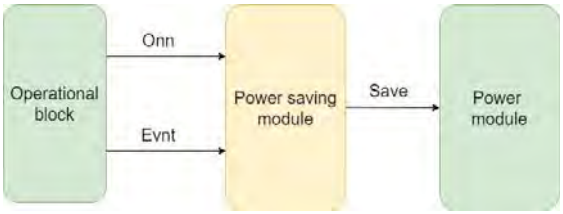


Figure 7 – Power saving module interface

The next step is to define the FSM states and timing constraints. The states set is the following:

– state a1 – the module is working in bypass mode, i.e. no external events are monitored;

– state a2 – power saving mode computation – in case no events occur during a fixed time period the FSM moves into the a3 state;

– state a3 – power saving mode.

The algorithm of the save power module is quite straightforward. Whenever this block is enabled by the *onn* signal it starts the *evnt* input signal monitoring. In case the *evnt* signal is not asserted during a fixed amount of time – the system must enter the power saving mode (output signal *save* is set high). The system exits the saving mode in case the *evnt* signal is asserted or the module is disabled via *onn* input signal with subsequent *save* signal de-assertion.

Thus, for the power saving module temporal state diagram of timed Moore FSM (Figure 8) with a proper HDL implementation can be built.

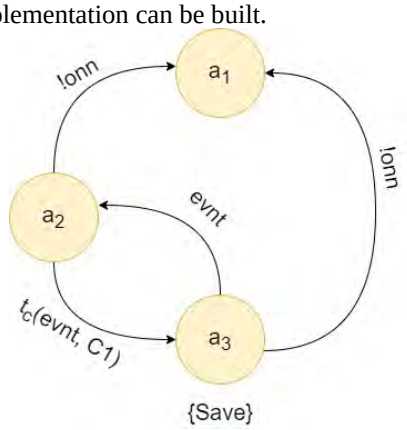


Figure 8 – Power saving module Moore timed FSM state diagram

Figure 9 represents FSM next-state Verilog process description.

Figure 10 shows the waveform of the module usage. Here t_c equals 5 clock cycles. The event calculation period is marked by red lines; blue lines denote signals change (*evnt* and *onn*) making FSM go into the event monitoring or bypass state.

CAD identified an FSM with 3 states for both XC3S500E-5fg320 and CPLD XC9572XL-53C44 chips. The state encoding is equal for both devices which is: a1=”00”, a2=”01”, a3=11”.

```
always @(state, evnt, onn, count)
  case (state)
    a1: if (onn) begin
          next_state = a2;
        end
       else begin
          next_state = a1;
       end

    a2: if (!onn) begin
          next_state = a1;
        end
        else if (count >= C1 - 1 && !evnt) begin
          next_state = a3;
          next_count = 3'd0;
        end
        else if (!evnt) begin
          next_state = a2;
          next_count = count + 1'b1;
        end
        else begin
          next_state = a2;
          next_count = 3'd0;
        end

    a3: if (!onn) begin
          next_state = a1;
        end
        else if (evnt) begin
          next_state = a2;
        end else begin
          next_state = a3;
        end

    default: begin
      next_state = a1;
      next_count = 3'd0;
    end
  endcase
```

Figure 9 – FSM next state process Verilog description

DOI 10.15588/1607-3274-2023-4-5

5 RESULTS

Xilinx ISE 14.7 CAD system was used for the design verification, implementation, and synthesis. Behavioral and post-implementation simulations for initial description were performed on the CPLD XC9572XL-10-TQ100 (Post-Fit Simulation) and on the FPGA XC3S500E-5fg320 (Post-Place & Route Simulation).

Figures 10 and 11 shows the post synthesis waveform for XC3S500E-5fg320 and XC9572XL-53C44 correspondingly.

Switching delay is 0 ns. Single short-term pulses do not occur. Thus, when implementing the device on FPGA and CPLD, its operation must comply with the original description (specification).

The expected minimum number of triggers is 9: 3 triggers for encoding 7 states, 6 triggers for the counter (since the maximum timeout is 40 clock cycles). RTL schematic report for both chips confirmed this. (since the maximum timeout is 40 clock cycles). RTL schematic report for both chips confirmed this.

Latch triggers are absent in the report. To implement functions of transitions and outputs, combinational circuits were synthesized. 52 combinational elements were synthesized on the XC9572XL-10-TQ100, and 21 combinational elements – on the XC3S500E-5fg320. Table 1 shows some totals from the synthesis report.

For FPGA: minimum clock period: 4.153 ns (Maximum Frequency: 240.790 MHz).

For CPLD: minimum clock period: 6.300 ns (Maximum Frequency: 158.730 MHz).

Table 1 – FPGA and CPLD based FSM synthesis results

PLD	Flip-Flops	Latches	BELS	Slices / LUTs
FPGA XC3S500E5fg320	8	0	21	11/21
CPLD XC9572XL-TQ100	8	0	52	–

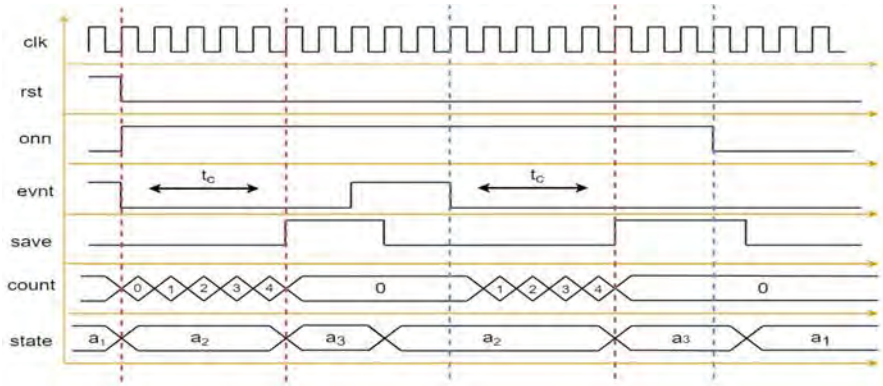


Figure 10 – Timing diagram of Moor FSM for power saving module

Figure 11 –Post-Place & Route Simulation of XC3S500E-5fg320

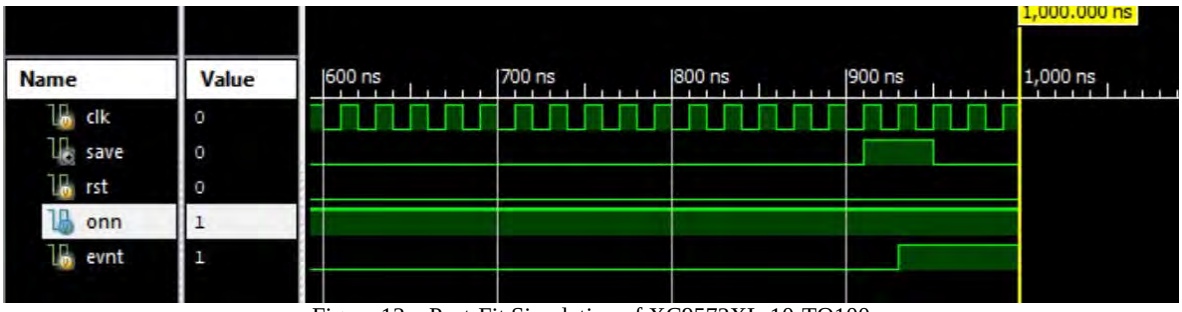


Figure 12 – Post-Fit Simulation of XC9572XL-10-TQ100

DOI 10.15588/1607-3274-2023-4-5

6 DISCUSSION

Moore FSM models with timing delays represent a traditional way for describing logic control systems. Timed FSM models with timeouts, constraints and output delays are used during control systems and telecommunication protocol testing [5–10]. In [13–15] a hardware implementation of such FSM models using hardware description languages was introduced. In [4] VHDL patterns for timed Moore control systems were presented. These patterns consider event processing external to the control system. However, no HDL patterns for processing external events that must satisfy certain timing constrains were mentioned.

In this research a general HDL pattern for processing temporal events was introduced. The model synthesizability was proven by implementing power saving model in Xilinx IDE on two chips – FPGA and CPLD. The results shown different hardware usage results. At the same time the behavioral simulation confirmed both devices satisfy specification.

Further works might be related to the verification of such HDL models, especially using assertion-based methods [2].

CONCLUSIONS

The issue of real-time devices modeling using the finite state machine pattern and hardware description languages was solved. The classification of events as models of finite state automaton interaction with the external environment is analyzed. Temporal events and a class of devices which functionality depends on such factors is are discussed. It is shown that such events are widespread in various digital devices and real-time systems. A new type of FSM transition is proposed for the temporal transitions state diagram as a digital systems design canonical model. The considered models are illustrated by simulations and timing diagrams analysis. The work solves the problem of designing time-controlled automata in real-time logic control systems. The issue of logical systems with external events processing that should last a certain time is considered.

An algorithm of a power saving unit was analyzed. Based on this algorithm, the temporal Moore state diagram was developed. The temporal state diagram was used to develop a three-process Verilog-model of the timed Moore FSM.

Verification, synthesis, and implementation of the developed Verilog model using Xilinx ISE environment were performed. Synthesis and simulation before and after implementation were performed for CPLD XC9572XL-10-TQ100 and FPGA XC3S500E-5fg320. Synthesis and simulation results of the circuit after implementation confirm the operability and correctness of the developed Verilog model.

The scientific novelty of the work lies in temporal state diagram improvement. This allowed to expand the class of event based real-time logical devices represented by FSM.

The practical significance of the obtained results lies in the following:

– investigate a class of real time control logic devices with corresponding FSM models and HDL description is introduced;

– the possibility of the proposed model is illustrated with examples that show various correlation between event duration and the time during which control FSM stays in a certain state;

– the theoretical basis is illustrated with the control unit for power saving device implementation on FPGA using Verilog HDL.

Prospects for further research are to use the proposed theoretical basis and results for event-based devices diagnostics systems [15].

ACKNOWLEDGEMENTS

The work was carried out within the framework of the state budget research topic of the Kharkov National University of Radioelectronics “Smart cyber university – Cloud-Mobile services for managing scientific and educational processes l” (state registration 0017U002524, order of the Ministry of Education of Ukraine) on the basis of the Design automation department of KNURE.

And out within the framework of the state budget research topic of the V. N. Karazin, Kharkiv National University “Modeling of information processes in complex and distributed systems” (state registration number 0121U109183, order of the Ministry of Education of Ukraine) on the basis of the of Theoretical and Applied Systems Engineering department of KNUK.

REFERENCES

1. Baranov S. Logic and System Design of Digital Systems. – Tallinn, TUT Press, 2008, 267 p. doi.org/10.1016/B978-0-7506-8397-5.00005-2.
2. Pedroni V. A. Finite state machines in hardware: theory and design (with VHDL and SystemVerilog). Cambridge, MA: MIT Press, 2013, 338 p. doi.org/10.7551/mitpress/9657.001.0001
3. Miroschnyk M., Shkil A., Kulak E., Rakhlis D., Filippenko I., Hoha M., Malakhov M., Serhiienko V. Design of real-time logic control system on FPGA, *Proceedings of 2019 IEEE East-West Design & Test Symposium (EWDTS’19), September 13–16.* Batumi, Georgia, 2019, pp. 488–491. doi.org/10.1109/ewdts.2019.8884387
4. Shalyto A. A. Software Automation Design: Algorithmization and Programming of Problems of Logical Control, *Journal of Computer and System Sciences International*, 2000, Vol. 39, No. 6, pp. 899–916. doi.org/10.1023/a:1012392927006
5. Alur R., Dill D. L. A theory of timed automata, *Theoretical Computer Science*, 1994, Vol. 126, No. 2, pp. 183–235. doi.org/10.1016/0304-3975(94)90010-8
6. Zhigulin M., Yevtushenko N., Maag S., Cavalli A. R. FSM-Based Test Derivation Strategies for Systems with Time-Outs, *Proceedings of the 11th International Conference on Quality Software (QSIC 2011)*. Madrid, 2011, pp. 141–149. doi.org/10.1109/qsic.2011.30
7. Solov’ev V. V., Klimowicz A. S., Structural models of finite-state machines for their implementation on

DOI 10.15588/1607-3274-2023-4-5

programmable logic devices and systems on chip, *Journal of Computer and Systems Sciences International*, 2015, Vol. 54, № 2, pp. 230–242. doi.org/10.1134/s1064230715010074
8. Shkil A. S., Miroschnyk M. A., Kulak E. N., Rakhlis D. Y., Miroshnyk A. M., Malahov N. V. Design timed FSM with VHDL Moore pattern, *Radio Electronics, Computer Science, Control*, 2020, № 2(53), pp. 137–148. /doi.org/10.15588/1607-3274-2020-2-14
9. Bresolin D., Tvardovskii A., Yevtushenko N., Villa T., Gromov M. Minimizing Deterministic Timed Finite State Machines, *In 14th IFAC Workshop on Discrete Event Systems WODES 2018. – IFAC-PapersOnLine*, 2018, Vol. 51, Issue 7, pp. 486–492. /doi.org/10.1016/j.ifacol.2018.06.344
10. Bresolin D., El-Fakih K., Villa T., Yevtushenko N. Equivalence checking and intersection of deterministic timed finite state machines, *Formal Methods in System Design*, 2022, № 7, pp. 1–26. doi.org/10.1007/s10703-022-00396-6
11. Tvardovskii A. S., Yevtushenko N. V. Deriving homing sequences for finite state machines with timed guard, *Automatic Control and Computer Sciences*, 2021, Vol. 55, № 7, pp. 738–750. doi.org/10.3103/s0146411621070154
12. André Etienne, Lime Didier, Ramparison Mathias TCTL Model Checking Lower/Upper-Bound Parametric Timed Automata Without Invariants, *Proc.International Conference Formal Modeling and Analysis of Timed Systems FORMATS 2018, September 4–6*. Biejing, China, 2018, pp. 37–52. doi.org/10.1007/978-3-030-00151-3_3
13. Wagner G. An abstract state machine semantics for discrete event simulation, *Proc. of the 2017 Winter Simulation Conference (WSC), 3–6 Dec. 2017*. Las Vegas, USA, 12 p. [Electronic resource] / IEEE Xplore Digital Library. Access mode: www / URL: https://ieeexplore.ieee.org/document/8247830. /doi.org/10.1109/wsc.2017.8247830
14. Shkil A., Miroschnyk M., Kulak E., Rakhlis D., Filippenko I., Malakhov M. Hardware implementation of timed logical control FSM, *Proc. of 2020 IEEE East-West Design & Test Symposium (EWDTS'20), Sept. 4–7*. Varna, Bulgaria, 2020.– 6 p. [Electronic resource] / IEEE Xplore Digital Library – Access mode: www / URL: https://ieeexplore.ieee.org/document/ 9225129. /doi.org/10.1109/ewdts50664.2020.9225129
15. Lamperti G. Zanella M., Zhao Xiangfu Introduction to Diagnosis of Active Systems. Springer, 2018, 353 p. /doi.org/10.1007/978-3-319-92733-6
16. André Etienne, Lime Didier, Ramparison Mathias TCTL Model Checking Lower/Upper-Bound Parametric Timed Automata Without Invariants, *Proc.International Conference Formal Modeling and Analysis of Timed Systems FORMATS 2018, September 4–6*. Biejing, China, 2018, pp. 37–52. doi.org/10.1007/978-3-030-00151-3_3
17. Stéphane Lafortune. Discrete Event Systems: Modeling, Observation, and Control, *Annual Review of Control, Robotics, and Autonomous Systems*, 2019, No. 2:1, pp. 141–159. doi.org/10.1146/annurev-control-053018-023659
18. Sabah Al-Fedaghi. Modeling Physical / Digital Systems: Formal Event-B vs. Diagrammatic Thinging Machine, *International Journal of Computer Science and Network Security*, 2020, No. 20 (4), pp. 208–220. hal-02614504 , version 1 (20-05-2020)
19. Karl Wiegers, Beatty Joy Software Requirements (Developer Best Practices) 3rd Edition, Developer Best Practices, 2013, 672 p. /doi.org/10.1109/9781118156629.ch3

Received 15.08.2023.
Accepted 04.11.2023.

УДК 004.93

МОДЕЛІ ТЕМПОРАЛЬНИХ ПОДІЙ У КІНЦЕВИХ АВТОМАТАХ

Мірошник М. А. – д-р техн. наук, професор, професор кафедри теоретичної та прикладної системотехніки Харківського національного університету імені В. Н. Каразіна, Харків, Україна.

Шматков С. І. – д-р техн. наук, професор, завідувач кафедри теоретичної та прикладної системотехніки Харківського національного університету імені В. Н. Каразіна, Харків, Україна.

Шкіль О. С. – канд. техн. наук, доцент, професор кафедри автоматизації проектування обчислювальної техніки Харківського національного університету радіоелектроніки, Харків, Україна.

Мірошник А. М. – асистент кафедри автоматизації проектування обчислювальної техніки Харківського національного університету радіоелектроніки, Харків, Україна.

Пшеничний К. Ю. – аспірант кафедри автоматизації проектування обчислювальної техніки Харківського національного університету радіоелектроніки, Харків, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто задачу розробки шаблонів кінцевих автоматів з обробкою зовнішніх темпоральних подій з використанням мов опису апаратури. Об'єктом роботи є питання моделювання зовнішніх подій у системах реального часу.

Мета роботи. Метою роботи є представити способи вираження темпоральних подій у на графі переходів кінцевого автомата, а також відповідні HDL шаблони обробки таких подій у системах управління.

Метод. Проаналізовано класифікацію зовнішніх подій у системах реального часу. Виділено клас пристроїв, у яких зміна стану відбувається внаслідок настання зовнішніх подій, що подовжені у часі (темпоральні події). Запропоновано спосіб вираження такого роду подій на темпоральному графі переходів кінцевого автомата. Проаналізовано різні сценарії поведінки запропонованої автоматної моделі в залежності від тривалості зовнішньої події. Розроблено HDL шаблони на мові опису апаратури Verilog для імплементації обробки темпоральних подій. Працездатність запропонованих методів доведено на прикладі розробки, верифікації та синтезу модуля збереження енергії на FPGA та CPLD у системі автоматизованого проектування Xilinx ISE. Отримані результати автоматизованого синтезу довели правильність запропонованої методології.

Результати. Запропоновано методи обробки зовнішніх темпоральних подій у моделях пристроїв реального часу. Представлено відповідні шаблони мові опису апаратури Verilog для імплементації запропонованої моделі.

Висновки. Вирішено задачу автоматизованого синтезу систем реального часу з зовнішніми темпоральними подіями. Для вирішення цієї проблеми були розроблені та протестовано модель пристроя на базі кінцевого автомата з використанням

DOI 10.15588/1607-3274-2023-4-5

мови Verilog. Наукова новизна полягає у представлені способу вираження темпоральних подій на графі переходів кінцевого автомата, а також за допомогою HDL конструкцій під час розробки систем керування на CPLD та FPGA у система автоматизованого синтезу.

КЛЮЧОВІ СЛОВА: автоматний шаблон, мова опису апаратури, пристрої реального часу, темпоральні події, системи автоматизованого синтезу.

ЛІТЕРАТУРА

1. Baranov S. Logic and System Design of Digital Systems / S. Baranov. – Tallinn : TUT Press, 2008. – 267 p. doi.org/10.1016/B978-0-7506-8397-5.00005-2.
2. Pedroni V. A. Finite state machines in hardware: theory and design (with VHDL and SystemVerilog) / Volnei A. Pedroni. – Cambridge, MA: MIT Press., 2013. – 338 p. doi.org/10.7551/mitpress/9657.001.0001 .
3. Design of real-time logic control system on FPGA / [M. Miroschnyk, A. Shkil, E. Kulak et al.] // Proceedings of 2019 IEEE East-West Design & Test Symposium (EWDTS'19), September 13–16, Batumi, Georgia, 2019. – P. 488–491. doi.org/10.1109/ewdts.2019.8884387
4. Shalyto A. A. Software Automation Design: Algorithmization and Programming of Problems of Logical Control / A. A. Shalyto // Journal of Computer and System Sciences International. – 2000. – Vol. 39, No. 6. – P. 899–916. doi.org/10.1023/a:1012392927006
5. Alur R. A theory of timed automata / R. Alur, D. L. Dill // Theoretical Computer Science. – 1994. – Vol. 126, No. 2. – P. 183–235. doi.org/10.1016/0304-3975(94)90010-8
6. FSM-Based Test Derivation Strategies for Systems with Time-Outs / [M. Zhigulin, N. Yevtushenko, S. Maag, A. R. Cavalli] // Proceedings of the 11th International Conference on Quality Software (QSIC 2011), Madrid, 2011. – P. 141–149. doi.org/10.1109/qsic.2011.30
7. Solov'ev V. V. Structural models of finite-state machines for their implementation on programmable logic devices and systems on chip / V. V. Solov'ev, A. S. Klimowicz // Journal of Computer and Systems Sciences International. – 2015. – Vol. 54, № 2. – P. 230–242. doi.org/10.1134/s1064230715010074
8. Design timed FSM with VHDL Moore pattern / [A. S. Shkil, M. A. Miroschnyk, E. N. Kulak et al.] // Radio Electronics, Computer Science, Control. – 2020. – №2(53). – P. 137–148. doi.org/10.15588/1607-3274-2020-2-14
9. Minimizing Deterministic Timed Finite State Machines / [D. Bresolin, A. Tvardovskii, N. Yevtushenko et al.] // In 14th IFAC Workshop on Discrete Event Systems WODES 2018. – IFAC-PapersOnLine, 2018. – Vol. 51, Issue 7. – P. 486–492. doi.org/10.1016/j.ifacol.2018.06.344
10. Equivalence checking and intersection of deterministic timed finite state machines / [D. Bresolin, K. El-Fakih, T. Villa, N. Yevtushenko] // Formal Methods in System Design – 2022. – № 7. – P. 1–26. doi.org/10.1007/s10703-022-00396-6
11. Tvardovskii A. S. Deriving homing sequences for finite state machines with timed guard / A. S. Tvardovskii, N. V. Yevtushenko // Automatic Control and Computer Sciences. – 2021. – Vol. 55, № 7. – P. 738–750. doi.org/10.3103/s0146411621070154
12. Ramparison M. TCTL Model Checking Lower/Upper-Bound Parametric Timed Automata Without Invariants / Ramparison Mathias, André Etienne, Lime Didier // Proc.International Conference Formal Modeling and Analysis of Timed Systems FORMATS 2018, September 4–6, Biejing, China, 2018. – P. 37–52. doi.org/10.1007/978-3-030-00151-3_3
13. Wagner G. An abstract state machine semantics for discrete event simulation / G. Wagner // Proc. of the 2017 Winter Simulation Conference (WSC), 3–6 Dec. 2017, Las Vegas, USA – 12 p. [Electronic resource] / IEEE Xplore Digital Library – Access mode: www / URL: https://ieeexplore.ieee.org/ document/8247830. doi.org/10.1109/wsc.2017.8247830
14. Hardware implementation of timed logical control FSM / [A. Shkil, M. Miroschnyk, E. Kulak et al.] // Proc. of 2020 IEEE East-West Design & Test Symposium (EWDTS'20), Sept. 4–7, Varna, Bulgaria, 2020. – 6 p. [Електронний ресурс] / IEEE Xplore Digital Library – Режим доступу: www / URL: https://ieeexplore. ieee.org/document/ 9225129. doi.org/10.1109/ewdts50664.2020.9225129
15. Lamperti G. Introduction to Diagnosis of Active Systems / G. Lamperti, M. Zanella, Xiangfu Zhao. – Springer, 2018. – 353 p. doi.org/10.1007/978-3-319-92733-6
16. Ramparison M. TCTL Model Checking Lower/Upper-Bound Parametric Timed Automata Without Invariants / Ramparison Mathias, André Etienne, Lime Didier // Proc.International Conference Formal Modeling and Analysis of Timed Systems FORMATS 2018, September 4–6, Biejing, China, 2018. – P. 37–52. doi.org/10.1007/978-3-030-00151-3_3
17. Stéphane Lafortune Discrete Event Systems: Modeling, Observation, and Control / Stéphane Lafortune // Annual Review of Control, Robotics, and Autonomous Systems. 2019. – No. 2:1. – P. 141–159. doi.org/10.1146/annurev-control-053018-023659 .
18. Sabah Al-Fedaghi Modeling Physical/Digital Systems: Formal Event-B vs. Diagrammatic Thinging Machine / Sabah Al-Fedaghi // International Journal of Computer Science and Network Security. – 2020. – No. 20 (4). – P. 208–220. hal-02614504 , version 1 (20-05-2020).
19. Karl Wiegers. Software Requirements (Developer Best Practices) 3rd Edition / Karl Wiegers, Joy Beatty. – Developer Best Practices, 2013. – 672 p. doi.org/10.1109/9781118156629.ch3

DOI 10.15588/1607-3274-2023-4-5

UDC 004.942:001.57

THE METHOD OF HYDRODYNAMIC MODELING USING A CONVOLUTIONAL NEURAL NETWORK

Novotarskyi M. A. – Dr. Sc., Professor of the Department of Computer Engineering, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

Kuzmych V. A. – Post-graduate student of the Department of Computer Engineering, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

ABSTRACT

Context. Solving hydrodynamic problems is associated with high computational complexity and therefore requires considerable computing resources and time. The proposed approach makes it possible to significantly reduce the time for solving such problems by applying a combination of two improved modeling methods.

Objective. The goal is to create a comprehensive hydrodynamic modeling method that requires significantly less time to determine the dynamics of the velocity field by using the modified lattice Boltzmann method and the pressure distribution by using a convolutional neural network.

Method. A method of hydrodynamic modeling is proposed, which realizes the synergistic effect arising from the combination of the improved lattice Boltzmann method and a convolutional neural network with a specially adapted structure. The essence of the method consists of implementing a sequence of iterations, each of which simulates the process of changing parameters when moving to the next time layer. Each iteration includes a predictor step and a corrector step. At the predictor step, the lattice Boltzmann method works, which allows us to obtain the field of fluid velocities in the working area at the next time layer using the field of velocities at the previous layer. At the corrector step, we apply an improved convolutional neural network trained on a previously created data set. Using a neural network allows us to determine the pressure distribution on a new time layer with a predetermined accuracy. After adding the fluid compressibility correction on the new time layer, we get a refined value of the velocity field, which can be used as initial data for applying the lattice Boltzmann method at the next iteration. Calculations stop when the specified number of iterations is reached.

Results. The operation of the proposed method was studied on the example of modeling fluid movement in a fragment of the human gastrointestinal tract. The simulation results showed that the time spent implementing the simulation process was reduced by 6–7 times while maintaining acceptable accuracy for practical tasks.

Conclusions. The proposed hydrodynamic modeling method with a convolutional neural network and the lattice Boltzmann method significantly reduces the time and computing resources required to implement the modeling process in areas with complex geometry. Further development of this method will make it possible to implement real-time hydrodynamic modeling in three-dimensional domains.

KEYWORDS: hydrodynamic modeling, convolutional neural network, lattice Boltzmann method.

ABBREVIATIONS

LBM is a lattice Boltzmann method;
CNN is a convolutional neural network;
BGK is a Bhatnagar-Gross-Krook model.

NOMENCLATURE

f is a distribution function;

x is a vector that specifies the position of the elementary volume of liquid in space;

k is a discretization index of the kinematic velocity;

v_k is a kinematic velocity vector;

w_k is weight factor for equilibrium function;

t is a parameter that specifies a point in time;

f^{eq} is an equilibrium distribution function;

η is a kinematic viscosity;

τ is a time of relaxation;

Δx is a grid spacing;

Δt is a time step;

c is a speed in the grid;

$\bar{\tau}$ is a viscous stress tensor;

$\otimes$ is the Kronecker product;

φ is a body forces;

μ is a dynamic viscosity;

$\bar{I}$ is an identity tensor.

