ISSN 2500-316X (Online)

RUSSIAN TECHNOLOGICAL JOURNAL

РОССИЙСКИЙ
ТЕХНОЛОГИЧЕСКИЙ
ЖУРНАЛ

Information systems.
Computer sciences.
Issues of information security

Multiple robots (robotic centers) and systems.
Remote sensing and non-destructive testing

Modern radio engineering and telecommunication systems

Micro- and nanoelectronics.
Condensed matter physics

Analytical instrument engineering and technology

Mathematical modeling

Economics of knowledge-intensive and high-tech enterprises and industries.
Management in organizational systems

Product quality management. Standardization

Philosophical foundations of technology and society

12+
www.rtj-mirea.ru

МИРЭА
Российский технологический университет

11(2) 2023

ISSN 2500-316X (Online)

RUSSIAN TECHNOLOGICAL JOURNAL

РОССИЙСКИЙ ТЕХНОЛОГИЧЕСКИЙ ЖУРНАЛ

- Information systems. Computer sciences. Issues of information security
- Multiple robots (robotic centers) and systems. Remote sensing and non-destructive testing
- Modern radio engineering and telecommunication systems
- Micro- and nanoelectronics. Condensed matter physics
- Analytical instrument engineering and technology
- Mathematical modeling
- Economics of knowledge-intensive and high-tech enterprises and industries. Management in organizational systems
- Product quality management. Standardization
- Philosophical foundations of technology and society

- Информационные системы. Информатика. Проблемы информационной безопасности
- Роботизированные комплексы и системы. Технологии дистанционного зондирования и неразрушающего контроля
- Современные радиотехнические и телекоммуникационные системы
- Микро- и наноэлектроника. Физика конденсированного состояния
- Аналитическое приборостроение и технологии
- Математическое моделирование
- Экономика наукоемких и высокотехнологичных предприятий и производств. Управление в организационных системах
- Управление качеством продукции. Стандартизация
- Мировоззренческие основы технологии и общества

Russian Technological Journal
2023, Vol. 11, No. 2

Russian Technological Journal
2023, том 11, № 2

https://www.rtj-mirea.ru

https://doi.org/10.32362/2500-316X ISSN 2500-316X (Online)

Russian Technological Journal
2023, Vol. 11, No. 2

Publication date March 31, 2023.

The peer-reviewed scientific and technical journal highlights the issues of complex development of radio engineering, telecommunication and information systems, electronics and informatics, as well as the results of fundamental and applied interdisciplinary researches, technological and economical developments aimed at the development and improvement of the modern technological base.

Periodicity: bimonthly.
The journal was founded in December 2013. The titles were «Herald of MSTU MIREA» until 2016 (ISSN 2313-5026) and «Rossiiskii tekhnologicheskii zhurnal» from January 2016 until July 2021 (ISSN 2500-316X).

Founder and Publisher:
Federal State Budget
Educational Institution of Higher Education
«MIREA – Russian Technological University»
78, Vernadskogo pr., Moscow, 119454 Russia.

The journal is included into the List of peer-reviewed science press of the State Commission for Academic Degrees and Titles of Russian Federation. The Journal is included in Russian State Library (RSL), Russian Science Citation Index, eLibrary, Socionet, Directory of Open Access Journals (DOAJ), Directory of Open Access Scholarly Resources (ROAD), Google Scholar, Ulrich's International Periodicals Directory.

The registration number ПИ № ФС 77 - 81733
was issued in August 19, 2021
by the Federal Service for Supervision
of Communications, Information Technology,
and Mass Media of Russia.

The subscription index of *Pressa Rossii*: **79641**.

Russian Technological Journal
2023, том 11, № 2

Дата опубликования 31 марта 2023 г.

Научно-технический рецензируемый журнал освещает вопросы комплексного развития радиотехнических, телекоммуникационных и информационных систем, электроники и информатики, а также результаты фундаментальных и прикладных междисциплинарных исследований, технологических и организационно-экономических разработок, направленных на развитие и совершенствование современной технологической базы.

Периодичность: один раз в два месяца.
Журнал основан в декабре 2013 года. До 2016 г. издавался под названием «Вестник МГТУ МИРЭА» (ISSN 2313-5026), а с января 2016 г. по июль 2021 г. под названием «Российский технологический журнал» (ISSN 2500-316X).

Учредитель и издатель:
федеральное государственное бюджетное
образовательное учреждение высшего образования
«МИРЭА – Российский технологический университет»
119454, РФ, г. Москва, пр-т Вернадского, д. 78.

Журнал входит в Перечень ведущих рецензируемых научных журналов ВАК РФ, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук и доктора наук, индексируется в РГБ, РИНЦ, eLibrary, Соционет, Directory of Open Access Journals (DOAJ), Directory of Open Access Scholarly Resources (ROAD), Google Scholar, Ulrich's International Periodicals Directory.

Регистрационный номер и дата принятия решения
о регистрации СМИ ПИ № ФС 77 - 81733 от 19.08.2021 г.
СМИ зарегистрировано Федеральной службой
по надзору в сфере связи, информационных технологий
и массовых коммуникаций (Роскомнадзор).

Индекс по объединенному каталогу «Пресса России» **79641**.

Editorial Board

Редакционная коллегия

Contents

Содержание

Information systems. Computer sciences. Issues of information security
Информационные системы. Информатика. Проблемы информационной безопасности

UDC 004.942
https://doi.org/10.32362/2500-316X-2023-11-2-7-19

RESEARCH ARTICLE

Robust neural network filtering in the tasks of building intelligent interfaces

Anton V. Vasiliev [@], Alexey O. Melnikov, Sergey A. Lesko

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding author, e-mail: bysslaev@gmail.com

Abstract

Objectives. In recent years, there has been growing scientific interest in the creation of intelligent interfaces for computer control based on biometric data, such as electromyography signals (EMGs), which can be used to classify human hand gestures to form the basis for organizing an intuitive human-computer interface. However, problems arising when using EMG signals for this purpose include the presence of nonlinear noise in the signal and the significant influence of individual human characteristics. The aim of the present study is to investigate the possibility of using neural networks to filter individual components of the EMG signal.

Methods. Mathematical signal processing techniques are used along with machine learning methods.

Results. The overview of the literature on the topic of EMG signal processing is carried out. The concept of intelligent processing of biological signals is proposed. The signal filtering model using a convolutional neural network structure based on Python 3, TensorFlow and Keras technologies was developed. Results of an experiment carried out on an EMG data set to filter individual signal components are presented and discussed.

Conclusions. The possibility of using artificial neural networks to identify and suppress individual human characteristics in biological signals is demonstrated. When training the network, the main emphasis was placed on individual features by testing the network on data received from subjects not involved in the learning process. The achieved average 5% reduction in individual noise will help to avoid retraining of the network when classifying EMG signals, as well as improving the accuracy of gesture classification for new users.

Keywords: digital signal processing, frequency filtering, electromyography, machine learning, neural networks, interfaces, gesture manipulation

• **Submitted:** 17.06.2022 • **Revised:** 22.09.2022 • **Accepted:** 09.02.2023

For citation: Vasiliev A.V., Melnikov A.O., Lesko S.A. Robust neural network filtering in the tasks of building intelligent interfaces. *Russ. Technol. J*. 2023;11(2):7–19. https://doi.org/10.32362/2500-316X-2023-11-2-7-19

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Применение робастной нейросетевой фильтрации в задачах построения интеллектуальных интерфейсов

А.В. Васильев [@], **А.О. Мельников, С.А. Лесько**

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: bysslaev@gmail.com*

Резюме

Цели. В последние годы возрос научный интерес к построению интеллектуальных интерфейсов для управления компьютером на основе биометрических данных. Одним из источников таких данных служит сигнал электромиографии (ЭМГ). Сигнал ЭМГ можно использовать для классификации жестов рук человека. Это позволяет организовать интуитивно понятный интерфейс «человек – компьютер». Основными проблемами при использовании сигналов ЭМГ являются наличие нелинейных шумов в сигнале и значительное влияние индивидуальных особенностей человека. Цель работы – исследование возможностей применения нейронных сетей для фильтрации индивидуальных компонент сигнала ЭМГ.

Методы. Использованы математические методы обработки сигналов и методы машинного обучения.

Результаты. Проведен анализ исследований по теме обработки ЭМГ-сигналов. Предложена концепция интеллектуальной обработки биологических сигналов. Разработана модель фильтрации сигнала, построена структура сверточной нейронной сети на основе технологий Python 3, TensorFlow и Keras. Проведен эксперимент на наборе данных ЭМГ по фильтрации индивидуальных компонент сигнала.

Выводы. Продемонстрирована возможность применения искусственных нейронных сетей для выявления и подавления индивидуальных особенностей человека в биологических сигналах. При обучении сети основной упор делался на индивидуальные особенности, тестируя сеть на данных, полученных от субъектов, не участвующих в процессе обучения. Достигнуто уменьшение индивидуального шума в среднем на 5%. Для решения задачи классификации сигнала ЭМГ данный результат поможет избежать переобучения сети и повысить точность классификации жестов для новых пользователей.

Ключевые слова: цифровая обработка сигнала, частотная фильтрация, электромиография, машинное обучение, нейронные сети, интерфейсы, управление жестами

• **Поступила:** 17.06.2022 • **Доработана:** 22.09.2022 • **Принята к опубликованию:** 09.02.2023

Для цитирования: Васильев А.В., Мельников А.О., Лесько С.А. Применение робастной нейросетевой фильтрации в задачах построения интеллектуальных интерфейсов. *Russ. Technol. J.* 2023;11(2):7–19. https://doi.org/10.32362/2500-316X-2023-11-2-7-19

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

One of the key steps in software design is the choice of a method to communicate with an individual. For this, unified structural, hardware, and software tools are used, which are necessary for the interaction of various functional elements of the system. A set of such elements is referred to as an “interface”. The interface between an individual and software is especially important since this is what standardizes the interaction and determines the boundaries of the functionality of working with the software. Thus, interface concept is closely related to the usability of software systems. First of all, the usability is associated with a graphical user interface. The interface is considered usable if the user needs the least amount of time to use the information system. The second parameter that affects the usability is, of course, the simplicity and time needed to train a new user to work

with the information system. A good interface should be intuitive and have as few hidden dependencies as possible, as well as a minimal learning curve. In addition to the graphical interface, an information system may also have some command-program interface, which is a set of messages (commands) that can be perceived by the software system and processed using the application programming interface (API). The usability of this type of interface is evaluated by the number of commands that need to be used to perform the targeted action on the system. At the same time, it is desirable that commands for different target actions not be repeated (duplicated). These requirements impose a serious responsibility on interface developers when designing interfaces for software products and systems.

In the modern digital space, a human (user) is transformed into an interactive system possessing a rapidly expanding set of its capabilities. However, the range of interactivity varies from system to system. For example, in the aviation or aerospace industry, where the working conditions of a user operating with an information system are constrained by physical conditions, interactivity can be described as limited. On the other hand, everyday interactive systems, such as multimedia devices and gaming complexes, do not impose significant physical restrictions on the set of interactions of an individual (user) with an information system.

In order to improve the efficiency and usability of information systems, researchers are constantly looking for new ways to organize interfaces. Among the factors that reduce the usability of information systems, we can distinguish between technical, physical, and informational varieties. Technical factors refer to the quality of technologies applied both in software development (network speed, amount of memory) and in hardware (for example, the quality of a computer monitor or camera). Physical factors mean the physical environmental conditions during the use of the software system, such as humidity, light, visibility, the possibility of physical movements, etc. Information factors are understood as the development of the interface of the software system that ensures ease of use in general, for example, the size of buttons in the graphical interface, the ability to enter text, the ability to save data, etc.

Interfaces can be divided into several categories: text, graphic, voice, video and hybrid. To improve usability, each of these approaches should be considered. Currently, interfaces based on audio and video information received from various external sensors are being actively developed. Among the difficulties of using video, the following factors can be distinguished: extraneous noise, poor visibility, physical obstacles between the camera and the subject, the lack of the appropriate angle for shooting, or a lack of verbal communication (silence mode). Nevertheless, interfaces based on audio or video can greatly expand the scope of software systems. Moreover, while a text-based or graphical interface necessarily requires an input console or a screen plus input devices (keyboard), an audio- or video-based interface requires only a microphone and a video camera. This allows a user to free his or her hands and improve the quality of the user experience when working with the system, using hands as an additional control channel. In order to overcome the discussed limitations of interfaces based on video information and preserve their advantages, it is required to use a new type of interface either hybrid or biological. Biological interfaces are widespread in medicine. It should be noted, however, that in medicine, an individual interacts with information systems mostly passively, allowing the device to retrieve and process the information received through the interface. At the same time, the potential of using biological interfaces is much wider. They can be used to build complete information systems with a high level of usability.

Measurements of biological signals, such as electromyography (EMG), electroencephalography (EEG), etc., can be used as additional information exchange channels. In recent years, innovative research has been carried out on the development and use of clothes containing various sensors and transducers [1, 2], which allow a person's physiological activities to be recorded. Such studies commonly use items of clothing that contain sensors for recording EMG signals [3, 4]. EMG allows you to record the electrical activity that occurs when the muscle fibers are excited. Clothes containing EMG sensors are in demand in many areas: from any physically active activity (e.g., construction work and sporting activities) to the calmest (e.g., office work).

EMG signals are used to diagnose neuromuscular diseases, in psychophysiology, in the study of motor activity, in studies of higher nervous activity, to evaluate the results of prosthetics and orthopedics, and in engineering psychology. Among other things, research into the possibility of organizing a silent interface, that is, an interface that does not require voice input and allows controlling the information system through articulation, has recently gained popularity.

The present work is focused on the use of EMG signals as a basis for a human-computer interface. Much research is aimed at analyzing the EMG signal for developing smart prosthetics systems [5, 6] and systems controlled by gestures [7]. Besides the EMG signal, ultrasonic scanners can be used to solve the problems of gesture classification [8]. A particular problem identified by researchers involves the difficulty of recovering specific control units in the EMG signal, along with a high dependence of classification accuracy on the specific person with whom the experiment is

carried out [9, 10]. To solve such problems, methods of decomposition [11] and clustering of EMG signals are used in order to identify the muscle groups involved in a particular hand gesture [3]. Signal conversion methods used to minimize noise include the method of principle components, auto-encoders, etc. [12, 13].

1. LITERATURE OVERVIEW

In a number of studies devoted to the classification of the EMG signal, the problem of its filtering is raised. In most of the publications only frequency filtering is used, but other approaches are researched as well. The most effective method is the preliminary clustering of EMG signals in order to isolate motor units. However, this approach leads to irreversible signal distortion and does not apply to filtering.

In [14], the authors developed a system for identifying muscles using needle EMG for prosthetics. The main characteristics of this model are the following: the use of needle EMG (16 sensors) and kinematic gloves, signal preprocessing with low (10 Hz) and high (100 Hz) frequency filters, and the use of artificial neural networks. The data set consists of 5 movements with 10 repetitions of each movement. The input data for the neural network are correlation matrices. The advantage of the model is the compactness of a fully connected neural network (3 hidden layers). The disadvantages of the model include relatively low accuracy (90.1% for the test set) and the need to use kinematic gloves. The negative impact in model evaluation is due to the low accuracy of signal recognition from a number of muscles for the data set. The authors note that classification errors may be due to insufficiently accurate labeling of reference data.

In [15], the authors conducted an experiment to compare statistical approaches to classify an EMG signal with machine learning models. The task of the researchers was the binary classification of the EMG signal. The goal was to recognize meal intervals in a person's daily activities. The comparison was made between SVM [16], RandomForest [17], LogisticRegression [18], XGBoost [19], LightGBM [20], LSTM [21], and Conv-LSTM models [22]. The advantages of the work include an extensive comparison of statistical methods and machine learning methods. An accuracy of 94.76% was achieved for the balanced dataset for statistical methods and 95.35% for the unbalanced dataset. XGBoost turned out to be the most effective statistical method for classifying the EMG signal. The use of LSTM-type neural networks has improved the classification accuracy up to 97%, however, the researchers note the problem associated with the need for a large amount of data to train this type of networks, as well as insufficient data for machine learning methods, data pollution by cumulative actions, poor Bluetooth connectivity, features of right-handers and left-handers.

In [3], the authors developed a device for reading an EMG signal powered by solar energy. The main characteristics of the developed model are ultra-low power consumption and an intelligent EMF sensor localization system on a user's wrist. Data for the experiment were collected from 20 people and included 15 unique hand movements. Accuracy of 95.3% was achieved when classifying 15 gestures. The position of sensors on a person's wrist is one of the problems for this kind of tasks. EMG signal analysis methods are highly sensitive to the position of sensors. Therefore, the model needs to be retrained every time the position of the sensors changes. This problem can be solved by locating sensors on the wrist using intelligent data processing from capacitive sensors. In order to analyze EMG signals, clustering of all wrist muscles into 8 groups was performed. To adjust the position of the bracelet on the wrist, calibration based on data from 15 sensors using convolution was used. To classify gestures, a convolutional neural network with two convolution layers with the rectified linear unit (ReLU) activation function was used. It is noted that it is such small number of layers allows solving the problem of retraining. The disadvantage of the method is a significant drop in the accuracy of the classification of movements in a static position (a decrease in accuracy by 3%). Also, gestures describing fine motor skills of fingers were not included in the work.

In [23], a model for classifying gestures based on ultrasound was developed. An ultrasound representation of the muscles of the forearm was used to classify gestures.

2. NATURE OF THE EMG SIGNAL

EMG is a method for studying the bioelectric potentials that arise in the skeletal muscles of humans and animals during excitation of muscle fibers.

There are three types of EMG:

1) EMG using needle electrodes that are inserted into the muscle;
2) stimulation EMG;
3) EMG using skin electrodes.

Needle EMG provides the most accurate representation of the electrical activity that occurs during muscle stimulation but requires physical penetration into human muscle tissue. The invasive nature of the method is a limitation for use as a basis for an information system interface.

Stimulation EMG is a type of non-invasive EMG that uses skin surface electrodes to assess the conduction of an impulse along peripheral nerves in response to stimulation with a low-intensity electrical current. This

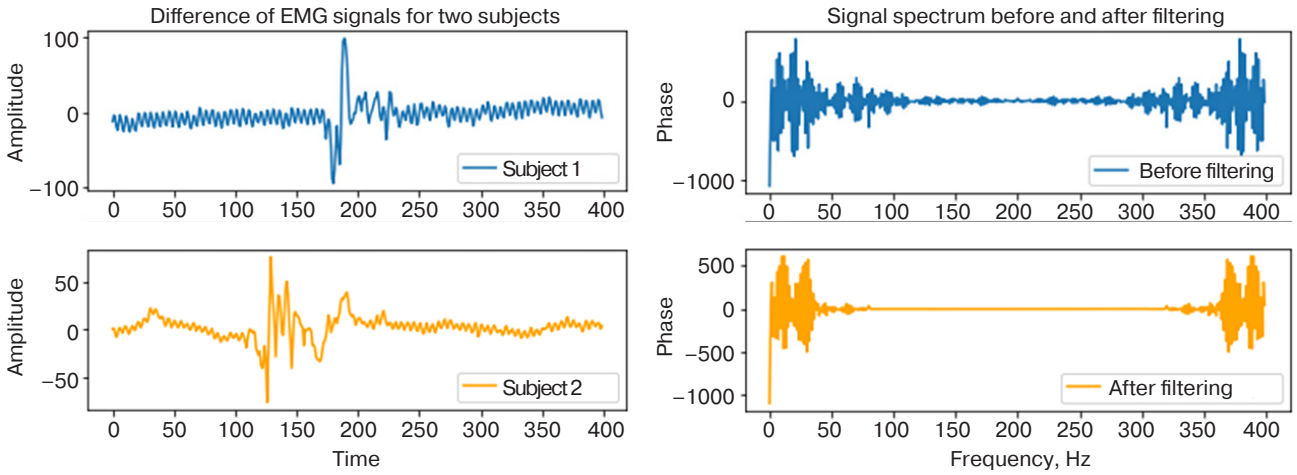

Fig. 1. EMG signals in the time and frequency domains

type of EMG is used, in particular, to diagnose diseases of the peripheral nervous system.

Surface EMG is a type of EMG in which skin surface electrodes are used. Unlike stimulation EMG, this type does not involve stimulation of the nervous system but, on the contrary, consists only in recording electrical activity during active excitation and relaxation of muscle tissues.

With weak muscle contraction, either the potential of a single motor unit or the potential of many motor units is recorded. With an average strength and strong muscle contraction, interference EMG is observed, in which it is almost impossible to identify the potentials of individual motor units. For people at rest, who do not have problems in the area of the nervous system, usually either no electrical potential activity is detected or electrical potentials of individual muscle fibers are recorded.

In a simple case, we will consider the following scenario: a muscle reacts to a single action with a single contraction. In this case, three phases can be distinguished:

- latent period (from 2–3 to 10 ms), lasting from the moment of applying irritation to the start of contraction;
- shortening or contraction phase (40–50 ms);
- relaxation phase (about 50 ms).

The device for recording an EMG signal includes electrodes that pick up muscle potentials, an amplifier of these potentials, and a recording device.

The main parameters of the EMG signal are:

- amplitude (1 μV – 50 mV),
- frequency (0.5–500 Hz).

To analyze the EMG signal in more detail, it is presented as a decomposition of frequencies and amplitudes obtained using the Fourier transform.