INTRODUCTION

The rapid growth in the popularity of artificial neural networks, methods of analyzing large volumes of data, and other alternative approaches has led to the discovery of several breakthrough technologies. In particular, significant progress is observed in studying complex physical processes that can be mathematically described by boundary value problems based on differential equations with partial derivatives. In the paper, we will consider applying this approach to determining the hydrodynamic parameters of liquids. The traditional way of modeling the corresponding physical process is the numerical solution of the boundary value problem, which includes the flow continuity equation and the Navier-Stokes equation. In the practical implementation of this approach, in the case of a modeling area with a complex shape, some difficulties always arise, which are manifested due to the problems of achieving convergence of the corresponding numerical method, which is inextricably linked with the need to use significant computing resources to ensure obtaining results with high accuracy. However, the high accuracy of

DOI 10.15588/1607-3274-2023-4-6

the results could be more accurate in many cases due to the impossibility of accurately determining the initial data for solving this or that boundary value problem. A natural approach to solving this problem is developing a toolkit to solve the given situation with some predetermined approximation.

The object of study is the process of modeling the physical phenomenon of the movement of liquids in objects with complex geometry under the influence of external forces.

The subject of the study The subject of the study is a modeling method that uses the synergistic effect of the joint use of the improved lattice Boltzmann method and the application of a convolutional neural network with a special structure.

The purpose of the work The purpose of the work is to shorten the time of modeling changes in fluid movement parameters and increase the accuracy of parameter determination by correcting the deviation from compressibility.

1 PROBLEM STATEMENT

In the general case, modeling of fluid movement is performed by solving a boundary value problem based on the Navier-Stokes equation. For this purpose, the discretization of the equation and domain is used to apply the appropriate numerical methods. The main problem that arises on this path is ensuring the convergence of numerical methods in the case of studying an area with complex geometry and the significant computational complexity of the corresponding algorithms, which leads to a considerable expenditure of time and computing resources.

An alternative approach is to use the lattice Boltzmann method. This method describes the movement of the liquid at the mesoscopic level through the interaction of the elementary volumes of the liquid, which is represented by the Boltzmann equation:

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} = \frac{f - f^{eq}}{\tau}. \qquad (1)$$

The right-hand side of this equation describes the collision process of elementary volumes. This work uses the so-called BGK model, which is often applied to liquids moving at a speed that does not exceed the Mach number for the given liquid. The disadvantage of this approach is the accumulation of errors when determining the hydrodynamic parameters due to density fluctuations.

To overcome the abovementioned shortcoming, we offer a method of specifying the fluid pressure on each time layer.

2 REVIEW OF THE LITERATURE

Applying the lattice Boltzmann method makes it possible to describe the dynamics of the velocity field by colliding elementary volumes of liquid at the mesoscopic level. The history of publications on LBM already dates back several decades[1]. Many practical applications with a severe theoretical basis were implemented [2]. The popularity of this method is due to a large number of advantages, which include relatively low computational complexity [3], the possibility of studying areas with complex geometry [4], and the possibility of parallelizing computational processes [5].

A significant drawback of this method is the increase in the error in determining the parameters of the velocity field with an increase in the number of iterations. This shortcoming does not allow modeling complex physical processes with a given accuracy on long time intervals, which are essential for describing operations in living nature.

A neural network is proposed as a tool to overcome this drawback.

The main advantages of using methods based on neural networks:

1. A neural network provides an analytically presented solution that can be applied repeatedly after preliminary training.

2. Methods based on neural networks can be applied to a broad class of differential equations.

3. These methods include a smaller set of parameters and therefore require fewer resources for implementation.

4. Have mechanisms for regulating the level of generalization of the solution.

Today, a relatively large number of varieties of artificial neural networks are known. Most of them correspond to already well-developed methods of solving differential equations. Historically, an approach focused on deep neural networks was one of the first to be proposed [7]. The application of a deep neural network with extended learning based on the backpropagation algorithm is also offered in [8]. The mentioned methods can be used to solve both ordinary differential equations and differential equations with partial derivatives. They can be used to explore two-dimensional and three-dimensional areas. The disadvantage of the methods is the high labor intensity at the neural network learning stage.

Reducing the complexity can be achieved by modifying the structure of the neural network by adding radial basis functions (RBF) and other facilities to the deep neural network [8]. For example, [9] describes a method that combines a deep neural network with an evolutionary algorithm. The paper shows that the proposed structure can approximate arbitrary functions and their derivatives.

The paper [10] proposed a combination of an artificial neural network with the Karhunen-Loev decomposition [11], which is focused on modeling the solution of the two-dimensional Navier-Stokes equation.

Cellular neural networks [10, 11], an analog computing paradigm, are considered alternative structures for implementing differential equation-solving processes. This concept of analog computing is used to solve complex nonlinear differential equations. This method is based on the Taylor series expansion.

The modern approach to solving partial differential equations is based on convolutional neural networks

DOI 10.15588/1607-3274-2023-4-6

(CNN). Compared to deep neural networks, we get a significantly shorter training time and higher accuracy of solution modeling [12]. At the same time, the main significant advantage of using CNN is the possibility of using complex boundary conditions since the geometry of the area can be included in the learning process.

In this work, the use of a convolutional neural network is proposed. The structure of the network and the method are adapted to solving boundary value problems based on the two-dimensional Poisson equation, which is an essential component of the complex way that models the process of movement of an incompressible fluid.

3 MATERIALS AND METHODS

The way of modeling the movement of fluids aims to determine the distribution of pressure and the velocity field in a given area and changes in these parameters over time. The proposed complex method is represented by a cyclically repeated sequence of iterations, each containing two steps: a predictor step and a corrector step. Fig. 1 shows the general algorithm of this method, in which the functions of the predictor are concentrated in block 2, and the corrector's functions are implemented by blocks 4 and 5.

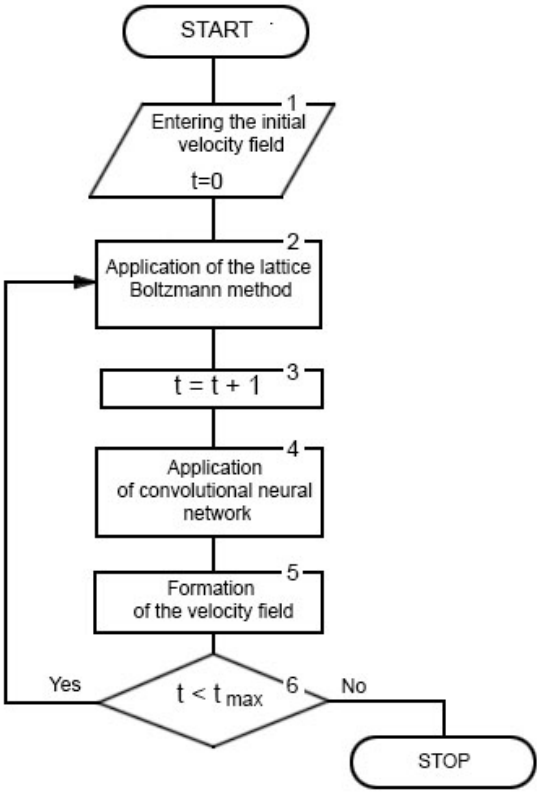


Figure 1 – Algorithm for implementing the complex method

Block 1. Within the framework of this block, the study area is discretized, and the initial data set is formed, which contains the parameters of the velocity field vectors specified in the discretization nodes. Block 2 uses the generated set as initial data to implement the lattice Boltzmann method. The lattice Boltzmann method is used to model the evolution of the parameters of the initial velocity field. Using the value of the velocity field for the time layer to perform one iteration within the framework of this model, we get an updated velocity field on the time layer. We determine the fluid pressure in the discretization nodes on the updated time layer with the value formed in block 3. The pressure distribution is calculated by a previously trained convolutional neural network, represented by block 4. Using the refined pressure value on the time layer in block 5, we determine the corrected velocity field on this time layer. If the current time layer is smaller than the maximum specified time layer, we use the refined velocity field as initial data for the lattice Boltzmann method.

This method is based on the use of the physical model of the Boltzmann equation, which corresponds to the behavior of the liquid flow at the macroscopic level [13]. To apply the Boltzmann equation, it is necessary to consider the fluid flow as a set of elementary volumes. Then the Boltzmann equation describes the evolution of one elementary volume in the form of a distribution function, where is the vector that specifies the position of the elementary volume of liquid in space, is the velocity vector of the elementary volume, and is the parameter that specifies the moment of time. In its general form, the Boltzmann equation can be represented by the expression (1).

We discretize the space of velocities at the mesoscopic level by applying a two-dimensional grid using nine bounce directions of elementary volumes, as shown in Fig. 2.

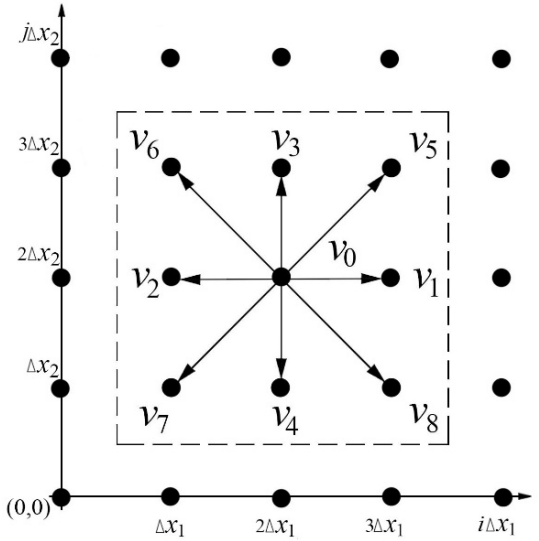


Figure 2 – Discretization of the domain and velocity field

Such a discretization scheme is called D2Q9. It provides sufficient modeling accuracy and is recommended, for example, in [16] for a similar class of problems. If we set $\Delta x = \Delta x_1 = \Delta x_2 = 1$, and the coordinates of the initial node $(0,0)$, then the coordinates of 9 vectors can be written as follows:

$$v_0 = (0,0), \quad v_1 = (1,0), \quad v_2 = (-1,0),$$

DOI 10.15588/1607-3274-2023-4-6

$$v_3 = (0,1), \quad v_4 = (0,-1), \quad v_5 = (1,1),$$
$$v_6 = (-1,1), \quad v_7 = (-1,-1), \quad v_8 = (1,-1).$$

For the region discretized in this way, we set the discretized Boltzmann equation:

$$\frac{\partial f_k}{\partial t} + v_k \frac{\partial f_k}{\partial x} = \frac{f_k - f_k^{eq}}{\tau}. \qquad (2)$$

The right-hand side of this equation describes the collision process of elementary volumes. This work uses the BGK model, which is often applied to liquids moving at a speed that does not exceed the Mach number for the given liquid.

Calculating the macroscopic parameters of speed and pressure for a liquid can be calculated as the result of the sequential execution of two steps for each of the discretization vectors v_k. In the first step, we will form a discrete kinetic equation for the distribution function of f_k:

$$f_k(x,t+\Delta t) = f_k(x,t) - \frac{1}{\tau}\left[f_k(x,t) - f_k^{eq}(x,t)\right], \qquad (3)$$

where $\tau = \frac{\eta}{c_s^2} + \frac{1}{2}$, $\Delta t = 1$.This equation determines the value of each distribution function at the domain discretization node after colliding the elementary fluid volumes at the corresponding node.

The second step determines the redistribution of the values of the distribution functions on the new time layer:

$$f_k(x + v_k \Delta t,\ t + \Delta t) = f_k(x, t + \Delta t). \qquad (4)$$

After determining the distribution functions at the given time level, we calculate the density and velocity of the fluid in the given discretization node using the formulas:

$$\rho = \sum_{k=0}^{8} f_k,\ u = \frac{1}{\rho}\sum_{k=0}^{8} f_k v_k\ . \qquad (5)$$

Equation (3) also includes the equilibrium distribution function $f_k^{eq}(x,t)$. We determine this function based on the calculations presented in [17]:

$$f_k^{eq} = \rho w_k\left(1 + \frac{3}{c^2} v_k u + \frac{9}{2c^4}(v_k u)^2 - \frac{3}{2c^2} u^2\right). \qquad (6)$$

where $c = \frac{\Delta x}{\Delta t}$, $w_0 = \frac{4}{9}$, $w_1 = w_2 = w_3 = w_4 = \frac{1}{9}$, $w_5 = w_6 = w_7 = w_8 = \frac{1}{36}$.

Based on [18], we can claim that the method remains stable under $\tau > 0.5$. But practical applications of the technique have shown that instability can also occur when tau goes to 0.5. Avoiding a significant growth of the lattice velocity helps maintain the method's stability. In the framework $c_s = \frac{c}{\sqrt{3}}$ the value of pressure is $p = \rho c_s^2$ for the ideal gas. To calculate the pressure in a region filled with a moving incompressible fluid, we must solve a boundary value problem based on the Poisson equation.

Block 4 of the complex method implementation algorithm (Fig. 1) includes solving the Poisson equation separately on each time layer to refine the pressure distribution in the liquid. We use the previously obtained velocity field on this time layer to do this. The transition to the next time layer at low values of the Mach number is performed using the lattice Boltzmann method, which allows us to simulate the change in the velocity field. Since the lattice Boltzmann method, when modeling the parameters of an incompressible fluid, can allow certain density fluctuations that affect the accuracy of determining the velocity field, we must consider these errors when creating the corresponding Poisson equation.

We modify the design scheme proposed in [19] to do this. This scheme consists of three steps. In the first step, the velocity field values are calculated using an explicit iterative scheme, which is based on a discrete representation of the convective and diffuse parts of the momentum equation. In the second step, the resulting velocity field is corrected by the pressure gradient determined after solving the Poisson equation for pressure. The third step is a correction step that ensures the convergence of the iterative determination of the velocity field in the first step.

The main difference of the approach proposed in this paper is that the determination of the velocity field in the next time layer occurs by using the modified lattice Boltzmann method, which is based on the strictly proven fact that at small values of the Mach number, we obtain a result that coincides with the solution in terms of the boundary value problem:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0, \\ \frac{\partial \rho u}{\partial t} + \nabla \cdot (u \otimes u) = \frac{\nabla p}{\rho} + \frac{1}{\rho}\nabla \cdot (\bar{\tau}) + \varphi. \end{cases} \qquad (7)$$

The viscous stress tensor for a compressible Newtonian fluid can be determined through the fluid velocity field as:

$$\bar{\tau} = 2\mu\left[\frac{\nabla u + (\nabla u)^T}{2} - \frac{1}{3}(\nabla u)\bar{I}\right]. \qquad (8)$$

For a small Mach number when moving to an incompressible boundary value problem, the value of the term can be simplified as follows [20]:

$$\frac{1}{\rho}\nabla \cdot (\bar{\tau}) = \nu \Delta u\ . \qquad (9)$$

DOI 10.15588/1607-3274-2023-4-6

To determine the pressure distribution on the new time layer, we define the Poisson equation for pressure [20]:

$$\nabla \cdot \left(\frac{\Delta t}{\rho} \nabla p \right) = \nabla u^*, \qquad (10)$$

where $\Delta x_1 = \Delta x_2 = \Delta x$, $\Delta t = 1$ are steps in spatial and temporal coordinates.

Modeling the solution of equation (10) takes place in two stages. The first step is to create a dataset for further use of the convolutional neural network.

We discretize the left and right parts of equation (10). For the left side of equation (10), we get:

$$\nabla\left(\frac{1}{\rho} \nabla p \right) = \frac{p_{i-1,j} - 2p_{i,j} + p_{i+1,j}}{\Delta x_1^2} + \frac{p_{i,j-1} - 2p_{i,j} + p_{i,j+1}}{\Delta x_2^2}.$$

Due to the fact that we neglect the compressibility of the investigated fluids, we can apply a linear discretization of the right-hand side of equation (10) using the central difference. Then:

$$\frac{\partial}{\partial t}\left(\frac{\partial \tilde{u}_{i,j}}{\partial x} \right) = \left(\frac{\tilde{u}_{i+1,j} - \tilde{u}_{i-1,j}}{2\Delta x_1} + \frac{\tilde{u}_{i,j+1} - \tilde{u}_{i,j-1}}{2\Delta x_2} \right) = S_{i,j}.$$

The resulting system of linear algebraic equations is solved using the iterative Jacobi method [13].

$$p_{i,j}^{n+1} = \frac{\left(p_{i-1,j}^n + p_{i+1,j}^n\right)\Delta x^2 + \left(p_{i,j-1}^n + p_{i,j+1}^n\right)\Delta x^2}{2\left(\Delta x^2 + \Delta x^2\right)} - \frac{\Delta x^2 \Delta x^2 S_{i,j}}{2\left(\Delta x^2 + \Delta x^2\right)} = \frac{1}{4}\left(p_{i-1,j}^n + p_{i+1,j}^n + p_{i,j+1}^n + p_{i,j-1}^n - \Delta x^2 S_{i,j}\right)$$

The obtained pressure distribution is considered and used as an object of the training data set. Training of a convolutional neural network takes place on the training data set formed in this way to solve this type of problem effectively. All problems of obtaining the pressure distribution on each time layer when solving applied problems will be solved in the future with the help of a convolutional network in a significantly shorter time.

After obtaining the values of the pressure field, we adjust the values of the velocity field:

$$u_{i,j} = \tilde{u}_{i,j} - \frac{1}{\rho_{i,j}}\left(\frac{p_{i+1,j} - p_{i-1,j} + p_{i,j+1} - p_{i,j-1}}{2\Delta x} \right).$$

Thus, we get the velocity field on the new time layer. We use these data as initial data when applying the Boltzmann lattice model. This process is repeated if the simulation process continues in time.

Let us briefly consider the development of the mentioned convolutional network and the results of its application.

Based on the previous work [21], a convolutional neural network was developed to model the solution of the boundary value problem based on the Poisson equation. Some changes were made to adapt the basic structure of the neural network to the actual problem. The general structure of the developed neural network is shown in Fig. 3.

First, the size of the input data was increased – from 96×96 to 128×128, which made it possible to simulate the movement of liquids with greater accuracy. The “bottom-up pathway” and “top-down pathway” parts of the network were also expanded by adding additional convolutional layer blocks. Also, the “output_conv” block was extended by increasing the convolutional layers from 2 to 4. Each layer now receives an additional input tensor “rho_input” of size 128×128 that contains the density value.

The output of the model is a two-dimensional array of size 128×128. The values of the previously generated solutions were normalized to a distribution with a mean value of 0 and a standard deviation of 1 to ensure the stability of the learning of the convolutional network. The process of obtaining the solution of the Poisson equation in physical units of measurement is as follows: the value of the output array of the neural network is inversely normalized, and thus the original distribution of the training data is restored.

To create a training dataset that will ensure the accuracy and efficiency of the neural network, 18 space geometries were prepared. They were used to simulate the movement of liquids using the LBM method. The numerical solution of equation (12) was used to calculate the values of the pressure field. Different random values of the initial fluid velocity, density, and relaxation time were used in the simulation of each geometry to ensure the variability of the dataset. In this case, the neural network features are the values of free members and density, and the target variable is the pressure value. In this way, a training dataset consisting of 75,000 objects was formed. A test dataset of 10,000 objects was created in the same way.

To implement the developed neural network, we used the following software: Python programming language, TensorFlow 2.4.1 machine learning framework. The Adam optimizer [22] was used to optimize the parameters of the neural network. The values of the optimizer parameters were as follows: learning_rate=0.0005, beta_1 = 0.95, beta_2 = 0.99, epsilon = 1e–7. Each of the parameters performs the following function: learning_rate is a parameter in the optimization algorithm that determines the step size of updating the coefficients of the neural network at each learning iteration; beta_1 is the forgetting

DOI 10.15588/1607-3274-2023-4-6

coefficient for the gradient; beta_2 is the forgetting factor for the second moment of the gradient; epsilon is a small constant introduced to ensure optimization stability. The average absolute error represents the loss function for this model. An MSI GeForce GTX 1660 Super Ventus OC 6GB GDDR6 GPU was used to train the neural network for 300 epochs. The value of the average absolute error on the test data set was 0.001021.

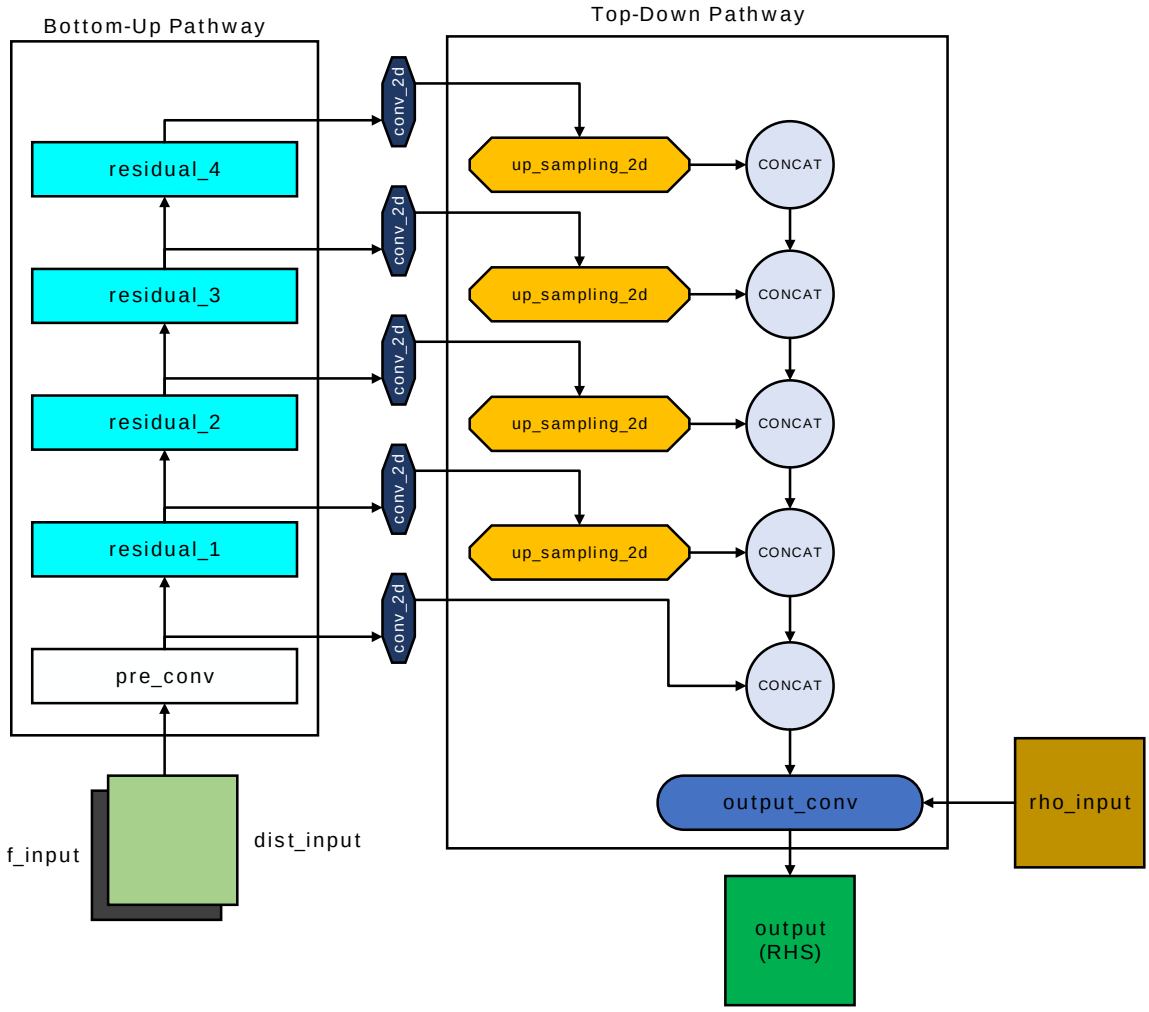


Figure 3 – The general structure of a convolutional neural network

4 EXPERIMENTS

To test the developed method, we used the model of the section of the human colon, which is shown in Fig. 4.

Black color indicates the working area, white color indicates the bounding surface.

The working area was discretized by a 128×128 grid. The parameters of the method are as follows: $\tau = 0.5012$, $\rho = 1000$. The boundary condition of fluid inflow from the right boundary was applied, equal to $v_0 = 0.01 \times \left|\sin\left(\frac{2\pi t}{T}\right)\right|$, where $T = 110$.

In the experiment, modeling was carried out using two methods: LBM and the LBM method with velocity field correction. The velocity field was measured at simulation iterations 300 and 900. The simulation results are shown in Fig. 5 and 6.

Figure 4 – Model of the section of the human colon

DOI 10.15588/1607-3274-2023-4-6

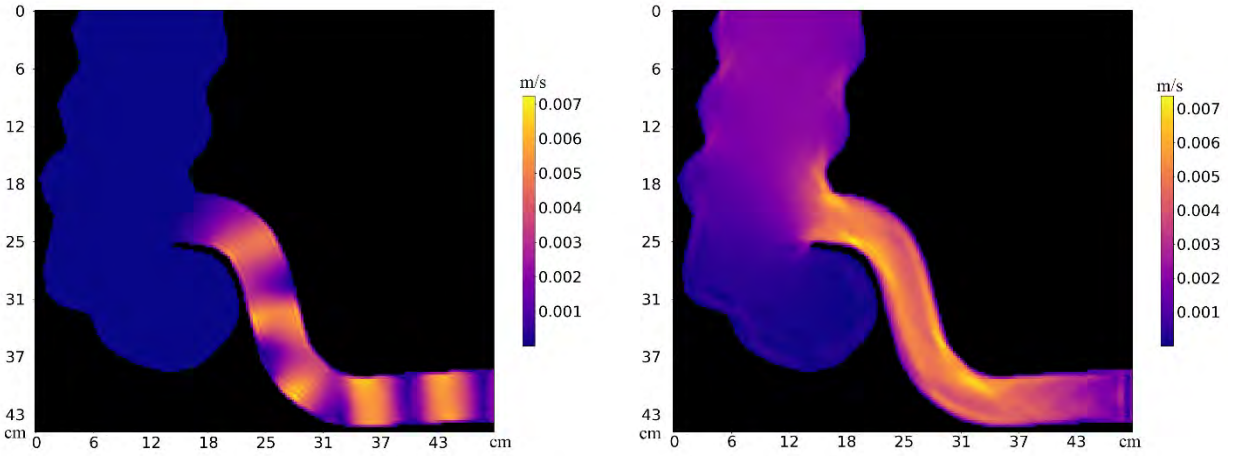


Figure 5 – Distribution of the velocity field at the 300th iteration. On the left is the LBM method, on the right is the proposed method with speed correction

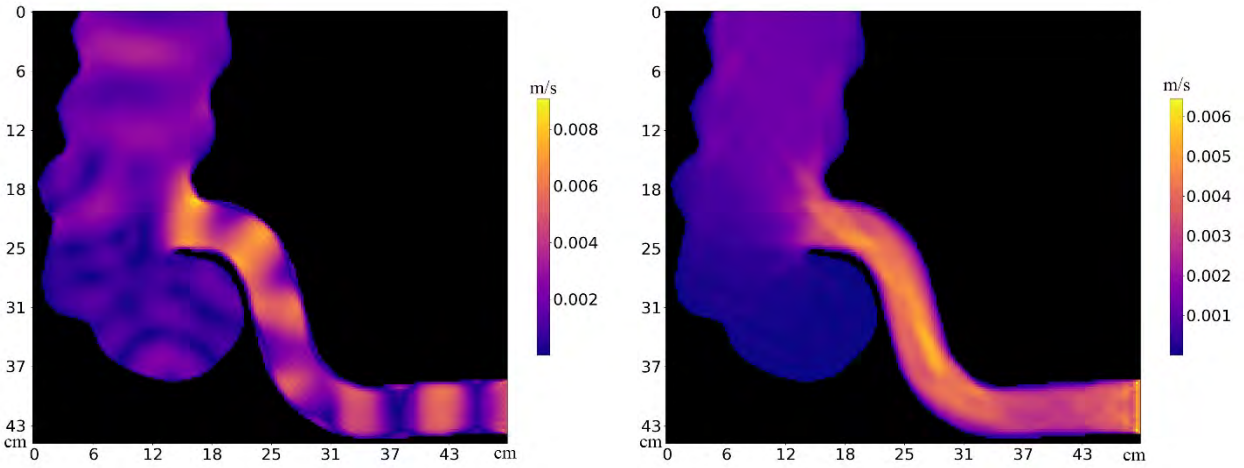


Figure 6 – Distribution of the velocity field at the 900th iteration. On the left is the LBM method, on the right is the proposed method with speed correction

Fig. 7 shows the results of measurements of the deviation of the average density values and from the initial density value in the computational domain after 1500 iterations.

When modeling with the developed method, the density values deviate less from the initial density value than the usual LBM method. Therefore, the obtained results indicate higher stability when modeling fluid movement using the combined method.

The computational speed of the developed method was measured in comparison with other methods. Three methods were used for comparison. The first method is the conventional LBM method without rate correction described in equation (14), the second method is the proposed rate-corrected LBM method using a convolutional neural network to solve equation (13), the third method is a velocity-corrected LBM method that uses the AMG numerical method [23] to solve equation (13). The comparative results of the experiments are shown in Table 1.

Table 1 – Comparison of modeling methods

Modeling method	**Time of one iteration, sec**
LBM	0.00301
LBM+ neural network	0.06087
LBM+ numerical method	0.54123

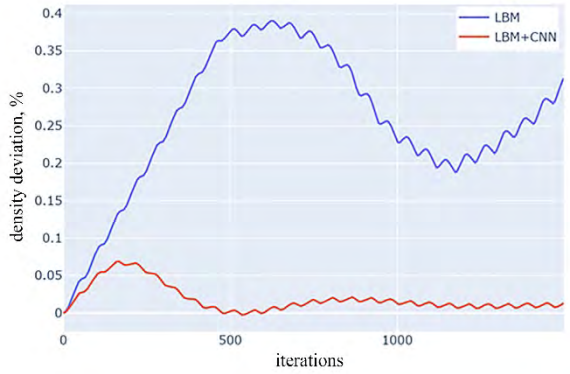


Figure 7 – The average value of the deviation of the density during the simulation

DOI 10.15588/1607-3274-2023-4-6

The experimental results show that the simulation speed of one iteration of the developed method is 6–7 times faster than the speed of the method that uses the numerical solution of the Poisson equation. Also, in comparison with the usual LBM method, the condition of fluid incompressibility is preserved.

5 RESULTS

The main result is that a new method of modeling two-dimensional hydrodynamic processes is proposed, which is characterized by increased accuracy and lower resource consumption compared to known methods. The sequence of actions for implementing the method is represented by the algorithm shown in Fig 1. The positive effect is achieved by combining the modified lattice Boltzmann method to determine the velocity field at the next time step. At each time step, we simulate the solution of the Poisson equation using a convolutional neural network to determine the pressure distribution. The obtained results make it possible to correct the deviation of the density to increase the simulation's accuracy. A parallel numerical method was used to reduce the time of preparing a dataset for neural network training. To practically confirm this approach's effectiveness, modeling the physical process of fluid movement in a fragment of the digestive tract was performed. All technical parameters of this experiment are given in this paper. Studies have shown that the simulation speed of one iteration of the developed method is 6–7 times faster than the speed of the method that uses the numerical solution of the Poisson equation. The value of the average absolute error on the set of test data is 0.001021.

6 DISCUSSION

The traditional approach to modeling hydrodynamic processes is based on solving a boundary value problem, which includes the Navier-Stokes equation and the flow continuity equation. This problem can be solved by one of the numerical methods, which causes significant difficulties in the case of the complex geometry of the surface area and requires significant resources since it is necessary to use excessive accuracy to ensure the convergence of the numerical method.

When modeling the movement of liquids in living nature, it is essential to get an estimate of the dynamics of the process in a short time with relatively low accuracy. Therefore, LBM is more often used for such tasks. This method allows us to get the desired solution in a significantly shorter time, but several disadvantages characterize it. One of the critical disadvantages is that this method is fundamentally oriented towards compressible liquids, and the dynamics of density change are challenging to control. Another significant drawback is the considerable laboriousness of the correct definition of the boundary conditions, given the complex geometry of the area.

Another approach to solving this problem includes using a predictor-corrector method. The predictor applies an explicit iterative scheme to determine the velocity field at the next time layer, and the corrector refines the value of the velocity field using the pressure distribution.

This work is devoted to developing this approach by improving both the predictor and corrector stages. To improve the accuracy of the simulation, a modified LBM is used in the predictor stage, and a pre-trained convolutional neural network is applied in the corrector stage to solve the Poisson equation. Combining these two approaches made it possible to improve the speed and accuracy of modeling in a region with complex geometry.

Further development of this approach consists of applying reinforcement learning mechanisms at the corrector stage, which will improve the accuracy of the obtained modeling results with significant changes in the geometry of the area without spending time retraining the convolutional network.

After certain refinements, the proposed method can be used to study hydrodynamic parameters in three-dimensional domains.

CONCLUSIONS

The urgent problem of developing a software system for mathematical modeling of the movement of liquids in areas with complex geometry, which is characteristic of living organisms, is being solved.

The scientific novelty of the obtained results lies in the fact that, for the first time, a method for modeling the movement of fluids in living organisms has been proposed, which reduces the modeling time and increases its accuracy compared to known approaches due to the synergistic effect obtained by improving the modeling parameters at each step of the iterative process, which includes a predictor stage and a corrector stage. The predictor stage implements a modified LBM in which a modified equilibrium distribution function is applied, which increases the accuracy of determining the distribution function in one iteration step by the BGK model. The LBM implementation time is reduced by parallelizing the calculation of discrete values of the distribution function during the collision of elementary liquid volumes at the mesoscopic level. The corrector stage realizes a reduction in modeling time due to the use of a previously trained convolutional network, the structure of which is adapted to the solution of a specific problem.

The practical significance of the obtained results is that a software system has been developed for simulating the movement of liquids in areas of complex shape, significantly reducing the simulation time and using computing resources, provided that the obtained results are of acceptable accuracy. Experiments were conducted using different workloads of the developed software simulation system. The obtained results made it possible to recommend using this software system when studying the movement of fluids in living nature, particularly when studying the digestive and cardiovascular systems of living organisms.

Prospects for further research consist in expanding the possibilities of this method for its application to three-dimensional areas of complex shape. Another direction

DOI 10.15588/1607-3274-2023-4-6

for improving this method is using reinforcement learning mechanisms to reduce the time it takes to reconfigure a convolutional neural network for an expanded range of tasks.

ACKNOWLEDGEMENTS

The work was supported by a NATO grant (program: “Science for peace and security”). State registration number 0123U101109 dated 02/14/2023.

REFERENCES

1. Chen S., Doolen G. D. Lattice Boltzmann method for fluid flows, *Annual Review of Fluid Mechanics*, 1998, Vol. 30, pp. 329–364. https://doi.org/10.1146/annurev.fluid.30.1.329
2. Guo Z., Shu Ch. Lattice Boltzmann Method and Its Applications in Engineering. Singapore, World Scientific, 2013, 420 p. https://doi.org/10.1142/8806
3. Tessarotto M., Fonda E., Tessarotto M. The computational complexity of traditional Lattice Boltzmann methods for incompressible fluids, *Rarefied Gas Dynamics: The 26th International Symposium, Kyoto (Japan), 20–25 July 2008 : proceedings*. New York, AIP Publishing, AIP Conference Proceedings, 2008, Vol. 1084, № 1, pp. 470–475, https://doi.org/10.1063/1.3076524
4. Zimny S., Masilamani K., Jain K., Roller S. P. Lattice Biltzmann Simulations on Complex Geometries [Electronic resource]. Access mode: https://www.researchgate.net/publication/289041300_Latice_Boltzmann_Simulations_on_Complex_Geometries
5. Korner C., Pohl T., Rude U., Thürey N., Zeiser T. Parallel Lattice Boltzmann Method for CFD Applications, *Lecture Notes in Computational Science and Engineering*. Berlin, Springer, 2006, Vol. 51, pp.439–466, https://doi.org/10.1007/3-540-31619-1_13
6. Xunliang L., Wen-Quan T., He Y. L. A simple method for improving the SIMPLER algorithm for numerical simulations of incompressible fluid flow and heat transfer problems, *Engineering Computations*. Berlin, Springer, 2005, Vol.22, pp. 921–939. https://doi.org/10.1108/02644400510626488
7. Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators, *Neural Networks*, 1989, Vol. 2, №5, pp. 359–366, https://doi.org/10.1016/0893-6080(89)90020-8
8. Lagaris I. E., Likas A. C., Papageorgiou D. G. Neural-network methods for boundary value problems with irregular boundaries, *IEEE Transfctions on Neural Networks*, 2000, Vol. 11, №5, pp. 1041–1049, https://doi.org/10.1109/72.870037
9. Aarts L. P., Veer P. V. Neural network method for solving partial differential equations, *Neural Processing Letters*. Berlin, Springer, 2001, Vol. 14, pp. 261–271.
10. Smaoui N., Al-Enezi S. Modeling the dynamics of nonlinear partial differential equations using neural networks, *Journal of Computational and Appliaed Mathematics*, 2004, Vol. 170, pp. 27–58, https://doi.org/10.1016/j.cam.2003.12.045
11. Deheuvels P., Martynov G. V. A Karhunen-Loeve decomposition of a Gaussian process generated by independent pairs of exponential random variables, *Journal of Functional Analysis*, 2008, Vol. 255, pp. 2363–2394
12. Chua L.O., Yang L. Cellular neural networks: theory, *IEEE Transactions on Circuits Systems*, 1988, Vol. 35, № 10, pp. 1257–1272. https://doi.org/10.1109/31.7600
13. Novotars'kij M. A., Nesterenko B. B. Shtuchni nejronni merezhi: obchis-lennya. Kyiv, In-t matematiki, 2004, 408 p.
14. He X., Luo L.-S. A Priori Derivation of the Lattice Boltzmann Equation, *Physical Review E*, 1997, Vol. 55, № 6, pp. R6333–R6336.
15. Eichinger M., Heinlein A., Klawonn A. Stationary flow predictions using convolutional neural networks, *Numerical Mathematics and Advanced Applications*. Berlin, Springer, 2021. – P.541–549
16. Koelman J.M.V.A. A simple lattice Boltzmann scheme for Navier-Stokes fluid flow, *Europhysics Letters*, 1991, Vol. 15, № 6, pp. 603–60. https://doi.org/10.1209/0295-5075/15/6/007
17. Qian Y. H., d’Humieres D., Lallemand P.] Lattice BGK models for Navier-Stokes equation, *Europhysics Letters*, 1992, Vol. 17, № 6, pp. 479–484, https://doi.org/10.1209/0295-5075/17/6/001
18. Sterling D. J., Chen S. Stability Analysis of Lattice Boltzmann Methods, *Journal of Computational Physics*, 1996, Vol. 123, № 1, pp. 196–206, https://doi.org/10.1006/jcph.1996.0016
19. Chorin A. J. Numerical solution of the Navier-Stokes equations, *Mathematics of computation*, 1968, Vol. 22, № 104, pp. 745–762. https://doi.org/10.2307/2004575
20. Vittoz G., Oger M., de Leffe M., Le Touze D.] Comparisons of weakly-compressible and truly incompressible approaches for viscous flow into a high-order Cartesian-grid finite volume framework, *Journal of Computational Physics*, 2019, Vol. 1, pp. 100015–100048. https://doi.org/10.1016/j.jcpx.2019.100015
21. Kuzmych V., Novotarskyi M. Accelerating simulation of the PDE solution by the structure of the convolutional neural network modifying : Lecture Notes on Data Engineering and Communications, *Artificial Intelligence and Logistics Engineering (ICAILE2022): The 2nd International Conference, 20–22 February 2022 : proceedings*. Berlin, Springer, 2022, pp. 3–15. https://doi.org/10.1007/978-3-031-04809-8_1
22. Duchi J., Hazan E., Singer Y. Adaptive subgradient methods for online learning and stochastic optimization, *Journal of Machine Learning Research*, 2011, Vol. 12, pp. 2121–2159
23. Brandt A., McCormick S., Ruge J. Algebraic multigrid (AMG) for sparse matrix equations [Electronic resource]. Access mode: https://www.researchgate.net/publication/243769570_Algebraic _multigrid_AMG_for_sparse_matrix_ equations

Received 16.10.2023.
Accepted 27.11.2023.

DOI 10.15588/1607-3274-2023-4-6

УДК 004.942:001.57

МЕТОД ГІДРОДИНАМІЧНОГО МОДЕЛЮВАННЯ З ВИКОРИСТАННЯМ ЗВЕРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ

Новотарський М. А. – д-р техн. наук, професор кафедри обчислювальної техніки Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Кузьмич В. А. – аспірант кафедри обчислювальної техніки Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

АНОТАЦІЯ

Актуальність. Розв'язування гідродинамічних задач пов'язане з високою обчислювальною складністю і тому вимагає значних обчислювальних ресурсів і часу. Запропонований підхід дозволяє суттєво скоротити час розв'язування таких задач шляхом застосування комбінації двох вдосконалених методів моделювання.

Мета. Метою є створення комплексного методу гідродинамічного моделювання, який вимагає значно менше часу для визначення динаміки поля швидкостей за рахунок використання модифікованого решітчастого методу Больцмана і розподілу тиску за рахунок використання згорткової нейронної мережі.

Метод. Запропоновано метод гідродинамічного моделювання, який реалізує синергетичний ефект, що виникає при поєднанні вдосконаленого решітчастого методу Больцмана та згорткової нейронної мережі з спеціально адаптованою структурою. Суть методу полягає у реалізації послідовності ітерацій, на кожній з яких відбувається моделювання процесу зміни параметрів при переході на наступний часовий шар. Кожна ітерація включає крок предиктора та крок коректора. На кроці предиктора працює решітчастий метод Больцмана, який дозволяє отримати поле швидкостей рідини в робочій зоні на наступному часовому шарі за допомогою поля швидкостей на попередньому шарі. На кроці коректора ми застосовуємо вдосконалену згорткову нейронну мережу, навчену на раніше створеному наборі даних. Використання нейронної мережі дозволяє визначити розподіл тиску на новому часовому шарі із заданою точністю. Після додавання поправки на стисливість рідини на новому часовому шарі ми отримуємо уточнені значення поля швидкостей, які можна використовувати як початкові дані для застосування решітчастого методу Больцмана на наступній ітерації. Обчислення припиняються при досягненні заданої кількості ітерацій.

Результати. Роботу запропонованого методу досліджено на прикладі моделювання руху рідини у фрагменті шлунково-кишкового тракту людини. Результати моделювання показали, що час, витрачений на реалізацію процесу моделювання, скоротився у 6–7 разів при збереженні прийнятної для практичних завдань точності.

Висновки. Запропонований метод гідродинамічного моделювання зі згортковою нейронною мережею та решітчастим методом Больцмана суттєво скорочує час та обчислювальні ресурси, необхідні для реалізації процесу моделювання в областях зі складною геометрією. Подальший розвиток цього методу дозволить реалізувати гідродинамічне моделювання в реальному часі в тривимірних областях.

КЛЮЧОВІ СЛОВА: гідродинамічне моделювання, згорткова нейронна мережа, решітчастий метод Больцмана.

ЛІТЕРАТУРА

1. Chen S. Lattice Boltzmann method for fluid flows/ S. Chen, G. D. Doolen // Annual Review of Fluid Mechanics. – 1998. – Vol. 30. – P. 329–364, https://doi.org/10.1146/annurev.fluid.30.1.329
2. Guo Z. Lattice Boltzmann Method and Its Applications in Engineering / Z. Guo, Ch. Shu. – Singapore : World Scientific, 2013. – 420 p. https://doi.org/10.1142/8806
3. Tessarotto M. The computational complexity of traditional Lattice Boltzmann methods for incompressible fluids / M. Tessarotto, E. Fonda, M. Tessarotto // Rarefied Gas Dynamics: The 26th International Symposium, Kyoto (Japan), 20–25 July 2008 : proceedings. – New York: AIP Publishing, AIP Conference Proceedings, 2008. – Vol. 1084, № 1. – P.470–475, https://doi.org/10.1063/1.3076524
4. Lattice Biltzmann Simulations on Complex Geometries [Electronic resource] / [S. Zimny, K. Masilamani, K. Jain, S. P. Roller] . – Access mode: https://www.researchgate.net/publication/289041300_Latice_Boltzmann_Simulations_on_Complex_Geometries
5. Parallel Lattice Boltzmann Method for CFD Applications / [C. Korner, T. Pohl, U. Rude et al.] // Lecture Notes in Computational Science and Engineering. – Berlin : Springer, 2006. – Vol. 51. – P.439–466, https://doi.org/10.1007/3-540-31619-1_13
6. Xunliang L. A simple method for improving the SIMPLER algorithm for numerical simulations of incompressible fluid flow and heat transfer problems / L. Xunliang, T. Wen-Quan, Y. L. He // Engineering Computations. – Berlin : Springer. – 2005 . – Vol. 22. – P. 921–939. https://doi.org/10.1108/02644400510626488
7. Hornik K. Multilayer feedforward networks are universal approximators / K. Hornik, M. Stinchcombe, H. White // Neural Networks. – 1989. – Vol. 2, № 5. – P. 359–366, https://doi.org/10.1016/0893-6080(89)90020-8
8. Lagaris I. E. Neural-network methods for boundary value problems with irregular boundaries / I. E. Lagaris, A. C. Likas, D. G. Papageorgiou // IEEE Transfctions on Neural Networks. – 2000. – Vol. 11, № 5. – P. 1041–1049, https://doi.org/10.1109/72.870037
9. Aarts L. P. Neural network method for solving partial differential equations / L. P. Aarts, P. V. Veer // Neural Processing Letters. – Berlin : Springer, 2001. – Vol. 14. – P. 261–271.
10. Smaoui N. Modeling the dynamics of nonlinear partial differential equations using neural networks / N. Smaoui, S. Al-Enezi // Journal of Computational and Appliaed Mathematics. – 2004. – Vol. 170. – P. 27–58, https://doi.org/10.1016/j.cam.2003.12.045
11. Deheuvels P. A Karhunen-Loeve decomposition of a Gaussian process generated by independent pairs of exponential random variables / P. Deheuvels, G. V. Martynov // Journal of Functional Analysis. – 2008. – Vol. 255. – P. 2363–2394.
12. Chua L. O. Cellular neural networks: theory / L. O. Chua, L. Yang // IEEE Transactions on Circuits Systems. – 1988. –

DOI 10.15588/1607-3274-2023-4-6

Vol. 35, № 10. – P. 1257–1272. https://doi.org/10.1109/31.7600
13. Новотарський М. А. Штучні нейронні мережі: обчислення / М. А. Новотарський, Б. Б. Нестеренко. – Київ: Ін-т математики, 2004. – 408 с.
14. He X. A Priori Derivation of the Lattice Boltzmann Equation / X. He, L.-S. Luo // Physical Review E. – 1997. – Vol. 55, № 6. – P. R6333–R6336.
15. Stationary flow predictions using convolutional neural networks / [M. Eichinger, A. Heinlein, A. Klawonn] // Numerical Mathematics and Advanced Applications. – Berlin: Springer. – 2021. – P.541–549
16. Koelman J.M.V.A. A simple lattice Boltzmann scheme for Navier-Stokes fluid flow / J.M.V.A. Koelman // Europhysics Letters. – 1991. – Vol. 15, № 6. – P. 603–607. https://doi.org/10.1209/0295-5075/15/6/007
17. Qian Y. H. Lattice BGK models for Navier-Stokes equation / [Y. H. Qian, D. d'Humieres, P. Lallemand] // Europhysics Letters. – 1992. – Vol. 17, №6. – P. 479–484, https://doi.org/10.1209/0295-5075/17/6/001
18. Sterling D. J. Stability Analysis of Lattice Boltzmann Methods / D. J. Sterling, S. Chen // Journal of Computational Physics. – 1996. – Vol. 123, № 1. – P. 196–206, https://doi.org/10.1006/jcph.1996.0016
19. Chorin A. J. Numerical solution of the Navier-Stokes equations / A. J. Chorin // Mathematics of computation. – 1968. – Vol. 22, № 104. – P.745–762. https://doi.org/10.2307/2004575
20. Comparisons of weakly-compressible and truly incompressible approaches for viscous flow into a high-order Cartesian-grid finite volume framework / [G. Vittoz, M. Oger, M. de Leffe, D. Le Touze] // Journal of Computational Physics. – 2019. – Vol. 1. – P. 100015–100048. https://doi.org/10.1016/j.jcpx.2019.100015
21. Kuzmych V. Accelerating simulation of the PDE solution by the structure of the convolutional neural network modifying : Lecture Notes on Data Engineering and Communications / V. Kuzmych, M. Novotarskyi // Artificial Intelligence and Logistics Engineering (ICAILE2022): The 2nd International Conference, 20–22 February 2022 : proceedings. – Berlin: Springer, 2022. – P. 3–15, https://doi.org/10.1007/978-3-031-04809-8_1
22. Duchi J. Adaptive subgradient methods for online learning and stochastic optimization / J. Duchi, E. Hazan, Y. Singer // Journal of Machine Learning Research, 2011. – Vol. 12. – P. 2121–2159.
23. Brandt A. Algebraic multigrid (AMG) for sparse matrix equations [Electronic resource] / A. Brandt, S. McCormick, J. Ruge. – Access mode: https://www.researchgate.net/publication/243769570_Algebraic_multigrid_AMG_for_sparse_matrix_ equations

DOI 10.15588/1607-3274-2023-4-6

p-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2023. № 4
e-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2023. № 4

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

UDC 004.852

NEURAL ORDINARY DIFFERENTIAL EQUATIONS FOR TIME SERIES RECONSTRUCTION

Androsov D. V. – Post-graduate student of the Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

ABSTRACT

Context. Neural Ordinary Differential Equations is a deep neural networks family that leverage numerical methods approaches for solving the problem of time series reconstruction, given small amount of unevenly distributed samples.

Objective. The goal of the following research is the synthesis of a deep neural network that is able to solve input signal reconstruction and time series extrapolation task.

Method. The proposed method exhibits the benefits of solving time series extrapolation task over forecasting one. A model that implements encoder-decoder architecture with differential equation solving in latent space, is proposed. The latter approach was proven to demonstrate outstanding performance in solving time series reconstruction task given a small percentage of noisy and uneven distributed input signals. The proposed Latent Ordinary Differential Equations Variational Autoencoder (LODE-VAE) model was benchmarked on synthetic non-stationary data with added white noise and randomly sampled with random intervals between each signal.

Results. The proposed method was implemented via deep neural network to solve time series extrapolation task.

Conclusions. The conducted experiments have confirmed that proposed model solves the given task effectively and is recommended to apply it to solving real-world problems that require reconstructing dynamics of non-stationary processes. The prospects for further research may include the process of computational optimization of proposed models, as well as conducting additional experiments involving different baselines, e. g. Generative Adversarial Networks or attention Networks.

KEYWORDS: neural ordinary differential equations, deep neural networks, variational autoencoders, recurrent neural networks, long term short memory networks

ABBREVIATIONS

NN is a neural network;

ODE is an ordinary differential equation;

LSTM is a long short-term memory network;

GRU is a gated recurrent unit network

ARMA is an autoregressive moving average model;

ARIMA is an autoregressive integrated moving average model;

GARCH is a generalized autoregressive conditional heteroskedasticity;

ELBO is an evidence lower bound function.

NOMENCLATURE

y is an unknown non-stationary non-negative continuous-time series;

t is a continuous time point, $t \in \mathrm{t}$;

$\lambda = \lambda(\tau)$ is an intensity parameter of time distribution, $\tau > 0$;

θ is a model parameter vector;

$\hat{y}$ is a observed sample from time series;

Y is a multivariate discrete-time or continuous-time process;

x_t is an input vector to neural network at time t, $x_t \in \mathrm{x}$;

E is a time series reconstruction loss;

N is a Normal distribution;

σ is a standard deviation of y;

α is a time series sampling percentage;

$y^{(rec)} = \hat{y} \bigcup f(x)$ is a reconstructed time series;

h_t is a hidden state of neural network at time t, $h_t \in \mathrm{h}$;

z_0 is a latent space initial parameter for decoder network;

ε is a moving average parameter for time series y;

b_h is a bias vector for hidden state of a recurrent neural network;

b_u is a bias vector for output state of a recurrent neural network;

DOI 10.15588/1607-3274-2023-4-7

b_r is a bias vector for reset state of a recurrent neural network;

W_h is a weight matrix for hidden state of a recurrent neural network;

$W_{.u}$ is a weight matrix for output state of a recurrent neural network;

$W_{.r}$ is a weight matrix for reset state of a recurrent neural network;

$h_t^{'}$ is a temporary hidden state at time t;

$h_t^{''}$ is a candidate hidden state at time t;

INTRODUCTION

Time series analysis nowadays is one of the most rapidly developing field of computational statistics. In recent years it gained a significant push towards applying machine learning methods to solve the problem of predicting real-world processes in various fields, e.g., physics or finances. The most popular approaches that are applied to overcome the given problems include autoregressive modeling via ARIMA and GARCH models [1–5]. These models completely rely on assumptions of autoregressive nature of given process, i.e., linear dependence between current state of process with previous ones, and stationarity, i.e., absence of mean/variance fluctuations through time of observation of given process. In order to process non-stationary data, ARIMA models leverage mechanisms of taking finite difference of given time series data to vanish trend curves and thus transform such data into stationary time series [6]. On the other hand, GARCH models perform conditional heteroskedasticity modeling via moving average estimation of variance [7]. The drawbacks of these approaches are implicated in assumption of linear dependence between lags of a given process.

To overcome these challenges recurrent neural networks (RNN) [8] were introduced. RNN rely on the same concept of "dependency" between time series states as in autoregressive models, except that it applies nonlinear transformations of input states and hidden states. These chains of operations allow RNN to model nonstationary series without performing reduction to stationarity. LSTM models are the most widely used RNN application, since they achieve ability to capture patterns in time-dependent data at large scale of observation [8, 9]. Since 2014, new family of recurrent neural networks was introduced – gated recurrent unit network (GRU) [8, 9]. GRU is, in essence, a lightweighted version of LSTM, offering reduced complexity whilst learning, both time and space. However, the weakest point of given approaches is that they are invariant to occurrence gaps, i.e., they built regarding the assumption that intervals between each sample are equal. However, for various cases that assumption is not a valid one, e.g., for tracking real-time financial data.

The object of study is the process of time series reconstruction from samples with uneven distribution regarding time. This data is difficult to predict since both autoregressive models, and recurrent neural networks are invariant to intra-sample gaps, Therefore, it is proposed to construct a new model, called latent ordinary differential equations variational autoencoders to improve the quality of predictions of given data and to solve process recreation task.

The subject of study is methods for time series prediction and recreation.

The purpose of the work is to create a machine learning model to solve time series reconstruction from small and unevenly distributed data samples.

1 PROBLEM STATEMENT

For a given sample $\hat{y}$ of time series y it is desired to create a reconstruction $y^{(rec)}$, such that $\forall \varepsilon > 0, d(y^{(rec)}, y) \le \varepsilon$, where $d(\cdot,\cdot)$ is some distance metric.

2 REVIEW OF THE LITERATURE

Autoregressive models, such as ARMA [9], presented in 1951, are the most used ones for the time series modeling task. They consider the time series to be in the form $y_t = \theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i} + \sum_{j=1}^{q} \theta_{p+j} \varepsilon_j$ [1]. ARMA models are suitable for learning behavior of stationary processes and hence, are widely and successfully used applied to this task [1–5, 9, 10].

In order to cope with non-stationary processes, ARIMA models were introduced [1, 11]. They perform numerical differentiation techniques, i.e. applying finite differences operator $\nabla_d(y_t) = \nabla_{d-1}(y_t \cdot - y_{t-1}), d \in \mathrm{N}$ to a given time series to vanish it's non-stationarity [1].

The other autoregressive approach is to model stationary processes in the form $y_t = \varepsilon \sqrt{\theta_0 + \sum_{i=1}^{p} \theta_i y_{t-i}^2 + \sum_{j=1}^{q} \theta_{p+j} \sigma_{t-j}^2}$ [1].

Since the breakthrough in the development of computational capacities, it became feasible to examine machine learning algorithms, and, in particular, artificial neural networks, on time series modeling and predictions tasks. One of the most successful approaches is to apply recurrent neural networks to discrete-time time series.