Any part of a muscle can contain muscle fibers belonging to 20–50 motor cells. As a result of movement, many motor units are excited. The cumulative action potential can be recorded using EMG equipment and will be presented in the time domain in the following form:

$$S(t) = \sum_j \mathrm{SAPMC}_j(t) + n(t) = \sum_j \sum_i k_j f\left(\frac{t-\theta_{ij}}{a_j}\right) + n(t)$$

where SAPMC_j is the sequence of the action potential of the motor cell; k_j is the amplitude factor for the muscle of the jth motor cell; f is the shape of the action potential; θ_{ij} is the time of occurrence of SAPMC; a_j is the scale change; $n(t)$ is the additional noise.

In this work, we use the signal obtained with a single-channel surface EMG. The use of a single-channel system makes it possible to simplify signal registration by ignoring the time synchronization of data from parallel EMG channels.

The main problem when using EMG signals as a control interface is their variability and instability, primarily due to external interference, electrode displacement, skin sweating, and muscle fatigue.

Attempts to eliminate the influence of muscle fatigue consist in the use of switching devices when the signal changes, or the use of static methods such as filtering.

The success of the implementation of the device control interface is determined by the degree of reliability of the decoding of muscle biopotentials in the registered EMG signal during the planned movement. An accurate determination of the motion type is hindered by the low signal-to-noise ratio in the measuring system.

Signal distortions can occur due to the side effect of the signals of the electrical activity of the heart, shifts of the electrodes relative to the designated position, changes in muscle biopotentials, noise from electronic devices, ambient electromagnetic radiation, and similar factors.

To date, a common method for determining the types of movement is the use of various classifiers.

3. PROBLEM STATEMENT

To build an efficient interface based on EMG signals, it is necessary to solve a number of problems. First, it is necessary to clean the signal from noise, which is recorded during excitation of human muscles. Secondly, the signal must be classified in some way so that the received actions or patterns can be used to create control actions by the information system. In this work, we will solve the problem of cleaning the signal from noise. The main problem is the non-linear nature of the noise. By noise, we mean some non-linear component of the signal, which depends on the parameters of recording the EMG signal from a biological object. The amount of noise depends on such parameters as the level of voltage in the laptop electrical network, the parameters of the signal amplifier, the quality of the electrodes, the quality of the preparation of the skin surface for installing EMG electrodes, etc.

A signal $\mathbf{X}$ is a sequence of samples x_i ($i = \overline{1, N}$). It is assumed that this signal may contain non-linear noise $\mathbf{Z}$, which can be suppressed using a filter:

$$\mathbf{I} = \mathbf{X} \times \mathbf{h},$$

where $\mathbf{I}$ is a useful signal, $\mathbf{X}$ is a noisy signal, $\mathbf{h}$ is a neural network filter.

We compared data for one type of gesture from two subjects (Fig. 2). As can be seen from the figure, these signals differ from each other not only in the phase of the signal, but also in the shape. Neural network filtering is used to minimize these differences.

The purpose of neural network filtering is to get rid of the individual component of the signal, which varies from person to person. This type of distortion is called individual noise. Individual noise is understood as a non-linear component of the signal, which can be defined as follows:

$$\mathbf{Z} = \mathbf{X} - \mathbf{I},$$

where $\mathbf{Z}$ is individual noise.

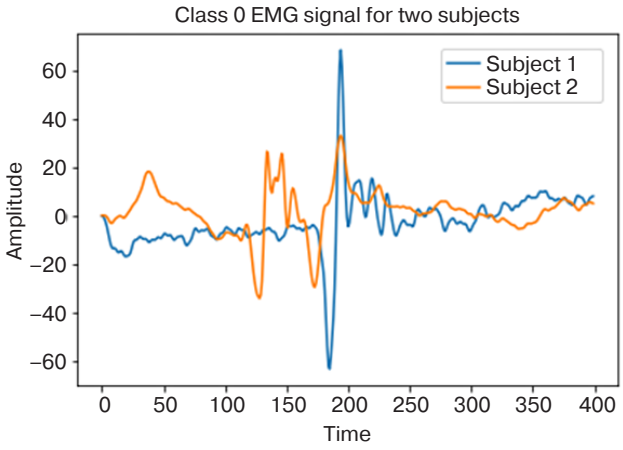

Fig. 2. EMG signal for one gesture received from two different subjects

The task is to find the filter parameters $\mathbf{h}$, which will minimize the difference in signals describing the same gesture class, but received from different subjects. Such a task can be described as:

$$(\mathbf{X}_{i,k} \times \mathbf{h} - \mathbf{X}_{j,k} \times \mathbf{h}) \rightarrow \min,$$

where i is the subject index, k is the gesture class number.

3.1. Signal processing block-diagram

After receiving the EMG signal from the sensors, filtering is carried out in two stages. At the first stage, a low-bandpass filter (up to 50 Hz) and a high-bandpass filter (more than 1.5 MHz) are used. These filters allow you to get rid of the noise generated by electronic equipment and external static electric field. The signal processing block-diagram is shown in Fig. 3.

3.2. Datasets

When planning the experiment, two databases containing EMG data were considered suitable for building an information system with gesture-based control. The experiment required a data set containing EMG signals received from the forearm region when

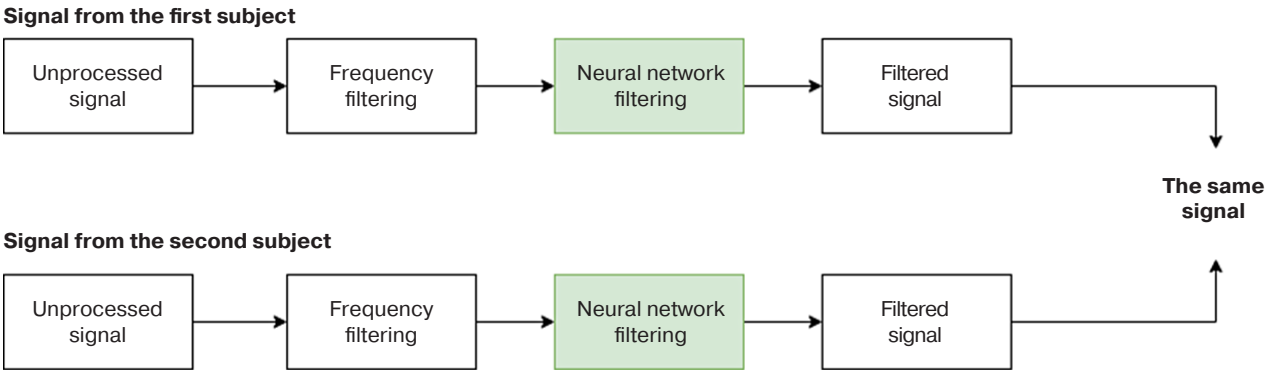

Fig. 3. Block-diagram of neural network filtering

making a set of hand gestures. When working with the information system, each of the gestures can be used as a control action when.

3.2.1. Ninapro Data

This database is available for academic purposes on a dedicated website[1] [24]. The goal of the project is to develop a family of algorithms that can significantly increase dexterity and reduce learning time for a controlled sEMG prosthesis. The project's challenge is how to provide patients with a cheap, easy to use prosthesis that can be controlled in a natural way.

The data set consists of bioelectrical muscle activities collected under special conditions using differential sEMG electrodes. Currently, data are available for 67 healthy subjects and 11 amputees.

The Ninapro data collection process was designed to be easily repeated for obtaining new data from different research groups.

3.2.2. RF-Lab. Digital Signal Processing Laboratory (DSP) RTU MIREA

The project database contains EMG signals sampled from the forearm area. Six subjects participated in data collection. Each subject consistently repeated one of the 9 hand movements (gesture) 79 times. The signals recorded for each gesture were written into a 400-sample vector. The total number of signals is 2820 [9]. The data set includes gestures of the following classes:

- wrist up (class 0);
- wrist down (class 1);
- clenching all fingers (class 2);
- clenching the index finger (class 3);
- clenching the middle finger (class 4);
- clenching the ring finger (class 5);
- clicking the thumb with the middle finger (class 6);
- unclenching all fingers (class 7);
- turning the hand to the left (class 8).

The following components were used to register the signals: Arduino Leonardo (Arduino AG, China), ECG-EMG Arduino Shield (OLIMEX LTD, Bulgaria), single-channel surface electrodes and USB Type-A / USB Micro-B.

As a result, 79 vectors of 400 units in length were used for each gesture which provided a window in which the action potential was captured. Thus, it includes only the most important data that are needed for the classification task, thereby reducing the consumed computing resources and increasing the accuracy.

As part of this work, the RF-Lab[2] dataset was used as it is focused on building a human-machine interface with gesture control.

[1] http://ninapro.hevs.ch/. Accessed June 15, 2022.
[2] https://rf-lab.org/. Accessed June 01, 2022.

3.3. Filtration quality assessment

To assess the quality of the signal, the standard deviation of the EMG signals is analyzed. For each type of movement, the value of the standard deviation of the signal for all subjects is calculated (Fig. 4).

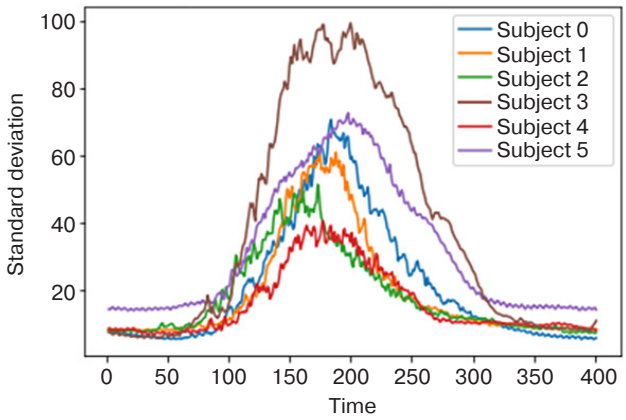

Fig. 4. Standard deviation for gesture classes for each subject

When filtering the signal, it is necessary to reduce the standard deviation of the EMG signal within each gesture class. As an example, the standard deviation was calculated for a class 0 gesture before and after filtering using a frequency filter (Fig. 5).

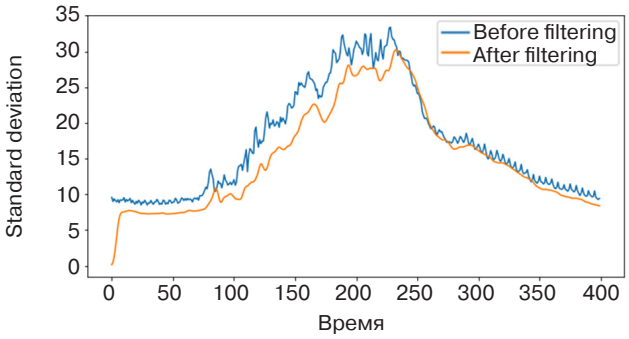

Fig. 5. Comparison of the standard deviation of the EMG signal amplitude before and after frequency filtering

It can be seen from the figure that frequency filtering reduces the standard deviation of the signal over the entire segment; therefore, this method is used to evaluate the efficiency of signal filtering in our experiment.

4. DEVELOPED MODEL OF A NEURAL NETWORK

Digital filters are widely used today in various areas of signal processing, both technical and biological, which include the EMG signal. Mathematical models of digital filters can be described using vectors and numerical matrices. For a binary signal, the numbers in the matrices can be binary. There are two types of filters: finite pulse

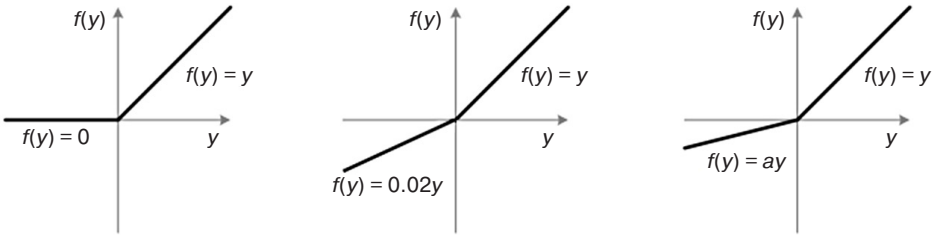

Fig. 6. Comparison of the activation function of ReLU (left), LeakyReLU (center), and PReLU (right). *a* is a parameter of the PReLU activation function; *y* is an input signal

response and infinite pulse response. The filter must suppress the harmonic components of the original signal in one frequency band (stop band) and pass them in another frequency band (pass band). In most cases, in extremely complex problems of signal analysis, classical techniques based on the Fourier transform and wavelet transform are used to construct the feature space. Due to the complexity of understanding the nature of the signal, the features of one task may be completely unsuitable for another task, and it is required to construct a feature space from scratch. The EMG signal falls into the category of signals with a complex nature, for which standard filters are not suitable for processing. It can be represented as a time series [25]. Therefore, to build a filter for the EMG signal, it may be appropriate to use intelligent models. One of the most effective intellectual models at the moment is neural networks.

To solve the problem, we will use a neural network type that includes convolution layers. Such networks are called convolutional networks. As an activation function, it is proposed to use parameterized ReLU (PReLU). The use of this activation function is an achievement in machine vision that has allowed for surpassing the human level in ImageNet[3] image recognition tasks. The error back propagation and update process for PReLUs is simple and similar to traditional ReLUs. The main difference between PReLU and ReLU is that this function does not zero out negative input signals. Instead, negative input signals are multiplied by some non-zero factor, which allows negative values to be taken into account in network training and signal processing. A comparison of the PReLU activation function with a regular ReLU is shown in Fig. 6.

As part of the National Data Science Bowl (NDSB) Kaggle competition, the PReLU activation function made it possible to reduce overtraining due to the element of randomness in the work. When comparing the classification accuracy of two convolutional artificial neural networks with different activation functions on data sets (images used to test the quality of pattern recognition algorithms) CIFAR-10, CIFAR-100[4], and NDSB[5], results were obtained that indicate that for all sets the modified functions ReLU family activations have surpassed traditional functions. RReLU is significantly superior to other activation functions on the NDSB dataset because on this dataset, the activation function avoids overtraining as this dataset contains less training data. To train machine learning models, modern cloud infrastructure tools such as Docker[6] and Amazon Azure[7] were used [26].

In the experiment, the Python 3.8 programming language and the Keras 2.9.0[8] library were used when building a neural network model. The architecture of the neural network developed for filtering the EMG signal is shown in Fig. 7.

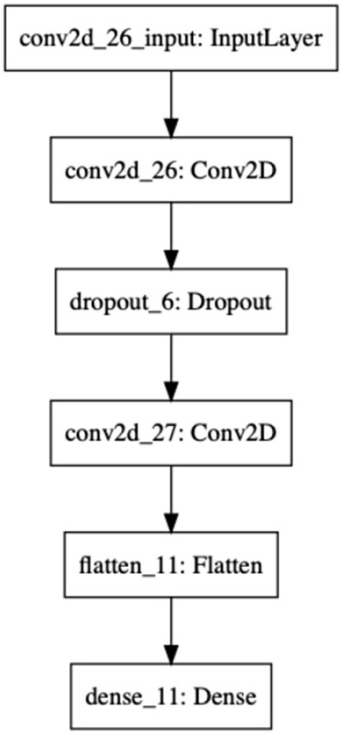

Fig. 7. Convolutional neural network architecture

The developed model contains two convolutional layers. A matrix of 20 × 20 serves as input data for the network, which is a raw EMG signal consisting of

[3] https://www.image-net.org/. Accessed June 09, 2022.

[4] https://www.cs.toronto.edu/~kriz/cifar.html. Accessed June 10, 2022.

[5] https://www.kaggle.com/competitions/datasciencebowl/overview/about-the-ndsb. Accessed June 10, 2022.

[6] https://www.docker.com/. Accessed June 10, 2022.

[7] https://azure.microsoft.com/en-us. Accessed June 10, 2022.

[8] https://github.com/keras-team/keras/releases/tag/v2.9.0. Accessed June 10, 2022.

400 samples. The first layer contains 64 feature maps of size 5 × 5 and the PReLU activation function (with parameter a = 0.02). The second convolutional layer contains 32 feature maps of size 3 × 3. Then there is a rectification layer and a fully connected output layer with a dimension of 400, which corresponds to the dimension of the input signal. Such a dimension at the output of the network allows the output signal to be used on a par with the input signal, expecting that the output signal will retain useful information about the subject's wrist movement pattern. The model was trained using Microsoft Azure[9] cloud computing power [26].

5. IMPLEMENTATION OF THE MODEL AND ASSESSMENT OF THE OBTAINED RESULTS

In the experiment, a data set was used collected by the team of the DSP Laboratory (RF-Lab) [9], containing the data of one channel of the EMG signal received from 6 subjects. Each subject performed 9 different hand gestures.

5.1. Experiment structure

To train the neural network, the dataset was transformed as follows. First, it was divided into 3 parts: training, validation and test sets. The training set comprised 60% of the total data and included data from four of the six subjects. The validation set contained 20% of the data, including data from the same four subjects. The test sample contained the remaining 20% of the total data and included data from two subjects not participating in training. This approach was used to demonstrate the validity of the resulting model on data from subjects that the model did not see during training. Each training example consisted of the original signal as input and the paired signal as the target value. The paired (target) signal was selected in such a way that it belonged to another subject. The neural network was trained using the Adam optimization algorithm [27], the number of training epochs was 25. The mean squared error [28] was used as the error function.

[9] https://azure.microsoft.com/en-us. Accessed June 01, 2022.

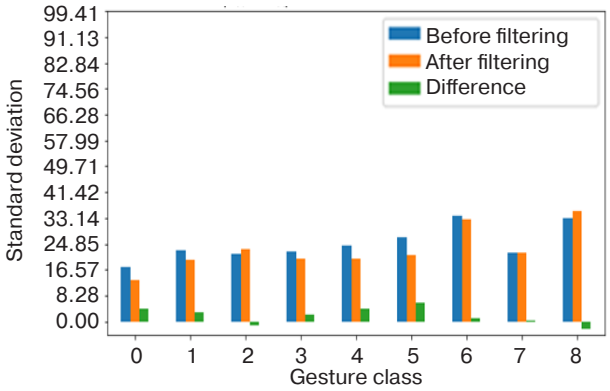

Fig. 8. Comparison of the standard deviation before and after neural network filtering for each signal class

5.2. Results

After training, the resulting model was evaluated on the remaining two subjects from the data set. Comparisons were made for each signal class separately (such as wrist movement). As an indicator of the effectiveness of the developed model, the standard deviation of the signal before and after filtering was measured. Its values before and after filtering are shown in Fig. 8. The measurements were carried out for signals of each class separately. As can be seen from the figure, the signals for gestures with class 6 (clenching the ring finger) and 8 (unclenching all fingers) have the greatest deviation.

As can be seen, on average, the reduction in the standard deviation is 5% for the signals received during the movement of the hand. The best result was obtained for movements with classes 4 and 5. An increase in the standard deviation was recorded for signals with classes 2 and 8. A decrease in the standard deviation was achieved for signals belonging to classes 0, 1, 3, 4, 5, 6, and 7.

The results obtained allow us to speak about the possibility of using neural network filtering in the tasks of cleaning individual signals. The filtering result for the class 0 gesture is shown in Fig. 9.