Consider sample of time series y_t of the form $D = \langle \mathrm{x}_t, y_t \rangle$. RNN in this case is a mapping function $f : \mathrm{x}_t \to y_t$, and that function is essentially a chain of non-linear transformations over affine transformations that are provided by state-space modeling of y_t [8]. Classic RNN models these chains in a following way:

$$h_t = \sigma(W_{xh} x_t + W_{hh} h_t + b_h),$$

DOI 10.15588/1607-3274-2023-4-7

$$y_t = u_t = \sigma(W_{hu} h_t + b_u).$$

RNN is trained to maximize logarithmic likelihood $\log p(y_t \mid \mathrm{x}, U, W, V, c)$ [8]. However, despite the ability to model non-stationary time series, RNN has a severe drawback – gradient vanishing – that is caused by it's architecture [8].

To overcome this challenge, LSTM model was introduced in 1997 [12]. LSTM offers more complex architecture yet greater precision of forecasts, introducing 3 "gates" – input, output and forget. The latter one implements the process of capturing short-term dependencies at a long scale, thus giving it's name to the method [8, 12].

In 2014, GRU model was firstly described [8, 13]. GRU model aims to achieve the same quality of forecasts but at a lower computational cost, reducing number of parameters to train. Despite LSTM-based models have become de facto standard in recent years, GRU models are also widely used [8, 13].

In recent years, new family of neural networks was introduced, called neural ordinary differential equations [14]. Neural ODE model interprets time series as a continuous process with unknown dynamics and thus solving a differential equation with respect to the hidden state and time: $\frac{d\mathrm{h}}{dt} = f(\mathrm{h}(t), t, \theta)$. Neural ODEs are proven to be effective for survival analysis [15] and weather data prediction [16].

3 MATERIALS AND METHODS

It is proposed to recreate time series structure from the latent space (i.e., some mapping of feature space) process dynamics. Let's add a mapping $f : \mathrm{P}^m \to \mathrm{P}^n$, such that:

$$\mathrm{H} = f(\mathrm{Y}), \tag{1}$$

where H represents hidden dynamics in latent space and thus is proposed to be modeled via Neural ODE:

$$\frac{d\mathrm{h}}{dt} = g(\mathrm{h}(t), t, \theta_{\mathrm{h}}). \tag{2}$$

Integrating (2) with respect to time allows forecasting multivariate continuous-time process (1) in latent space. Then, to retrieve forecast of Y it is necessary to add an inverse mapping $f^{-1} : \mathrm{P}^n \to \mathrm{P}^m$. However, the inverse mapping could not exist under certain conditions, e.g., f is not bijective. In that case, to overcome this restriction, it is proposed to use encoder-decoder approach, i.e., defining another mapping $g : \mathrm{P}^n \to \mathrm{P}^m$, such that:

$$\hat{\mathrm{Y}} = g(\mathrm{H}). \tag{3}$$

By substituting (1) and (4) as a parameters to the given reconstruction loss, an optimization problem is defined:

$$E\left(\mathrm{Y}, \hat{\mathrm{Y}}\right) \to \min. \tag{4}$$

There are multiple approaches to define exact task in the form of problem (4). Let's consider that process Y is drawn from unknown random vector of distribution $\rho(\mathrm{x})$. Then chain of transformations defined by (1) and (4) produce the process $\hat{\mathrm{Y}}$, drawn from random vector of distribution $\hat{\rho}(\mathrm{h})$. Then to measure difference between these two distributions it is proposed to use Kullback-Leibler divergence, defined as:

$$KL(\rho \mid \hat{\rho}) = \int_{-\infty}^{+\infty} \rho(\mathrm{x}) \log \frac{\rho(\mathrm{x})}{\hat{\rho}(\mathrm{x})} d\mathrm{x}. \tag{5}$$

However, since $\rho(\mathrm{x})$ is an unknown distribution, mentioning that distribution proposed approach should define both mappings (1) and (3), it is proposed to define a joint probability distribution $p(\mathrm{x}, \mathrm{h})$, and applying defining it as:

$$p(\mathrm{x}, \mathrm{h}) = p(\mathrm{h} \mid \mathrm{x}) p(\mathrm{x}) = p(\mathrm{h}) p(\mathrm{x} \mid \mathrm{h}). \tag{6}$$

Taking a logarithm of (7):

$$\begin{aligned} &\log p(\mathrm{x}, \mathrm{h}) = \log p(\mathrm{h} \mid \mathrm{x}) + \log p(\mathrm{x}) = \\ &= \log p(\mathrm{h}) + \log p(\mathrm{x} \mid \mathrm{h}). \end{aligned} \tag{7}$$

To achieve deterministic measure of "fitness" of distributions $p(\mathrm{h}) = p(\mathrm{h} \mid \theta)$ and $p(\mathrm{x}) = p(\mathrm{x} \mid \theta)$, where θ is a parameters vector of the desired model, let's apply a mathematical expectation operator with respect to latent state h to (8):

$$\begin{aligned} &\int_{\mathrm{h}} q(\mathrm{h}) \log p(\mathrm{x} \mid \theta) d\mathrm{h} = \\ &= \int_{\mathrm{h}} q(\mathrm{h}) \log p(\mathrm{x}, \mathrm{h} \mid \theta) d\mathrm{h} - \\ &- \int_{\mathrm{h}} q(\mathrm{h}) \log p(\mathrm{h} \mid \mathrm{x}, \theta) d\mathrm{h} \end{aligned} \tag{8}$$

Left-hand side of (8) is simply equals to $p(\mathrm{x} \mid \theta)$. Then let's add and subtract mathematical expectation of logarithmic probability of hidden state:

DOI 10.15588/1607-3274-2023-4-7

$$p(\mathrm{x}\mid\theta)= \\ =\int_{\mathrm{h}} q(\mathrm{h})(\log p(\mathrm{x},\mathrm{h}\mid\theta)-\log p(\mathrm{h}))d\mathrm{h}- \\ -\int_{\mathrm{h}} q(\mathrm{h})(\log p(\mathrm{h}\mid\mathrm{x},\theta)-\log p(\mathrm{h}))d\mathrm{h}. \quad (9)$$

Last component of right-hand side of the equation (9) is Kullback-Leibler divergence of $q(\mathrm{h})$ and $p(\mathrm{h}\mid\mathrm{x},\theta)$ distributions, and the first one is ELBO. ELBO is proposed to use as a reconstruction loss of a proposed model.

Let's define the mapping functions in the scope of problem (1)-(9). It is proposed to use a recurrent neural network, in particular, GRU as an encoder mapping, i.e. mapping (1) is defined as:

$$h_t^{''}=\hat{f}(y_\theta,h_{t-1},t), \quad (10)$$

$$u_t=\sigma(W_{xu}x_t+W_{hu}h_t^{''}+b_u), \quad (11)$$

$$r_t=\sigma(W_{xr}x_t+W_{hr}h_t^{''}+b_r), \quad (12)$$

$$h_t^{'}=\sigma(W_{xh'}x_t+W_{hh'}h_t^{''}+b_{h'}), \quad (13)$$

$$h_t=u_th_t^{''}+(1-u_t)h_t^{'}, \quad (14)$$

Hence it is proposed to interpret latent space features as the dynamics of given process, defined by (1), (10) is defined as:

$$h_t^{''}=Sol(f_\theta,h_{t-1},t), \quad (15)$$

where Sol is a numeric ODE solver, e.g., Runge-Kutta method.

Let's add a layer that produces parameter for latent parameter z_0 and define mapping (4) as:

$$\forall t\in\mathrm{t},\ z_t=Sol(z_0,\theta_f,t), \quad (16)$$

$$x\sim p(\mathrm{x}\mid\mathrm{z}_t,\theta). \quad (17)$$

For achieving time sensitivity, it is feasible to model probability distributions of time spots using non-stationary Poisson processes. By adding and modeling intensity parameter λ, (16)–(17) can be augmented in the following way:

$$\forall t\in\mathrm{t},\ t\sim\mathrm{PoissProcess}(\lambda(\tau)). \quad (18)$$

Then (10) could be augmented by adding:

$$\log(\mathrm{t}\mid t_{\min},t_{\max},\lambda))=\sum_{t\in\mathrm{t}}\log\lambda(t)-\int_{t_{\min}}^{t_{\max}}\lambda(t)dt. \quad (19)$$

By applying task (10)–(17) to input samples, drawn from Y and minimizing loss, defined in (9) and (19), task of time series reconstruction was achieved.

4 EXPERIMENTS

To solve previously defined task and measure the effectiveness of proposed approach the computer program that implements time series reconstruction was developed.

For time series reconstruction experiment a following synthetic dataset was chosen:

$$y=\sin(25\pi t)+\varepsilon,\ \varepsilon\sim\mathrm{N}(0,\sigma). \quad (20)$$

Multiple samples were drawn from (20) with the following setups:

1. $\sigma=[0.1,0.5,1]$;
2. $\alpha=[0.15,0.35,0.55]$.

For model there were chosen 2 options – LODE-VAE without modeling distribution of time points and LODE-VAE with modeling distribution of time points using (18). For reference, first model is called LODE-VAE-N and LODE-VAE-P.

For both models the next parameters were chosen:

1. Dimension size of latent state dynamics process is 6.
2. Dimension size of integrated state vector is 6.
3. Dimension size of decoded vector is 1.
4. Number of epochs is 200.
5. Learning rate is adaptive with exponential decay with start rate at 0.01.
6. ODEs are solved using Dormand-Prince method.

Metrics for benchmarking are mean squared error (MSE) and coefficient of determination R^2.

Results of LODE-VAE-N model benchmarking by MSE metric are shown in Table 1.

5 RESULTS

In the following Tables 1 – 4 results of benchmarking of LODE-VAE-N and LODE-VAE-P models by MSE and R^2 metrics are provided. Since MSE and R^2 metrics are both used for validating model adequacy for forecasting time series, their optimization objectives are opposite – MSE needs to be minimized and R^2 needs to be maximized.

6 DISCUSSION

As follows from Tables 1–4, MSE and R^2 metric values of benchmarking of LODE-VAE-N and LODE-VAE-P differs slightly. Despite the difference, both models are well suitable for time series reconstruction and forecasting from obtained unevenly distributed samples.

Tables 1–4 show the same tendencies for both metrics and both models – the more data is available the better the quality of predictions.

DOI 10.15588/1607-3274-2023-4-7

Table 1 – Results of LODE-VAE-N model benchmarking by MSE metric

α\σ	0.1	0.5	1
0.15	0.4309	0.7043	1.4491
0.35	0.4081	0.5921	1.4104
0.55	0.3802	0.6757	1.2195

As expected, model performed reconstruction task well, and results the better the noise level is lower

Results of LODE-VAE-P model benchmarking by MSE metric are shown in Table 2.

Table 2 – Results of LODE-VAE-P model benchmarking by MSE metric

α\σ	0.1	0.5	1
0.15	0.4011	0.583	1.4583
0.35	0.3301	0.6047	1.3876
0.55	0.1065	0.429	1.0456

As is shown above, learning time distribution better model forecasts, and the gap between two models is increasing with more dense samples.

Results of LODE-VAE-N model benchmarking by R^2 metric are shown in Table 3. Coefficient of determination is stable and increasing slowly with increasing density of sampling and decreasing noise level. The same is true for LODE-VAE-P.

Results of LODE-VAE-P model benchmarking by R^2 metric are shown in Table 4.

Table 3 – Results of LODE-VAE-N model benchmarking by R^2 metric

α\σ	0.1	0.5	1
0.15	0.7343	0.5943	0.5523
0.35	0.7529	0.6303	0.571
0.55	0.827	0.6988	0.61

Table 4 – Results of LODE-VAE-P model benchmarking by R^2 metric

α\σ	0.1	0.5	1
0.15	0.843	0.5413	0.5612
0.35	0.8501	0.6171	0.6001
0.55	0.9146	0.7307	0.658

LODE-VAE-P model is demonstrating better results for all the metrics and all the experiment setups. By leveraging separate model for learning the distribution of time points in the sample, the latter model can better approximate the ground truth distribution of the sample.

CONCLUSIONS

The problem of continuous-time processes reconstruction from noised and unevenly distributed samples is solved in this work.

The scientific novelty of obtained results shows that neural ordinary differential equations models could be embedded into variational autoencoders framework for reconstructing dynamic of given unknown but observed process. Combining numerical integration techniques with stochastic generative models is a valid and effective approach for modeling and forecasting non-stationary time series.

The practical significance of current work and its' results is that implemented models could be applied to forecast non-stationary processes from real world, such as climate-related processes or simplifying simulations of physical processes.

Prospects for further research are to study different approaches to use as a decoder network, replacing variational autoencoders with different stochastic generative models.

ACKNOWLEDGEMENTS

The author of the following research wants to appreciate his scientific advisor, Nadezhda I. Nedashkovskaya, – Dr. Tech. Sc., Associate Professor at the Department of Mathematical Methods of System Analysis, Institute for Applied Systems Analysis at National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

REFERENCES

1. Bidyuk P. I., Romanenko V. D., Timoshchuk O. L. Time series analysis. Kyiv, Polytechnika, 2013, 230 p. (In Ukrainian)
2. Parfenenko Y. V., Shendryk V. V., Kholiavka Y. P., Pavlenko P. M. Comparison of short-term forecasting methods of electricity consumption in microgrids, *Radio Electronics, Computer Science, Control*, 2023, № 1, pp. 14–23. DOI: https://doi.org/10.15588/1607-3274-2023-1-2
3. Terence C. M. Chapter 3 – ARMA Models for Stationary Time Series, *Applied Time Series Analysis. A Practical Guide to Modeling and Forecasting*. Cambridge, Academic Press, 2019, Ch. 3, pp. 31–56. DOI: https://doi.org/10.1016/B978-0-12-813117-6.00003-X.
4. Terence. C. M. Chapter 4 – ARIMA Models for Nonstationary Time Series, *Applied Time Series Analysis. A Practical Guide to Modeling and Forecasting*. Cambridge, Academic Press, 2019, Ch. 4, pp. 57–69. DOI: https://doi.org/10.1016/B978-0-12-813117-6.00004-1.
5. Charles A., Darné O. The accuracy of asymmetric GARCH model estimation, *International Economics*, 2019, Vol. 157, pp. 179–202. DOI: https://doi.org/10.1016/j.inteco.2018.11.001
6. Bidyuk P., Prosyankina-Zharova T., Terentiev O. Modeling nonlinear nonstationary processes in macroeconomy and finances, *Advances in Computer Science for Engineering and Education*, 2019, Vol. 754, pp. 735–745. DOI: https://doi.org/10.1007/978-3-319-91008-6_72.

DOI 10.15588/1607-3274-2023-4-7

7. Kumar A. S., Anandarao S. Volatility spillover in cryptocurrency markets: Some evidences from GARCH and wavelet analysis, *Physica A: Statistical Mechanics and its Applications*, 2019, Vol. 524, pp. 448–458. DOI: https://doi.org/10.1016/j.physa.2019.04.154.
8. Geron A. Hands-On Machine Learning with Scikit-Learn and TensorFlow. Sebastopol: O'Reilly Media Inc., 2017, 760 p.
9. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge, The MIT Press, 2016, 802 p.
10. Whittle P. Hypothesis Testing in Time Series Analysis. Stockholm, Almquist and Wicksell, 1951, 187 p.
11. Box G. E.P., Jenkins G. M. Time Series Analysis: Forecasting and Control. San-Francisco, Holden-Day, 1976, 575 p.
12. Hochreiter S., Schmidhuber J. Long short-term memory, Neural computation, 1997, Vol. 9, № 8, pp. 1735–1780.
13. Cho K., Merrienboer B. van, Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *EMNLP 2014: Conference on Empirical Methods in Natural Language Processing, Doha, 25–29 October 2014: proceedings.* Doha, Association for Computational Linguistics, 2014, pp. 1724–1734. DOI: https://doi.org/10.48550/arXiv.1406.1078
14. Chen R. T.Q., Rubanova Y., Bettencourt J., Duvenaud D. Neural ordinary differential equations [Electronic resource]. *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems, Vancouver, 3–8 December 2018: proceedings*. Access mode: https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf
15. Lu J., Deng K., Zhang X., Liu G., Guan Y. Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens, *Science*, 2021, Vol. 24, Issue 7, pp. 1–13. DOI: https://doi.org/10.1016/j.isci.2021.102804.
16. De Brouwer E., Simm J., Arany A., Moreau Y. GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series [Electronic resource], *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems, Vancouver, 8–14 December 2019: proceedings.* Access mode: https://proceedings.neurips.cc/paper_files/paper/2019/file/455cb2657aaa59e32fad80cb0b65b9dc-Paper.pdf

Received 06.09.2023.
Accepted 15.11.2023.

УДК 004.93

НЕЙРОННІ ЗВИЧАЙНІ ДИФЕРЕНЦІАЛЬНІ РІВНЯННЯ ДЛЯ РЕКОНСТРУКЦІЇ ЧАСОВИХ РЯДІВ

Андросов Д. В. – аспірант Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто задачу реконструкції нестаціонарних часових рядів на основі моделей кодувальник-декодувальник за допомогою нейронних звичайних диференціальних рівнянь. Об'єктом дослідження є задача відновлення та прогнозування нестаціонарних часовиї рядів та процесів в неперевному часі. Мета роботи – синтез моделі на основі архітектури кодувальник-декодувальник та з використанням моделей типу нейронних звичайних диференційних рівнянь для реконструкції часових рядів по зашумленими, нерівномірно розподіленими у час, вхідними сигналами.

Метод. Запропоновано метод, що реалізує архітектуру кодувальника-декодувальника та аппарат штучних нейронних мереж з розв'язанням диференціальних рівнянь у латентному просторі. Було встановлено, що даний підхід демонструє високу ефективність та якість прогнозів при вирішенні задачі реконструкції часових рядів по зашумленим вхідним сигналам з випадковими інтервалами між сигналами. Запропонована модель варіаційного автокодувальника на з використанням апарату нейронних мереж була протестована на синтетичних нестаціонарних даних з додаванням білим шумом і семплінгом з випадковими інтервалами між кожним сигналом.

Результати. Розроблені показники реалізовані програмно і досліджені при вирішенні задачі реконструкції нестацонарного ряду з сезонністю.

Висновки. Проведені експерименти підтвердили, що запропонована модель ефективно вирішує задану задачу і рекомендується застосовувати її для вирішення реальних завдань, що вимагають реконструкції динаміки нестаціонарних процесів. Перспективи включають в себе подальші дослідження різних архітектур нейронних мереж, окрім рекурентних нейронних мереж та архітектур автокодувальників. Зокрема пропонується використовувати інші підходи генеративного нейромережевого моделювання, як генеративно-змагальні мережі у контексті відновлення структури часового ряду

КЛЮЧОВІ СЛОВА: нейронні звичайні диференціальні рівняння, глибокі нейронні мережі, варіаційні автокодувальники, рекурентні нейронні мережі, мережі довгострокової короткої пам'яті.

ЛІТЕРАТУРА

1. Bidyuk P. I. Time series analysis / P. I. Bidyuk, V. D. Romanenko, O. L. Timoshchuk. – Kyiv : Polytechnika, 2013. – 230 p. (In Ukrainian)
2. Comparison of short-term forecasting methods of electricity consumption in microgrids / [Y. V. Parfenenko, V. V. Shendryk, Y. P. Kholiavka, P. M. Pavlenko] // Radio Electronics, Computer Science, Control. – 2023. – № 1. – P. 14–23. DOI: https://doi.org/10.15588/1607-3274-2023-1-2
3. Terence. C. M. Chapter 3 - ARMA Models for Stationary Time Series / C. M. Terence // Applied Time Series Analysis. A Practical Guide to Modeling and Forecasting. – Cambridge: Academic Press, 2019. – Ch. 3. – P. 31–56. DOI: https://doi.org/10.1016/B978-0-12-813117-6.00003-X.
4. Terence. C. M. Chapter 4 – ARIMA Models for Nonstationary Time Series / C. M. Terence // Applied Time Series Analysis. A Practical Guide to Modeling and Forecasting. – Cambridge : Academic Press, 2019. – Ch. 4. – P. 57–69. DOI: https://doi.org/10.1016/B978-0-12-813117-6.00004-1.

DOI 10.15588/1607-3274-2023-4-7

5. Charles A. The accuracy of asymmetric GARCH model estimation. / A. Charles, O. Darné // International Economics. – 2019. – Vol. 157. – P. 179–202. DOI: https://doi.org/10.1016/j.inteco.2018.11.001
6. Bidyuk P. Modeling nonlinear nonstationary processes in macroeconomy and finances. / P. Bidyuk, T. Prosyankina-Zharova, O. Terentiev // Advances in Computer Science for Engineering and Education. – 2019. – Vol. 754. – P. 735–745. DOI: https://doi.org/10.1007/978-3-319-91008-6_72.
7. Kumar A. S. Volatility spillover in crypto-currency markets: Some evidences from GARCH and wavelet analysis. / A. S. Kumar, S. Anandarao // Physica A: Statistical Mechanics and its Applications. – 2019. – Vol. 524. – P. 448–458. DOI: https://doi.org/10.1016/j.physa.2019.04.154.
8. Geron A. Hands-On Machine Learning with Scikit-Learn and TensorFlow / A. Geron. – Sebastopol : O'Reilly Media Inc., 2017. – 760 p.
9. Goodfellow I. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. – Cambridge : The MIT Press, 2016. – 802 p.
10. Whittle P. Hypothesis Testing in Time Series Analysis / P. Whittle. – Stockholm : Almquist and Wicksell, 1951. – 187 p.
11. Box G. E.P., Jenkins G. M. Time Series Analysis: Forecasting and Control / G. E. P. Box, G. M. Jenkins. – San-Francisco: Holden-Day, 1976. – 575 p.
12. Hochreiter S. Long short-term memory / S. Hochreiter, J. Schmidhuber // Neural computation. – 1997. – Vol. 9, № 8. – P. 1735–1780.
13. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. / [K. Cho, B. van Merrienboer, C. Gulcehre et al.] // EMNLP 2014: Conference on Empirical Methods in Natural Language Processing, Doha, 25–29 October 2014: proceedings. – Doha: Association for Computational Linguistics, 2014. – P. 1724–1734. DOI: https://doi.org/10.48550/arXiv.1406.1078
14. Neural ordinary differential equations [Electronic resource] / [R. T. Q. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud, // Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems, Vancouver, 3–8 December 2018: proceedings. Access mode: https://proceedings.neurips.cc/paper_files/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf
15. Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens / [J. Lu, K. Deng, X. Zhang et al.] // Science. – 2021. – Vol. 24, Issue 7. – P. 1–13. DOI: https://doi.org/10.1016/j.isci.2021.102804.
16. GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series [Electronic resource] / [E. De Brouwer, J. Simm, A. Arany, Y. Moreau] // Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems, Vancouver, 8–14 December 2019: proceedings. Access mode: https://proceedings.neurips.cc/paper_files/paper/2019/file/455cb2657aaa59e32fad80cb0b65b9dc-Paper.pdf

DOI 10.15588/1607-3274-2023-4-7

UDC 004.932

DEEP NETWORK-BASED METHOD AND SOFTWARE FOR SMALL SAMPLE BIOMEDICAL IMAGE GENERATION AND CLASSIFICATION

Berezsky O. M. – Dr. Sc., Professor, Professor of the Department of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

Liashchynskyi P. B. – Post-graduate student of the Department of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

Pitsun O. Y. – PhD, Associate Professor, Associate Professor of the Department of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

Melnyk G. M. – PhD, Associate Professor of the Department of Computer Engineering, West Ukrainian National University, Ternopil, Ukraine.

ABSTRACT

Context. The authors of the article investigated the problem of generating and classifying breast cancer histological images. The widespread incidence of breast cancer explains the problem's relevance. The automated diagnosing procedure saves time and eliminates the subjective aspect. The study's findings can be applied to cancer CAD systems.

Objective. The purpose of the study is to develop a deep neural network-based method and software tool for generating and classifying histological images in order to increase classification accuracy.

Method. The method of histological image generation and classification was developed in the research study. This method employs CNN and GAN. To improve the classification accuracy, the initial image sample was expanded using GAN.

Results. The computer research of the developed method of image generation and classification was conducted on the basis of the dataset located on the Zenodo platform. Light microscopy served as the basis for obtaining the image. The dataset contained three classes of G1, G2, and G3 breast cancer histological images. Based on the developed method, the accuracy of image classification was 96%. This is a higher classification accuracy compared to existing models such as AlexNet, LeNet5, and VGG16. The software module can be integrated into CAD.

Conclusions. The developed method of generating and classifying images is the basis of the software module. The software module can be integrated into CAD.

KEYWORDS: computer-aided diagnosis system, breast cancer, deep neural networks, generative competitive networks, convolutional neural networks.

ABBREVIATIONS

GAN is a generative adversarial network;

CNN is a convolutional neural network;

DNN is a deep neural network;

CAD is a computer-aided diagnosis;

AWS is a Amazon Web Services;

Zenodo is a general-purpose open repository developed under the European OpenAIRE program and operated by CERN;

ReLU is a rectified linear unit;

Batch Norm is batch normalization;

ROC is a receiver operating characteristic;

CLI is a command line interface;

IS metric is a metric based on the Google Inception V3 image classification model;

FID is Fréchet inception distance;

AlexNet is a name of a convolutional neural network architecture designed by Alex Krizhevsky;

LeNet5 is a name of a convolutional neural network structure proposed by LeCun;

VGG16 is a name of a Visual Geometry Group convolutional neural network;

ResNet50 is a name of a 50-layer Residual Network;

DenseNet201 is a name of 201-layer Densely Connected Network;

CSAResnet is a channel and spatial attention embedded Resnet network;

DAMCNN is a dual attention multiscale convolutional neural network;

DSS is a decision support system;

DeepGrade is a histological grade model;

DenseNet is a Densely Connected Network;

CA-BreastNet is a Coordinated Attention Breast Network;

DHE-Mit-Classifier is a Deep Heterogeneous Ensemble mitotic Classifier;

SVM is a Support Vector Machine;

SSDHO is a Shuffled Shepherd Deer Hunting Optimization;

LR is a Logistic Regression;

MLP is a Multilayer Perceptron;

U-Net is a Network with U-shaped structure;

PyTorch is a open source machine learning framework Python Torch;

AWS SageMaker is an Amazon Web Services Sage Maker machine learning service;

AWS S3 is an Amazon Web Services Simple Storage Service;

URL is an Uniform Resource Locator;

CrossEntropyLoss is a cross entropy loss;

DOI 10.15588/1607-3274-2023-4-8

RAM is a Random Access Memory;

vCPU is a virtual Central processing unit;

GPU is a graphics processing unit;

TFLOPS is a Tera FLoating-point OPerations per Second;

Adam optimizer is an adaptive moment stochastic gradient descent method.

NOMENCLATURE

I_{inp} is a set of input images;

I_l is a training set of images;

I_t is a test set of images;

L is a set of CNN layers;

A_{CNN} is a CNN architecture;

L_{DEV} is a set of developed CNN layers;

O_{DEV} is a set of developed CNN operations;

O_{CNN} is a set of CNN operations;

P_{CNN} is a set of CNN parameters;

Gmax is a pooling function with maximum element definition in the scan window;

Gavg is a pooling function with average element definition in the scan window;

Gad_avg is a pooling function with adaptive definition of the average element in the scan window;

i is a layer index;

n is a number of CNN layers;

TP is a true positive;

TN is a true negative;

FP is a false positive;

FN is a false negative;

AC_{ALEX} is a accuracy of AlexNet original image classification;

AC_{LE} is a LeNet5 original image classification accuracy;

AC_{VGG} is a VGG16 original image classification accuracy;

I_{inp}^{g} is a set of images on the basis of GAN;

I_{l}^{g} is a training images on the basis of GAN;

I_{t}^{g} is a test images on the basis of GAN;

AC_{ALEX}^{g} is a AlexNet classification accuracy based on the extended sample;

AC_{LE}^{g} is a LeNet5 classification accuracy based on the extended sample;

AC_{VGG}^{g} is a VGG16 classification accuracy based on the extended sample;

AC_{DEV}^{g} is a classification accuracy for developed CNN architecture;

$G1$ is a Nottingham grading for breast cancer type number 1;

$G2$ is a Nottingham grading for breast cancer type number 2;

$G3$ is a Nottingham grading for breast cancer type number 3;

$\langle C \rangle$ is a set of convolution functions;

$\langle G \rangle$ is a set of pooling functions;

$\langle K \rangle$ is a set of activation functions;

V is a batch normalization operation;

$\langle B \rangle$ is a set of functions of a fully connected network;

c_i is an activation function in layer i;

v_{c_i} is a parameters of activation function in layer i.

INTRODUCTION

Breast cancer in women is still a major medical and social problem that demands immediate attention. According to recent statistics, breast cancer continues to be the most common kind of malignant neoplasm in women. In 2020, the incidence rate of breast cancer in European Union nations was 13.3% of all new cases [1]. According to the American Cancer Society, breast cancer is also the most common among American women as of 2021 [2].

When analyzing morbidity data in Ukraine for the years 2021–2022, it is important to take into account that both periods were characterized by specific conditions in the country: the long course of the COVID-19 pandemic and military operations, which affected the work of both medical institutions and the cancer registration system. The report's data for 2022 cannot accurately represent the country's real onco-epidemiological process [3].

In 2021, women's oncological incidence was dominated by breast cancer, skin cancer, and neoplasms of the body and cervix, accounting for 54.5% of identified diseases. Deaths from breast cancer, colon cancer, esophageal cancer, and ovarian cancer accounted for the majority of the overall structure of mortality among women (48.8%) [3].

Biomedical images are widely used to diagnose diseases in oncology. Let us define biomedical images.

A biomedical image is a structural and functional image of human and animal organs, used to diagnose diseases and study the anatomy and physiology of the human body [4].

Cytological, histological, and immunohistochemical images are used to diagnose [4] oncological diseases.

Accurate cancer diagnosis involves histological analysis of materials. Histopathology is the microscopic examination of thin slices of damaged tissue. Histopathologists examine tissues and offer diagnostic information based on their findings.

Histological stains are frequently used to improve the capacity to visualize or differentiate microscopic structures. Chemical fixatives are used to protect tissues against destruction while also preserving the structure of cells and subcellular components.

Cytopathological, histopathological, and immunohistochemical examinations are used to learn about the features of the tumor, its degree of dissemination, and the best treatment option [5].

DOI 10.15588/1607-3274-2023-4-8

The diagnostic process normally begins with cytopathological investigation, which might reveal the existence of abnormalities in the cells. Following that, a histological analysis of the resected tumor is undertaken for a more in-depth investigation. Immunohistochemical tests can supplement these findings by giving further information about the tumor's biological characteristics.

Cytopathological investigations involve the examination of cells obtained during a tumor biopsy or puncture. They enable the assessment of anomalies in cellular structure and the identification of a suspected cancer process. Histopathological tests involve a detailed analysis of the removed tumor and adjacent tissues under a microscope. This helps determine the type of cancer, its aggressiveness, and its penetration into adjacent tissues.

Immunohistochemical investigations employ antibodies to identify specific proteins in tissues. They make it possible to more precisely detect the subtype of cancer and examine the presence of particular chemicals that might suggest prognosis and treatment alternatives. The molecular genetic subtype of the tumor is evaluated by immunohistochemistry expression of estrogen, progesterone, and oncoprotein HER-2/neu receptors, as well as detection of tumor cell proliferation using Ki-67.

The subject of research is the process of histological image generation and classification.

The object of research is deep neural networks used for image synthesis and classification.

The purpose of the research is to develop a method and software tool for breast cancer automatic diagnosis based on histological image analysis.

1 PROBLEM STATEMENT

Let the given set of original images is I_{inp}. Let us divide this set into two subsets: I_l and I_t, and $I_{inp} = I_l \cup I_t$. In addition, the architecture of A_{CNN} CNN is given. The CNN architecture can be represented through multiple layers:

$$A_{CNN} = \left\{L_i, i = \overline{1,n}\right\}.$$

The classification accuracy is determined by the accuracy measure:

$$AC = \frac{TP + TN}{TP + TN + FP + FN}.$$

Classification accuracy depends on the number of layers and their parameters. The classification accuracy function is then presented in the following form:

$$AC = f(L, P_{CNN}).$$

For the known AlexNet, LeNet5 and VGG16 architectures, based on the original images, we get the following classification accuracies:

$$AC_{ALEX}, AC_{LE}, AC_{VGG}.$$

Then, we perform affine distortions on the input original images and generate I_{inp}^g on the basis of GAN. Let us divide these images into I_l^g and I_t^g ones, that is:

$$I_{inp}^g = I_l^g \cup I_t^g.$$

Based on the extended sample, we obtain the following classification accuracies for the known architectures AlexNet, LeNet5 and VGG16, respectively:

$$AC_{ALEX}^g, AC_{LE}^g, AC_{VGG}^g.$$

For the developed CNN architecture, we have the following classification accuracy: AC_{DEV}^g.

Classification accuracy of the developed architecture depends on the following parameters:

$$AC_{DEV}^g = f\left(L_{DEV}^g, P_{DEV}^g\right)$$

and

$$AC_{DEV}^g > \max\left(AC_{ALEX}^g, AC_{LE}^g, AC_{VGG}^g\right).$$

Therefore, it is necessary to find a CNN architecture with a certain number of layers and parameters to satisfy the following condition:

$$A_{CNN} = \underset{L_{DEV}, O_{DEV}}{\arg\max}\ AC_{DEV}^g\left(P_{DEV}^g, L_{DEV}, I_l^g\right).$$

2 REVIEW OF THE LITERATURE

The use of deep neural networks for cancer detection based on histological image processing has been widely addressed in research studies. For example, in the article [6], an ensemble of CNNs was investigated for malignancy identification using histological and cytological images. In [7], a multi-scale deep learning model was developed for breast cancer classification. In [8], the authors explored multi CNNs (VGG16, ResNet50, and DenseNet201) to detect mitotic cells.

Some researchers worked on the development of a deep learning-based approach for breast cancer image classification [9]. Others focused on the comparison of different CNN architectures for breast cancer classification [10]. Besides, segmentation and classification of cell nuclei in histology was discussed in [11]. The authors of [12] proposed a modified residual neural network-based method for breast cancer detection based on histological images. And the authors of [13] used an ensemble of deep multiscale CNN networks, namely CSAResnet and DAMCNN.

DOI 10.15588/1607-3274-2023-4-8

Additionally, a DSS for diagnosis of oncopathologies using histological images was developed by researchers in [14]. A CNN-based classifier for the automatic classification of breast cancer histological images was developed by the authors of [15]. A new DeepGrade histological class model was developed in [16]. The authors employed deep learning to analyze histological images. In [17], researchers combined two CNN architectures with the use of fractal geometry to improve histological image classification accuracy. In [18], the authors reviewed histological image classification methods for diagnosing breast cancer.

Thus, many researchers focused on the classification of breast cancer histological images. For example, the article [19] is devoted to the detection of breast cancer based on the CNN ensemble using histological images. In [20], the authors developed a method based on the combination of convolutional and recurrent deep neural networks for the classification of breast cancer histological images. In [21], the authors improved the DenseNet network and synthesized the CA-BreastNet model for the classification of breast cancer histological images. In [22], the authors analyzed modern algorithms for the automatic classification of breast cancer based on histological images. In [23], a heterogeneous ensemble based on CNN "DHE-Mit-Classifier" was developed. This ensemble was used to analyze mitotic nuclei in breast cancer histological images.

Besides, in [24], the authors analyzed the color and texture features of histological image slides. These features were used to count breast cancer mitosis. Article [25] was also devoted to the detection of mitosis in breast cancer. SVM, Naive Bayes, and Random Forest classifiers were used in the research study. The authors of the study [26] developed a DNN neural network based on the SSDHO optimization method to classify six classes of breast cancer images. In [27], the authors analyzed two methods of machine learning SVM and LR. The study also considered combinations of CNN + LR and CNN + SVM.

In the article [28], a review of machine learning methods for breast cancer diagnosis was carried out. In [29], the authors also used CNN to classify histological images of breast cancer. In the research study [30], the authors developed a 3-tier CNN model, which was used to classify breast cancer histological images. And researchers in [31] analyzed VGG16, VGG19, and ResNet50 networks for the classification of breast cancer histological images. A comparison of MLP and CNN networks was made in [32]. These networks have been used in breast cancer detection. The authors of [33] used a cascade deep learning network with U-Net architecture for segmentation and a ResNet network for breast cancer classification.

The authors of the article [34] developed a method of manual feature selection and applied a DNN to classify breast cancer. The authors of the article [35] analyzed the use of artificial intelligence methods in DSS for diagnosing breast cancer. In the article [36], a method and a software tool for diagnosing skin diseases were developed.

These authors have experience in the development of methods, algorithms, and software tools for diagnosing oncological diseases based on the analysis of cytological, histological, and immunohistochemical images. For example, in works [36–42] technology and software systems for the analysis of biomedical images for diagnosis were developed. A number of publications [42–45] are devoted to the development of methods and algorithms for the analysis of biomedical images.

3 MATERIALS AND METHODS

The developed method of generating and classifying histological images consists of the following steps:

1. Formation of the initial dataset of histological images of three classes: G1, G2, and G3 based on affine distortions.
2. Experimental study of known architectures of neural networks for histological image classification and evaluation of classification accuracy on a given sample.
3. Expanding the histological image sample based on GAN networks.
4. Determining the improved neural network accuracy on the extended sample.
5. Designing new neural network architecture.
6. Comparison of accuracy of neural network architectures for histological images.

This section details steps 1, 3, and 5. Steps 2, 4, and 6 are described in section 5.

Let's create an initial data set. Diagnosticians determine the type of breast cancer using a histological examination. The Nottingham scale assesses the degree of difference in study findings. The following types of breast cancer are distinguished by the Nottingham grading: G1, G2, and G3. Breast tumor differentiation is determined by the degree of differentiation between pathological and normal cells, as well as the tumor cell growth rate. The authors of this article used cytological and histological images of breast cancer [46] from a private image database on the Zenodo platform. All images are anonymized, which complies with European standards [47].

The original sample contains images by class: G1 – 9 images, G2 – 100 images, and G3 – 76 images. This sample was expanded to 100 images in each class by applying affine distortions [48]

An example of histological images of different cancer types is shown in Fig. 1.

Based on the initial data set, we will form an extended data set using GAN.

The generator and discriminator architectures are based on the ResNet Block, borrowed from the ResNet . [49].

The generator takes a noise vector with a Gaussian distribution of dimensions 1x100 as input and produces a 64x64x3 image as output. The generator's architecture may be generally divided into three levels, as shown in Fig. 2.

DOI 10.15588/1607-3274-2023-4-8

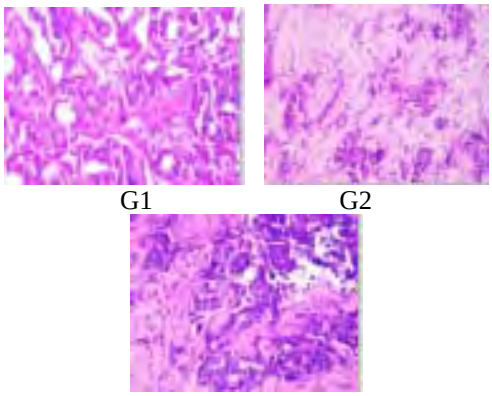


Figure 1 – Examples of histological images of type G1, G2, G3

The first layer is the pre-processing layer. In this layer, a one-dimensional input noise vector is read and sent to the linear layer for further conversion into a three-dimensional array.

The second layer is the main computing layer. It consists of four ResNet units. After each block, nearest-neighbor interpolation was used to enlarge the image by two times. In addition, following the third ResNet block, the Self-Attention block was used [50]. All convolutional layers use step 1. ReLU is also applied as the activation function.

The last layer is the output layer. Batch normalization, activation, another convolution layer, and the final Tanh activation function are all used here.

The discriminator is a convolutional neural network that takes an image of 64x64x4 pixels as input. The discriminator consists of five ResNet blocks. A Self-Attention mechanism is applied after the first block.

To reduce image dimensionality, the Average Pooling operation with a kernel size of 2x2 and a step of 2 is used in each ResNet block. However, dimensionality reduction is not used in the final block. Convolution layers use step 1.

ReLU is also applied as an activation function.

The output of the discriminator is two linear layers. The first has 1280 neurons, whereas the second has only one. The architecture of the discriminator is shown in Fig. 3.

IS and *FID* metrics were used to evaluate the quality of synthesized images [51, 52].

The next step is to design the new CNN architecture.

The CNN architecture will be presented in set of layers:

$$A_{CNN} = \{L_i, i = \overline{1,n}\}.$$

Let us also define a set of CNN operations:

$$O_{CNN} = \{\langle C\rangle, \langle P\rangle, \langle K\rangle, V, \langle B\rangle\}.$$

Each CNN layer is a separate operation with parameters. For example, we detail a set of convolution operations:

$$C = \{\langle c_1\rangle\langle v_{c_1}\rangle, ..., \langle c_i\rangle\langle v_{c_i}\rangle, ..., \langle c_n\rangle\langle v_{c_n}\rangle\}.$$

Other conversion operations are determined similarly.

The architecture of the developed CNN is presented in Figure 4.

The CNN architecture consists of nine convolutional layers, four pooling layers, and one fully connected (linear) layer.

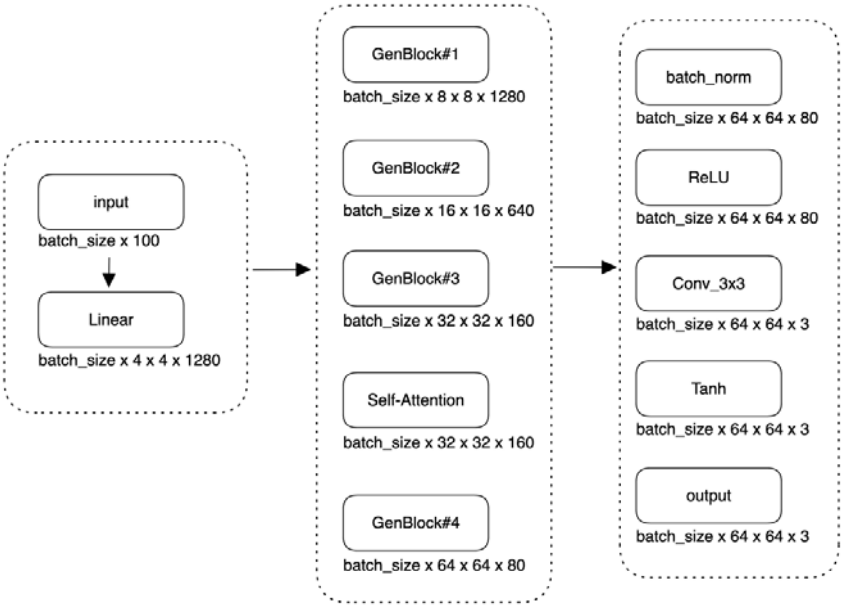


Figure 2 – Generator

DOI 10.15588/1607-3274-2023-4-8

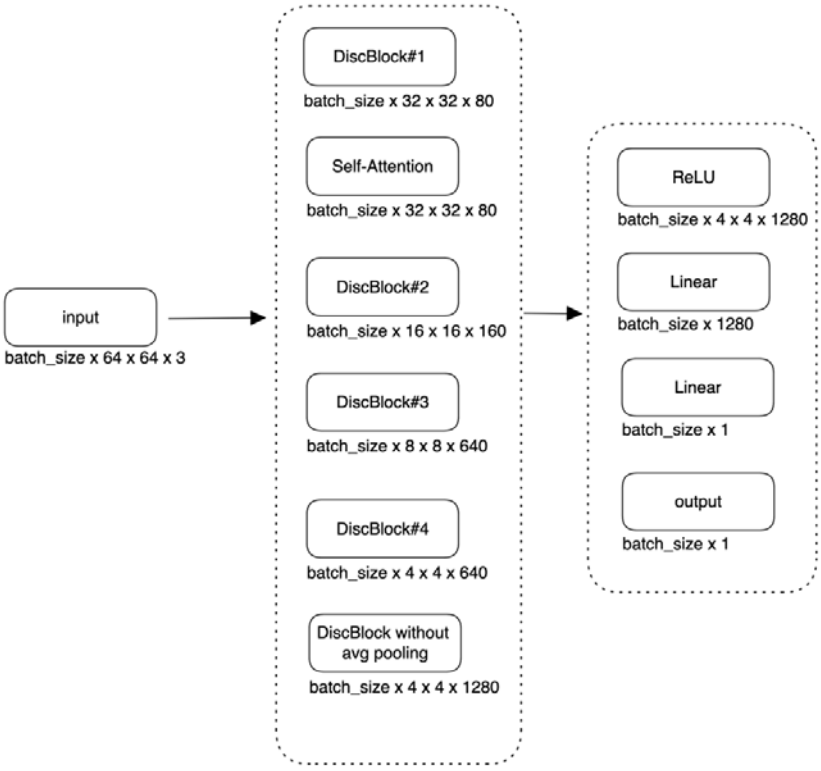


Figure 1 – Discriminator

All convolutional layers use step 1. Maximum pooling layers with a 3x3 kernel and step 2 are used to minimize the image's dimensionality.

An image with a size of 64x64x3 pixels is sent to the network input. The image is then passed to the first convolutional layer, which uses 64 feature maps with a 3x3 kernel.

The next two layers are convolutional blocks, which consist of successive layers of 3x3 kernel convolution, ReLU activation, and batch normalization. The first block employs 64 feature maps, whereas the second employs 128. A maximum pooling layer is then applied after these blocks. Accordingly, now the image size is 32x32x128.

Then there are two convolutional blocks with a maximum pooling layer at the end. However, here the convolution layers use a 1x1 kernel and the same number of feature maps – 128. The size of the image after these layers is 16x16x128.

The next two convolution blocks are identical to the first two and also use the same number of feature maps – 128. At the end, a maximum pooling layer is applied. The image size is 8x8x128.

Next, there are two convolution blocks, identical to the previous ones, but the pooling layer is no longer applied after them. The size of the image remains the same – 8x8x128.

The output layer consists of sequential batch normalization layers, an adaptive average pooling layer with the number of output nodes 1, and a linear layer with 3 nodes.

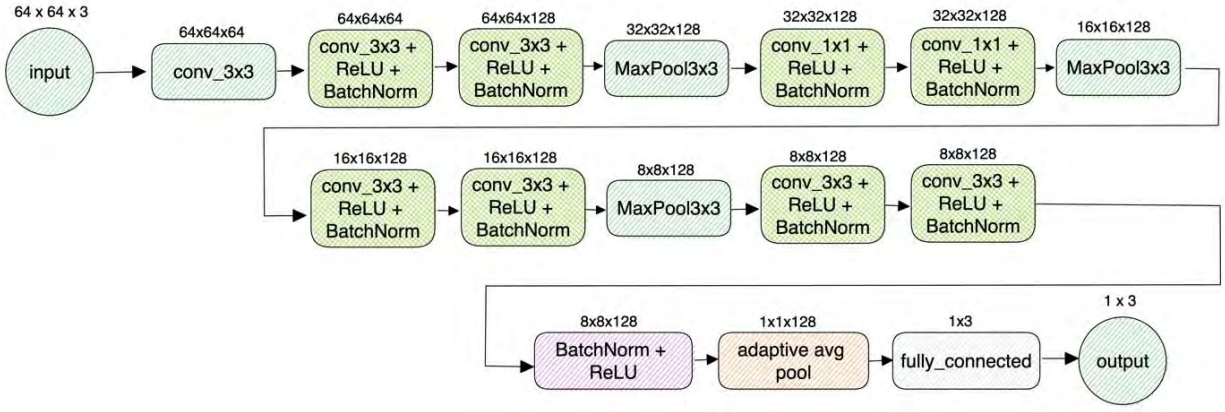


Figure 4 – Developed CNN Architecture

DOI 10.15588/1607-3274-2023-4-8

The CNN parameters are shown in Table 1. The formalized description of the developed CNN is as follows:

$$A_{CNN} = \left\{ \left\langle I_{inp_1} \right\rangle \left\langle 64 \times 64 \times 3 \right\rangle \right\};$$

$$\left\langle c_2 \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 1, \text{padding} = 1 \right\rangle;$$

$$\left\langle \left\langle c_3 \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 1, \text{padding} = 1 \right\rangle; \right.$$

$$\left. \left\langle k_3 \right\rangle \left\langle \text{ReLu} \right\rangle; \left\langle v_3 \right\rangle \left\langle \text{Batch Norm} \right\rangle \right\rangle;$$

$$\left\langle \left\langle c_4 \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 1, \text{padding} = 1 \right\rangle; \right.$$

$$\left. \left\langle k_4 \right\rangle \left\langle \text{ReLu} \right\rangle; \left\langle v_4 \right\rangle \left\langle \text{Batch Norm} \right\rangle \right\rangle;$$

$$\left\langle g_{\max_5} \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 2 \right\rangle;$$

$$\left\langle \left\langle c_6 \right\rangle \left\langle \text{kernel size} = 1 \times 1, \text{stride} = 1, \text{padding} = 1 \right\rangle; \right.$$

$$\left. \left\langle k_6 \right\rangle \left\langle \text{ReLu} \right\rangle; \left\langle v_6 \right\rangle \left\langle \text{Batch Norm} \right\rangle \right\rangle;$$

$$\left\langle \left\langle c_7 \right\rangle \left\langle \text{kernel size} = 1 \times 1, \text{stride} = 1, \text{padding} = 1 \right\rangle; \right.$$

$$\left. \left\langle k_7 \right\rangle \left\langle \text{ReLu} \right\rangle; \left\langle v_7 \right\rangle \left\langle \text{Batch Norm} \right\rangle \right\rangle;$$

$$\left\langle g_{\max_8} \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 2, \text{padding} = 1 \right\rangle;$$

$$\left\langle \left\langle c_9 \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 1, \text{padding} = 1 \right\rangle; \right.$$

$$\left. \left\langle k_9 \right\rangle \left\langle \text{ReLu} \right\rangle; \left\langle v_9 \right\rangle \left\langle \text{Batch Norm} \right\rangle \right\rangle;$$

$$\left\langle \left\langle c_{10} \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 1, \text{padding} = 1 \right\rangle; \right.$$

$$\left. \left\langle k_{10} \right\rangle \left\langle \text{ReLu} \right\rangle; \left\langle v_{10} \right\rangle \left\langle \text{Batch Norm} \right\rangle \right\rangle;$$

$$\left\langle g_{\max_{11}} \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 2, \text{padding} = 1 \right\rangle;$$

$$\left\langle \left\langle c_{12} \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 1, \text{padding} = 1 \right\rangle; \right.$$

$$\left. \left\langle k_{12} \right\rangle \left\langle \text{ReLu} \right\rangle; \left\langle v_{12} \right\rangle \left\langle \text{Batch Norm} \right\rangle \right\rangle;$$

$$\left\langle \left\langle c_{13} \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 1, \text{padding} = 1 \right\rangle; \right.$$

$$\left. \left\langle k_{13} \right\rangle \left\langle \text{ReLu} \right\rangle; \left\langle v_{13} \right\rangle \left\langle \text{Batch Norm} \right\rangle \right\rangle;$$

$$\left\langle \left\langle v_{14} \right\rangle \left\langle \text{Batch norm} \right\rangle; \left\langle k_{14} \right\rangle \left\langle \text{ReLu} \right\rangle \right.;$$

$$\left\langle g_{\max_{15}} \right\rangle \left\langle \text{kernel size} = 3 \times 3, \text{stride} = 1, \text{padding} = 1 \right\rangle;$$

$$\left. \left\langle f_{16} \right\rangle \left\langle clauses = 3 \right\rangle \right\}.$$

Table 1 – Developed CNN parameters

Layer number	Layer type	Options
1	Image input	64x64x3 image
2	Convolution	3x3 kernel convolution with stride 1 and same padding
3	Convolution	ReLU followed by 3x3 kernel convolution with stride 1 and padding 1 followed by Batch Normalization
4	Convolution	ReLU followed by 3x3 kernel convolution with stride 1 and padding 1 followed by Batch Normalization
5	MaxPooling	3x3 kernel max pooling with stride 2
6	Convolution	ReLU followed by 1x1 kernel convolution with stride 1 and padding 1 followed by Batch Normalization
7	Convolution	ReLU followed by 1x1 kernel convolution with stride 1 and padding 1 followed by Batch Normalization
8	MaxPooling	3x3 kernel max pooling with stride 2
9	Convolution	ReLU followed by 3x3 kernel convolution with stride 1 and padding 1 followed by Batch Normalization
10	Convolution	ReLU followed by 3x3 kernel convolution with stride 1 and padding 1 followed by Batch Normalization
11	MaxPooling	3x3 kernel max pooling with stride 2
12	Convolution	ReLU followed by 3x3 kernel convolution with stride 1 and padding 1 followed by Batch Normalization
13	Convolution	ReLU followed by 3x3 kernel convolution with stride 1 and padding 1 followed by Batch Normalization
14	BatchNorm	Batch normalization followed by ReLU
15	Adaptive Average Pooling	3x3 kernel adaptive average pooling with output nodes
16	Output	Linear layer with 3 output nodes

4 EXPERIMENTS

For computer experiments, special software has been developed. The software implementation is based on the reliability and scalability of Amazon Web Services cloud infrastructure, allowing efficient use of cloud computing resources. The program focuses on the development and deployment of a PyTorch-based CNN model for histological image classification. The infrastructure of the software is shown in Figure 5.

The developed software is made up of two different Python files: *train.py* and *predict.py*, each of which serves a specific purpose in the entire workflow.

In the *train.py* file, we describe a convolutional neural network model. We define the architecture by specifying parameters such as the number of convolution layers, filters, activation functions, and loss functions. This file also outlines the model training process.

To train the CNN model, we use AWS SageMaker and a cloud-based machine learning service. This service offers a scalable and controlled environment for efficient GPU training.

After the training process is completed, the model is downloaded and stored in the AWS S3 service. A URL is generated in AWS SageMaker to make the model available for use. The URL allows you to utilize the model in both the AWS cloud architecture and other web apps.

The software *predict.py* allows the use of the trained model to classify new data. This application is designed to be a CLI tool. To obtain classification results for the images, just call the file, specifying the path of the directory containing the images as the first parameter. The model analyses images and produces classification findings that may be further analyzed or integrated into other research procedures. The classification results are output to the *output.txt* file for convenience, with the data separated into two columns – the image file and the class.

DOI 10.15588/1607-3274-2023-4-8

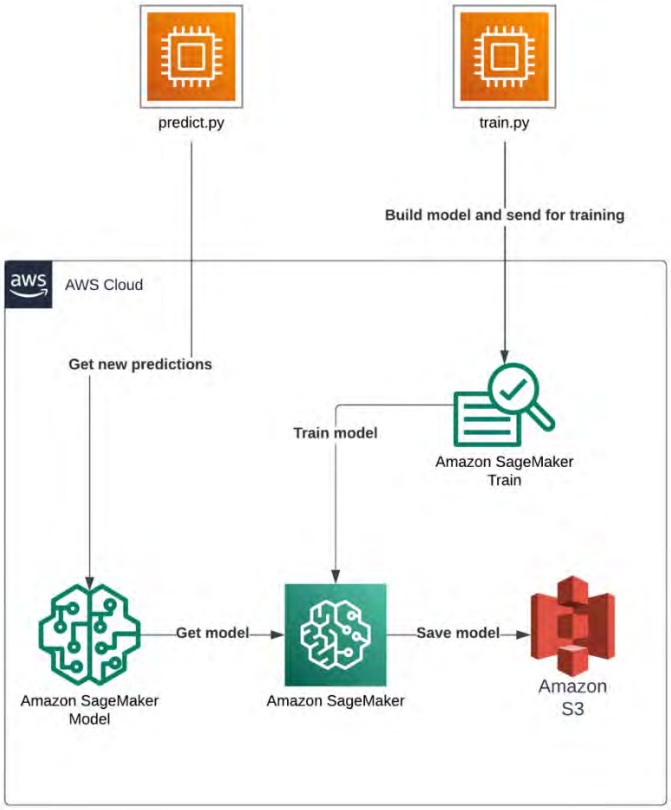


Figure 5 – Software infrastructure

5 RESULTS

Computer experiments were performed on the original and expanded samples for AlexNet, LeNet5, and VGG16. Experiments have also been carried out for the developed CNN architecture on the expanded sample.