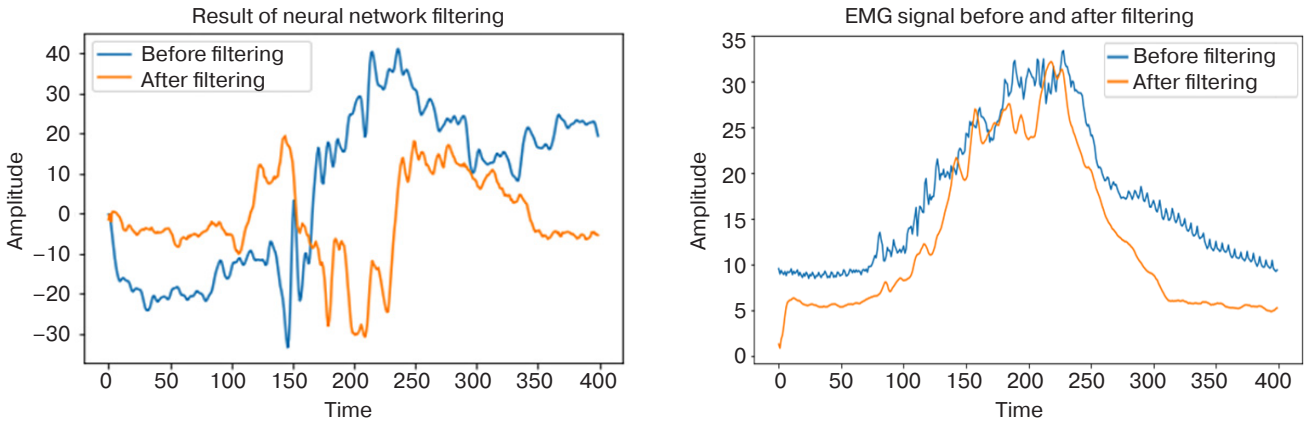

Fig. 9. Signal standard deviation before and after neural network filtering

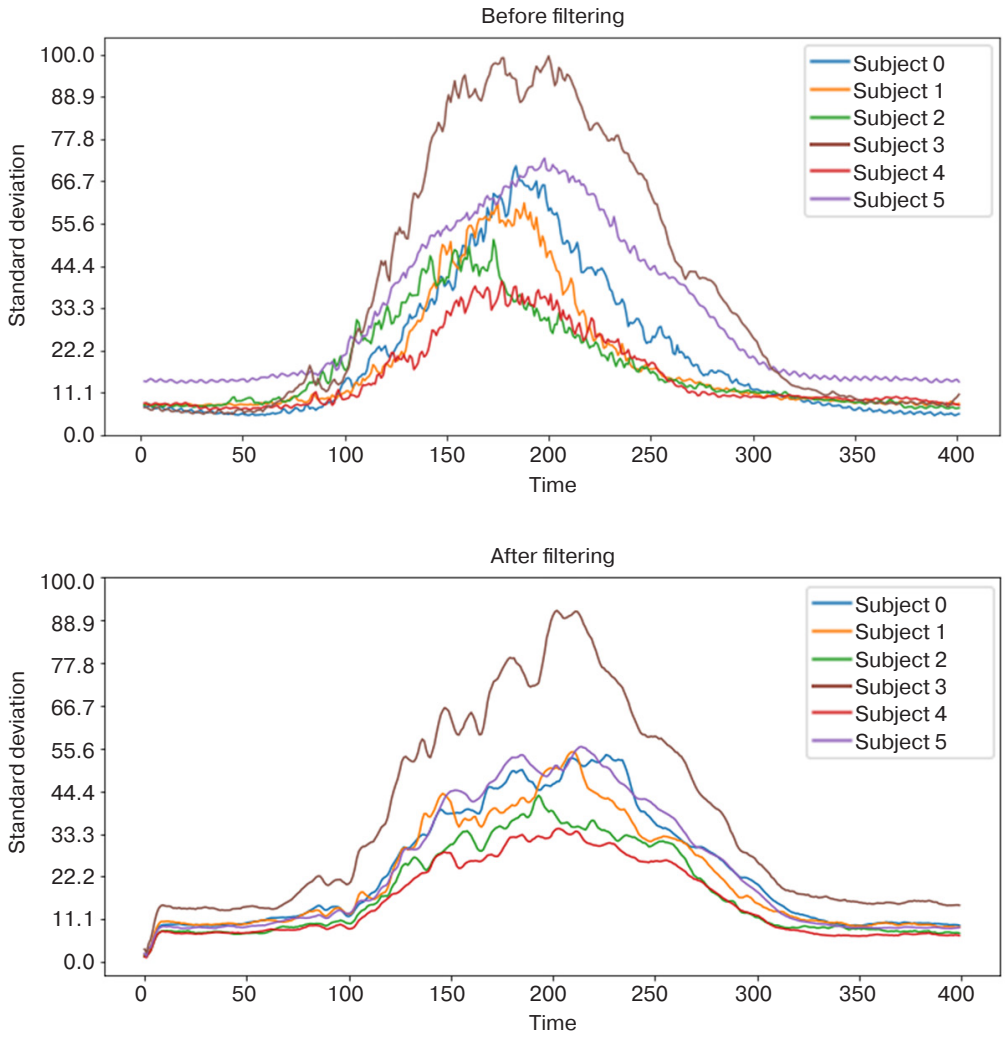

Fig. 10. Standard deviation for different gestures for each subject from the sample

Table 1. Comparison of the filtering result for different signal classes

Signal class	0	1	2	3	4	5	6	7	8
Change in the value of standard deviation after filtering	3.93	3.1	−1.33	2.16	3.95	6.03	0.99	0.07	−2.41

We also compared the standard deviation between different gestures for the same subject before and after filtering. The data obtained show that the difference between gestures remained almost at the same level (Fig. 10).

As can be seen from the results, the developed neural network filtering model is able to compensate for individual components in the EMG signal. The obtained indicators of signal quality improvement are shown in Table 1.

CONCLUSIONS

In this work, a study of approaches and methods for the development of neural network filters for biological signals was carried out. The proposed scheme for filtering biological signals takes the presence of individual signal components into account. A model was developed and a convolutional neural network for intelligent filtering was trained. Over the course of the study, an efficient convolutional neural network architecture for filtering the EMG signal was identified.

An experiment on filtering a single-channel EMG signal showed the effectiveness of the proposed model. By using neural network filtering, the influence of individual noise in the EMG signal can be reduced by an average of 5%.

In further studies, it is planned to evaluate the effect of neural network filtering on the accuracy of gesture classification using an EMG signal.

Authors' contributions

A.V. Vasiliev—preparing algorithms, data collection, conducting research, and writing the text of the article.

A.O. Melnikov—the research idea, developing objectives and aims, and formulating conclusions.

S.A. Lesko—consultations on research issues, scientific editing of the article.

REFERENCES

1. Arruda L.M., Calado A., Boldt R.S., Yu.Y., Carvalho H., Carvalho M.A., Soares F., Matos D. Design and testing of a textile EMG sensor for prosthetic control. In: Garcia N.M., Rires I.M., Goleva R. (Eds.). *IoT Technologies for HealthCare: 6th EAI International Conference, HealthyIoT 2019*. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Springer; 2020;341:37–51. https://doi.org/10.1007/978-3-030-42029-1_3
2. Hu Y., Wang H., Sheikhnejad O., Xiong Y., Gu H., Zhu P., Sun R., Wong C.P. Stretchable and printable medical dry electrode arrays on textile for electrophysiological monitoring. In: *IEEE 69th Electronic Components and Technology Conference* (*ECTC*). 2019;243–248. https://doi.org/10.1109/ECTC.2019.00043
3. Truong H., Zhang S., Muncuk U., Nguyen P., Bui N., Nguyen A., Dinh T.N., Vu T. CapBand: Battery-free successive capacitance sensing wristband for hand gesture recognition. In: *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems* (*SenSys '18*). 2018;54–67. https://doi.org/10.1145/3274783.3274854
4. Goto D., Shiozawa N. Can textile electrode for ECG apply to EMG measurement? In: *World Congress on Medical Physics and Biomedical Engineering*. 2018;431–434. https://doi.org/10.1007/978-981-10-9038-7_81
5. Samuel O.W., Asogbon M.G., Geng Y., Al-Timemy A.H., Pirbhulal S., Ji N., Chen S., Li G. Intelligent EMG pattern recognition control method for upper-limb multifunctional prostheses: advances, current challenges, and future prospects. *IEEE Access*. 2019;7:10150–10165. https://doi.org/10.1109/ACCESS.2019.2891350
6. Raheema M.N., Hussain J.S., Al-Khazzar A.M. Design of an intelligent controller for myoelectric prostheses based on multilayer perceptron neural network. In: *IOP Conf. Ser.: Mater. Sci. Eng*. 2020;671(1):012064. https://doi.org/10.1088/1757-899X/671/1/012064
7. Sosa M., Oviedo G., Fontana J.M., O'Brien R., Laciar E., Molisani L. Development of a serious game controlled by myoelectric signals. In: *The 8th Latin American Conference on Biomedical Engineering and The 42nd National Conference on Biomedical Engineering. CLAIB 2019. IFMBE Proceedings*. 2019;75:1171–1177. https://doi.org/10.1007/978-3-030-30648-9_152
8. McIntosh J., Marzo A., Fraser M., Phillips C. EchoFlex: Hand gesture recognition using ultrasound imaging. In: *Proceedings of The 2017 CHI Conference on Human Factors in Computing Systems*. (*CHI '17*). 2017;1923–1934. https://doi.org/10.1145/3025453.3025807
9. Lukyanchikov A.I., Melnikov A.O., Lukyanchikov O.I. Algorithms for classification of a single channel EMG signal for human-computer interaction. In: *ITM Web of Conferences*. 2018;18:02001. https://doi.org/10.1051/itmconf/20181802001
10. Tavakoli M., Benussi C., Lopes P.A., Osorio L.B., de Almeida A.T. Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier. *Biomed. Signal Process. Control*. 2018;46:121–130. https://doi.org/10.1016/j.bspc.2018.07.010
11. Chen C., Ma S., Sheng X., Zhu X. Continuous estimation of grasp kinematics with real-time surface EMG decomposition. In: *International Conference on Intelligent Robotics and Applications*. 2019;11744:108–119. https://doi.org/10.1007/978-3-030-27541-9_10
12. Wang Y., Wang C., Wang Z., Wang X., Li Y. Hand gesture recognition using sparse autoencoder-based deep neural network based on electromyography measurements. In: *Nano-, Bio-, Info-Tech Sensors, and 3D Systems II*. 2018;105971D:163–169. https://doi.org/10.1117/12.2296382
13. Qi J., Jiang G., Li G., Sun Y., Tao B. Surface EMG hand gesture recognition system based on PCA and GRNN. *Neural Comput. Appl.* 2020;32(10):6343–6351. https://doi.org/10.1007/s00521-019-04142-8
14. Cappellari P., Gaunt R., Beringer C., Mansouri M., Novelli M. Identifying electromyography sensor placement using dense neural networks. In: *Proceedings of The 7th International Conference on Data Science, Technology and Applications*. 2018:130–141. http://dx.doi.org/10.5220/0006912501300141
15. Pal K.K., Banerjee P., Choudhuri S., Sampat S. *Activity classification using Myo Gesture Control Armband data through machine learning*. 2019. Available from URL: https://kuntalkumarpal.github.io/files/MC_Report.pdf
16. Noble W. What is a support vector machine? *Nat. Biotechnol.* 2006;24:1565–1567. https://doi.org/10.1038/nbt1206-1565
17. Breiman L. Random forests. *Machine learning*. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324

18. Wright R.E. Logistic regression. In: Grimm L.G., Yarnold P.R. (Eds.). *Reading and understanding multivariate statistics*. American Psychological Association; 1995. P. 217–244. https://psycnet.apa.org/record/1995-97110-007
19. Chen T., Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (*KDD '16*). New York, NY, USA: Association for Computing Machinery; 2016. P. 785–794. https://doi.org/10.1145/2939672.2939785
20. Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.Y. LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems* (*NIPS 2017*). Long Beach, CA, USA: 2017;30. Available from URL: https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
21. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*. 2020 Mar. 1;404:132306. https://doi.org/10.1016/j.physd.2019.132306
22. Shi X., Chen Z., Wang H., Yeung D.Y., Wong W.K., Woo W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems* (*NIPS 2015*). 2015;28. Available from URL: https://papers.nips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html
23. Chen C., Yu Y., Ma S., Sheng X., Lin C., Farina D., Zhu X. Hand gesture recognition based on motor unit spike trains decoded from high-density electromyography. *Biomed. Signal Process. Control*. 2020;55:101637. https://doi.org/10.1016/j.bspc.2019.101637
24. Atzori M., Müller H. The Ninapro database: A resource for sEMG naturally controlled robotic hand prosthetics. In: *2015 The 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (*EMBC*). 2015:7151–7154. https://doi.org/10.1109/EMBC.2015.7320041
25. Andrianova E.G., Golovin S.A., Zykov S.V., Lesko S.A., Chukalina E.R. Review of modern models and methods of analysis of time series of dynamics of processes in social, economic and socio-technical systems. *Rossiiskii tekhnologicheskii zhurnal = Russian Technological Journal*. 2020;8(4):7–45 (in Russ.). https://doi.org/10.32362/2500-316X-2020-8-4-7-45
26. Nikonov V.V., Gorchakov A.V. Train machine learning models using modern containerization and cloud Infrastructure. *Promyshlennye ASU I kontrollery = Industrial Automated Control Systems and Controllers*. 2021;6:33–43 (in Russ.). https://doi.org/10.25791/asu.6.2021.1288
27. Kingma D.P., Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv.1412.6980.* 2014. https://doi.org/10.48550/arXiv.1412.6980
28. Wang Z., Bovik A.C. Mean squared error: Love it or leave it? A new look at signal fidelity measures. In: *IEEE Signal Processing Magazine*. 2009;26(1):98–117. https://doi.org/10.1109/MSP.2008.930649

About the authors

Anton V. Vasiliev, Postgraduate Student, Department of Applied Information Technologies, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: bysslaev@gmail.com. RSCI SPIN-code 4562-5628, https://orcid.org/0000-0001-6712-0072

Alexey O. Melnikov, Cand. Sci. (Eng.), Associate Professor, Department of Applied Information Technologies, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: melnikov.aleksey@gmail.com. https://orcid.org/0000-0003-1980-2727

Sergey A. Lesko, Cand. Sci. (Eng.), Associate Professor, Department of Applied Information Technologies, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: sergey@testor.ru. Scopus Author ID 57189664364, https://orcid.org/0000-0002-6641-1609

Об авторах

Васильев Антон Владимирович, аспирант кафедры «Прикладные информационные технологии» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: bysslaev@gmail.com. SPIN-код РИНЦ 4562-5628, https://orcid.org/0000-0001-6712-0072

Мельников Алексей Олегович, к.т.н., доцент кафедры «Прикладные информационные технологии» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: melnikov.aleksey@gmail.com. https://orcid.org/0000-0003-1980-2727

Лесько Сергей Александрович, к.т.н., доцент, доцент кафедры «Прикладные информационные технологии» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: sergey@testor.ru. Scopus Author ID 57189664364, https://orcid.org/0000-0002-6641-1609

Translated from Russian into English by Evgenii I. Shklovskii
Edited for English language and spelling by Thomas A. Beavitt

Information systems. Computer sciences. Issues of information security
Информационные системы. Информатика. Проблемы информационной безопасности

UDC 311.21+681.7.06
https://doi.org/10.32362/2500-316X-2023-11-2-20-32

RESEARCH ARTICLE

Bibliometric analysis of holographic data storage literature

Kutty Kumar [1, @], R. Parameswaran [2]

[1] *Library and Information Science, College of Veterinary Science, Sri Venkateswara Veterinary University, Proddatur-516360, India*
[2] *Central Library, Banaras Hindu University, Varanasi-221005, India*
[@] *Corresponding author, e-mail: kumarkkutty@gmail.com*

Abstract

Objectives. Snapshots of data can be stored in a holographic medium at varying depths. Data can be written via a spiral data channel in spinning holographic media in the form of circular disks like CDs or DVDs. This data is then read by shining a reference beam through the refraction following writing. However, holographic storage is distinct from CD/DVD media in the sense that information is encoded in all three dimensions. Two-dimensional data is written using a single laser beam that spirals around the material. Prototype holographic storage solutions use minuscule cones formed by individual snapshots or pages to store one million pixels. As compared with magnetic disks and tapes, which have a finite lifespan of 50 years at most, the longevity and dependability of optical media storage is advantageous for long-term archiving. Holographic technology allows for the portability of data-intensive media such as broadcast or high-definition video. However, the shelf life of holographic media remains low due to its sensitivity to light. The primary goals of most storage devices are more storage space and faster data transport. Holographic storage devices have the potential to outperform traditional optical storage devices both in terms of capacity and performance. The present paper aims to evaluate the current international research trends in Holographic Data Storage (HDS) and produce a graphical mapping of co-authorship and countries.

Methods. The major outputs of the dataset were authors, document type, publication, institution, nation, and citations. After exporting 1052 data sources, *HistCite* software was used to analyze the citations; visualization mapping was carried out using *VOSviewer* software and R programming language for the analysis of the author-country-title association on Holographic Storage Devices.

Results. The most prominent authors, papers, journals, organizations, and nations in the field of HDS were identified in *HistCite*. Then, four clusters were investigated using *VOSviewer* based on author keywords, citation collaboration networks among different organizations, countries, and the HDS co-authorship network.

Conclusions. During the study period from 2000–2020 (21 years), 4636 authors contributed to 1052 publications. The highest number of publications was in 2009, with a linear adjustment of $R^2 = 0.0136$. The most prolific author, Lee J., published 3.14% of the articles on this subject. In terms of country distribution, Japan took first-place ranking, claiming 16.54% of the total number of articles. The “holographic” keyword was used in 62.55% of the articles.

Keywords: holographic, data, storage, bibliometric, *HistCite*, *VOSviewer*

• **Submitted:** 22.09.2022 • **Revised:** 27.10.2022 • **Accepted:** 13.01.2023

For citation: Kumar K., Parameswaran R. Bibliometric analysis of holographic data storage literature. *Russ. Technol. J.* 2023;11(2):20–32. https://doi.org/10.32362/2500-316X-2023-11-2-20-32

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Библиометрический анализ литературы по голографическому хранению данных

K. Kumar [1, @], R. Parameswaran [2]

[1] *Кафедра библиотечно-информационных наук, Колледж ветеринарных наук, Ветеринарный университет Шри Венкатешвары, Проддатур-516360, Индия*
[2] *Центральная библиотека, Бенаресский индуистский университет, Варанаси-221005, Индия*
[@] *Автор для переписки, e-mail: kumarkkutty@gmail.com*

Резюме

Цели. Моментальные снимки данных можно хранить на голографических носителях на различной глубине. Они могут быть записаны по спиральному каналу передачи данных на вращающиеся голографические носители в виде круглых дисков, похожих на CD или DVD. После записи данные можно считать через просвечивание опорным лучом при помощи рефракции. В отличие от CD/DVD носителей, в голографических запоминающих устройствах информация кодируется во всех трех измерениях. Двумерные данные записываются с помощью одного лазерного луча, который закручивается по спирали вокруг материала. Для того чтобы сохранить один миллион пикселей, прототипы решений для голографического хранения данных использовали крошечные конусы, образованные отдельными снимками данных или страницами. По сравнению с магнитными дисками и кассетами, срок службы которых ограничен максимум 50 годами, долговечность и надежность оптических носителей информации имеет явное преимущество при долгосрочном архивировании. Голографическая технология обеспечивает переносимость носителей с большим объемом данных, таких как телепрограммы или видео высокой четкости. Однако срок годности голографических носителей остается низким из-за их чувствительности к свету. Основными целями использования большинства устройств хранения данных являются увеличение объема памяти и более быстрая передача данных. Голографические запоминающие устройства потенциально могут превзойти традиционные оптические устройства как по емкости, так и по производительности. Цель настоящей работы – оценить актуальные международные тенденции исследований в области голографического хранения данных и составить графическое отображение соавторства и стран.

Методы. Для анализа была осуществлена выборка данных, в которую вошли авторы, тип, количество публикаций, учреждение, страна, количество и место цитирований. После экспорта 1052 источников данных для анализа цитат использовалось программное обеспечение *HistCite*; визуализация была выполнена с использованием программного обеспечения *VOSviewer* и языка программирования R для анализа ассоциации «автор – страна – название» о голографическом хранении данных.

Результаты. При помощи *HistCite* были определены наиболее значимые авторы, статьи, журналы, организации и страны в области голографического хранения данных. Затем, используя *VOSviewer*, мы исследовали четыре кластера, основанных на авторских ключевых словах, сетях сотрудничества по цитированию между различными организациями, странами, а также сетями соавторов, пишущих о голографическом хранении данных.

Выводы. За период исследования с 2000 по 2020 гг. (21 год) 4636 авторов написали 1052 публикации. Наибольшее количество публикаций было издано в 2009 г. с коэффициентом детерминации $R^2 = 0.0136$. Наиболее продуктивный автор, Джей Ли, опубликовал 3.14% статей по голографическому хранению данных. С точки зрения распространения по странам первое место в рейтинге заняла Япония с 16.54% от общего количества статей. Ключевое слово «голографический» использовалось в 62.55% статей.

Ключевые слова: голографический, данные, хранилище, библиометрический, *HistCite*, *VOSviewer*

• **Поступила:** 22.09.2022 • **Доработана:** 27.10.2022 • **Принята к опубликованию:** 13.01.2023

Для цитирования: Kumar K., Parameswaran R. Bibliometric analysis of holographic data storage literature. *Russ. Technol. J*. 2023;11(2):20–32. https://doi.org/10.32362/2500-316X-2023-11-2-20-32

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Since the advent of the digital age, the wealth of knowledge available to the general public has grown exponentially. This shift has been largely due to declining data storage costs and the increasing storage capacities of smaller devices. Although capable of satisfying current storage needs, the data storage industry will need to invest in new technologies if it wants to keep up with rising demand. Magnetic and classical optical data storage technologies store bits of information as individual magnetic or optical changes on a recording media. However, both approaches are starting to reach physical limits beyond which it becomes impossible to encode and thus store individual bits. A novel and promising high-capacity option is optical data storage that is distributed through the bulk of a medium as opposed to being limited to its two-dimensional surface. Significant progress toward practical utilization of holographic data storage (HDS) technology due to the recent emergence of cheaper supporting technologies and major findings from various research initiatives, which has generated conceptual breakthroughs [1]. The high density of holographic 3D memories is obtained by superimposing numerous holograms inside the same volume of the recording material. Thus, data storage across three dimensions becomes a practical option for next-generation memory storage [2]. Given the expected development of such systems at a comparable cost with current technology, along with the optimization and standardization of storage media, HDS could one day overtake magnetic and traditional optical data storage solutions as the industry standard for high-capacity data storage [3]. A comparable HDS system can store the equivalent of data from over 1000 CDs, as well as offering benefits over and above those of a traditional storage system. Based on the research collated here, HDS represents a novel three-dimensional data storage system offering significant advantages over conventional read/write memory systems. Some fundamental characteristics of HDS will be discussed along with an examination of potential uses for HDS in modern computing systems. Since the primary focus of this paper is to examine the current state of HDS research around the world, we present a visualization network map to show relationships between authors and their home countries, along with their most frequently used terms, as well as referenced sources and the authors who most frequently cited them.