AlexNet, LeNet5, and VGG16 original sample experiments.

100 images were used for each class. The training sample was divided into 80/20. The same training parameters were used for all networks: the Adam optimizer (1e–4), the number of epochs was 100, the batch size was 100, and the loss function was CrossEntropyLoss.

Computer experiments were carried out using the three most prominent architectures: AlexNet, LeNet5, and VGG16. The results of the experiments are as follows: AlexNet classification accuracy was 74%, LeNet5 classification accuracy was 57%, and VGG16 classification accuracy was 70%.

ROC curves for these architectures are shown in Fig. 6–8.

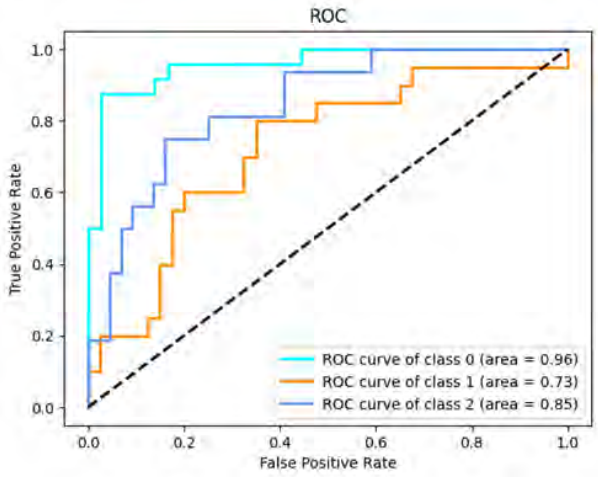


Figure 6 – ROC curve of AlexNet

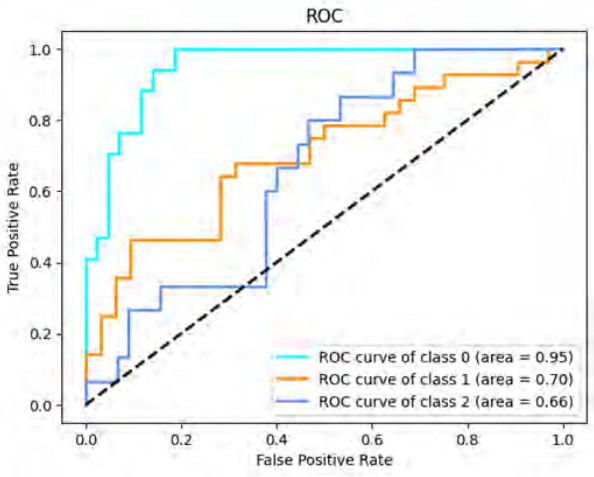


Figure 7 – ROC curve of LeNet5

DOI 10.15588/1607-3274-2023-4-8

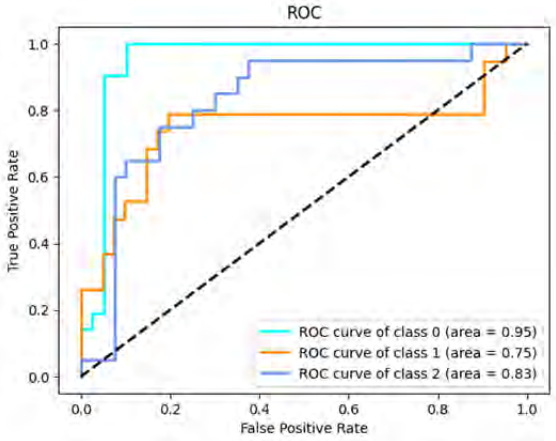


Figure 8 – ROC curve of VGG16

The original sample of images (100 images per class) was used as the training dataset. The Python programming language and the Pytorch framework were used to write the code. The GAN network training parameters were as follows: Adam optimizer, generator learning rate – 1e–4, discriminator – 4e–4, number of epochs – 100000, batch size – 96, and loss function – HingeLoss.

A virtual machine with the following configuration was used to run the experiments: 16 GB RAM, 10 vCPU x 2.2 GHz, Nvidia Tesla V100 GPU 16 GB (13.2 TFLOPS). The GAN network was trained for 11 hours. As a result of the experiments, the values of the metrics for the network were as follows: IS – 3.024, FID – 68.

Examples of synthesized images are shown in Fig. 9.

AlexNet, LeNet5, and VGG16 extended sample experiments. The training sample was increased to 3000 images per class using a generative-competitive network. The sample has 9,000 images in total. The training sample was 80/20 divided. Three architectures were tested on computers: AlexNet, LeNet5, and VGG16. The same training parameters were used for all networks – the Adam optimizer (1e–4), the number of epochs was 10, and the batch size was 100. The results of the experiments are as follows: AlexNet classification accuracy was 85%, LeNet5 – 90%, and VGG16 – 91%.

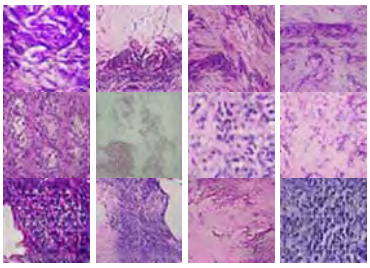


Figure 9 – Examples of synthesized images

The ROC curves for these architectures are shown in Fig. 10–12.

Experiments on extended samples for the developed CNN. The number of images per class and training parameters was similar to those for experiments with classical architectures. As a result of the experiments, the classification accuracy was 96%. The ROC curve is shown in Fig. 13.

A comparison of neural network architectures is shown in Table 2.

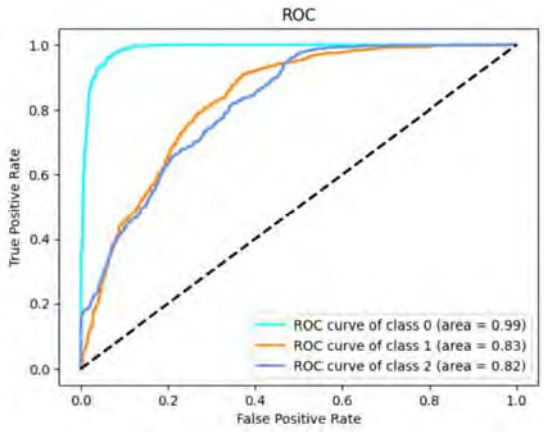


Figure 10 – ROC curve of AlexNet

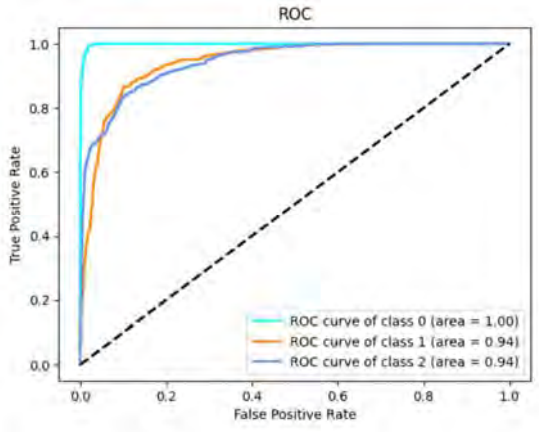


Figure 11 – ROC curve of LeNet5

DOI 10.15588/1607-3274-2023-4-8

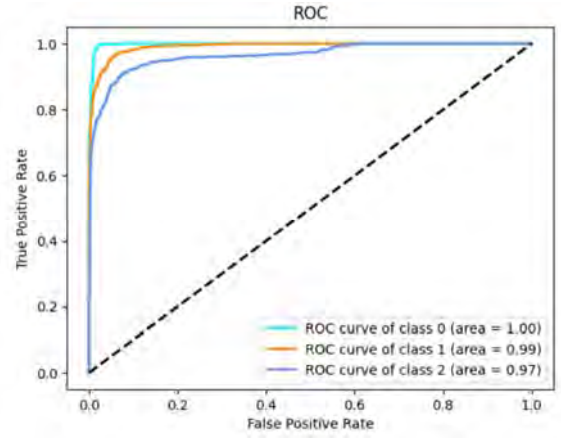


Figure 12 – ROC curve of VGG16

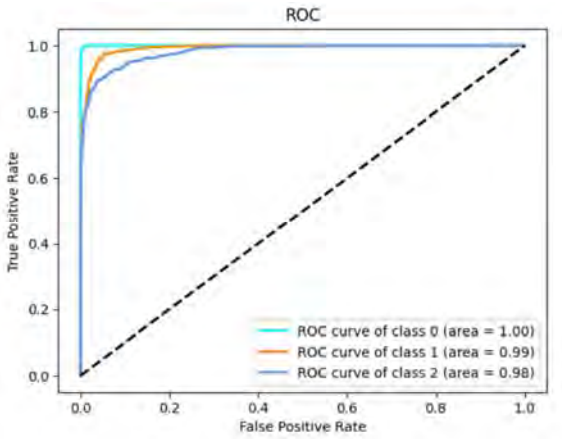


Figure 13 – ROC curve of the developed CNN

Table 2 – ANN comparison

Aricle	ANN	Accuracy %
[53]	AlexNet	85
[54]	LeNet5	90
[55]	VGG16	91
Proposed CNN	Custom CNN	96

6 DISCUSSION

The limitation of using histological images to diagnose different forms of breast cancer is the limited initial sample size.

According to studies of CNN applications, every CNN model performs better on larger datasets. Large datasets offer a more diversified set of samples to train on, allowing the model to generalize the data more effectively. CNNs learn a broader range of characteristics and patterns when trained on a huge dataset, making them more resistant to variations in the data. Overtraining is more likely with smaller datasets. Simultaneously, the model learns to recall rather than generalize training data. Because of the higher variety in the data, large datasets make retraining a model challenging. The model has the capacity to handle large datasets. In large datasets, the model has the ability to learn complex and hierarchical features. This is especially important for deep CNNs like VGG16, which have many layers. Large datasets allow these models to extract meaningful features at different levels of abstraction. With a large dataset, the optimization process (e.g., gradient descent) usually works more efficiently.

GAN was used to expand the initial sample.

The initial original sample contained the following number of images per class: G1 – 9 images, G2 – 100 images, and G3 – 76 images. Based on affine distortions, the sample is expanded to 100 images in each class. 7200 artificial images were generated using GAN. The GAN network has been trained for 11 hours. At the same time, it was possible to obtain the value of IS metrics – 3.024, and FID – 42.552.

For the three architectures of AlexNet, LeNet5, and VGG16, the following results were obtained for the initial samples: AlexNet classification accuracy was 74%, LeNet5 classification accuracy was 57%, and VGG16 classification accuracy was 70%. On the expanded sample, the results of the experiments are as follows: AlexNet classification accuracy was 85%, LeNet5 – 90%, and VGG16 – 91%.

The developed CNN architecture showed an accuracy of 96% on extended samples.

The developed program can be used in CAD. This will allow accurate histological image classification when making a diagnosis.

CONCLUSIONS

The problem of histological image classification to identify different breast cancer types was examined in this article. The authors used CNN in the research study. Low classification accuracy was revealed by analysis of known CNN based on original samples. The initial GAN-based image sample was enlarged for this purpose. On extended samples, AlexNet, LeNet5, and VGG16 showed a significant increase in classification accuracy.

In comparison to the well-known AlexNet, LeNet5, and VGG16 architectures, the newly developed CNN architecture demonstrated higher classification accuracy.

Scalability, dependability, and cost-effectiveness were provided for model training and classification using AWS SageMaker and AWS S3. Separating the training and prediction steps into two files (train.py and predict.py) ensured modularity and ease of maintenance.

With a user-friendly and simple interface, the CLI tool (predict.py) simplified the prediction process.

The cloud architecture ensures that the trained model is stored in a safe and accessible place and can be used for prediction in any application.

The architecture of the software system allows efficient training of models and classification of new data in a cloud environment, making it suitable for scalable machine learning applications.

DOI 10.15588/1607-3274-2023-4-8

The scientific novelty of the article is the development of a method of small initial sample image generation and classification.

The practical value of the article is the development of software for image generation and classification, which can be used as a separate module in CAD.

Prospects for further research are the investigation of methods of automatic design of convolutional networks and generative-competitive networks and the development of CADs for automatic diagnosis in oncology.

REFERENCES

1. Breast cancer burden in EU-27. [Electronic resource]. Access mode: https://ecis.jrc.ec.europa.eu/pdf/Breast_cancer_factsheet-Oct_2020.pdf
2. Cancer Facts for Women. American Cancer Society. [Electronic resource]. Access mode: https://www.cancer.org/cancer/risk-prevention/understanding-cancer-risk/cancer-facts/cancer-facts-for-women.html
3. Cancer in Ukraine 2021–2022: Incidence, mortality, prevalence and other relevant statistics. Bulletin of the National Cancer Registry of Ukraine № 24, 2021–2022. [Electronic resource]. Access mode: http://www.ncru.inf.ua/publications/BULL_24/PDF_E/bull_eng_24.pdf
4. Berezsky O., Pitsun O., Verbovyy S. et al. Computer diagnostic tools based on biomedical image analysis 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), 21–25 Feb. 2017. Lviv, Ukraine, 2017, pp. 388–391.
5. Aswathy M. A., Jagannath M. Detection of breast cancer on digital histopathology images: Present status and future possibilities, *Informatics in Medicine Unlocked*, 2017, Vol. 8, pp. 74–79. DOI: 10.1016/j.imu.2016.11.001
6. Dey S., Mitra S., Chakrabortyc S. et al. GC-EnC: A Copula based ensemble of CNNs for malignancy identification in breast histopathology and cytology images, *Computers in Biology and Medicine*, 2023, Vol. 152, P. 106329. DOI: 10.1016/j.compbiomed.2022.106329
7. Chattopadhyay S., Dey A., Singh P. K. et al. MTRRE-Net: A deep learning model for detection of breast cancer from histopathological images, *Computers in Biology and Medicine*, 2022, Vol. 150, P. 106155. DOI: 10.1016/j.compbiomed.2022.106155
8. Shihabuddin A. R., Sabeena K. S. Multi CNN based automatic detection of mitotic nuclei in breast histopathological images, *Computers in Biology and Medicine*, 2023, Vol. 158, P. 106815. DOI: 10.1016/j.compbiomed.2023.106815
9. Vo D. M., Nguyen N.-Q., Lee S.-W. Classification of breast cancer histology images using incremental boosting convolution networks, *Information Sciences*, 2019, Vol. 482, pp. 123–138. DOI: 10.1016/j.ins.2018.12.089
10. Saikia A. R., Bora K., Mahanta L. B. et al. Comparative assessment of CNN architectures for classification of breast FNAC images, *Tissue Cell*, 2019, Vol. 57, pp. 8–14. DOI: 10.1016/j.tice.2019.02.001
11. Dogar G. M., Shahzad M., Fraz M. M. Attention augmented distance regression and classification network for nuclei instance segmentation and type classification in histology images, *Biomedical Signal Processing and Control*, 2023, Vol. 79, P. 104199. DOI: 10.1016/j.bspc.2022.104199
12. Gupta V., Vasudev M., Doegar A. et al. Breast cancer detection from histopathology images using modified residual neural networks, *Biocybernetics and biomedical engineering*, 2021, Vol. 41, pp. 1272–1287. DOI: 10.1016/j.bbe.2021.08.011
13. Karthik R., Menaka R., Siddharth M. V. Classification of breast cancer from histopathology images using an ensemble of deep multiscale networks, *Biocybernetics and biomedical engineering*, 2022, Vol. 42, pp. 963–976. DOI: 10.1016/j.bbe.2022.07.006
14. Dovbysh A., Shelehov I., Romaniuk A. et al. Decision-making support system for diagnosis of oncopathologies by histological images, *Journal of Pathology Informatics*, 2023, Vol. 14, P. 100193. DOI: 10.1016/j.jpi.2023.100193
15. Roy K., Banik D., Bhattacharjee D. et al. Patch-based system for Classification of Breast Histology images usingdeep learning, *Computerized Medical Imaging and Graphics*, 2019, Vol. 71, pp. 90–103. DOI: 10.1016/j.compmedimag.2018.11.003
16. Wang Y., Acs B., Robertson S. et al. Improved breast cancer histological grading using deep learning, Annals of oncology, 2021, Vol. 33, № 1, pp. 89–98. DOI: 10.1016/j.annonc.2021.09.007
17. Roberto G. F., Lumini A., Neves L. A. et al. Fractal Neural Network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images, *Expert Systems With Applications*, 2021, Vol. 166, P. 114103. DOI: 10.1016/j.eswa.2020.114103
18. Houssein E. H., Emama M. M., Ali A. A. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review, *Expert Systems with Applications*, 2021, Vol. 167, P. 114161. DOI:10.1016/j.eswa.2020.114161
19. Majumdar S., Pramanik P., Ram S. Gamma function based ensemble of CNN models for breast cancer detection in histopathology images, *Expert Systems With Applications*, 2023, Vol. 213, P. 119022. DOI: 10.1016/j.eswa.2022.119022
20. Yana R., Ren F., Wangb Z. et al. Breast cancer histopathological image classification using a hybrid deep neural network, *Methods*, 2020, Vol. 173, pp. 52–60. DOI: 10.1016/j.ymeth.2019.06.014
21. Dianzhi Y., Jianwu L., Tengbao C. et al. SECS: An effective CNN joint construction strategy for breast cancer histopathological image classification, *Journal of King Saud University – Computer and Information Sciences*, 2023, Vol. 35, pp. 810–820. DOI: 10.1016/j.jksuci.2023.01.017
22. Aresta G., Araújo T., Kwok S. et al. BACH: Grand challenge on breast cancer histology images, *Medical Image Analysis*, 2019, Vol. 56, pp. 122–139. DOI: 10.1016/j.media.2019.05.010
23. Sohail A., Khan A., Nisar H. et al. Mitotic nuclei analysis in breast cancer histopathology images using deep ensemble classifier, *Medical Image Analysis*, 2021, Vol. 72, P. 102121. DOI: 10.1016/j.media.2021.102121
24. Wahab N., Khan A. Multifaceted fused-CNN based scoring of breast cancer whole-slide histopathology images, *Applied Soft Computing Journal*, 2020, Vol. 97, P. 106808. DOI: 10.1016/j.asoc.2020.106808
25. Rehman M. U., Akhtar S., Zakwan M. et al. Novel architecture with selected feature vector for effective classification of mitotic and non-mitotic cells in breast cancer histology

DOI 10.15588/1607-3274-2023-4-8

images, *Biomedical Signal Processing and Control*, 2022, Vol. 71, P. 103212. DOI:10.1016/j.bspc.2021.103212
26. Bhausaheb D. P., Kashyap K. L. Shuffled shepherd deer hunting optimization based deep neural network for breast cancer classification using breast histopathology images, *Biomedical Signal Processing and Control*, 2023, Vol. 83, P. 104570. DOI: 10.1016/j.bspc.2023.104570
27. Guptaa K., Chawlaa N. Analysis of Histopathological Images for Prediction of Breast Cancer Using Traditional Classifiers with Pre-Trained CNN, *Procedia Computer Science*, 2020, №167, pp. 878–889. DOI: 10.1016/j.procs.2020.03.427
28. Yadav R. K., Singh P., Kashtriya P. Diagnosis of Breast Cancer using Machine Learning Techniques – A Survey, *Procedia Computer Science*, 2023, Vol. 218, pp. 1434–1443. DOI: 10.1016/j.procs.2023.01.122
29. Dabeer S., Khan M. M., Islam S. Cancer diagnosis in histopathological image: CNN based approach, *Informatics in Medicine Unlocked*, 2019, Vol. 16, P. 100231. DOI: 10.1016/j.imu.2019.100231
30. Kate V., Shukla P. A 3 Tier CNN model with deep discriminative feature extraction for discovering malignant growth in multi-scale histopathology images, *Informatics in Medicine Unlocked*, 2021, Vol. 24, P. 100616. DOI: 10.1016/j.imu.2021.100616
31. Shallu S., Mehra R. Breast cancer histology images classification: Training from scratch or transfer learning?, *ICT Express*, 2018, Vol. 4, pp. 247–254. DOI: 10.1016/j.icte.2018.10.007
32. Desai M., Shah M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN), *Clinical eHealth*, 2021, Vol. 4, pp. 1–11. DOI: 10.1016/j.ceh.2020.11.002
33. Asadi B., Memon Q. Efficient breast cancer detection via cascade deep learning network, *International Journal of Intelligent Networks*, 2023, Vol. 4, pp. 46–52. DOI: 10.1016/j.ijin.2023.02.001
34. Joseph A. A., Abdullahi M., Junaidu S. B. et al. Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer), *Intelligent Systems with Applications*, 2022, Vol. 14, P. 200066. DOI: 10.1016/j.iswa.2022.200066
35. Chaudhury S., Sau K. A BERT encoding with Recurrent Neural Network and Long-Short Term Memory for breast cancer image classification, *Decision Analytics Journal*, 2023, Vol. 6, P. 100177. DOI: 10.1016/j.dajour.2023.100177
36. Lovkin V. M., Subbotin S. A., Oliinyk A. O. et al. Method and software component model for skin disease diagnosis, *Radio Electronics, Computer Science, Control*, 2023, № 1, P. 40. DOI: 10.15588/1607-3274-2023-1-4
37. Berezsky O. M., Berezska K. M., Melnyk G. M. et al. Design of computer systems for biomedical image analysis, *Proceedings of the X th International Conference «The Experience of Designing and Application of CAD Systems in Microelectronics» CADSM 2009, 24–28 February 2009, Lviv-Polyana, Ukraine*. Lviv, Publishing House Vezha&CoC, P. 186–191.
38. Berezsky O., Datsko T., Verbovyy S. The intelligent system for diagnosing breast cancers based on image analysis, *Proceedings of Information Technologies in Innovation Business* (ITIB). Kharkiv, Ukraine, 7–9 October, 2015, pp. 27–30. DOI: 10.1109/ITIB.2015.7355067.
39. Berezsky O., Verbovyy S., Dubchak L. et al. Fuzzy system diagnosing of precancerous and cancerous conditions of the breast cancer, *Proceedings of the XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT'2016)*. Lviv, Ukraine, 6–10 September, 2016, pp. 200–203. DOI: 10.1109/STC-CSIT.2016.7589906.
40. Berezsky O., Dubchak L., Pitsun O. Access Distribution in Automated Microscopy System / O. Berezsky, // *Proceedings of the 14 th International Conference «The Experience of Designing and Application of CAD Systems in Microelectronics» CADSM 2017, 21–25 February 2017, Lviv, Ukraine.* Lviv, 2017, pp. 241–243. DOI: 10.1109/CADSM.2017.7916125
41. Berezsky O., Verbovyy S., Pitsun O. Hybrid Intelligent information techology for biomedical image processing, *Proceedings of the IEEE International Conference "Computer Science and Information Technologies" CSIT'2018, Lviv, Ukraine, 11–14 September, 2018*. Lviv, 2018, pp. 420–423. DOI: 10.1109/STC-CSIT.2018.8526711.
42. Berezsky O., Pitsun O., Batryn N., et al. Modern automated microscopy systems in oncology, *Proceedings of the 1st International Workshop on Informatics & Data-Driven Medicine, Lviv, Ukraine, 28–30 november 2018.* Lviv, 2018, pp. 311–325.
43. Berezsky O. M., Pitsun O. Y. Evaluation methods of image segmentation quality, *Radio Electronics, Computer Science, Control*, 2018, Vol. 1, pp. 119–128. DOI:10.15588/1607-3274-2018-1-14
44. Berezsky O., Dubchak L., Batryn N., et al. Fuzzy System For Breast Disease Diagnosing Based On Image Analysis, *Proceedings of the II International Workshop Informatics & Data-Driven Medicine (IDDM 2019).* Lviv, Ukraine, 11–13 November, 2019. Lviv, 2019, pp. 69–83.
45. Berezsky O., Pitsun O., Melnyk G. et al. An Approach toward Automatic Specifics Diagnosis of Breast Cancer Based on an Immunohistochemical Image, *Journal of Imaging*, 2023, Vol. 9, № 1, P. 12. DOI: 10.3390/jimaging9010012
46. Berezsky O., Datsko T., Melnyk G. Cytological and histological images of breast cancer, [Electronic resource]. Access mode: https://zenodo.org/records/7890874. DOI: 10.5281/zenodo.7890874
47. Intresoft Consulting. General Data Protection Regulation (GDPR). Recital 26. [Electronic resource]. Access mode: https://gdpr-info.eu/recitals/no-26/
48. Liashchynskyi P. Rudi. Lightweight image converter and dataset augmentor, [Electronic resource]. Access mode: https://github.com/liashchynskyi/rudi
49. He K., Zhang X., Ren S. et al. Deep Residual Learning for Image Recognition, 2015. [Electronic resource]. Access mode: https://arxiv.org/pdf/1512.03385.pdf
50. Lim J. H., Ye J. C. Geometric GAN, 2017. [Electronic resource]. Access mode: https://arxiv.org/pdf/1705.02894v2.pdf
51. Salimans T., Goodfellow I., Zaremba W. et al. Improved techniques for training GANs, *Advances in neural information processing systems: Annual Conference on Neural Information Processing Systems 2016, December 5–10, 2016, Barcelona.* Spain, Barcelona, 2016, pp. 2226–2234 DOI: 10.48550/arXiv.1606.03498
52. Heusel M., Ramsauer H., Unterthiner T. et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium, *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NY,

DOI 10.15588/1607-3274-2023-4-8

United States, Curran Associates Inc., 2017, pp. 6629–6640. Access mode: https://dl.acm.org/doi/10.5555/3295222.3295408
53. Nawaz W., Ahmed S., Tahir A. et al. Classification Of Breast Cancer Histology Images Using ALEXNET. In: Campilho, A., Karray, F., ter Haar Romeny, B. (eds) Image Analysis and Recognition. ICIAR 2018, *Lecture Notes in Computer Science*. Springer, Cham, 2018, Vol. 10882, pp. 869–876. DOI: 10.1007/978–3-319-93000-8_99
54. Verdhan V. Image Classification Using LeNet. In: Computer Vision Using Deep Learning. Apress, Berkeley, CA., 2021, pp. 67–101. DOI: 10.1007/978-1-4842-6616-8_3
55. Tammina S. Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images, *International Journal of Scientific and Research Publications*, 2019, Vol. 9, Issue 10, pp. 143–150. DOI: 10.29322/IJSRP.9.10.2019.p9420

Received 08.10.2023.
Accepted 26.11.2023.

УДК 004.932

МЕТОД І ПРОГРАМНИЙ ЗАСІБ ГЕНЕРУВАННЯ І КЛАСИФІКАЦІЇ БІОМЕДИЧНИХ ЗОБРАЖЕНЬ НА ОСНОВІ ГЛИБОКИХ МЕРЕЖ ІЗ МАЛОЮ ВИБІРКОЮ

Березький О. М. – д-р техн. наук, професор, професор кафедри комп'ютерної інженерії Західноукраїнського національного університету, Тернопіль, Україна.
Лящинський П. Б. – аспірант кафедри комп'ютерної інженерії Західноукраїнського національного університету, Тернопіль, Україна.
Піцун О. Й. – канд. техн. наук, доцент, доцент кафедри комп'ютерної інженерії Західноукраїнського національного університету, Тернопіль, Україна.
Мельник Г. М. – канд. техн. наук, доцент кафедри комп'ютерної інженерії Західноукраїнського національного університету, Тернопіль, Україна.

АНОТАЦІЯ

Актуальність. У статті досліджено проблему класифікації гістологічних зображень раку молочної залози. Актуальність проблеми пояснюється широкою розповсюдженістю хвороби – раку молочної залози . Автоматизація процесу постановки діагнозу дає можливість зменшити час і виключити суб'єктивний фактор. Результати дослідження можуть бути використані в CAD в онкології.

Мета роботи – розробка методу та програмного засобу генерування і класифікації гістологічних зображень на основі глибоких нейронних мереж для підвищення точності класифікації.

Метод. У роботі розроблено метод генерування і класифікації гістологічних зображень. Цей метод базується на основі використання CNN і GAN. Для підвищення точності класифікації початкову вибірку зображень розширено за допомогою GAN.