Brief introduction to holographic data storage

Holographic data storage comprises a high-volume data storage technology that creates holograms of each data instance. Like traditional optical storage devices, it stores high volumes of data in single volume also called 3D Storage [4]. Both write-once and rewriteable holographic media are possible (changes are reversible), the latter using crystal photorefractive impact. The memory architecture is comprised of a blue-green argon laser, beam splitters, reflectors, LCD board, lenses, lithium-niobite crystal, and a charge-coupled device camera. The blue-green argon laser shaft splits into two beams: a signal beam, which goes straight ahead, and a reference beam, which is controlled by a beam splitter. LCD screens reflect signals into lithium-niobite crystals. The reference beam showed crystal from a new angle. When the two beams meet, the signal beam would hologram information [1, 5, 6].

Uses of HDS

Data mining: HDS can be used to identify patterns more rapidly. Large databases with hidden patterns benefit from data mining. However, data mining on PCs places a heavy burden on data storage systems, whereas holographic memory could speed up data mining by improving access and storage [7].

Petaflop operations: a computer's processing speed with HDS is 1000 trillion floating point operations/s. Holographic memory frameworks speed up data access. Holographic memory can handle massive data requirements and be used instead of 10ns DRAM, hard drives, CD ROMs, and rock-mounted petabytes of storage [8].

Benefits of HDS

One terabyte can be stored in holographic memory. Unlike the fastest hard disk, which typically offer data access times of 5–10 ms, data recovery takes place in the microsecond range. For example, HDS can transport

Table 1. HistCite indicators

Ser. No.	Indicator	Definition
1	TLCS/TGCS	All records from an author, source, or other category, or all records, have a total score. TLCS = Total Local Citation Scores. TGCS = Total Global Citation Scores
2	Recs	The number of records where a given item is found is shown in the "number of records" column
3	LCS	The number of citations to work within the collection is represented in the Local Citation Score
4	GCS	The Global Citation Score shows all Web of Science Core Collection citations
5	LCR	Local Cited References shows a paper's reference list's number of collection-wide citations

a film in under a minute and two pages can be compared optically in holographic capacity without data recovery. HDS is also motionless. Thus, mechanical constraints like erosion can be removed [2]. Another advantage of HDS is the ability to recover data from damaged media.

A variety of research frameworks have been proposed for HDS [9, 10]. Various linked applications of HDS analyzed by researchers include digital holograms [11], deep learning [12, 13] modulation code [14], digital watermarking [15], convolutional neural networks [16], and data compression methods [17]. Other research sets out to explore such aspects of HDS as fundamental issues [18], holographic memory [19], and bit error prediction [20], as well as focusing on holographic grating [21], fluid dynamics [22], and optical storage [23]. To the best of the present authors' knowledge, there is no published bibliometric analysis of the holographic data storage literature. As a result, the purpose of the present work is to present a comprehensive bibliometric analysis of HDS studies, indicating current research trends by emphasizing significant research contributions. Our analysis is based on publishing data, citation distributions and statistics, regional and institutional productivity, research topics, impact journals, and keyword frequency. This work will pave the way for future HDS research by outlining research gaps and obstacles in extended storage.

METHODOLOGY

On June 17, 2021, the Web of Science core collection database was accessed alongside a comprehensive web search using the phrases "holographic data storage", "high capacity", "magnetic and optical data storage systems"[1]. The dataset was refined by document type, language, and duplication criteria. Authors, document type, publication, institution, nation, and citations were the major outputs of the dataset. After extracting data from 1052 sources, we used *HistCite*[2] to examine references, *VOSviewer* software[3] to create maps for analysis, and the R programming language to examine the connections between authors, nations, and titles in the Holographic archive. A visualization mapping was produced on the basis of co-authorship country, co-occurrence author keywords, and co-citation cited sources. The process of selecting publications is depicted in Fig. 1.

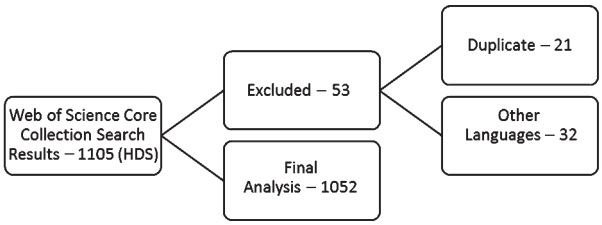

Fig. 1. Holographic data storage publications

Indicators in *HistCite*

The *HistCite* program is used in historiography analysis to organize bibliographic collections obtained by querying the Web of Science's Science Citation Index (WoS). It presents a visual display of the most influential publications on a subject chronology, as well as the evolution of articles, authors, and journals. The current contribution is based on version 12.3. Table 1 lists the *HistCite* indicators [24] employed in this study.

RESULTS

Over the previous 20 years, the number of published papers related to HDS has increased worldwide as shown in Fig. 2. The highest number of publications was 8.56% in 2009; the largest number of worldwide

[1] Clarivate. Web of Science Core Collection. Web of Science Group. 2021. https://clarivate.com/webofsciencegroup/solutions/web-of-science-core-collection/. Accessed July 05, 2021.

[2] Garfield E. HistCite®. *Bibliographic Analysis and Visualization Software*. https://garfield.library.upenn.edu/histcomp/. Accessed July 05, 2021.

[3] Van Eck N.J., Waltman L. *VOSViewer*. (*Visualizing Scientific Landscapes*). 2010. https://www.vosviewer.com. Accessed October 06, 2021.

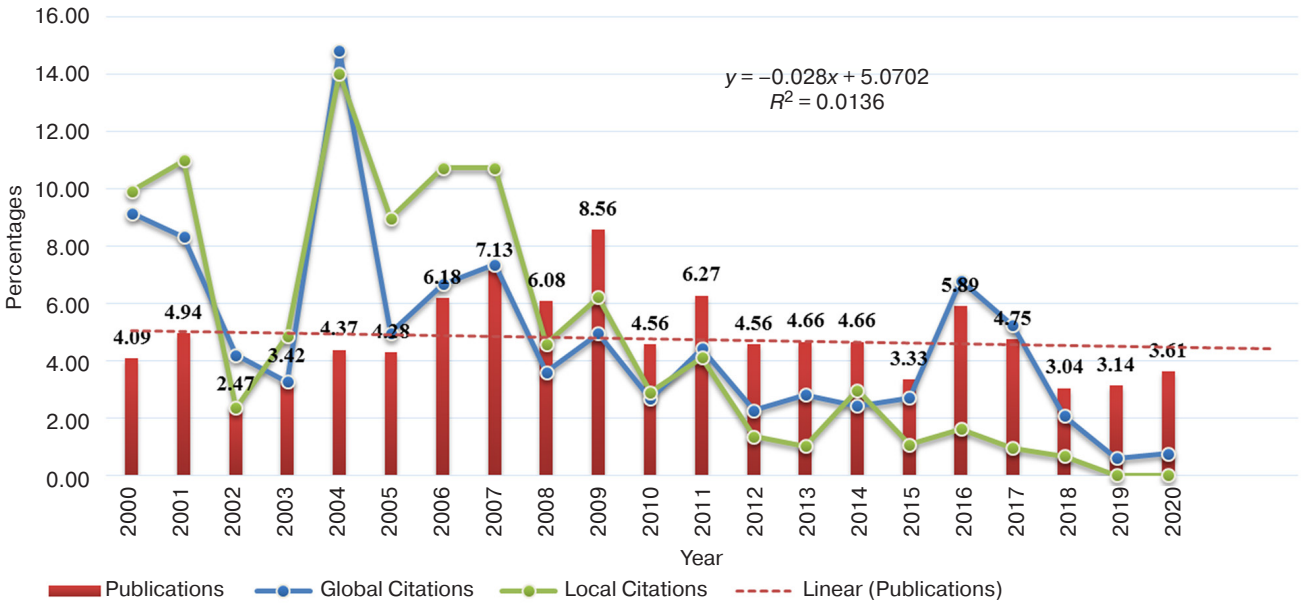

Fig. 2. Growth of HDS publications with citations

citations (14.82%) and local citations (14.03%) occurred in 2004. A linear adjustment of the observed variables $R^2 = 0.0136$ yielded a percentage of average years from the publication of 11.2%, average citations per document of 16.59%, and average citations per year of 1.49%.

As can be seen in Table 2, the 1052 publications had a total of 2236 authors. There were 51 single-authored works and 2185 collaborative works. The most cited author, Lee J., has 33 records (3.14%), a total citation score of 16, and a total citation score of 180 from all over the world. There were two researchers with the highest overall local citation scores: Jin G.F. (29 records, 2.76%) and Yang H. (27 records, 2.57%); their global citation totals were 28 and 260, respectively.

The authorship-collaboration network generated by *VOSviewer* is depicted in Fig. 3. Minimum documents with five authors were chosen for the co-authorship graph approach; of those, 157 met the criteria. A total of 17 authors are evenly distributed throughout four groups (red, green, blue, and yellow). Tan X. has 13 links, 61 total link strength, and 20 documents, making up the majority of cluster-1 co-authorship patterns.

Figure 4 depicts a Sankey diagram of the Authors (left) between Countries (middle) and Titles (right) relationship in HDS literature. The study revealed which titles HDC authors had published most frequently, along with the specific HDS research areas (keywords). An analysis of top authors, titles, and keywords reveal that three authors Lin S.H., Belendez A., and Gallego S., representing the three nations Japan, China, and the United States, had a strong association with the HDS research keywords “holographic”, “storage”, and “data”.

Table 2. Highly productive ten authors on HDS

Ranking	Author	Recs	%	TLCS	TGCS
1	Lee J.	33	3.14	16	180
2	Jin G.F.	29	2.76	28	260
3	Yang H.	27	2.57	4	11
4	Tan X.D.	26	2.47	104	734
5	He Q.S.	25	2.38	25	205
6	Pascual I.	24	2.28	71	387
7	Belendez A.	22	2.09	71	392
8	Cao L.C.	22	2.09	25	210
9	Sheridan J.T.	22	2.09	78	798
10	Ishii N.	20	1.90	10	80

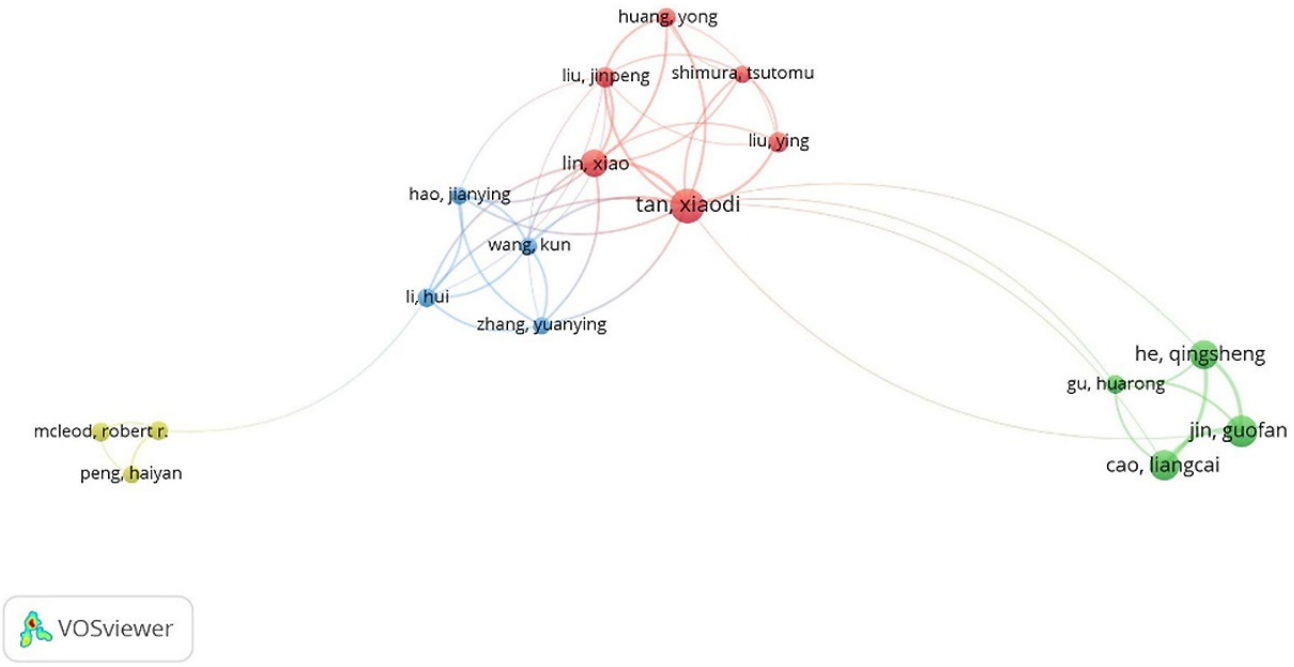

Fig. 3. Co-Authorship network of HDS

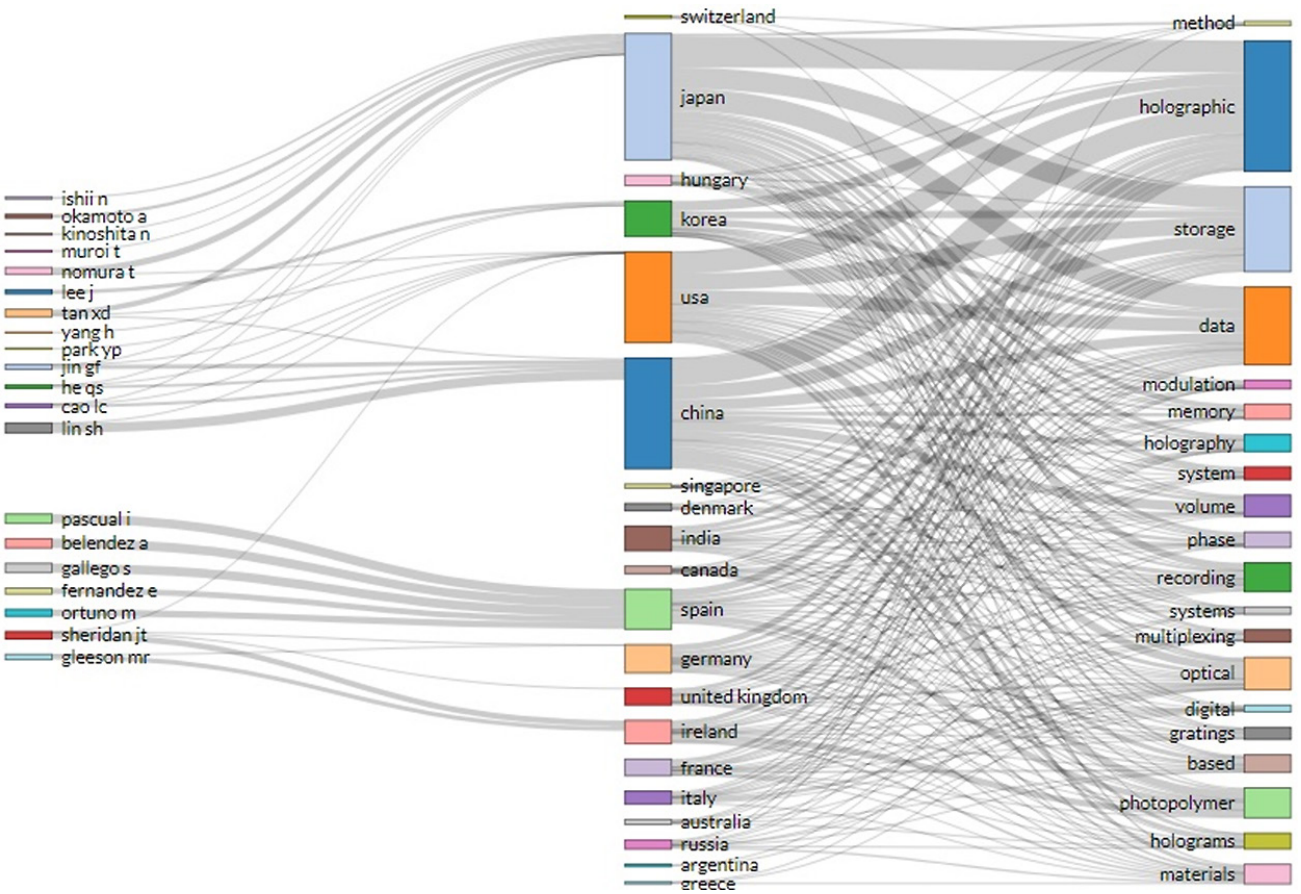

Fig. 4. Sankey diagram displaying the connection between authors (left), countries (center), and titles (right)

Table 3. Distribution of HDS document types

Ser. No.	Document type	Recs	%	TLCS	TGCS
1	Article	960	91.25	1226	13160
2	Proceedings Paper	60	5.70	134	1668
3	Review	27	2.57	100	2460
4	Editorial Material	2	0.19	5	22
5	Review; Book Chapter	2	0.19	2	64
6	News Item	1	0.10	16	79

Table 4. Distribution of HDS records by country

Ranking	Country	Recs	%	TLCS	TGCS
1	Japan	174	16.54	382	3528
2	USA	123	11.69	401	4735
3	China	96	9.13	87	2530
4	Germany	51	4.85	120	2538
5	South Korea	38	3.61	55	789
6	Ireland	31	2.95	108	1169
7	Spain	28	2.66	81	1207
8	UK	27	2.57	67	1431
9	Taiwan	23	2.19	106	548
10	Canada	18	1.71	31	1074

According to Table 3, HDS documents are broken down into six distinct article types, with journals making up the vast majority (91.25%). In total, 1226 HDS documents have a local citation score, while the total number of HDS documents having a global citation score is 13160. There are a total of 16 citations in the local area, and 79 citations in the global arena, making this item the one with the lowest citation score.

HDS publications featured contributions by researchers from 40 different countries. Table 4 covers all countries that contribute to the effectiveness of HDS research publications. Japan tops the list with 174 (16.54%) publications, 382 total local citation scores, and 3528 total global citation scores. It is followed by the United States in second place with 123 (11.69%), 401 total local citation scores, and a total global citation score of 4735, while China is in third place with 96 (9.13%), a total local citation score of 87, and a total global citation score of 2530.

Figure 5 depicts a citation of the countries network graph created using the full counting approach and a minimum of five countries' papers. Only 27 of the 50 countries fit the criteria. Six clusters grouped 26 countries as follows: Cluster 1 (red color, 8 countries)—Belgium, Canada, Denmark, England, Latvia, Netherlands, Poland, and Russia; Cluster 2 (green color, 6 countries)—Bulgaria, Germany, Ireland, Italy, Spain, and Ukraine; Cluster 3 (blue color, 5 countries)—Australia, South Korea, Switzerland, Türkiye, and USA; Cluster 4 (yellow color, 3 countries)—Hungary, India, and Japan; Cluster 5 (purple color, 3 countries)—France, China, and Singapore; Cluster 6 (light blue color, 1 country)—Taiwan.

Overall, 1343 different keywords were used by researchers who contributed to 42 different HDS research publications. Table 5 has 658 (62.55%) documents, the majority of which are holographic studies; the total local citation score for these records is 1031; the total global citation score for these records was 7717; the last position is held by 'System' with 96 (9.13%) records, which have a local score of 130 and a global score of 740. The number of occurrences of the minimum set of five keywords is shown in Fig. 6. In Cluster 7 (orange), the term "Holographic Data Storage" appeared 105 times, was linked 30 times, and had an average link strength of 76.

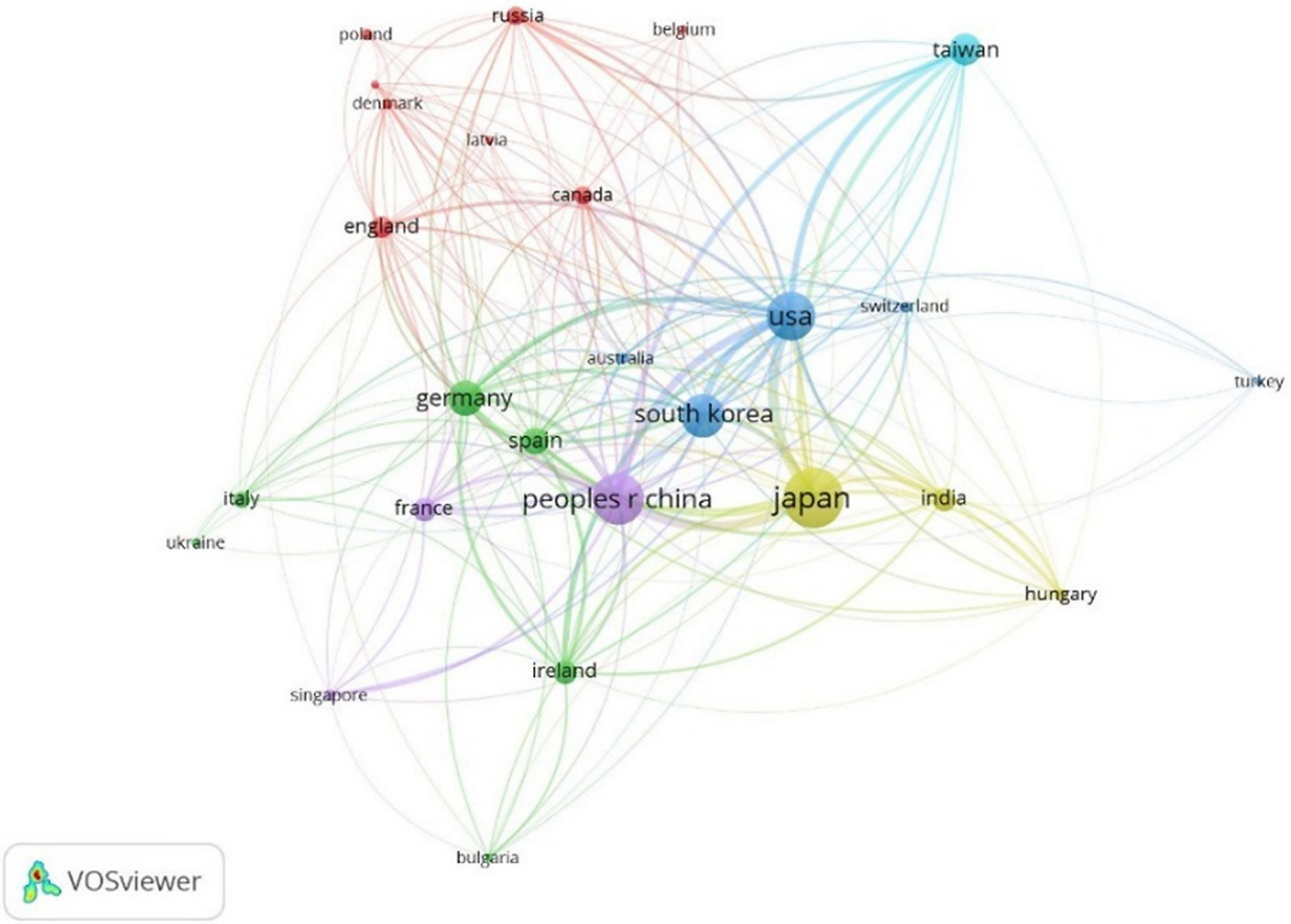

Fig. 5. Citation from different countries on HDS research

Table 5. Predominant author keyword in HDS research

Ranking	Word	Recs	%	TLCS	TGCS
1	Holographic	658	62.55	1031	7717
2	Storage	428	40.68	866	5501
3	Data	397	37.74	775	4722
4	Optical	139	13.21	206	2765
5	Recording	129	12.26	176	1364
6	Using	114	10.84	135	1231
7	Based	104	9.89	113	1763
8	Photopolymer	102	9.70	285	2030
9	Phase	101	9.60	159	1246
10	System	96	9.13	130	740

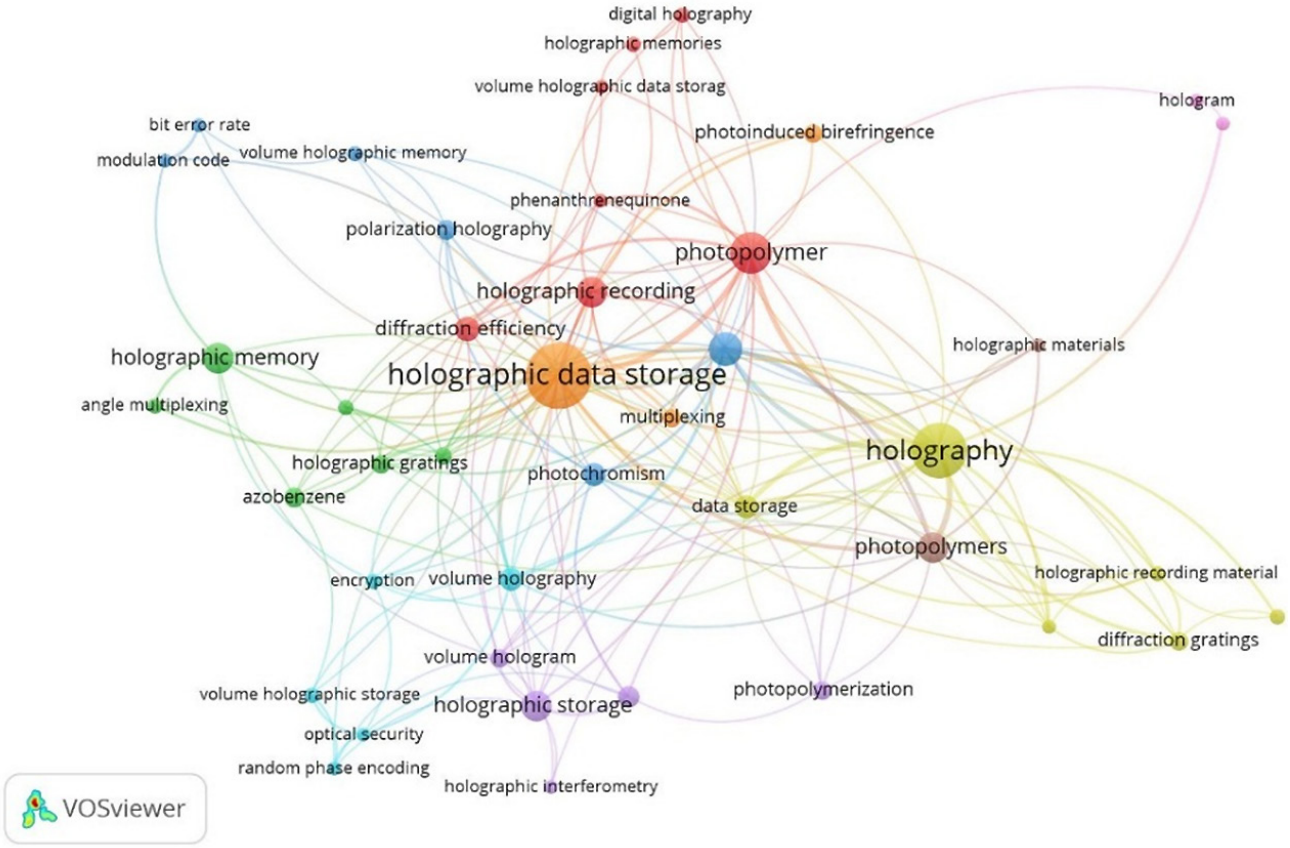

Fig. 6. Author keywords in HDS research

Table 6. Most productive institutions in HDS related publications

Ranking	Institution	Country	Recs	%	TLCS	TGCS
1	University of Alicante	Spain	18	1.71	70	408
2	University College Dublin	Ireland	18	1.71	63	698
3	National Chiao Tung University	China	14	1.33	97	418
4	Tsinghua University	China	14	1.33	18	226
5	Harbin Institute of Technology	China	13	1.24	23	334
6	Sony Corporation	Japan	13	1.24	83	312
7	Beijing Institute of Technology	China	11	1.05	10	258
8	University of Bayreuth	Germany	11	1.05	40	501
9	University of Birmingham	United Kingdom	11	1.05	10	660
10	Wakayama University	Japan	11	1.05	21	516

Table 6 lists the names of 630 different institutions, each of which has contributed to at least five different HDS publications. The list was topped by the University of Alicante in Spain, which had 18 records (1.71%), 70 total local citation scores, and 408 total worldwide citation ratings; University College Dublin was second, with 18 records, 63 total citation scores, and 698 total global citation ratings. Figure 7 depicts the HDS citation cooperation network, which organizes 83 organizations into 7 clusters; here, the “National Chiao Tung University” (Cluster 5, purple) is the most productive citation collaboration organization with 31 publications, 38 links, and a total link strength of 211.

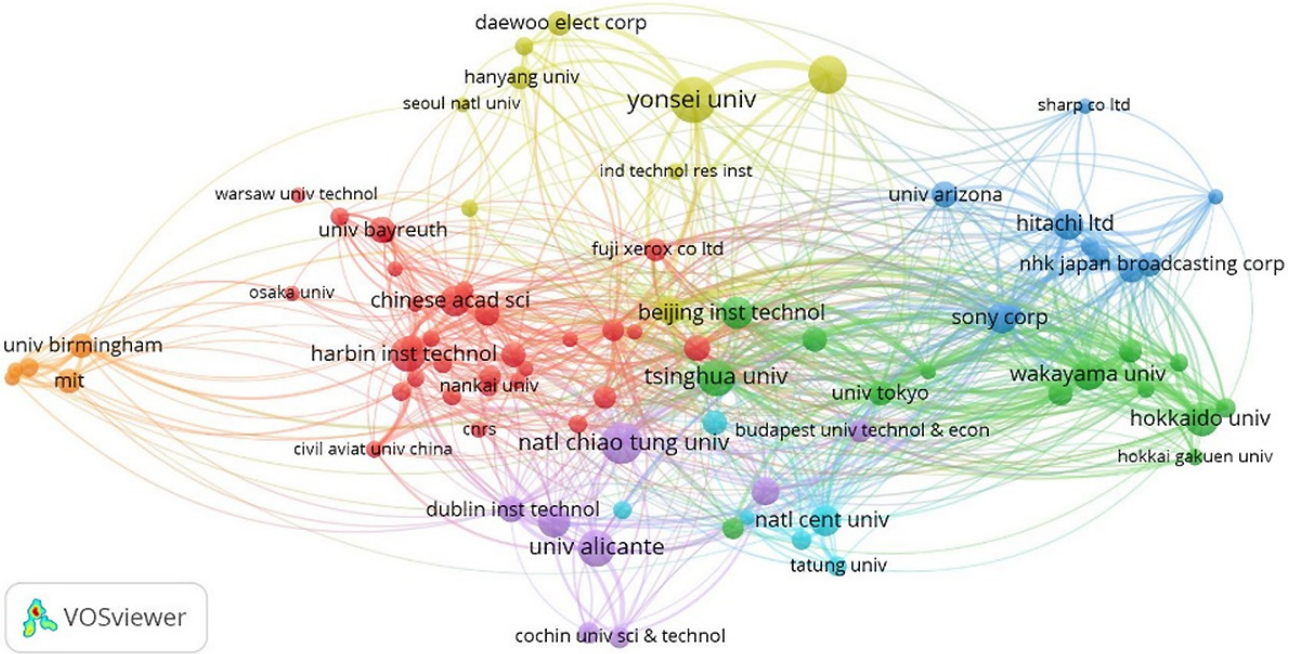

Fig. 7. Citation Collaboration Network of Different Organization in HDS related publications

Table 7. Journal productivity in HDS

Ranking	Journal	Recs	%	TLCS	TGCS
1	Japanese Journal of Applied Physics	119	11.31	22	375
2	Applied Optics	112	10.65	350	2151
3	Optics Express	65	6.18	98	846
4	Optics Communications	48	4.56	67	488
5	Optics Letters	47	4.47	186	1295
6	Japanese Journal of Applied Physics Part 1-Regular Papers Brief Communications & Review Papers	32	3.04	54	213
7	Optical Engineering	30	2.85	41	324
8	Journal of Optics A-Pure and Applied Optics	20	1.90	21	270
9	Optical Review	20	1.90	13	61
10	Microsystem Technologies-Micro-and Nanosystems-Information Storage and Processing Systems	17	1.62	3	12

Table 7 is a list of all peer-reviewed journals that have published at least 15 articles in the last two years. There are a total of 119 published papers (11.31%) published in the Japanese Journal of Applied Physics (1.471 impact factor), which have been cited a total of 22 times locally and 375 times globally. The Applied Optics journal (1.961 impact factor) is in second position with 112 (10.65%) publications, 350 total local citations, and 2151 global citations; Optics Express Journal (3.669 impact factor) is in third place with 65 (6.18%) publications and a total of 98846 global citations.

HistCite graph marker (Fig. 8) exhibited the most cited 50 articles (nodes) with 82 links (relationship among articles by local citation score) with a maximum of 62 and a minimum of 7 citations. These publications had well-integrated citation mapping, implying the notability of published HDS works that referencing them. It can be noted that most citations were made during the first twelve years of the study period (2000 to 2011). A strong separation between one inspired research work (number 182) and multiple linkages was revealed by the *HistCite* citation mapping (numbers 194, 226, 25, and 46). The top five citations are detailed in Table 8.

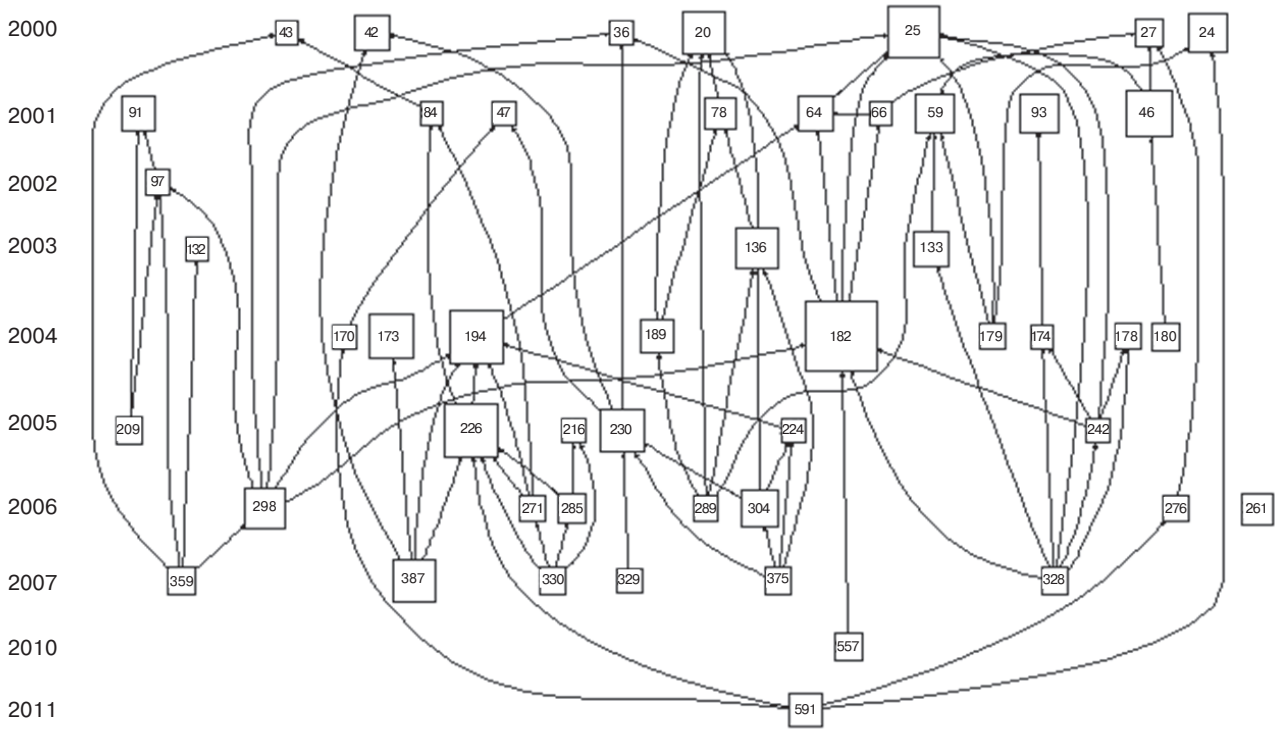

Fig. 8. Total local citation score mapping for HDS

Table 8. Top five citations in HDS

Record	Author	Title	Source	Year	DOI	LCS	GCS
182	Hesselink L., Orlov S.S., Bashaw M.C.	Holographic data storage systems	Proceedings of the IEEE	2004	http://doi.org/10.1109/JPROC.2004.831212	62	258
194	Orlov S.S., Phillips W., Bjornson E., Takashima Y., Sundaram P., Hesselink L., Okas R., Kwan D., Snyder R.	High-transfer-rate high-capacity holographic disk data-storage system	Applied Optics	2004	http://doi.org/10.1364/AO.43.004902	36	135
226	Horimai H., Tan X.D., Li J.	Collinear holography	Applied Optics	2005	http://doi.org/10.1364/AO.44.002575	36	221
25	Ashley J., Bernal M.P., Burr G.W., Coufal H., Guenther H., Hoffnagle J.A., Jefferson C.M., Marcus B., Macfarlane R.M., Shelby R.M., Sincerbox G.T.	Holographic data storage	IBM Journal of Research and Development	2000	http://doi.org/10.1147/rd.443.0341	33	219
46	Lawrence J.R., O'Neill F.T., Sheridan J.T.	Photopolymer holographic recording material	Optik	2001	http://doi.org/10.1078/0030-4026-00091	27	186

The present research employed Local Cited References (LCR), which displays the number of citations in a paper's reference list to other papers in the collection, to predict where the field of holographic data storage is headed. Article rankings according to LCR are displayed in Table 9. A good example is the 26 articles cited by "Holographic polymer materials with diffusion development: principles, arrangement, investigation, and applications " by A.V. Veniaminov and U.V. Mahilny. This suggests that articles using similar data are both highly relevant to the topic and likely to be current articles; moreover, in recent years there have been numerous publications on the topic, resulting in more frequent mentions.

Table 9. Top future direction articles in HDS

Ranking	Record	Author	Title	Source	Year	DOI	LCR
1	750	Veniaminov A.V., Mahilny U.V.	Holographic polymer materials with diffusion development: principles, arrangement, investigation, and applications	Optics and Spectroscopy	2013	http://doi.org/10.1134/S0030400X13120199	26
2	524	Das B., Joseph J., Singh K.	Phase-image-based sparse-gray-level data pages for holographic data storage	Applied Optics	2009	http://doi.org/10.1364/AO.48.005240	19
3	939	Nobukawa T., Barada D., Nomura T., Fukuda T.	Orthogonal polarization encoding for reduction of interpixel cross talk in holographic data storage	Optics Express	2017	http://doi.org/10.1364/OE.25.022425	17
4	921	Malallah R., Li H.Y., Kelly D.P., Healy J.J., Sheridan J.T.	A review of hologram storage and self-written waveguides formation in photopolymer media	Polymers	2017	http://doi.org/10.3390/polym9080337	16
5	591	Bruder F.K., Hagen R., Rolle T., Weiser M.S., Facke T.	From the surface to volume: concepts for the next generation of optical-holographic data-storage materials	Angewandte Chemie-International Edition	2011	http://doi.org/10.1002/anie.201002085	15

CONCLUSIONS

The present article contributes to an understanding of the HDS literature by grouping publications into clusters and identifying new research streams. Using *HistCite*, the study reveals the most prominent authors, papers, journals, organizations, and nations in the field of HDS. Then, using *VOSviewer*, we investigated four clusters based on author keywords, citation collaboration networks among different organizations, countries, and the co-authorship network of HDS. Further, the relationship between authors, countries and titles on HDS literature was established using the R programming language. Finally, the study analyzed the Total Local Citation Score Mapping for HDS, Top 5 citations and subjects of future research on HDS using *HistCite* software. HDS is a new concept that has become a big phenomenon in today's digital world. The most significant issue in the study was that during citation mapping and cluster analysis, only publications having a minimum of seven and a maximum of sixty-two citations were included. Regardless of their proportional contribution, recent papers could not demonstrate their true potential in this manner. Since more than 4.75% of the publications in this study (50 of 1052) were published within the last decade (2000–2011), the bibliometric analysis should be carried out again in the future to recognize new structures having made a mark in the field. Second, non-WoS source databases are not supported by *HistCite*. As a result, this investigation concentrated entirely on WoS publications, yielding articles from respected journals. As a result, there may be a bias against non-WoS journal papers and high-quality publications may present information not found in our study that may sway HDS opinions. It is feasible that future bibliometric software will be created to permit the inclusion of smaller journal publications. Using bibliometric meta-analyses, researchers will be able to compare the construct influences in WoS and non-WoS works, as well as applying them to a wider range of fields.

Authors' contribution

All authors equally contributed to the research work.

All authors approved the final text of the manuscript for publication.