Результати. Комп'ютерне дослідження розроблено методу генерування і класифікації зображень проводилося на основі dataset, який знаходиться платформі Zenodo . Зображення отримано на основі світлової мікроскопії. Dataset містить три класи G1, G2,G3 гістологічних зображень раку молочної залози. На основі розробленого методу отримано точність класифікації зображень 96% . Це краща точність класифікації порівняно з існуючим моделями типу AlexNet, LeNet5 і VGG16 . Програмний модуль може буди інтегрований у CAD .

Висновки. Розроблений метод генерування і класифікації зображень є основою програмного модуля. Програмний модуль може бути інтегрований у CAD.

КЛЮЧОВІ СЛОВА: система автоматизованої діагностики, рак молочної залози, глибокі нейронні мережі, генеративні змагальні мережі, згорткові нейронні мережі.

ЛІТЕРАТУРА

1. Breast cancer burden in EU-27. [Electronic resource]. – Access mode: https://ecis.jrc.ec.europa.eu/pdf/Breast_cancer_factsheet-Oct_2020.pdf
2. Cancer Facts for Women. American Cancer Society. [Electronic resource]. – Access mode: https://www.cancer.org/cancer/risk-prevention/understanding-cancer-risk/cancer-facts/cancer-facts-for-women.html
3. Cancer in Ukraine 2021–2022: Incidence, mortality, prevalence and other relevant statistics. Bulletin of the National Cancer Registry of Ukraine № 24, 2021–2022. [Electronic resource]. – Access mode: http://www.ncru.inf.ua/publications/BULL_24/PDF_E/bull_eng_24.pdf
4. Computer diagnostic tools based on biomedical image analysis / [O. Berezsky, O. Pitsun, S. Verbovyy et al.] // 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM), 21–25 Feb. 2017 – Lviv, Ukraine, 2017. – P. 388–391.
5. Aswathy M. A. Detection of breast cancer on digital histopathology images: Present status and future possibilities / M. A. Aswathy, M. Jagannath // Informatics in Medicine Unlocked. – 2017. – Vol. 8. – P. 74–79. DOI: 10.1016/j.imu.2016.11.001
6. GC-EnC: A Copula based ensemble of CNNs for malignancy identification in breast histopathology and cytology images/ [S. Dey, S. Mitra, S. Chakrabortyc et al.] // Computers in Biology and Medicine. – 2023. – Vol. 152. – P. 106329. DOI: 10.1016/j.compbiomed.2022.106329
7. MTRRE-Net: A deep learning model for detection of breast cancer from histopathological images / [S. Chattopadhyay, A. Dey, P. K. Singh et al.] // Computers in Biology and Medicine. – 2022. – Vol. 150. – P. 106155. DOI: 10.1016/j.compbiomed.2022.106155

DOI 10.15588/1607-3274-2023-4-8

8. Shihabuddin A. R. Multi CNN based automatic detection of mitotic nuclei in breast histopathological images / A. R. Shihabuddin, K. S. Sabeena // Computers in Biology and Medicine. – 2023. – Vol. 158. – P. 106815. DOI: 10.1016/j.compbiomed.2023.106815
9. Vo D. M. Classification of breast cancer histology images using incremental boosting convolution networks / D. M. Vo, N.-Q. Nguyen, S.-W. Lee // Information Sciences. – 2019. – Vol. 482. – P. 123–138. DOI: 10.1016/j.ins.2018.12.089
10. Comparative assessment of CNN architectures for classification of breast FNAC images / [A. R. Saikia, K. Bora, L. B. Mahanta et al.] // Tissue Cell. – 2019. – Vol.57. – P. 8–14. DOI: 10.1016/j.tice.2019.02.001
11. Dogar G. M. Attention augmented distance regression and classification network for nuclei instance segmentation and type classification in histology images / G. M. Dogar, M. Shahzad, M. M. Fraz // Biomedical Signal Processing and Control. – 2023. – Vol. 79. – P. 104199. DOI: 10.1016/j.bspc.2022.104199
12. Breast cancer detection from histopathology images using modified residual neural networks/ [V. Gupta, M. Vasudev, A. Doegar et al.] // Biocybernetics and biomedical engineering. – 2021. – Vol. 41. – P. 1272–1287. DOI: 10.1016/j.bbe.2021.08.011
13. Karthik R. Classification of breast cancer from histopathology images using an ensemble of deep multiscale networks / R. Karthik, R. Menaka, M. V. Siddharth // Biocybernetics and biomedical engineering. – 2022. – Vol. 42. – P. 963–976. DOI: 10.1016/j.bbe.2022.07.006
14. Decision-making support system for diagnosis of oncopathologies by histological images. / [A. Dovbysh, I. Shelehov, A. Romaniuk et al.] // Journal of Pathology Informatics. – 2023. – Vol. 14. – P. 100193. DOI: 10.1016/j.jpi.2023.100193
15. Patch-based system for Classification of Breast Histology images usingdeep learning / [K. Roy, D. Banik, D. Bhattacharjee et al.] // Computerized Medical Imaging and Graphics. – 2019. – Vol. 71. – P. 90–103. DOI: 10.1016/j.compmedimag.2018.11.003
16. Improved breast cancer histological grading using deep learning / [Y. Wang, B. Acs, S. Robertson et al.] // Annals of oncology. – 2021. – Vol. 33, № 1. – P. 89–98. DOI: 10.1016/j.annonc.2021.09.007
17. Fractal Neural Network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images / [Guilherme Freire Roberto, Alessandra Lumini, Leandro Alves Neves et al.] // Expert Systems With Applications. – 2021. – Vol. 166. – P. 114103. DOI: 10.1016/j.eswa.2020.114103
18. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review / [Essam H. Houssein, Marwa M. Emama, Abdelmgeid A. Ali] // Expert Systems with Applications. – 2021. – Vol. 167. – P. 114161. DOI:10.1016/j.eswa.2020.114161
19. Majumdar S. Gamma function based ensemble of CNN models for breast cancer detection in histopathology images / Samriddha Majumdar, Payel Pramanik, Ram Sarkar // Expert Systems With Applications. – 2023. – Vol. 213. – P. 119022. DOI: 10.1016/j.eswa.2022.119022
20. Breast cancer histopathological image classification using a hybrid deep neural network / [Rui Yana, Fei Ren, Zihao Wangb et al.] // Methods. – 2020. – Vol. 173. – P. 52–60. DOI: 10.1016/j.ymeth.2019.06.014
21. SECS: An effective CNN joint construction strategy for breast cancer histopathological image classification / [Dianzhi Yu , Jianwu Lin, Tengbao Cao et al.] // Journal of King Saud University – Computer and Information Sciences. – 2023. – Vol. 35. – P. 810–820. DOI: 10.1016/j.jksuci.2023.01.017
22. BACH: Grand challenge on breast cancer histology images / [G. Aresta, T. Araújo, S. Kwok et al.] // Medical Image Analysis. – 2019. – Vol. 56. – P. 122–139. DOI: 10.1016/j.media.2019.05.010
23. Mitotic nuclei analysis in breast cancer histopathology images using deep ensemble classifier / [A. Sohail, A. Khan, H. Nisar et al.] // Medical Image Analysis. – 2021. – Vol. 72. – P. 102121. DOI: 10.1016/j.media.2021.102121
24. Wahab N. Multifaceted fused-CNN based scoring of breast cancer whole-slide histopathology images / Noorul Wahab, Asifullah Khan // Applied Soft Computing Journal. – 2020. – Vol. 97. – P. 106808. DOI: 10.1016/j.asoc.2020.106808
25. Novel architecture with selected feature vector for effective classification of mitotic and non-mitotic cells in breast cancer histology images/ [M. U. Rehman, S. Akhtar, M. Zakwan et al.] // Biomedical Signal Processing and Control. – 2022. – Vol. 71. – P. 103212. DOI:10.1016/j.bspc.2021.103212
26. Bhausaheb D. P. Shuffled shepherd deer hunting optimization based deep neural network for breast cancer classification using breast histopathology images / D. P. Bhausaheb, K. L. Kashyap // Biomedical Signal Processing and Control. – 2023. – Vol. 83. – P. 104570. DOI: 10.1016/j.bspc.2023.104570
27. Guptaa K. Analysis of Histopathological Images for Prediction of Breast Cancer Using Traditional Classifiers with Pre-Trained CNN / K. Guptaa, N. Chawlaa // Procedia Computer Science. – 2020. – №167. – P. 878–889. DOI: 10.1016/j.procs.2020.03.427
28. Yadav R. K. Diagnosis of Breast Cancer using Machine Learning Techniques – A Survey / R. K. Yadav, P. Singh, P. Kashtriya // Procedia Computer Science. – 2023. – Vol. 218. – P. 1434–1443. DOI: 10.1016/j.procs.2023.01.122
29. Dabeer S. Cancer diagnosis in histopathological image: CNN based approach / S. Dabeer, M. M. Khan, S. Islam // Informatics in Medicine Unlocked. – 2019. – Vol. 16. – P. 100231. DOI: 10.1016/j.imu.2019.100231
30. Kate V. A 3 Tier CNN model with deep discriminative feature extraction for discovering malignant growth in multi-scale histopathology images / V. Kate, P. Shukla // Informatics in Medicine Unlocked. – 2021. – Vol. 24. – P. 100616. DOI: 10.1016/j.imu.2021.100616
31. Shallu S. Breast cancer histology images classification: Training from scratch or transfer learning? / S. Shallu, R. Mehra // ICT Express. – 2018. – Vol. 4. – P. 247–254. DOI: 10.1016/j.icte.2018.10.007
32. Desai M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN) / Meha Desai, Manan Shah // Clinical eHealth. – 2021. – Vol. 4. – P. 1–11. DOI: 10.1016/j.ceh.2020.11.002
33. Asadi B. Efficient breast cancer detection via cascade deep learning network / B. Asadi, Q. Memon // International Journal of Intelligent Networks. – 2023. – Vol. 4. – P. 46–52. DOI: 10.1016/j.ijin.2023.02.001

DOI 10.15588/1607-3274-2023-4-8

34. Improved multi-classification of breast cancer histopathological images using handcrafted features and deep neural network (dense layer) / [A. A. Joseph, M. Abdullahi, S. B. Junaidu et al.] // Intelligent Systems with Applications. – 2022. – Vol. 14. – P. 200066. DOI: 10.1016/j.iswa.2022.200066
35. Chaudhury S. A BERT encoding with Recurrent Neural Network and Long-Short Term Memory for breast cancer image classification / S. Chaudhury, K. Sau // Decision Analytics Journal. – 2023. – Vol. 6. – P.100177. DOI: 10.1016/j.dajour.2023.100177
36. Method and software component model for skin disease diagnosis / [V. M. Lovkin, S. A. Subbotin, A. O. Oliinyk et al.] // Radio Electronics, Computer Science, Control. – 2023. – № 1. – P. 40. DOI: 10.15588/1607-3274-2023-1-4
37. Berezsky O. M. Design of computer systems for biomedical image analysis / [O. M. Berezsky, K. M. Berezska, G. M. Melnyk et al.] // Proceedings of the X th International Conference «The Experience of Designing and Application of CAD Systems in Microelectronics» CADSM 2009, 24–28 February 2009, Lviv-Polyana, Ukraine. – Lviv: Publishing House Vezha&CoC. – P.186–191.
38. Berezsky O. The intelligent system for diagnosing breast cancers based on image analysis / O. Berezsky, T. Datsko, S. Verbovyy // Proceedings of Information Technologies in Innovation Business (ITIB). Kharkiv: Ukraine. 7–9 October, 2015. p. 27–30. DOI: 10.1109/ITIB.2015.7355067.
39. Fuzzy system diagnosing of precancerous and cancerous conditions of the breast cancer / [O. Berezsky, S. Verbovyy, L. Dubchak et al.] // Proceedings of the XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT'2016). Lviv: Ukraine. 6–10 September, 2016. – P. 200–203. DOI: 10.1109/STC-CSIT.2016.7589906.
40. Berezsky O. Access Distribution in Automated Microscopy System / O. Berezsky, L. Dubchak, O. Pitsun // Proceedings of the 14 th International Conference «The Experience of Designing and Application of CAD Systems in Microelectronics» CADSM 2017, 21–25 February 2017, Lviv, Ukraine. – Lviv, 2017. – P. 241–243. DOI: 10.1109/CADSM.2017.7916125
41. Berezsky O. Hybrid Intelligent information techology for biomedical image processing / O. Berezsky, S. Verbovyy, O. Pitsun // Proceedings of the IEEE International Conference "Computer Science and Information Technologies" CSIT'2018. Lviv: Ukraine. 11–14 September, 2018. – Lviv, 2018. – P. 420–423. DOI: 10.1109/STC-CSIT.2018.8526711.
42. Modern automated microscopy systems in oncology / [O. Berezsky, O. Pitsun, N. Batryn, et al.] // Proceedings of the 1st International Workshop on Informatics & Data-Driven Medicine, Lviv, Ukraine, 28–30 november 2018. – Lviv, 2018. – P. 311–325.
43. Berezsky O. M. Evaluation methods of image segmentation quality. / O. M. Berezsky, O. Y. Pitsun // Radio Electronics, Computer Science, Control. – 2018. – Vol. 1. – P. 119–128. DOI:10.15588/1607-3274-2018-1-14
44. Fuzzy System For Breast Disease Diagnosing Based On Image Analysis / [O. Berezsky, L. Dubchak, N. Batryn et al.] // Proceedings of the II International Workshop Informatics & Data-Driven Medicine (IDDM 2019). Lviv, Ukraineб 11–13 November, 2019. – Lviv, 2019. – P. 69–83.
45. An Approach toward Automatic Specifics Diagnosis of Breast Cancer Based on an Immunohistochemical Image. / [O. Berezsky, O. Pitsun, G. Melnyk et al.] // Journal of Imaging. – 2023. – Vol. 9, № 1. – P. 12. DOI: 10.3390/jimaging9010012
46. Berezsky O. Cytological and histological images of breast cancer. / Oleh Berezsky, Tamara Datsko, Grygoriy Melnyk. [Electronic resource] – Access mode: https://zenodo.org/records/7890874. DOI: 10.5281/zenodo.7890874
47. Intresoft Consulting. General Data Protection Regulation (GDPR). Recital 26. [Electronic resource]. – Access mode: https://gdpr-info.eu/recitals/no-26/
48. Liashchynskyi P. Rudi. Lightweight image converter and dataset augmentor / Petro Liashchynskyi [Electronic resource]. – Access mode: https://github.com/liashchynskyi/rudi
49. Deep Residual Learning for Image Recognition / [Kaiming He, Xiangyu Zhang, Shaoqing Ren et al.] – 2015. [Electronic resource]. – Access mode: https://arxiv.org/pdf/1512.03385.pdf
50. Lim J.H. Geometric GAN / Jae Hyun Lim, Jong Chul Ye. – 2017. [Electronic resource]. – Access mode: https://arxiv.org/pdf/1705.02894v2.pdf
51. Improved techniques for training GANs. / [T. Salimans, I. Goodfellow, W. Zaremba et al.] // Advances in neural information processing systems: Annual Conference on Neural Information Processing Systems 2016, December 5–10, 2016, Barcelona, Spain – Barcelona, 2016. – P: 2226–2234 DOI: 10.48550/arXiv.1606.03498
52. GANs trained by a two time-scale update rule converge to a local nash equilibrium / M. Heusel, H. Ramsauer, T. Unterthiner et al.] // Proceedings of the 31st International Conference on Neural Information Processing Systems – NY, United States: Curran Associates Inc., 2017 – P. 6629–6640. Access mode: https://dl.acm.org/doi/10.5555/3295222.3295408
53. Classification Of Breast Cancer Histology Images Using ALEXNET. In: Campilho, A., Karray, F., ter Haar Romeny, B. (eds) Image Analysis and Recognition. ICIAR 2018. / [W. Nawaz, S. Ahmed, A. Tahir et al.] // Lecture Notes in Computer Science. – 2018. – Vol. 10882. – P. 869–876. Springer, Cham. DOI: 10.1007/978-3-319-93000-8_99
54. Verdhan V. Image Classification Using LeNet. In: Computer Vision Using Deep Learning. / V. Verdhan. – Apress, Berkeley, CA., 2021. – P. 67–101. DOI: 10.1007/978-1-4842-6616-8_3
55. Tammina S. Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images / Srikanth Tammina // International Journal of Scientific and Research Publications. – 2019. – Vol. 9, Issue 10. – P. 143–150. DOI: 10.29322/IJSRP.9.10.2019.p9420

DOI 10.15588/1607-3274-2023-4-8

UDC 004.93

ENSEMBLE OF ADAPTIVE PREDICTORS FOR MULTIVARIATE NONSTATIONARY SEQUENCES AND ITS ONLINE LEARNING

Bodyanskiy Ye. V. – Dr. Sc., Professor, Professor of the Artificial Intelligence Department, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Lipianina-Honcharenko Kh. V. – PhD, Associate Professor, Associate Professor of the Department for Information Computer Systems and Control, West Ukrainian National University, Ternopil, Ukraine.

Sachenko A. O. – Dr. Sc., Professor, Director of the Research Institute for Intelligent Computer Systems, West Ukrainian National University, Ternopil, Ukraine.

ABSTRACT

Context. In this research, we explore an ensemble of metamodels that utilizes multivariate signals to generate forecasts. The ensemble includes various traditional forecasting models such as multivariate regression, exponential smoothing, ARIMAX, as well as nonlinear structures based on artificial neural networks, ranging from simple feedforward networks to deep architectures like LSTM and transformers.

Objective. A goal of this research is to develop an effective method for combining forecasts from multiple models forming metamodels to create a unified forecast that surpasses the accuracy of individual models. We are aimed to investigate the effectiveness of the proposed ensemble in the context of forecasting tasks with nonstationary signals.

Method. The proposed ensemble of metamodels employs the method of Lagrange multipliers to estimate the parameters of the metamodel. The Kuhn-Tucker system of equations is solved to obtain unbiased estimates using the least squares method. Additionally, we introduce a recurrent form of the least squares algorithm for adaptive processing of nonstationary signals.

Results. The evaluation of the proposed ensemble method is conducted on a dataset of time series. Metamodels formed by combining various individual models demonstrate improved forecast accuracy compared to individual models. The approach shows effectiveness in capturing nonstationary patterns and enhancing overall forecasting accuracy.

Conclusions. The ensemble of metamodels, which utilizes multivariate signals for forecast generation, offers a promising approach to achieve better forecasting accuracy. By combining diverse models, the ensemble exhibits robustness to nonstationarity and improves the reliability of forecasts.

KEYWORDS: ensemble, metamodels, boosting, bagging, multivariate signals, nonstationarity, forecasting.

ABBREVIATIONS

MP – Multi-dimensional Predictors;

ARIMAX – AutoRegressive Integrated Moving Average with eXogenous inputs;

LSTM – Long Short-Term Memory;

RF – Random Forest;

NB – Naive Bayes;

SVM – Support Vector Machine;

LR – Logistic Regression;

AdaBoost – Adaptive Boosting;

AUC – Area Under the Curve;

DT – Decision Tree;

MIMO – Multiple-Input Multiple-Output;

LSTM – Long Short-Term Memory;

SMOTE – Synthetic Minority Over-sampling Technique;

ADASYN – Adaptive Synthetic Sampling.

NOMENCLATURE

$x(\tau)$ – multivariate signal with time index τ;

MPj – member of the ensemble of models with index j;

$\hat{x}_j(T)$ – estimate obtained at the output of member MPj of the ensemble;

$x^*(\tau)$ – combined forecast of the metamodel at time τ;

c – metamodel parameters, a vector of estimates forming the combined forecast;

λ – Lagrange multiplier used in optimization;

$D(T)$ – matrix used for estimating metamo del parameters;

$d(T)$ – vector that incorporates estimates at the previous time step;

α – regularization parameter that ensures the method's operation for nonstationary data;

s – size of the "sliding window", determining the number of recent observations considered in the estimation;

$v^2(T)$ – squared error of the estimate at the last time step.

INTRODUCTION

Forecasting multivariate nonstationary signals is a relevant and challenging problem in various domains. To achieve reliable and accurate results, different forecasting models such as ARIMAX, LSTM, SVM, and many others are used.

In this work, we consider the ensemble of metamodels method for forecasting, which is based on combining forecasts from different forecasting models. The metamodel helps to merge information from various models to improve forecasting accuracy and ensure more robust results.

DOI 10.15588/1607-3274-2023-4-9

The object of study is an ensemble of multivariate predictors used for forecasting multivariate signals.

The subject of study is the ensemble of metamodels method for combining forecasts from different forecasting models to improve forecasting accuracy based on nonstationary signals.

The purpose of the work of this research is to develop and evaluate the effective method based on ensemble of metamodels for forecasting multivariate nonstationary signals. We aim to investigate how combining forecasts from individual models can enhance the quality of forecasting and provide more reliable results.

Forecasting tasks based on multivariate nonstationary signals find broad applications in various fields, including finance, economics, medicine, and engineering. An efficient ensemble of metamodels can become a powerful tool for addressing these tasks and ensuring accurate and reliable forecasts.

1 PROBLEM STATEMENT

Let's consider an ensemble of multivariate predictors, $MP1,..MPj,..MPh$, each of which processes the same multivariate signal $x(\tau)=(x_1(\tau)\ldots x_i(\tau)\ldots)^{\mathrm{T}}, \tau=1,2,\ldots,T.$ The estimate that appears at the output of each member of the ensemble will be denoted as $\hat{x}_j(\tau), j=1,2\ldots h.$ It is worth noting that traditional forecasting models based on multivariate regression, exponential smoothing, ARIMAXs-MIMO models (Box-Jenkins), as well as nonlinear structures based on artificial neural networks, ranging from simple shallow recurrent networks to deep architectures like LSTM or transformers, can be used as members of the ensemble.

The estimates $\hat{x}_j(\tau)$ are input to the metamodel, which forms the combined forecast of the metamodel:

$$x^*(\tau)=\sum_{j=1}^{h} c_j\hat{x}_j(\tau)=\hat{x}(\tau)c,$$

here

$c=(c_1,..,c_j,..,c_h)^{\tau}, \hat{x}(\tau)=(\hat{x}_1(\tau),..,\hat{x}_j(\tau),..,\hat{x}_h(\tau))-$

$-(n\times h)-$ matrix formed by the signals at the outputs of individual models, where metamodel parameters satisfy the condition of unbiasedness:

$$\sum_{j=1}^{h} c_j = c^T E_h = 1,$$

here $E_h-(h\times 1)$ – is a vector formed by ones.

To solve this problem, methods of Lagrange multipliers are used, leading to the estimation of the metamodel parameters c defined in a recursive form. The case where estimation is carried out based on a "sliding window" of size s is also considered, allowing for consideration of only the last s observations from the training dataset. To choose the best metamodel, a second-level metamodel is introduced, which processes the outputs of the first-level metamodels using a meta-algorithm.

Thus, the formal mathematical formulation of the problem involves defining an ensemble of predictors, computing estimates $\hat{x}_j(\tau)$ for each member of the ensemble, constructing a combined forecast $x^*(\tau)$, determining the parameters of the metamodel c using the method of Lagrange multipliers, and the ability to work with different "sliding window" sizes and second-level metamodels for selecting the optimal solution.

2 REVIEW OF THE LITERATURE

This approach has gained the most popularity in classification tasks, such as image recognition, where the AdaBoost algorithm and its various modifications [1–8] are very popular. The underlying idea of this algorithm is stacked generalization, where the results of each member of the ensemble (stack) are combined within a metamodel, whose parameters are tuned using metalearning procedures. Typically, this involves weighted averaging, where each member of the ensemble (committee) is assigned a weight obtained through optimization of the adopted learning criterion.

The foundation of AdaBoost lies in the ideas of Bayesian estimation, logistic regression, and support vector machines. Interestingly, these ideas also form the basis of several artificial neural networks, where ensemble approaches [9–11] are also utilized to obtain optimal forecasts. In this case, weights for each member of the ensemble are estimated using an optimization procedure implemented in batch mode, making the use of known approaches for solving Data Stream Mining tasks practically impossible. Recurrent procedures for metamodel parameter tuning were introduced in [12, 13], generalizing the output signals of predictor neural networks based on the optimization of the standard least squares criterion under certain constraints. Although these procedures are designed for online evaluation, they are not adapted to work with nonstationary time series, where parameters change unpredictably at any moment.

Therefore, it is worthwhile to introduce adaptive recurrent metalearning procedures for a generalizing metamodel that combines the output signals of a neural predictor ensemble, each of which can have its own architecture and its own algorithm for tuning-learning its synaptic weights.

3 MATERIALS AND METHODS

Metamodel parameters (vector of estimates c) can be determined using the classical method of Lagrange multipliers, for which the Lagrange function is introduced:

$$L(c,\lambda)=\mathrm{Sp}(V^T(T)V(T))+\lambda(c^T E_h-1)=$$
$$=\mathrm{Sp}(X(T)-\hat{X}(T)E_{nn}\otimes c)^{\mathrm{T}}(X(T)-\hat{X}(T)E_{nn}\otimes c)+$$
$$+\lambda(c^T E_h-1)=$$
$$\sum_{\tau=1}\left\|x(\tau)-\hat{x}(\tau)c\right\|^2+\alpha(c^T(\tau)E_h-1),$$

DOI 10.15588/1607-3274-2023-4-9

where

$$x(T)=\begin{pmatrix} x^T(1) \\ x^T(\tau) \\ x^T(T) \end{pmatrix},$$

$$\hat{x}(T)=\begin{pmatrix} \hat{x}_1^T(1) & \hat{x}_2^T(1) & \dots & \hat{x}_h^T(1) \\ \dots & \dots & \dots & \dots \\ \hat{x}_1^T(\tau) & \hat{x}_2^T(\tau) & \dots & \hat{x}_h^T(\tau) \\ \dots & \dots & \dots & \dots \\ \hat{x}_1^T(T) & \hat{x}_2^T(T) & \dots & \hat{x}_h^T(T) \end{pmatrix}.$$

$$V(T)=X(T)-\hat{X}(T)\ E_{nn}\otimes \text{c}.$$

– the matrix of errors is the training sample, $E_{nn}-(n\times n)$ is the identity matrix, $\otimes$ – denotes the tensor product, Sp(•) – denotes the trace of a matrix, λ – Lagrange multiplier.

Solving the Kuhn-Tucker system of equations leads to the estimate [12]:

$$c=c^*+D(T)\ \frac{1-E_h^T\ c^*}{E_h^T D(T)E_h}E_h, \qquad (1)$$

where

$$D(T)=(\sum_{\tau=1}\hat{x}^T(\tau)\hat{x}(\tau))^{-1}, \qquad (2)$$

$$c^*=D(T)\sum_{\tau=1}\hat{x}^T(\tau)x(\tau)=D(T)d(T)$$

the regular estimate of the standard least squares method.

In [13], the optimality of this estimate is proven over the entire training sample, meaning that the output of the metamodels $x^*(\tau)$, does not compromise accuracy compared to any of the individual ensemble models $\hat{x}_j(\tau)$ in the interval from $\tau=1$ to $\tau=T$.

Equations (1) and (2) can be easily rewritten in a recursive form similar to the recursive least squares method:

$$\begin{aligned} D(T+1)&=D(T)-D(T)\ \hat{x}^T\ (T+1)\times \\ &\times(E_{nn}+\hat{x}(T+1)D(T)\ x^T\ (T+1))^{-1}\times \\ &\times\hat{x}(T+1)D(T), \\ d(T+1)&=d(T)+\hat{x}^T(T+1)\ \hat{x}(T+1), \\ c^*\ (T+1)&=D(T+1)d(T+1), \\ c(T+1)&=c^*\ (T+1)+D(T+1)\times \\ &\times(E_h^T D(T+1)E_h)^{-1}(1-E_h^T c^*(T+))E_h. \end{aligned} \qquad (3)$$

The use of the least squares criterion is associated with the assumption of stationarity in the processed sequences, as all observations from $x(1)$ to $x(T)$ are assigned equal weights. Since we assume non-stationarity in the controlled signals, including abrupt changes in the forecasting model, the estimates based on the least squares method are found to be inefficient. In such situations, more suitable predictors are those synthesized using "sliding window" estimation procedures that consider not the entire training sample but only the last s (window size) observations from $x(T-s+1)$ to $x(T)$ When the value $x(T+1)$ arrives, the observation $x(T-s+1)$ is excluded from consideration, and the estimate is calculated over the interval from $x(T-s+2)$ to $x(T+1)$. In this case, the procedure takes the form:

$$\begin{aligned} D(T+1)&=D(T)-D(T)\ \hat{x}^T\ (T+1)\times \\ &\times(E_{nn}+\hat{x}(T+1)D(T)\ \hat{x}^T\ (T+1))^{-1}\ \hat{x}(T+1)D(T), \\ D^S\ (T+1)&=D(T+1)+D(T+1)\ \hat{x}^T\ (T+S-1)\times \\ &\times(E_{nn}-\hat{x}(T+S-1)D(T+1)\ \hat{x}^T\ (T+S-1))^{-1}\times \\ &\times\hat{x}(T+S-1)D(T+1), \\ d^S(T+1)&=d(T)+\hat{x}^T(T+1)\ \hat{x}(T+1)- \\ &-\hat{x}^T(T+S-1)\ \hat{x}(T+S-1), \\ &c^{*S}\ (T+1) \\ &=D^S\ (T+1)\ d^S\ (T+1). \end{aligned} \qquad (4)$$

An interesting situation arises when the estimation is performed under the assumption of s=1, meaning that the optimization criterion (learning) is based on the square error of estimation at the last observation timestep.

$$v^2\ (T)=\left\|x(T)-\hat{x}(T)c\right\|^2.$$

In this case, the procedures (1), (2), and (4) take on a simple form:

$$\begin{aligned} D(T+1)&=(\hat{x}^T\ (T+1)\ \hat{x}(T+1)+E_{hh})^{-1}, \\ d^s\ (T+1)&=\hat{x}^T\ (T+1)\ \hat{x}(T+1), \\ c^{*S}(E+1)&=D^\alpha\ (T+1)\ d^s\ (T+1), \\ c^*(T+1)&=C^{*\alpha}\ (T+1)+D^1\ (T+1)\times \\ &\times(E_h^T\ D(T+1)E_h\ )^{-1}\ (1-E_h^T\ c^{*1}\ (T+1))\ E_h. \end{aligned} \qquad (5)$$

This is a generalization for the case under consideration, an adaptive identification algorithm of Kachmazh-Uidro-Hoff, where $\alpha>0$ is a regularization parameter that ensures the possibility of inversion during the calculation of $D(T+1)$.

The most challenging issue here remains the choice of the "window" size, s, which is usually done based on purely empirical considerations since the nature of possi-

DOI 10.15588/1607-3274-2023-4-9

ble changes in the controlled signal $x(\tau)$ is unknown a priori. In this case, it is advisable to use not a single metamodel but a set of such structures built at different values of the "sliding window".

To select the best metamodel from such a set, it is appropriate to introduce metamodels of the second level that process the outputs of first-level metamodels using the metaalgorithm (3), covering the entire training sample at $\tau = 1,2,\ldots,T,T+1,\ldots$.

The method of constructing an ensemble of metamodels that use multidimensional signals for forecasting can be presented in the following steps 1–9 (Figure 1):

Step 1: Data Collection: Gather a large dataset of multidimensional data to be used in the analysis.

Step 2: Input Data Formation: The outputs (predictions) of each predictor are used as inputs for the metamodel.

Step 3: Data Processing: Each of the multidimensional predictors in the ensemble processes the same input data in various ways. Each predictor may include different machine learning methods, such as neural networks, support vector machines, gradient boosting, etc.

Step 4: Sliding Window Evaluation of Random Values: Model parameters are re-estimated for each new data point using only the last s observations. This ensures that the model is continuously updated with the most recent data.

Step 5: Metamodel Synthesis: Develop metamodels that use the method of Lagrange multipliers to determine their parameters. This means that the metamodel utilizes weights from different forecasts to form a single forecast. The weights of these forecasts are determined through the optimization of the Lagrange function.

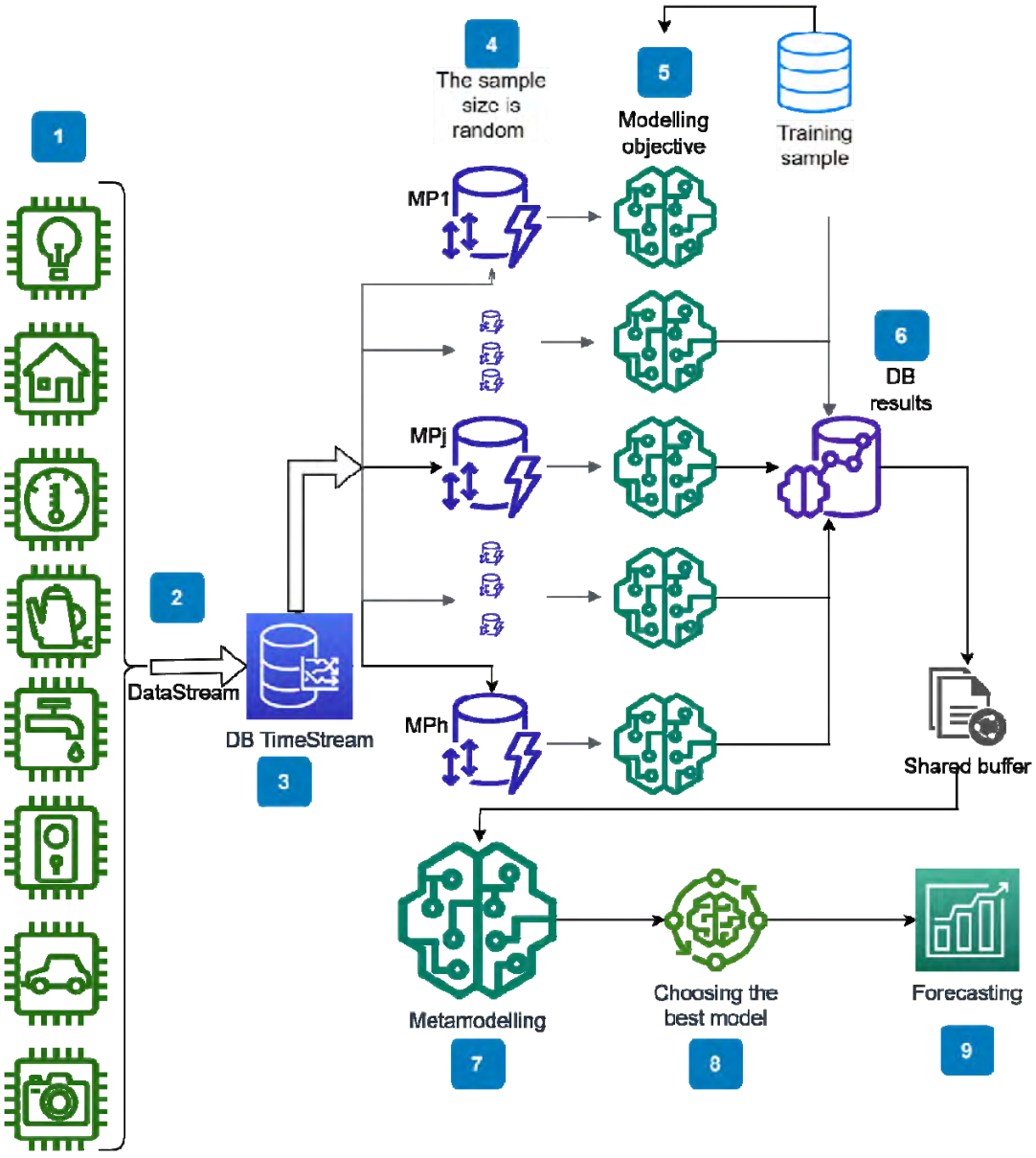


Figure 1 – Structural and Logical Diagram of the Method

DOI 10.15588/1607-3274-2023-4-9

Step 6: Base Results Formation: Metamodel estimates are stored in a database for further analysis.

Step 7: Synthesis of Second-level Metamodel: Develop another metamodel that processes the outputs of the first-level metamodels. This can help gather information from different metamodels and make a more accurate forecast.

Step 8: Selection of the Best Metamodel: The second-level metamodel is used to select the best metamodel among the ensemble based on their performances.

Step 9: Forecasting: The final metamodel is used to produce forecasts based on the input data. This allows using a single, optimally weighted forecast instead of independent forecasts from each predictor.

This method employs the model ensembles to work with multidimensional data and produce forecasts based on a combination of predictions from each predictor.

5 EXPERIMENTS

In the previous study [14], an intelligent method for identifying fraudulent websites was proposed. This method was implemented using various machine learning classification methods, including Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT). Additionally, each classification method was modeled using different approaches, including addressing imbalanced data, undersampling, oversampling, SMOTE, and ADASYN.

5 RESULTS

The method was applied to a dataset of websites operating in Ukraine, consisting of 67 sites, out of which 45% were identified as fraudulent. The results showed that the DTADASYN and RF Oversampling models achieved the highest accuracy (1.0), AUC (1.0), precision (1.0), recall (1.0), and F1-score (1.0).

Using the same intelligent method for an updated dataset consisting of 1039 websites, of which 68% were identified as fraudulent, slightly different results were obtained (Table 1). The SVM Undersampling model showed an accuracy of 0.93, AUC of 0.87, precision of 0.88, recall of 0.78, and F1-score of 0.82. The KNN Undersampling model demonstrated an accuracy of 0.90, AUC of 0.94, precision of 0.69, recall of 1.0, and F1-score of 0.82. These results indicate that although accuracy and other metrics may vary depending on the dataset and methods used, the proposed intelligent method still achieves high accuracy in identifying fraudulent websites.

The proposed ensemble metamodel, utilizing multidimensional signals for forecasting, was implemented. In this case, the metamodel was constructed based on the predictions of logistic regression (LR), decision tree (DT), K-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), and naive Bayes (NB) models, which were selected from the previous study [14].

Table 1 – Modeling Results without Metamodel for the New Dataset

	Model	Accuracy_Test	AUC_Test	PrecisionScore_Test	RecallScore_Test	F1Score_Test
1	LR Undersampling	0.939394	0.950000	0.866667	1.000000	0.928571
2	LR Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
3	LR SMOTE	0.939394	0.950000	0.866667	1.000000	0.928571
5	DT Undersampling	0.939394	0.950000	0.866667	1.000000	0.928571
10	KNN Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
11	KNN SMOTE	0.939394	0.950000	0.866667	1.000000	0.928571
12	SVM imbalance	0.939394	0.950000	0.866667	1.000000	0.928571
13	SVM Undersampling	0.939394	0.950000	0.866667	1.000000	0.928571
14	SVM Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
15	SVM SMOTE	0.939394	0.950000	0.866667	1.000000	0.928571
16	RF imbalance	0.939394	0.950000	0.866667	1.000000	0.928571
17	RF Undersampling	0.939394	0.950000	0.866667	1.000000	0.928571
18	RF Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
19	RF SMOTE	0.939394	0.950000	0.866667	1.000000	0.928571
22	NB Oversampling	0.939394	0.950000	0.866667	1.000000	0.928571
0	LR imbalance	0.909091	0.925000	0.812500	1.000000	0.896552
8	KNN imbalance	0.909091	0.925000	0.812500	1.000000	0.896552
20	NB imbalance	0.909091	0.925000	0.812500	1.000000	0.896552
21	NB Undersampling	0.909091	0.925000	0.812500	1.000000	0.896552
23	NB SMOTE	0.909091	0.925000	0.812500	1.000000	0.896552
4	DT imbalance	0.909091	0.911538	0.857143	0.923077	0.888889
7	DT SMOTE	0.909091	0.911538	0.857143	0.923077	0.888889
9	KNN Undersampling	0.909091	0.884615	1.000000	0.769231	0.869565
6	DT Oversampling	0.878788	0.886538	0.800000	0.923077	0.857143

DOI 10.15588/1607-3274-2023-4-9

The metamodel was built using AdaBoostClassifier, an adaptive boosting algorithm that combines several weak models to create a strong one.

The results of the metamodel were as follows (Fig. 3):

– Accuracy: 0.98. This indicates that the metamodel correctly classified 98% of the websites.

– Recall: For class 0 (non-fraudulent websites), it was 0.97, and for class 1 (fraudulent websites), it was 1.00. This means that the metamodel identified 97% of non-fraudulent websites and 100% of fraudulent websites.

– F1-score: For class 0, it was 0.98, and for class 1, it was 0.95. The F1-score is the harmonic mean between precision and recall, providing an overall evaluation of the model.

These results demonstrate improvement compared to the previous individual models trained separately. The metamodel delivers more accurate and consistent website classification, making it an effective tool for detecting fraudulent websites.

Next, we examine an example of using the metamodel to forecast the label for the 16th observation in the test dataset (Figure 3). Firstly, we obtain this observation and its true label. The true label for this observation is 0, indicating that the website is not fraudulent. Then, we get the predicted labels for this observation from each model, including logistic regression (LR), decision tree (DT), K-nearest neighbors (KNN), support vector machine (SVM), random forest (RF), and naive Bayes (NB) models.

Meta-Model Performance:

	precision	recall	f1-score
0	1.00	0.97	0.98
1	0.90	1.00	0.95
accuracy			0.98
macro avg	0.95	0.98	0.97
weighted avg	0.98	0.98	0.98

Figure 2 – Metamodeling Results

```
True label: 0
Predicted label from each model: [0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0]
Predicted label from meta-model: 0
```

Figure 3 – Example of Applying the Metamodel to the 16th Row of the Dataset

The predicted labels from these models range from 0 to 1, reflecting different predictions from different models. Finally, we obtain the predicted label from the metamodel for this observation. The metamodel predicts a label of 0, which aligns with the true label. This demonstrates that the metamodel can correctly classify this observation, despite varying predictions from individual models. This result underscores the effectiveness of the metamodel in combining forecasts from different models to improve overall prediction accuracy.

The metamodel exhibited high accuracy in classifying websites, achieving an accuracy of 0.98. This means that the metamodel correctly classified 98% of the websites in the test dataset. Additionally, the metamodel demonstrated high precision (0.95 for class 0 and 0.90 for class 1), recall (0.97 for class 0 and 1.00 for class 1), and F1-score (0.98 for class 0 and 0.95 for class 1). These metrics indicate that the metamodel performed well in classifying both fraudulent and non-fraudulent websites. The example prediction for the 16th observation also showed that the metamodel can accurately classify websites, despite diverse predictions from individual models. This confirms that the metamodel can effectively leverage predictions from different models to enhance the overall prediction accuracy.

Thus, these results confirm that using a metamodel can be an effective approach to improve the accuracy of classification in fraud detection tasks for websites.

6 DISCUSSION

In this study, we investigated the ensemble metamodel approach for forecasting multi-dimensional non-stationary signals. The proposed approach allows us to combine predictions from different forecasting models to obtain more accurate and reliable forecasts based on multiple sources of information.

Firstly, we conducted a literature review and explored various approaches to forecasting multi-dimensional non-stationary signals. Traditional models such as ARIMAX and exponential smoothing may be insufficiently effective in non-stationary conditions. On the other hand, neural networks such as LSTM and transformers exhibit high adaptability and the ability to work with changing conditions, making them attractive candidates for use in metamodel ensembles.

Next, we performed experiments with different forecasting models, such as ARIMAX, LSTM, SVM, Random Forest, etc., and collected their forecasts as input data for the metamodel. Using the method of Lagrange multipliers, we found the optimal parameters for the metamodel to achieve the best forecasting accuracy.

The results of the experiments showed that the proposed ensemble of metamodels indeed helps improve forecast accuracy. The metamodel based on combined forecasts from different models demonstrated higher accuracy compared to individual models. This approach allows for balancing forecasts and reducing the risk of overfitting or underfitting.

DOI 10.15588/1607-3274-2023-4-9

Moreover, we compared various approaches to synthesizing the second-level metamodel and found that utilizing the adaptive Kachmazh-Widrow-Hoff (KWH) identification algorithm helps provide more accurate forecasts based on the forecasts from the first-level models.

Overall, the research results confirmed the effectiveness of the ensemble metamodel method for forecasting multi-dimensional non-stationary signals. Using the ensemble approach helps achieve more accurate results.

CONCLUSIONS

This research addressed the problem of adaptive forecasting of multi-dimensional non-stationary sequences, considering the prior uncertainty regarding their structure, through an ensemble approach. We developed the ensemble metamodel method, where each ensemble member processes predictions from different first-level forecasting models. Then, by collecting the results of individual models' forecasts, we applied a second-level metamodel to obtain the optimal forecast.

The scientific novelty of this study lies in the development and application of ensemble metamodels for forecasting multi-dimensional non-stationary signals. The use of ensembles allows obtaining more accurate and reliable forecasts based on multiple sources of information, reducing the impact of limitations of individual models.

The practical significance of our research is that the proposed approach can be applied in various domains where forecasting multi-dimensional non-stationary signals plays a crucial role. For example, this approach can be used in financial analysis, weather forecasting, medical diagnostics, and other fields where forecast accuracy and reliability are essential.

The conducted research confirms that the proposed ensemble metamodel has high accuracy in detecting fraudulent websites. The metamodel demonstrated high precision in website classification, correctly classifying 98% of websites in the test dataset. This demonstrates that the proposed method can be an effective tool for identifying fraudulent websites and can find practical applications in the field of cybersecurity and combating online fraudulent activities.

Regarding the prospects of this research, further improvement of the method can be achieved by expanding the set of first-level forecasting models and using more sophisticated learning algorithms for the second-level metmodel. Additionally, this approach can be applied to other types of non-stationary signals and forecasting tasks in various domains of science and technology.

ACKNOWLEDGEMENTS

n/a.

REFERENCES

1. Freund Y., Robert E. Schapire A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting [Electronic resource], *Journal of Computer and System Sciences*, 1997, Vol. 55, No. 1, P. 119–139. Mode of access: https://doi.org/10.1006/jcss.1997.1504
2. Schwenk H., Bengio Y. Boosting Neural Networks [Electronic resource], *Neural Computation*, 2000, Vol. 12, No. 8, pp. 1869–1887. Mode of access: https://doi.org/10.1162/089976600300015178 .
3. Jiang W. Process consistency for AdaBoost [Electronic resource], *The Annals of Statistics*, 2003, Vol. 32, No. 1, pp. 13–29. Mode of access: https://doi.org/10.1214/aos/1079120128
4. Li X., Wang L., Sung E. AdaBoost with SVM-based component classifiers [Electronic resource], *Engineering Applications of Artificial Intelligence*, 2008, Vol. 21, No. 5, pp. 785–795. Mode of access: https://doi.org/10.1016/j.engappai.2007.07.001 .
5. Hastie T., Rosset S., Zhu J., & Zou H. Multi-class AdaBoost [Electronic resource], *Statistics and Its Interface*, 2009, Vol. 2, No. 3, pp. 349–360. Mode of access: https://doi.org/10.4310/sii.2009.v2.n3.a8
6. Zhang C., Ma Y. ed. by Ensemble Machine Learning [Electronic resource]. Boston, MA, Springer US, 2012. Mode of access: https://doi.org/10.1007/978-1-4419-9326-7
7. Ying C., Qi-Guang M., Jia-Chen L., & Lin G. Advance and Prospects of AdaBoost Algorithm [Electronic resource], *Acta Automatica Sinica*, 2013, Vol. 39, No. 6, pp. 745–758. Mode of access: https://doi.org/10.1016/s1874-1029(13)60052-x .
8. Zhou Z.-H. Ensemble Learning [Electronic resource], *Machine Learning*. Singapore, 2021, pp. 181–210. Mode of access: https://doi.org/10.1007/978-981-15-1967-3_8 .
9. Hansen L. K., Salamon P. Neural network ensembles [Electronic resource], *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, Vol. 12, No. 10, pp. 993–1001. Mode of access: https://doi.org/10.1109/34.58871
10. Wolpert D. H. Stacked generalization [Electronic resource], *Neural Networks*, 1992, Vol. 5, No. 2, pp. 241–259. Mode of access: https://doi.org/10.1016/s0893-6080(05)80023-1 .
11. Sharkey A. J. C. On Combining Artificial Neural Nets [Electronic resource], *Connection Science*, 1996, Vol. 8, No. 3–4, pp. 299–314. Mode of access: https://doi.org/10.1080/095400996116785
12. Bodyanskiy Y., Otto P., Pliss I., & Popov S. An Optimal Algorithm for Combining Multivariate Forecasts in Hybrid Systems [Electronic resource], *Lecture Notes in Computer Science*. Berlin, Heidelberg, 2003, pp. 967–972. Mode of access: https://doi.org/10.1007/978-3-540-45226-3_132 .
13. Bodyanskiy Y., Popov S. Fuzzy Selection Mechanism for Multimodel Prediction [Electronic resource], *Lecture Notes in Computer Science*. Berlin, Heidelberg, 2004, pp. 772–778. Mode of access: https://doi.org/10.1007/978-3-540-30133-2_101 .
14. Lipianina-Honcharenko K., Lukasevych-Krutnyk I., Butryn-Boka N., Sachenko A., & Grodskyi S. Intelligent Method for Identifying the Fraudulent Online Stores [Electronic resource], *2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*. Kharkiv, Ukraine, 5–7 October 2021, [S. l.], 2021. Mode of access: https://doi.org/10.1109/picst54195.2021.9772195

Received 11.09.2023.
Accepted 15.11.2023.

DOI 10.15588/1607-3274-2023-4-9

УДК 004.93

АНСАМБЛЬ АДАПТИВНИХ ПРЕДИКТОРІВ ДЛЯ БАГАТОВИМІРНИХ НЕСТАЦІОНАРНИХ ПОСЛІДОВНОСТЕЙ ТА ЙОГО ОНЛАЙН-НАВЧАННЯ

Бодянський Є.В. – д-р техн. наук, професор, професор кафедри штучного інтелекту, Харківський національний університет радіоелектроніки, Харків, Україна.

Ліпяніна-Гончаренко Х. В. – канд. техн. наук, доцент, доцент кафедри інформаційно-комп'ютерних систем та управління, Західноукраїнський національний університет, Тернопіль, Україна.

Сачeнко А. О. – д-р техн. наук, професор, директор Науково-дослідного інституту інтелектуальних комп'ютерних систем, Західноукраїнський національний університет, Тернопіль, Україна.

АНОТАЦІЯ

Актуальність. У даному дослідженні ми розглядаємо ансамбль метамоделей, який використовує багатовимірні сигнали для генерації прогнозів. Ансамбль включає різні традиційні моделі прогнозування, такі як багатовимірна регресія, експоненційне згладжування, ARIMAX, а також нелінійні структури на основі штучних нейронних мереж, від простих поверхневих рекурентних мереж до глибоких архітектур, таких як LSTM і трансформери.

Мета роботи. Основною метою цього дослідження є розробка ефективного методу поєднання прогнозів декількох моделей, що утворюють метамоделі, для створення єдиного прогнозу, який перевищує точність окремих моделей. Ми прагнемо дослідити ефективність запропонованого ансамблю в контексті задач прогнозування з нестаціонарними сигналами.

Метод. Запропонований ансамбль метамоделей використовує метод множників Лагранжа для оцінки параметрів метамоделі. Система рівнянь Куна-Таккера розв'язується для отримання незміщених оцінок за допомогою методу найменших квадратів. Крім того, ми вводимо рекурентну форму алгоритму найменших квадратів для адаптивної обробки нестаціонарних сигналів.

Результати. Оцінка запропонованого ансамблю методу здійснюється на наборі даних часових рядів. Метамоделі, утворені шляхом поєднання різних окремих моделей, демонструють покращену точність прогнозу порівняно з індивідуальними моделями. Підхід проявляє ефективність в утриманні нестаціонарних шаблонів та покращенні загальної точності прогнозування.

Висновки. Ансамбль метамоделей, який використовує багатовимірні сигнали для формування прогнозів, пропонує перспективний підхід для досягнення кращої точності прогнозування. Шляхом поєднання різноманітних моделей, ансамбль проявляє стійкість до нестаціонарності та покращує надійність прогнозів.

КЛЮЧОВІ СЛОВА: ансамбль, метамоделі, бустінг, бегінг, багатовимірні сигнали, нестаціонарність, прогнозування.

ЛІТЕРАТУРА

1. Freund Y. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting [Electronic resource] / Yoav Freund, Robert E. Schapire // Journal of Computer and System Sciences. – 1997. – Vol. 55, No. 1. – P. 119–139. Mode of access: https://doi.org/10.1006/jcss.1997.1504
2. Schwenk H. Boosting Neural Networks [Electronic resource] / Holger Schwenk, Yoshua Bengio // Neural Computation. – 2000. – Vol. 12, No. 8. – P. 1869–1887. – Mode of access: https://doi.org/10.1162/089976600300015178 .
3. Jiang W. Process consistency for AdaBoost [Electronic resource] / Wenxin Jiang // The Annals of Statistics. – 2003. – Vol. 32, No. 1. – P. 13–29. Mode of access: https://doi.org/10.1214/aos/1079120128
4. Li X. AdaBoost with SVM-based component classifiers [Electronic resource] / Xuchun Li, Lei Wang, Eric Sung // Engineering Applications of Artificial Intelligence. – 2008. – Vol. 21, No. 5. – P. 785–795. Mode of access: https://doi.org/10.1016/j.engappai.2007.07.001 .
5. Multi-class AdaBoost [Electronic resource] / T. Hastie, S. Rosset, J. Zhu, & H. Zou // Statistics and Its Interface. – 2009. – Vol. 2, No. 3. – P. 349–360. Mode of access: https://doi.org/10.4310/sii.2009.v2.n3.a8
6. Ensemble Machine Learning [Electronic resource] / ed. by C. Zhang, Y. Ma. – Boston, MA : Springer US, 2012. – Mode of access: https://doi.org/10.1007/978-1-4419-9326-7
7. Advance and Prospects of AdaBoost Algorithm [Electronic resource] / [C. Ying, M. Qi-Guang, L. Jia-Chen, & G. Lin] // Acta Automatica Sinica. – 2013. – Vol. 39, No. 6. – P. 745–758. Mode of access: https://doi.org/10.1016/s1874-1029(13)60052-x .
8. Zhou Z.-H. Ensemble Learning [Electronic resource] / Zhi-Hua Zhou // Machine Learning. – Singapore, 2021. – P. 181–210. Mode of access: https://doi.org/10.1007/978-981-15-1967-3_8 .
9. Hansen L. K. Neural network ensembles [Electronic resource] / L. K. Hansen, P. Salamon // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1990. – Vol. 12, No. 10. – P. 993–1001. Mode of access: https://doi.org/10.1109/34.58871
10. Wolpert D. H. Stacked generalization [Electronic resource] / David H. Wolpert // Neural Networks. – 1992. – Vol. 5, No. 2. – P. 241–259. Mode of access: https://doi.org/10.1016/s0893-6080(05)80023-1 .
11. Sharkey A. J. C. On Combining Artificial Neural Nets [Electronic resource] / Amanda J. C. Sharkey // Connection Science. – 1996. – Vol. 8, No. 3–4. – P. 299–314. Mode of access: https://doi.org/10.1080/095400996116785
12. An Optimal Algorithm for Combining Multivariate Forecasts in Hybrid Systems [Electronic resource] / [Y. Bodyanskiy, P. Otto, I. Pliss, & S. Popov] // Lecture Notes in Computer Science. – Berlin, Heidelberg, 2003. – P. 967–972. Mode of access: https://doi.org/10.1007/978-3-540-45226-3_132 .
13. Bodyanskiy Y. Fuzzy Selection Mechanism for Multimodel Prediction [Electronic resource] / Y. Bodyanskiy, S. Popov // Lecture Notes in Computer Science. – Berlin, Heidelberg, 2004. – P. 772–778. Mode of access: https://doi.org/10.1007/978-3-540-30133-2_101 .
14. Intelligent Method for Identifying the Fraudulent Online Stores [Electronic resource] / K. Lipianina-Honcharenko, I. Lukasevych-Krutnyk, N. Butryn-Boka et al] // 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), Kharkiv, Ukraine, 5–7 October 2021. – [S. l.], 2021. Mode of access: https://doi.org/10.1109/picst54195.2021.9772195

DOI 10.15588/1607-3274-2023-4-9