REFERENCES

1. Ashley J., Bernal M.P., Burr G.W., Coufal H., Guenther H., Hoffnagle J.A., et al. Holographic data storage technology. *IBM J. Res. Dev*. 2000;44(3):341–368. https://doi.org/10.1147/rd.443.0341
2. Psaltis D., Burr G.W. Holographic data storage. *Computer*. 1998;31(2):52–60. https://doi.org/10.1109/2.652917
3. Azami M., Farha T.A., Taj A.S., Punitha B. Holographic mass storage system. *Journal of Emerging Technologies and Innovative Research* (*JETIR*). 2019;6(9):47–49. Available from URL: https://jetir.org/papers/JETIRDD06009.pdf
4. Leith E.N., Kozma A., Upatnieks J., Marks J., Massey N. Holographic data storage in three-dimensional media. *Appl. Opt*. 1966;5(8):1303–1311.
5. Burr G.W., Coufal H., Hoffnagle J.A., Michael C., Jurich M., Macfarlane R.M., et al. High–density and high–capacity holographic data storage. *Asian Journal of Physics, Special Issue on Optical Information Technology*. 2001;10(1):28 p. Available from URL: http://geoffreyburr.org/papers/ajp_review.pdf
6. Timucin D.A., Downie J.D. Holographic optical data storage. *IEEE Potentials*. 2000;19(4):32–36. https://doi.org/10.1109/45.877865
7. Nakamura Y. Magnetic holography and its application to data storage. *Photonics*. 2021;8(6):187. https://doi.org/10.3390/photonics8060187

8. Ebisuzaki T., Germain R., Taiji M. PetaFLOPS computing. *Commun. ACM*. 2004;47(11):42–45. https://doi.org/10.1145/1029496.1029524
9. Hesselink L., Orlov S.S., Bashaw M.C. Holographic data storage systems. In: *Proc. IEEE*. 2004;92(8):1231–1280. https://doi.org/10.1109/JPROC.2004.831212
10. Daiya K., Chouhan B., Rathi P. Holographic data storage. *Int. J. Eng. Res. Technol*. (*IJERT*). 2014;2(10):179–182. Available from URL: https://www.ijert.org/research/holographic-data-storage-IJERTCONV2IS10050.pdf
11. Cheremkhin P.A., Kurbatova E.A. Optical dynamic reconstruction of quantized digital and computer-generated holograms. In: *2018 International Conference Laser Optics* (*ICLO*). St. Petersburg: IEEE; 2018. P. 203–203. https://doi.org/10.1109/LO.2018.8435412
12. Ko H., Kim H.Y. Deep learning-based compression for phase-only hologram. *IEEE Access*. 2021;9:79735–79751. https://doi.org/10.1109/ACCESS.2021.3084800
13. Kim E., Park S., Hwang S., Moon I., Javidi B. Deep learning-based phenotypic assessment of red cell storage lesions for safe transfusions. *IEEE J. Biomed. Health Inform*. 2022;26(3):1318–1328. https://doi.org/10.1109/JBHI.2021.3104650
14. Nguyen C.D., Xuan Pham N., Duong C.C., Cong Nguyen L. Multilevel modulation coding for four-level holographic data storage systems. In: *2020 International Conference on Advanced Technologies for Communications* (*ATC*). IEEE; 2020. P. 30–34. https://doi.org/10.1109/ATC50776.2020.9255459
15. Qiu S., Yuan H., Zhu G., Dai D., Xu H., Zhang X.F. Network privacy information storage model based on holographic digital watermarking technology. In: *2021 6th International Conference on Smart Grid and Electrical Automation* (*ICSGEA*). IEEE; 2021. P. 392–396. Available from URL: https://doi.ieeecomputersociety.org/10.1109/ICSGEA53208.2021.00095
16. Katano Y., Muroi T., Kinoshita N., Ishii N. Demodulation of multi-level data using convolutional neural network in holographic data storage. In: *2018 Digital Image Computing: Techniques and Applications* (*DICTA*). IEEE; 2018. P. 1–5. https://doi.org/10.1109/DICTA.2018.8615863
17. Youssef A., Heshmat S. 3D holographic compression methods for real time applications. In: *2018 International Conference on Innovative Trends in Computer Engineering* (*ITCE*). IEEE; 2018. P. 136–139. https://doi.org/10.1109/ITCE.2018.8316612
18. Hesselink L. Fundamental issues related to digital holographic data storage. In: *1999 IEEE LEOS Annual Meeting Conference Proceedings. LEOS'99. 12th Annual Meeting.* IEEE Lasers and Electro-Optics Society 1999 Annual Meeting (Cat No99CH37009). IEEE; 1999. P. 327–328. https://doi.org/10.1109/LEOS.1999.813615
19. Tien-Hsin Chao. Compact digital holographic memory using blue diode laser. In: *Symposium Non-Volatile Memory Technology* 2005. IEEE; 2005 P. 72–74. https://doi.org/10.1109/NVMT.2005.1541403
20. Trelewicz J.Q., Cochran D. Bit error prediction for digital image data. In: *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP '98 (Cat No98CH36181). IEEE; 1998. P. 2645–2648. https://doi.org/10.1109/ICASSP.1998.678066
21. Sreebha A.B., Suresh S., Sreekala C.O., Pillai V.P.M. Volume holographic gratings in acrylamide-based photopolymer to provide selective light as an added input for improving the performance of dye-sensitized solar cells. *Curr. Sci*. 2018;114(11):2267–2272. Available from URL: https://www.currentscience.ac.in/Volumes/114/11/2267.pdf
22. Orlov S.S., Abarzhi S.I., Oh S.B., Barbastathis G., Sreenivasan K.R. High-performance holographic technologies for fluid-dynamics experiments. *Philos. Trans. A Math. Phys. Eng. Sci*. 2010;368(1916):1705–1737. Available from URL: http://www.jstor.org/stable/25663339
23. Tchalakov I. The history of holographic optical storage at the both sides of the iron curtain—1969–1989. *The Journal of the International Committee for the History of Technology* (*ICON*). 2005;11:95–119. Available from URL: http://www.jstor.org/stable/23787025
24. Barreiro E.W. *Using HistCite software to identify significant articles in subject searches of the Web of Science*. ArXiv151207069 Cs [Internet]. 2015 Dec 22 [cited 2021 Jul 5]; Available from URL: https://arxiv.org/ftp/arxiv/papers/1512/1512.07069.pdf

About the authors

Kutty Kumar, PhD, Assistant Professor, Department of Library and Information Science, College of Veterinary Science, Sri Venkateswara Veterinary University (New Building, Gopavaram Village, Korapadu Road, Proddatur-516360, Andhra Pradesh, India). E-mail: kumarkkutty@gmail.com. Scopus Author ID 55040539500, https://orcid.org/0000-0002-3510-5924

R. Parameswaran, PhD, Deputy Librarian, Central Library, Banaras Hindu University (Ajagara, Varanasi, Uttar Pradesh, 221005 India). E-mail: parameshwaranblu@gmail.com. Scopus Author ID 56009308100, https://orcid.org/0000-0001-5799-1472

Об авторах

K. Kumar, PhD, доцент кафедры библиотечно-информационных наук, Колледж ветеринарных наук, Ветеринарный университет Шри Венкатешвары (Проддатур, штат Андхра Прадеш, 516360 Индия). E-mail: kumarkkutty@gmail.com. Scopus Author ID 55040539500, https://orcid.org/0000-0002-3510-5924

R. Parameswaran, PhD, заместитель библиотекаря, Центральная библиотека, Бенаресский индуистский университет (Аджагара, Варанаси, штат Уттар Прадеш, 221005 Индия). E-mail: parameshwaranblu@gmail.com. Scopus Author ID 56009308100, https://orcid.org/0000-0001-5799-1472

The text was submitted by the authors in English
Edited for English language and spelling by Thomas A. Beavitt

Russian Technological Journal. 2023;11(2):33–49 ISSN 2500-316X (Online)

UDC 004.652
https://doi.org/10.32362/2500-316X-2023-11-2-33-49

REVIEW ARTICLE

Models and methods for analyzing complex networks and social network structures

Julia P. Perova [@], Vitaly R. Grigoriev, Dmitry O. Zhukov

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding author, e-mail: perova_yu@mirea.ru

Abstract

Objectives. The study aimed to investigate contemporary models, methods, and tools used for analyzing complex social network structures, both on the basis of ready-made solutions in the form of services and software, as well as proprietary applications developed using the Python programming language. Such studies make it possible not only to predict the dynamics of social processes (changes in social attitudes), but also to identify trends in socioeconomic development by monitoring users' opinions on important economic and social issues, both at the level of individual territorial entities (for example, districts, settlements of small towns, etc.) and wider regions.

Methods. Dynamic models and stochastic dynamics analysis methods, which take into account the possibility of self-organization and the presence of memory, are used along with user deanonymization methods and recommendation systems, as well as statistical methods for analyzing profiles in social networks. Numerical modeling methods for analyzing complex networks and processes occurring in them are considered and described in detail. Special attention is paid to data processing in complex network structures using the Python language and its various available libraries.

Results. The specifics of the tasks to be solved in the study of complex network structures and their interdisciplinarity associated with the use of methods of system analysis are described in terms of the theory of complex networks, text analytics, and computational linguistics. In particular, the dynamic models of processes observed in complex social network systems, as well as the structural characteristics of such networks and their relationship with the observed dynamic processes including using the theory of constructing dynamic graphs are studied. The use of neural networks to predict the evolution of dynamic processes and structure of complex social systems is investigated. When creating models describing the observed processes, attention is focused on the use of computational linguistics methods to extract knowledge from text messages of users of social networks.

Conclusions. Network analysis can be used to structure models of interaction between social units: people, collectives, organizations, etc. Compared with other methods, the network approach has the undeniable advantage of operating with data at different levels of research to ensure its continuity. Since communication in social networks almost entirely consists of text messages and various publications, almost all relevant studies use textual analysis methods in conjunction with machine learning and artificial intelligence technologies. Of these, convolutional neural networks demonstrated the best results. However, the use of support vector and decision tree methods should also be mentioned, since these contributed considerably to accuracy. In addition, statistical methods are used to compile data samples and analyze obtained results.

Keywords: social networks, modeling of social processes, oriented graphs, multilayer convolutional neural network, computational linguistics, clustering

• **Submitted:** 07.12.2021 • **Revised:** 23.12.2022 • **Accepted:** 09.02.2023

For citation: Perova J.P., Grigoriev V.R., Zhukov D.O. Models and methods for analyzing complex networks and social network structures. *Russ. Technol. J*. 2023;11(2):33–49. https://doi.org/10.32362/2500-316X-2023-11-2-33-49

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

ОБЗОР

Модели и методы анализа сложных сетей и социальных сетевых структур

Ю.П. Перова [@], В.Р. Григорьев, Д.О. Жуков

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] Автор для переписки, e-mail: perova_yu@mirea.ru

Резюме

Цели. Целью статьи является исследование современных моделей и методов анализа сложных социальных сетевых структур и применяемых для этого инструментов, как на основе готовых решений в виде сервисов и программного обеспечения, так и средств разработки собственных приложений с использованием языка программирования Python. Такие исследования позволяют прогнозировать не только динамику общественных процессов (изменение социальных настроений), но и тенденции социально-экономического развития за счет мониторинга мнений пользователей по важным экономическим и социальным вопросам на уровне отдельных территориальных образований (районов, поселений небольших городов и т.д.) и регионов.

Методы. Рассмотрены и подробно описаны динамические модели и методы анализа стохастической динамики изменения состояний, учитывающие процессы самоорганизации и наличие памяти; методы деанонимизации пользователей; рекомендательные системы; статистические исследования, использующие методы анализа профилей в социальных сетях; методы численного моделирования для анализа сложных сетей и протекающих в них процессов. Особое внимание уделено обработке данных в сложных сетевых структурах средствами языка Python и применению его библиотек.

Результаты. Описана специфика решаемых задач при исследовании сложных сетевых структур и их междисциплинарность, связанная с использованием методов системного анализа, теории сложных сетей, текстовой аналитики и компьютерной лингвистики. В частности, исследованы динамические модели процессов, наблюдаемых в сложных социальных сетевых системах, структурные характеристики таких сетей и их взаимосвязь с наблюдаемыми динамическими процессами, в т.ч., с использованием теории построения динамических графов. Исследовано применение нейронных сетей для прогнозирования эволюции динамических процессов, наблюдаемых в сложных социальных системах, и их структуры. Значительное внимание уделено применению методов компьютерной лингвистики, что необходимо для извлечения знаний из текстовых сообщений пользователей социальных сетей при создании моделей, описывающих наблюдаемые процессы.

Выводы. Сетевой анализ помогает структурировать модели взаимодействия между социальными единицами: людьми, коллективами, организациями и т.д. По сравнению с другими методами сетевой подход имеет одно неоспоримое преимущество: он позволяет оперировать данными на разных уровнях исследования – от микро- до макроуровня, обеспечивает преемственность этих данных. Установлено, что практически все исследования используют методы работы с текстом, т.к. общение в социальных сетях почти полностью состоит из текстовых сообщений и публикаций. В большинстве исследований используются технологии машинного обучения и искусственного интеллекта. Лучший результат показали сверточные нейронные сети. Из используемых методов также следует выделить метод опорных векторов и дерево решений, т.к. именно они показывали самую высокую точность. Для составления выборок данных и правильного анализа полученных результатов применялись статистические методы.

Ключевые слова: социальные сети, моделирование социальных процессов, ориентированные графы, многослойная сверточная нейронная сеть, компьютерная лингвистика, кластеризация

• **Поступила:** 07.12.2021 • **Доработана:** 23.12.2022 • **Принята к опубликованию:** 09.02.2023

Для цитирования: Перова Ю.П., Григорьев В.Р., Жуков Д.О. Модели и методы анализа сложных сетей и социальных сетевых структур. *Russ. Technol. J.* 2023;11(2):33–49. https://doi.org/10.32362/2500-316X-2023-11-2-33-49

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The study of social networks and the modeling of the processes observed in them is a very important scientific and practical task, since it allows predicting the dynamics of changes in the sentiments of their users and thereby ensuring the management of social processes in the interests of stable economic development.

The present review sets out to characterize the specifics of the research area, formulate its main objectives, indicate any links with other sciences, and give a brief overview of the main approaches and resources used. Dynamic models of complex social systems and network structures are discussed along with the characteristics of complex networks and observed processes, including those based on graph construction and data analysis using the Python programming language. In addition, issues involved in the use of neural networks to make the necessary evolution forecasts processes observed in complex social systems and network structures are discussed. Considerable attention in the review is paid to methods of computational linguistics used for extracting knowledge from text messages of social network users when creating models that describe the observed processes.

SOCIAL NETWORKS AND THEIR GENERAL PROPERTIES

A social network comprises a structure together with a set of objects and relations defined in relation to it. In terms of the number of users, the largest social networks include Facebook[1] (banned in Russia), VKontakte[2], Odnoklassniki.ru[3], YouTube[4], etc. The term "social network" refers to the concentration of social objects that can be considered in terms of a network (or graph), whose nodes are objects, and whose links are social relations [1]. Today, the term "social network" denotes a wider concept than that implied by its original purely social aspect: the term covers, for example, many information networks including the world-wide web itself. Formally, any complex social network structure can be represented by a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the set of graph vertices, and $\mathbf{E}$ is the set of graph edges. In a social network graph, the vertices are the participants (or actors), while the edges indicate the existence of relationships between them. Relations can be either directed (directed graph) or undirected. In the theory of complex networks, there are three main areas: the study of statistical properties that characterize networks; the creation of network models; predicting the behavior of networks and the processes observed in them when their structural properties change, including, as a result of destructive impacts on them.

Social network analysis (SNA) is widely used in a number of applications and disciplines. Some common applications of network analysis include data collection and accumulation, network propagation modeling, network and sample modeling, feature and user behavior analysis, community-provided resource support, location-based interaction analysis, social sharing and selection, recommendation systems development, as well as link prediction and object analysis. In the private sector, firms use social media analysis to support activities such as customer interaction and analysis, marketing, and business intelligence. The public sector's use of SNA includes the development of leadership participation strategies, analysis of individual and group participation, use of the media, and community-based problem solving.

SNA thus represents an efficient system for discovering and interpreting public online connections. These can be explored using a range of analytical techniques, ranging from simple centrality measures to multilevel modeling. If data collection formerly represented a task that required a lot of effort and time, today's electronic networks have somewhat simplified this task. This happened through the use of passive data (such as web pages and mail store data). However, due to the increase in efficiency leading to a limitation in data collection, it became necessary to determine criteria for determining relationship significance. Solving such problems requires high-level technical skills, in particular, knowledge of programming languages or related programs.

Given these issues and limitations, studies propose more efficient and reliable data collection methods in such networks. In addition, issues such as spoof node

[1] https://www.facebook.com/. Accessed December 07, 2021.
[2] https://vk.com/. Accessed September 20, 2022 (in Russ.).
[3] https://ok.ru/. Accessed September 20, 2022 (in Russ.).
[4] https://www.youtube.com/. Accessed September 20, 2022.

detection, as well as fake nodes and links, need to be studied.

In order to solve many applied problems currently in practice, sets of ready-made SNA tools are typically used. However, these have some limitations; in particular, they do not allow the development of new approaches and models for studying the observed processes. A detailed description of the tools and methods used in SNA can be found in the review [2].

When analyzing the structure of complex networks, as in graph theory, individual node-, entire network-, and network substructure parameters are studied. Nevertheless, some questions, such as the planarity of a graph for the theory of complex networks, are of little practical interest. Among topical problems in the study of complex networks, the following can be distinguished: determination of cliques in the network (cliques are subgroups or clusters in which nodes are more strongly interconnected than with members of other cliques); selection of components (parts of the network) that are not interconnected, but whose nodes are connected within these components; identifying blocks and jumpers (a node is called a jumper if, when it is removed, the network breaks up into unconnected parts); selection of groupings comprising clusters of equivalent nodes having the most similar link profiles.

KEY AREAS OF RESEARCH IN COMPLEX NETWORKS AND APPLIED TOOLS

The theory of complex networks is a complex scientific area at the intersection of such sciences as discrete mathematics, graph theory, algorithm theory, nonlinear dynamics, the theory of phase transitions, the theory of percolation, and many others. Therefore, in order to successfully analyze and model complex networks, basic knowledge from all these areas is required. There are a large number of publications, including, for example, textbooks[5], which cover theoretical aspects of complex networks: characteristics, algorithms, models, search and ranking problems. The publications also provide information necessary for mathematical and computer modeling and analysis of complex networks.

The theory of complex networks covers the following problems:

1) study of standard characteristics of graphs of complex networks of different nature—random graphs, scale-free networks, small world networks, etc.;
2) determination and study of new characteristics of complex networks (for example, elasticity and survivability under destructive influences);
3) study of various "physical" processes on complex networks—diffusion, epidemic processes in society, the spread of various flows (for example, traffic in computer networks or vehicle flows in transport networks);
4) a very important direction in terms of application—methods for restoring, protecting and destroying networks, and solving issues of their optimization;
5) search for implicit or latent connections between participants, which can be very important for identifying members of criminal communities.

It should be noted that for the study of complex networks and identification of the main patterns of the processes occurring in them, methods are used that were first created for the study of natural science problems, in particular, methods of theoretical physics.

The theory of complex networks as a field of discrete mathematics studies the characteristics of networks, taking into account not only their topology, but also statistical phenomena, the distribution of weights of individual nodes and edges, the effects of leakage, percolation, and conductivity in such networks of current, liquid, information, etc. It turned out that the properties of many real networks differ significantly from the properties of classical random graphs.

The study of such parameters of complex networks as clustering, mediation, or vulnerability is directly related to the theory of survivability, since these are the properties that determine the ability of networks to maintain their operability in the event of a destructive effect on their individual nodes or edges (connections). Despite the fact that the theory of complex networks includes various networks—electrical, transport, information, the greatest contribution to the development of this theory was made by the study of social networks.

Due to a significant increase in the volume of textual information generated by Internet users and the need for automatic processing of texts in natural language in order to determine the state of the nodes of complex social networks (for example, opposition or loyalty), computational linguistics has now received a significant impetus for its development.

The task of computational linguistics can be formulated as the development of computer programs for automatic processing of texts in natural languages in order to extract knowledge, cluster texts into semantic groups, annotate, etc.

The source material for extracting the necessary linguistic information can be collections and corpora of texts. A corpus of texts is a collection of texts collected according to a certain principle of representativeness (by genre, authorship, etc.), in which all texts are marked up, i.e., are equipped with some linguistic markup (annotations)—morphological, accent,

[5] Snarsky A.A., Lande D.V. *Modeling of complex networks.* Textbook. Kyiv: NTUU KPI; 2015. 212 p. (in Russ.).

syntactic, etc.[6] Currently, there are at least a hundred different corpora—for different natural languages and with different markup. In Russia, the most famous is the National Corpus of the Russian Language[7]. Labeled corpora are created by linguists and are used both for linguistic research and for tuning (training) models and processors used in computational linguistics using well-known mathematical methods of machine learning.

Computational linguistics shows quite tangible results in various applications for automatic processing of texts in natural languages. Its further development depends on both the emergence of new applications and the independent development of various models of the language, in which many problems have not yet been solved. The most developed are the models of morphological analysis and synthesis. Syntax models have not yet been brought to the level of stable and efficient modules, despite the large number of proposed formalisms and methods. Even less studied and formalized are models of the level of semantics and pragmatics, although automatic processing of discourse is already required in a number of applications. Despite this, the already existing tools of computational linguistics itself, the use of machine learning and text corpora can significantly advance the solution of these problems.

It should be noted that computational linguistics is not only used to analyze information in complex social systems in order to determine the state of nodes, but also itself uses the achievements of the theory of complex networks. The first step in applying the theory of complex networks to text analysis is to represent this text as a collection of nodes and links, thereby building a language network. There are different ways of interpreting nodes and links, which leads, respectively, to different representations of the network of the language. Nodes can be connected to each other if the words corresponding to them are next to each other in the text, belong to the same sentence, are connected syntactically or semantically. The preservation of syntactic links between words leads to the image of the text in the form of a directed network, where the direction of the link corresponds to the subordination of the word.

The study of graph properties of complex networks is becoming increasingly popular due to the growing availability of scientific and social data presented in graph form. Because of this, many researchers have focused on developing improved graph neural network models. One of the main components of a graph neural network is the aggregation operator required to generate a graph-level representation from a set of node-level embeddings. The aggregation operator is of crucial importance, since it should, in principle, provide an isomorphism-invariant representation of the graph, that is, the representation of the graph must be a function of the nodes of the graph, considered as a set.

In [3], the DeepSets aggregation operator based on self-organizing maps (SOM) is considered to transform a set of node-level representations into a single graph-level. The adoption of SOM allows computation of representations of nodes that embed information about their mutual similarity. Experimental results on several real datasets show that the proposed approach provides improved predictive performance compared to conventional summing aggregation and many modern graph neural network architectures presented in the literature.

In the framework of paper [4], the architecture of convolutional neural networks was considered, including the types of layers used and the principles of their operation, settings, and training features. The possibility of searching and preventing information leaks from corporate information systems on the Internet is described. The architecture of convolutional neural networks for the primary processing of information on Internet pages is proposed: the types of layers that make up the network, their purpose and mathematical representation, as well as the hyperparameters used are described. The paper [4] presents the architecture of the network and the model of its training. The possibility of using networks of this type for solving problems of detecting leaks of confidential data is described, as well as existing solutions and approaches are analyzed. Approaches are considered that enable using convolutional neural networks to solve the problems of classifying web pages containing news and information sources, navigation and information sources based on their text content.

As an example of a finished system, we can refer to Bidirectional Encoder Representation Transformers (BERT) [5]—language representation model, which is designed for preliminary training of deep bidirectional representations on simple unmarked texts by combining the left and right contexts in all layers. This allows you to tune a pre-trained BERT model with just one additional output layer and get the most up-to-date results for a wide range of tasks.

Standard language representation models that existed before BERT, such as the OpenAI GPT (Generative Pre-Trained Transformer)[8] were unidirectional. This limited the choice of architectures that could be used for pre-training. For example, in OpenAI GPT, each token could only serve the previous token (from left to right) in the internal attention layer of the model. Tokens are intended

[6] Boyarsky K.K. *Introduction to computational linguistics*. Textbook. St. Petersburg: NIU ITMO; 2013. 72 p. (in Russ.).

[7] https://ruscorpora.ru/. Accessed September 20, 2022 (in Russ.).

[8] https://openai.com/api/. Accessed September 20, 2022 (in Russ.).

for electronic identification, which are provided to the user after successful authorization. In a sense, a token is an electronic key to access something.

This approach creates a number of limitations, therefore, for pre-training BERT, a masked language model is used, in which a certain number of tokens in the input data are randomly masked. The model then has to predict the original meaning of the masked words based on the context. This provides the ability to combine left and right contexts, which in turn allows a bidirectional view model to be pre-trained.

There are two stages in using BERT.

1. Preliminary training. The model is trained on unlabeled data by performing various tasks.
2. Fine tune. The model is loaded with pre-trained parameters and trained on labeled data from subsequent tasks.

Besides theoretical methods, numerical simulation is often used to analyze complex networks and the processes occurring in them. In addition, one of the most powerful and widely used tools for analyzing complex networks is data processing using the Python language and the libraries available for it [7–12]. A special package (Python-networkX) has been developed for the Python programming language—a toolkit for creating, manipulating and studying complex networks, which allows you to determine many of their characteristics. Here we can also mention the NATASHA[9] tools—an open library for the Python programming language, which allows you to extract structured information from texts in Russian. NATASHA has a concise interface and includes extractors for names, addresses, amounts of money, dates, and some other entities.

The Python language and the libraries written for it can be used to effectively solve a wide range of tasks for analyzing various data:

- multidimensional lists (matrices);
- tabular data, when data in different columns can be of different types (strings, numbers, dates, etc.). This includes data that are typically stored in relational databases or in files with commas as separators;
- data presented in the form of several tables interconnected by key columns (what in SQL is called primary and foreign keys);
- equally spaced and not equally spaced time series.

This list is far from complete. A significant portion of datasets can be converted to a structured form more suitable for analysis and modeling. In cases where this fails, it is possible to extract a structured set of features from the data set. For example, a selection of news articles can be converted into a word frequency table, to which sentiment analysis can then be applied.

[9] https://pypi.org/project/natasha/. Accessed September 20, 2022.

ANALYSIS OF NETWORK STRUCTURES AND FORECASTING THE DYNAMICS OF SOCIAL PROCESSES

A review of published papers shows that SNA methods are useful tools for creating a complete picture of public sentiment. These methods are cheaper to implement than population survey methods and provide more data, since in surveys not all people express their real point of view. Based on this, it is possible to study the behavior of modern society in the era of the spread of social networks.

Considering the dynamic approach, namely, the direction in the study of social networks, in which the objects of research are changes in the network structure over time, it can be noted that structural analysis and analysis of the behavior of connections in social networks is necessary in order to determine the most important peaks, connections, communities and emerging regions of the network. Such an analysis allows an overview of the global evolutionary behavior of the network.

Community discovery in dynamic networks no longer requires complex mathematical heuristics. Using a simple comparison of time slices, it is possible to determine dynamically changing temporary communities of users of social network structures. The study of these dynamic communities makes it possible to significantly simplify the analysis of the dynamics of a complex system of social interactions as it develops over time.

In [13], the authors present the fundamental structures of dynamic social networks based on a high-resolution dataset describing a densely connected population of 1000 first-year students at a large European university. The authors look at physically short interactions measured using Bluetooth, supplemented with information from telecommunications networks (phone calls and text messages), online social networks, and geolocation and demographic data.

Human social communities are overlapped by individuals participating in several communities (in complex network theory, such nodes are called jumpers). During the week, meetings of the subjects of the created structure are held (such structures are called kernels). It can be both a meeting of friends outside the university, and all students. In a network of short physical interactions, meetings require that all participants be present at the same time and that they be in physical contact.

The location of members of kernels can also be predicted. The object that helps to do this is the kernels themselves. By observing the usual routes of the people who make up the kernel and their behavioral habits, it is possible to predict the geographical location of a person in the next time interval with high accuracy (on average

in 93% of cases). This high accuracy proves that human mobility patterns are regular. It is also worth noting that kernel members have fewer location states than other individuals, resulting in lower values of information entropy on average.

The fact that geospatial exploration occurs as part of a social group but limited to specific time frames shows the complex interplay between time, location, and social context, and thus supports the hypothesis that sometimes, when people are most unpredictable in the geospatial domain, they exhibit predictable social behavior.

Linking the results of paper [13] to the literature on dynamic community detection, it can be noted that there are many methods that would allow the discovery of collections in everyday life, but in paper [13] it is used a simple matching of graph components to emphasize that emerging social structures are so obvious that these complicated methods are not needed.

Thus, the authors of [13] give a quantitative assessment of long-term patterns encoded in the microdynamics of a large system of interacting individuals, characterized by a high degree of order and predictability.

Let us consider one more work devoted to dynamic models [14]. Recent developments in the field of social networks have shifted the focus from static to dynamic representations, requiring new methods for their analysis and modeling. While social networks are shaped by a variety of processes, two specific mechanisms have been found to play a central role in their emergence and evolution. The first is the strategy of activation of social ties, that is the selection process leading to the creation of a new connection or the activation of an old one. It is clear that the activation of social ties is not accidental. Empirical observations show that people tend to allocate most of their social activities towards pre-existing strong ties, while allocating fewer interactions to create new social relationships or maintain weak ties. In other words, over time, some connections are used frequently in repetitive interactions, while others are not. The second mechanism is a surge of activity, that is, the activity of separate individuals develops through heterogeneous distributions of time between events. In addition, the propensity of individuals to participate in a social act per unit of time is also heterogeneous. In fact, empirical measurements on real datasets capturing various types of social dynamics show that activity is not uniformly distributed among people. In other words, not only do individuals exhibit heterogeneous propensities for social activity, but their activation is also explosive, and this explosive activity can significantly affect the evolution of networks. Although the study of these mechanisms has been the focus of a number of publications, a general framework for modeling is still lacking. Such a structure would make it possible to give an analytical characterization of how the interaction of heterogeneous patterns of activity and the mechanisms of selection of connections shape the evolution of social networks and, in turn, the processes occurring in them. To do this, the authors introduce a model of time-varying networks, which allows you to simultaneously control the relative strength of the burst of activity and the strategy for activating the connection. The asymptotic behavior of the model is solved analytically and a non-trivial phase diagram is found that governs the interaction of two processes. In particular, one observes the regime in which the surge controls the evolution of the network, and another area where the dynamics is completely determined by the process of selection of links. If the reuse of previously activated connections is strong enough and people tend to preferentially contact the same social circle, the spike leads to an amplification mechanism even in the presence of divergent time intervals between events, without having any effect on the evolution of the network. Thus, the structure proposed by the authors can be used to classify the temporal features of real networks and can give a new idea of the influence of social mechanisms on the processes of distribution in social networks.

In the paper [15] recommender systems are discussed. There are decision-making situations in the context of Internet information overload, when people have an overwhelming number of choices available, such as products to buy on an e-commerce site or restaurants to visit in a big city. Recommender systems (reciprocal recommender systems, RRS) emerged as a data-driven personalized decision support tool. They are able to process user related data, filter and recommend items based on the user's preferences, needs and/or behavior. Unlike most traditional recommendation approaches, where items are inanimate objects recommended to users and success is determined solely by the end user's response to the recommendation received, in RRS users become objects recommended to other users. Therefore, both the end user and the recommended user must accept the compliance recommendation in order to ensure successful RRS performance. The operation of RRS not only makes it possible to predict accurate preference estimates based on user interaction data, but also makes it possible to calculate mutual compatibility between pairs of users by applying processes for combining one-way preference information of each user.

In [16], the assessment of public opinion and public sentiment is carried out using a method based on a lexicon inherited from the classical approach to the analysis of public sentiment. The neural network

determines the keywords which are later verified by subject matter experts. The program first analyzes the articles and documents and finds how often different words appear in the articles. After that, the program highlights the most frequently occurring words and makes them keywords. Based on them, the program builds a lexicon that is characteristic of the public mood based on news articles.

The program described in [17] uses the method of analyzing topics from a social network, which, in addition to collecting, processing and sorting information, also measures the time elapsed between publications so that later, based on these data, to create a time scale. Thus, as a result of the work of the program, a graph is obtained, according to which one can trace the growth and decline in the popularity of certain topics discussed in social networks. You can also trace what moods in society accompany these events and what is the time period of active discussion of certain topics.

The authors of paper [18] discuss a method for studying political climate in society using SNA, which was carried out using the search for keywords in the text that were previously entered into the program database. The main purpose of creating this program is to trace what made certain political parties popular and what topics are discussed the most. Also, using the program, you can find out how many people support a particular political party.

The SNA technique using neural networks is presented in [19]. It was used during presidential campaigns in order to trace the mood in society. This technique can be used as a replacement for traditional methods of public emotions analysis, because it has the ability to find and analyze radical opinions that is impossible to do with traditional methods.

The authors of [20] use the method of collecting and processing data from Twitter[10] (banned in Russia) accounts to determine gender, age, political preferences, and approximate place of residence. Machine learning is used to process the data, with the help of which the authors of the paper were able to collect information from users of the Twitter (banned in Russia) social network based on their posts, subscriptions and account information.

In [21], the authors discuss a method for determining a user's political preferences by analyzing user records belonging to different political groups. Through voluntary surveys of people belonging to different political groups, the program analyzes the language that is inherent in each of the groups and highlights its keywords for each of the groups. Based on these words, the program will later analyze the user's profile on the Twitter (banned in Russia) social network and, on this basis, determine which of the political groups the user belongs to. Based on the survey, in addition to political preferences, the gender and age of a person are determined, so that later it would be possible to compile statistics by comparing the gender, age and political preferences of users. After the survey, accounts left by users are analyzed to remove accounts belonging to other people who did not take part in the survey. Then, the last 3200 records are analyzed from the users' page. Based on these records, databases of keywords specific to a particular political group are compiled. Based on these databases, charts are created that show how often they occur in the records of people belonging to this group. Also, when the program analyzes people's records, their mood is revealed.

The paper [22] is a comprehensive analysis of the trace that each of us leaves daily on the Internet. We make purchases, communicate; many familiar things have long gone online. Every action that takes place on social networks does not go unnoticed. Each of us has a so-called digital footprint—the actions that we perform on the Internet and which remain there. This can be both our public information, which we ourselves leave on our pages on social networks and our non-public actions, information about which still remains on the network and can be extracted from there. The authors argue that this digital footprint reveals a lot about the user. In total, they identify 14 different demographic characteristics that they were able to establish using such social tracking.

Data is easy to collect and use when the user posts it on their profile, but if they prefer to remain secretive, there are so many ways to find out this information. The authors cite a few of them: you can analyze smartphone log files, likes on a social network, browser search history, frequency of hashtags, and many more things.

The authors also inform that according to many studies, people tend to communicate with those with whom they are in the same social group, and all members of this group often have similar behavioral traits and manner of communication. This opens up a huge scope for study.

Social networks and other services that we use collect our personal information. In most cases, we agree to this, but can a social network collect any information about a user who is not registered with it? The paper [23] considers the shadow profile hypothesis, according to which a social network can collect, based on the public information of users of this network and data from their phone books (if, for example, the user himself provides access to the social network) information about those people who are not registered in this social network. The paper proves the fact that a shadow profile as a structure can be created, and the larger the social network, the more accurate the shadow profile data will be.

[10] https://twitter.com/. Accessed December 07, 2021.

In the study [24], the authors use a large amount of information about the user account: the time of publication of each of the entries and the frequency of their publication; the number of publications containing geodata; reposts of records of other users; the number of liked publications; the number of responses to publications of other users; the number of mentions of other users; the number of publications containing media data and the average amount of media data per publication; the date the account was created; the number of followings and followers and many other data and ratios related to account information. This is an example of a comprehensive and complete study of a profile in a social network, which does not include the analysis of the media data itself and the analysis of the user's friends/followers. It should most accurately predict the user's age based on a large amount of information. This paper also describes in more detail the algorithms for creating a sample, training a neural network, and directly analyzing user profiles.

Since, as the network grows, the search for similarity between nodes in the network is a time-consuming process for optimization, the researchers in [25] use swarm algorithms to solve the problems of link prediction and community detection. Swarm-based optimization techniques used in SNA are compared in this paper with community analysis and connection analysis. As a future area of application, swarm-based optimization methods can be extended to the use of deep learning neural networks, especially for updating gradients when creating models of such networks.

A social network is a social structure with a set of social actors and social interactions between them. The study of the dynamics of these structures can be used to explain local and global economic patterns that are important for development. In [26], the discussed topic is the development and analysis of automatic control systems for making decisions about oil projects. Before deciding to invest in an oil project, engineers describe the project by providing sufficient economic data. Based on this data, a decision can be made to conduct a professional analysis to determine if the project is feasible. To automate the manual process and overcome the shortcomings of traditional evaluation methods (for example, expert evaluation depends on the quality of the choice of experts themselves), a back propagation neural network is used in the economic evaluation of oil projects.

A huge amount of work was carried out using analysis from the social network Twitter (banned in Russia). In the paper [27], the authors discuss the analysis of a community with extremist views using SNA and neural networks. According to the results of the study, it is proved that it is possible to analyze the community and find people associated with it, and possibly to predict the plans of this community in order to prevent terrorist activities.

The paper [28] refers to the definition of the personal qualities of users of the social network Twitter (banned in Russia) based on the records they made, as well as on the basis of their subscriptions. The program takes into account gender, age, education and political preferences to obtain a more accurate result of the study. Thus, on the basis of users' records and subscriptions, it is possible to determine the area of his or her interests, after which it is possible to draw up a chart showing the dependence of gender, age, etc. on the area of interest, and track which user groups prevail in a particular group of interests.

The study [29] again uses the collection of user data from about 1500 sites and the comparison of these data with the data of their accounts on the Twitter (banned in Russia) social network. Based on these data, a more accurate demographic model of users is built. The program analyzes the interests of users, using not only the data of their social networks, but also the data of the Quantcast.com[11] service, which allows collecting more accurate information. As a result, a table of user demographics and communities of interest to which they belong is made up.

The authors of [30] compare the data of users of the social network Twitter (banned in Russia) with political preferences. The program, based on demographic data and political preferences of a person, compiles statistics that show which groups of the population support which parties. This can be judged by the combination of such variables as gender, age, income, race, etc.

In [31–34], SNA is carried out in order to track the political mood of the population. Thus, thanks to SNA, it is possible to map support for various political parties, track public mood and find out the rating of political parties in different periods before and after elections, while associating these levels of support with various events that took place around the party.

The authors of [35, 36] carry out analysis using machine learning and text sentiment, which are one of the main tools of SNA, especially for the restoration of demographic characteristics, which requires knowledge in the field of machine learning and computational linguistics. When studying computational linguistics, one can find many different methods of analyzing written text besides sentiment analysis. Many of them, perhaps, will expand the toolkit for SNA or improve existing algorithms. With the help of machine learning methods, it is possible to automate the analysis process and make it much more convenient.

[11] https://www.quantcast.com/. Accessed September 20, 2022.

The studies described in [37–41] are aimed at analyzing profiles in social networks using gender classification, which solved the problem of face recognition using neural networks, algorithms that work using emoji emoticons in the source text, and determining the age and gender from a photo.

In [42], the authors describe a method for analyzing SMS messages in order to classify senders by gender and age. For the study, the authors used several algorithms with different structures of neural networks, as well as various methods of working with natural language, trying to achieve the best result. Ultimately, the best result for determining the age was shown by the support vector machine—an accuracy of about 71%. The best result of gender determination accuracy—almost 80%, was shown by decision tree J48. It is worth noting that various methods and filters of natural language processing only slightly affected the results, practically not improving either the accuracy or the speed of the algorithms.

The subject of research [43] is microblogs. The authors used the keyword analysis method. This approach fits well with the environment, in which the analysis is carried out, because people in microblogs often discuss some events, news, or discuss a certain topic. Using this method and machine learning methods, they were able to divide the initial sample into six age groups and identify the topic that participants in each age group most often discuss and express their thoughts on most often. It turned out that teenagers under 18 most often discuss sports; young people aged 18–25 talk about entertainment the most; people aged 25 to 30 see other goals, they want to firmly settle in life, so they mainly discuss family and business; older people (31–36 years old) are most interested in technology; further, users aged 26–40 begin to worry about their health and speak out more about it, while those over 40 like to discuss politics the most. Thus, the most frequent topic for discussion was determined for each age group. This does not mean that every member of the group necessarily discusses this topic, but it is more likely that the person discussing this topic belongs to this age group.

We can consider several statistical studies [44–46] that have widely used the method of analyzing profiles in social networks. Their purpose is to identify the social mobility of people based on their publications together by geodata. The authors found a large number of such publications; based on them an approximate map of the user's locations was created, the main centers of activity were identified, and the person's place of residence was established. Based on the place of residence, the names of people were found out. Further, using a database of names distributed by gender, it was possible to determine the gender of more than half of all the studied accounts. Using the last names, the researchers tried to determine information about the race and age of users—successfully in 38% and 14% of cases, respectively. This study showed that it is possible to establish some demographic characteristics, only by tracking a person or knowing his or her first and last name.

The paper [47] describes a method for determining the gender and age using the voice message function. The results are given, according to which the authors managed to achieve an accuracy of 80% in determining the gender and age of the speaker. This technology can be quite successfully used to analyze voice messages in social networks, if such a need arises, but in reality, such an algorithm is unlikely to find wide application in this area, since voice messages are sent personally to the recipient, and social network analysis is usually carried out publicly based on available information. In other industries, the value of such technology is difficult to overestimate: it should be useful in forensics, biometrics and for designing a system for speech recognition or recreation.

In addition to a regular text and the previously discussed additions to it—reposts, pictures, emoticons and subscriptions to other users, links and hashtags are also very often used on social networks. Links are used to share some content, be it a picture or news, and hashtags are used to indicate the topic of the publication and make it easier for other users to find this publication in the search. The authors of [48] suggest that the content found on the links shared by users and the hashtags they use can tell a lot about the age of these users. The researchers decided to analyze the posts of various users to determine their age group, but unlike many similar studies, they used not only the posts themselves, but also the content located on the links shared by users as well as recent posts with the same hashtag that the user mentions.

The paper [49] proposes a new method for predicting changes in complex social network structures based on the application of percolation theory and approaches adopted in stochastic dynamics. New results of computer modeling of the influence of the density of a social network on the threshold of its penetration are discussed. Percolation thresholds are calculated for various network densities and can be used in models that describe the stochastic dynamics of a system's transition from one state to another. The stochastic model presented in this paper provides the possibility of an abrupt transition of moods (states) of people in a social network for a very short period of time without any external influences, which is determined by the features of the system's self-organization and the memory of its nodes about previous states.

The developed model allows you to create an algorithm for monitoring social conditions based on the theory of percolation and stochastic dynamics, which can be easily applied in practice. The essence of this algorithm is as follows.

1. With the help of sociological monitoring, the average number of connections per person in a given social network is determined; then the proportion of negatively inclined people at a given time is determined ($t = 0$). The average density makes it possible to calculate the percolation threshold of the network structure, i.e., the share of network participants who have certain views, which allow these views to be freely distributed in the network.
2. After a fixed unit of time (day, week, etc.), the share of participants with certain views that are currently being investigated is determined. The change in this share compared to the previous share allows you to determine the value of the upward and downward trends.
3. Further, it is possible to use the received information about the trends, the percolation threshold and the initial share of network participants with certain representations to track and manage network participants.

The results obtained can be applied to the management of social processes. In one of the approaches first the sentiment in a social network is analyzed. Further, modern methods of psycholinguistic analysis based on artificial network technologies make it possible to attribute each user to a specific target group in accordance with his or her moods and views. The network nodes that can be identified as users that violate certain laws, such as those that spread extremist views and sentiments, are blocked. However, there are groups of nodes that do not violate any laws, but can potentially move into a group with extremist views. Since they do not break the law, they cannot be blocked, but it is still possible to limit their communication abilities using technical approaches, for example, by reducing the data transfer rate and reducing the number of other nodes or connections available to them. At the same time, the penetration threshold for information that can be freely transmitted over the network increases.

The authors[12, 13] of [50–52] propose a method for evaluating media in several modalities (topics, evaluation criteria/properties, classes), combining thematic modeling of text corpora and making multi-criteria decisions. The evaluation is based on the analysis of corpora in the following way: the conditional probability distribution of carriers by topics, properties and classes is calculated after the formation of a thematic model of corpora. Several approaches are used to obtain weights that describe how each topic relates to each evaluation criterion and to each class described in the document, including manual tagging, multi-enterprise approach, and automatic approach. The proposed multi-corporate approach involves assessing the thematic asymmetry of the corpora to obtain weights that describe the relationship of each topic to a certain criterion. These weights, in combination with the topic model, can be applied to evaluate each document in the corpora according to each of the considered criteria and classes. The proposed method was applied to a corpus of 804829 news publications from 40 Kazakhstan sources published from January 01, 2018 to December 31, 2019. The BigARTM model (200 topics) was obtained. The experiments confirm the general capability of media estimation using the topical text corpora model, as the classification task achieved a receiver performance area under the curve (ROC AUC) estimate of 0.81, which is comparable to the results obtained for the same task using the BERT model.

The developed system, in which the proposed model is integrated, allows solving classical problems such as simple reports or sentiment analysis. In addition, it also has a number of unique use cases that distinguish it from existing solutions: automatic analysis by topic, significant event and object without the need to generate queries based on keywords; analysis according to an arbitrary list of criteria not limited to sentiment, but also including social significance, popularity, manipulativeness, propaganda content, attitude to a certain country, attitude to a certain area, etc.; analysis of the dynamic behavior of topics; predictive analysis at the topic level.

In [53–55], the KroMFac method was proposed, which is used to detect a community by the regularized non-negative matrix factorization method based on the Kronecker graph model. KroMFac combines network analysis and community discovery techniques in a single unified framework.

The paper [56] is devoted to SNA and the development of methods for deanonymizing their users. Deanonymization refers to the identification of a user on the network or the true place of access to the network. After a comparative analysis of existing methods and models of user deanonymization [57–59], the authors propose a modified method based on the algorithm for combining selected vertices to defragment the deanonymization statement problem into smaller sub-problems that can

[12] Bushman B., Whitaker J. Media influence on behavior. Reference module. In: *Neuroscience and Biobehavioral Psychology*. 2017. http://scitechconnect.elsevier.com/neurorefmod/. Accessed November 24, 2020.

[13] Bandari R., Asur S., Huberman B.A. *The pulse of news in social media: Forecasting popularity*. https://arxiv.org/pdf/1202.0332.pdf. Accessed September 20, 2020.

be solved using existing methods. The main result of this work is the development of a new approach to optimization methods for identifying users of social networks based on the pairwise partitioning algorithm. The proposed algorithm improves the characteristics of existing deanonymization technologies, and is of theoretical and practical importance for the development of systems for modeling information actions in social networks.

The papers [60, 61] present the models developed by the authors for describing the stochastic dynamics of state changes in complex social systems, which take into account the processes of self-organization and the presence of memory. To create a model, graphical diagrams of the transition probabilities between the possible states of the described systems are considered, taking into account previous states that make it possible to take into account memory and describe not only Markovian, but also non-Markovian processes. Based on this approach, a nonlinear second-order differential equation is derived that allows one to formulate and solve boundary problems for determining the probability density function of the amplitude of parameter deviations that describe the observed processes of non-stationary time series depending on the values of the time interval of its determination and the depth of memory accounting. The resulting differential equation contains not only terms responsible for random change (diffusion) and ordered change (drift), but also contains a term responsible for the possibility of self-organization.

CONCLUDING REMARKS

Network analysis can be used to structure models of interaction between social units: people, teams, organizations, etc. The network approach has an indisputable advantage as compared with other methods in terms of operating with data obtained at different research levels—from the micro to the macro level, to ensure data continuity. Network methods can also increase understanding by describing processes both theoretically and quantitatively. The relevance of network analysis is growing, since at the moment there is a globalization of the world processes, and above all, in the form of global networking.

Almost all studies use different methods of working with text, since communication in social networks consists almost entirely of text messages and publications. To apply these methods, knowledge in the field of computational linguistics is required. Here, sentiment analysis, lexical analysis, and keyword extraction are among the most commonly used approaches.

Having studied the results of research in this area, we can assert that almost all studies use various combinations of machine learning and artificial intelligence technologies. Although there many architectures and methods of neural networks, convolutional neural networks demonstrate the best results. Of the methods used, the support vector and decision tree approaches can be singled out as delivering the highest accuracy.

In order to work effectively with neural networks, it is necessary to compile a sample correctly. For this purpose, it is first necessary to determine what characteristics and initial data are required to classify users. Since results are predicted on the basis of statistical data, knowledge of statistics and their application is also necessary both for sampling and correctly analyzing the results.

CONCLUSIONS

This review set out to familiarize a wide range of readers with contemporary models and methods for analyzing complex social network structures, as well as the tools used for this purpose, which include those based on ready-made solutions in the form of services and software, as well as research applications developed using the Python programming language. When setting up and conducting further investigations by a wide range of researchers, it is very important to consider the advantages and disadvantages of existing models and methods. It can be concluded that SNA methods may serve as a very useful tool for creating a complete picture of public mood.

The review describes the specifics of the main tasks solved in the study of complex network structures. The interdisciplinary discipline of system analysis includes the theory of complex networks, text analysis and computational linguistics, neural networks, and many related areas. In particular, the work explores dynamic models of processes observed in complex social network systems, as well as the structural characteristics of such networks and their relationship with observed dynamic processes, including using dynamic graph construction theory. The use of neural networks to predict the evolution of dynamic processes observed in complex social systems and their structure (for example, how activity, the number of users and the structure of their connections in social network communities change) is also analyzed. When creating models to describe the observed processes, considerable attention is focused on the use of computational linguistics methods to extract knowledge from text messages of social network users.

By monitoring the opinions of users on important economic and social issues, both at the level of individual territorial entities (for example, districts, settlements of small towns, etc.), and at the regional level, such studies can be used to help predict not only the dynamics of social processes (changes in social sentiment), but also socioeconomic development trends.

ACKNOWLEDGMENTS

This research was supported by the Russian Science Foundation, grant No. 22-21-00109 "Development of the dynamics forecasting models of social moods based on the analysis of text content time series of social networks using the Fokker–Planck and nonlinear diffusion equations."

Authors' contributions

J.P. Perova—collection and analysis of information for the review.

V.R. Grigoriev—processing of materials for the review.

D.O. Zhukov—conceptual idea and discussion of the obtained results.

REFERENCES

1. Gubanov D.A., Novikov D.A., Chkhartishvili A.G. *Sotsial'nye seti: modeli informatsionnogo vliyaniya, upravleniya i protivoborstva* (*Social networks: models of informational influence, management and confrontation*). Moscow: MTsNMO; 2018. 223 p. ISBN 978-5-4439-1302-5 (in Russ.).
2. Batura T.V. Methods of social networks analysis. *Vestnik NGU. Seriya: Informatsionnye tekhnologii = Vestnik NSU. Series: Information Technologies*. 2012;10(4):13–28 (in Russ.). Available from URL: https://lib.nsu.ru/xmlui/handle/nsu/250
3. Pasa L., Navarin N., Sperdut A. SOM-based aggregation for graph convolutional neural networks. *Neural Comput. & Applic.* 2022;34(1):5–24. https://doi.org/10.1007/s00521-020-05484-4
4. Zhukov D.O., Akimov D.A., Red'kin O.K., Los' V.P. Application of convolutional neural networks for preventing information leakage in open internet resources. *Aut. Control Sci*. 2017;51(8):888–893. https://doi.org/10.3103/S0146411617080314
5. Zhang Z., Wu S., Jiang D., Chen G. BERT-JAM: Maximizing the utilization of BERT for neural machine translation. *Neurocomputing*. 2021;460:84–94. https://doi.org/10.1016/j.neucom.2021.07.002
6. McKinney W. *Python i analiz dannykh* (*Python and Data Analysis*): transl. from Eng. Moscow: DMK Press; 2020. 540 p. (in Russ.). ISBN 978-5-94074-590-5
   [McKinney W. *Python for Data Analysis*: 2nd ed. US: O'Reilly Media, Inc.; 2017. 541 p. ISBN 978-1-491-95766-0. Available from URL: https://www.programmer-books.com/wp-content/uploads/2019/04/Python-for-Data-Analysis-2nd-Edition.pdf]
7. Nikolenko S., Kadurin A., Arkhangel'skaya E. *Glubokoe obuchenie. Pogruzhenie v mir neironnykh setei* (*Deep Learning. Immersion in the World of Neural Networks*). St. Petersburg: Piter; 2021. 476 p. (in Russ.). ISBN 978-5-4461-1537-2
8. Kan K. *Neironnye seti. Evolyutsiya* (*Neural Networks. Evolution*). LitRes; 2018. 380 p. (in Russ.).
9. Rashid T. *Sozdaem neironnuyu set'* (*Make Your Own Neural Network*): transl. from Eng. St. Petersburg: Al'fa-kniga; 2017. 272 p. (in Russ.). ISBN 978-5-9909445-7-2
   [Rashid T. *Make Your Own Neural Network*. 1st ed. CreateSpace Independent Publishing Platform; 2016. 222 p. ISBN-13 ‍978-1530826605]
10. Galushkin A.I. *Neironnye seti: osnovy teorii* (*Neural Networks: Fundamentals of Theory*). Moscow: Goryachaya liniya-Telekom; 2012. 496 p. (in Russ.).
11. Savel'ev A.V. The philosophy of methodology of neuro-modeling: Sense and prospects. *Filosofiya nauki = Philosophy of Sciences*. 2003;1(16):46–59 (in Russ.).
12. Alekseev A.Yu., Kuznetsov V.G., Petrunin Yu.Yu., Savel'ev A.V., Yankovskaya E.A. Neurophilosophy as a conceptual basis for neurocomputing. *Neirokomp'yutery: razrabotka, primenenie = Neurocomputers: Development, Application*. 2015;5:69–77 (in Russ).
13. Sekara V., Stopczynski A., Lehmann S. Fundamental structures of dynamic social networks. *Proc. Natl Acad. Sci. USA*. 2016;113(36):9977–9982. https://doi.org/10.1073/pnas.1602803113
14. Ubaldi E., Vezzani A., Karsai M., Perra N., Burioni R. Burstiness and tie activation strategies in time-varying social networks. *Sci. Rep*. 2017;7:46225. https://doi.org/10.1038/srep46225
15. Palomares I., Porcel C., Pizzato L., Guy I., Herrera-Viedma E. Reciprocal recommender systems: analysis of state-of-art literature, challenges and opportunities towards social recommendation. *Information Fusion*. 2021;69(16):103–127. https://doi.org/10.1016/j.inffus.2020.12.001
16. Yatim Md.A.F., Wardhana Y., Kamal A., Soroinda A.A.R., Rachim F., Wonggo M.I. A corpus-based lexicon building in Indonesian political context through Indonesian online news media. In: *2016 International Conference on Advanced Computer Science and Information Systems* (*ICACSIS*). IEEE. https://doi.org/10.1109/ICACSIS.2016.7872794

17. Kirn S.L., Hinders M.K. Dynamic wavelet fingerprint for differentiation of tweet storm types. *Soc. Netw. Anal. Min*. 2020;10(1):4. https://doi.org/10.1007/s13278-019-0617-3
18. Karami A., Elkouri A. Political Popularity Analysis in Social Media. In: Taylor N., Christian-Lamb C., Martin M., Nardi B. (Eds.). *Information in Contemporary Society*. Part of: *Lecture Notes in Computer Science* (including subseries *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*). 2019. V. 11420. P. 456–465. https://doi.org/10.1007/978-3-030-15742-5_44
19. Belcastro L., Cantini R., Marozzo F., Talia D., Trunfi P. Learning political polarization on social media using neural networks. *IEEE Access*. 2020;8:47177–47187. https://doi.org/10.1109/ACCESS.2020.2978950
20. Vijayaraghavan P., Vosoughi S., Roy D. Twitter demographic classification using deep multi-modal multi-task learning. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. 2017;2(Short Papers):478–483. https://doi.org/10.18653/v1/P17-2076
21. Preoţiuc-Pietro D., Liu Y., Hopkins D., Ungar L. Beyond binary labels: political ideology prediction of Twitter users. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. 2017;1(Long Papers):729–740. https://doi.org/10.18653/v1/P17-1068
22. Hinds J., Joinson A.N. What demographic attributes do our digital footprints reveal? A systematic review. *PLoS One*. 2018;13(11):e0207112. https://doi.org/10.1371/journal.pone.0207112
23. García D. Leaking privacy and shadow profiles in online social networks. *Sci. Adv*. 2017;3(8):e1701172. https://doi.org/10.1126/sciadv.1701172
24. Pandya A., Oussalah M., Monachesi P., Kostakos P. On the use of distributed semantics of tweet metadata for user age prediction. *Future Generation Computer Systems*. 2020;102(5915):437–452. https://doi.org/10.1016/j.future.2019.08.018
25. Pulipati S., Somula R., Parvathala B.R. Nature inspired link prediction and community detection algorithms for social networks: a survey. *Int. J. Syst. Assur. Eng. Manag*. 2021. https://doi.org/10.1007/s13198-021-01125-8
26. Li H., Mao X., Wu C., Yang F. Design and analysis of a general data evaluation system based on social networks. *EURASIP J. Wireless Com. Network*. 2018;1:109. https://doi.org/10.1186/s13638-018-1095-4
27. Xu F., Sun D., Li Z., Li B. Research on online supporting community of extreme organization by AI-SNA based method. In: *Proceedings of the 8th IEEE International Conference on Software Engineering and Service Sciences* (*ICSESS*). 2018. V. 2017. P. 546–551. https://doi.org/10.1109/ICSESS.2017.8342974
28. Volkova S., Bachrach Y., Van Durme B. Mining user interests to predict perceived psycho-demographic traits on Twitter. In: *2016 IEEE Second International Conference on Big Data Computing Service and Applications* (*BigDataService*). IEEE. 2016. P. 36–43. https://doi.org/10.1109/BigDataService.2016.28
29. Culotta A., Ravi N.K., Cutler J. Predicting Twitter user demographics using distant supervision from website traffic data. *J. Artif. Intell. Res*. 2016;55:389–408. https://doi.org/10.1613/jair.4935
30. Barberá P. Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data. *Working Paper*. Available from URL: http://pablobarbera.com/static/less-is-more.pdf
31. Ardehaly E.M., Culotta A. Learning from noisy label proportions for classifying online social data. *Soc. Netw. Anal. Min.* 2018;8:2. https://doi.org/10.1007/s13278-017-0478-6
32. Franco-Riquelme J.N., Bello-Garcia A., Ordieres-Meré J. Indicator proposal for measuring regional political support for the electoral process on Twitter: The case of Spain's 2015 and 2016 general elections. *IEEE Access*. 2019;7:62545–62560. https://doi.org/10.1109/ACCESS.2019.2917398
33. Jungherr A., Schoen H., Posegga O., Jürgens P. Digital trace data in the study of public opinion: an indicator of attention toward politics rather than political support. *Soc. Sci. Comput. Rev*. 2016;35(3):336–356. https://doi.org/10.1177/0894439316631043
34. Mwanza S., Suleman H. Measuring network structure metrics as a proxy for socio-political activity in social media. In: *IEEE International Conference on Data Mining Workshops* (*ICDMW*). IEEE. 2017. P. 878–883. https://doi.org/10.1109/ICDMW.2017.120
35. Al-Agha I., Abu-Dahrooj O. Multi-level analysis of political sentiments using Twitter data: A case study of the Palestinian-Israeli conflict. *Jordanian Journal of Computers and Information Technology*. 2019;5(3):195–215. https://doi.org/10.5455/jjcit.71-1562700251
36. Basil M., Gaikwad S., Salim A.S. Deep learning approach based dominant age group based classification for social network. In: Khalaf M., Al-Jumeily D., Lisitsa A. (Eds.). *Applied Computing to Support Industry*: *Innovation and Technology. ACRIT 2019*. *Communications in Computer and Information Science.* 2020;1174:148–156. https://doi.org/10.1007/978-3-030-38752-5_12
37. Guimaraes R., Renata R., De Gaetano D., Rodriguez D.Z., Bressan G. Age groups classification in social network using deep learning. *IEEE Access*. 2017;5:10805–10816. https://doi.org/10.1109/ACCESS.2017.2706674
38. Bhat S.F., Lone A.W., Dar T.A. Gender prediction from images using deep learning techniques. *In: 2019 International Artificial Intelligence and Data Processing Symposium* (*IDAP*). IEEE. 2019. https://doi.org/10.1109/IDAP.2019.8875934

39. Bulut İ., Erdoğan M., Gönülal B., Baş R., Kılıç Ö. Using short texts and emojis to predict the gender of a texter in Turkish. In: *2019 4th International Conference on Computer Science and Engineering* (*UBMK*). IEEE. 2019. P. 435–438. https://doi.org/10.1109/UBMK.2019.8907198
40. Dileep M.R., Danti A. Multiple hierarchical decision on neural network to predict human age and gender. In: *2016 International Conference on Emerging Trends in Engineering*, *Technology and Science* (*ICETETS*). IEEE. 2016. https://doi.org/10.1109/ICETETS.2016.7603026
41. Gupta R., Kumar S., Yadav P., Shrivastava S. Identification of age, gender, & race SMT (scare, marks, tattoos) from unconstrained facial images using statistical techniques. In: *2018 International Conference on Smart Computing and Electronic Enterprise* (*ICSCEE*). IEEE. 2018. https://doi.org/10.1109/ICSCEE.2018.8538423
42. Khdr J., Varol C. Age and gender identification by SMS text messages. In: *2018 International Conference on Artificial Intelligence and Data Processing* (*IDAP*). IEEE. 2018. https://doi.org/10.1109/IDAP.2018.8620780
43. Koti P., Pothula S., Dhavachelvan P. Age forecasting analysis – over microblogs. In: *2017 Second International Conference on Recent Trends and Challenges in Computational Models* (*ICRTCCM*). IEEE. 2017. P. 83–86. https://doi.org/10.1109/ICRTCCM.2017.38
44. López-Santamaría L.-M., Almanza-Ojeda D.-L., Gomez J.C., Ibarra-Manzano M. Age and gender identification in unbalanced social media. In: *2019 International Conference on Electronics, Communications and Computers* (*CONIELECOMP*). IEEE. 2019. https://doi.org/10.1109/CONIELECOMP.2019.8673125
45. Luo F., Cao G., Mulligan K., Li X. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*. 2015;70(3):11–25. https://doi.org/10.1016/j.apgeog.2016.03.001
46. Sánchez-Hevia H.A., Gil-Pita R., Utrilla-Manso M., Rosa-Zurera M. Convolutional-recurrent neural network for age and gender prediction from speech. In: *2019 Signal Processing Symposium* (*SPSympo*). IEEE. 2019. P. 242–245. https://doi.org/10.1109/SPS.2019.8881961
47. Wang Y., Song W., Liu L. Age prediction based on feature selection. In: *2017 2nd IEEE International Conference on Computational Intelligence and Applications* (*ICCIA*). IEEE. 2017. P. 359–363. https://doi.org/10.1109/CIAPP.2017.8167239
48. Pandya A., Oussalah M., Monachesi P., Kostakos P., Lovén L. On the use of URLs and hashtags in age prediction of Twitter users. In: *2018 IEEE International Conference on Information Reuse and Integration* (*IRI*). IEEE. 2018. P. 62–69. https://doi.org/10.1109/IRI.2018.00017
49. Zhukov D.O., Zaltcman A.D., Khvatova T.Yu. Forecasting changes in states in social networks and sentiment security using the principles of percolation theory and stochastic dynamics. In: *Proceedings of the 2019 IEEE International Conference* "*Quality Management, Transport and Information Security, Information Technologies"* (*IT&QM&IS*). IEEE. 2019. Article number 8928295. P. 149–153. https://doi.org/10.1109/ITQMIS.2019.8928295
50. Mukhamediev R.I., Yakunin K., Mussabayev R., Buldybayev T., Kuchin Y., Murzakhmetov S., Yelis M. Classification of negative information on socially significant topics in mass media. *Symmetry*. 2020;12(12):1945. https://doi.org/10.3390/sym12121945
51. Ko H., Jong Y., Sangheon K., Libor M. Human-machine interaction: A case study on fake news detection using a backtracking based on a cognitive system. *Cogn. Syst. Res*. 2019;55:77–81. https://doi.org/10.1016/j.cogsys.2018.12.018
52. Willaert T., Van Eecke P., Beuls K., Steels L. Building social media observatories for monitoring online opinion dynamics. *Soc. Media Soc*. 2020;6(2):205630511989877.
53. Tran C., Shin W.-Y., Spitz A. Community detection in partially observable social networks. *ACM Transactions on Knowledge Discovery from Data.* 2022;16(2):1–24. https://doi.org/10.1145/3461339
54. Chen Z., Li L., Bruna J. Supervised community detection with line graph neural networks. In *Proceedings of the 7th International Conference on Learning Representations* (*ICLR*). ACM. 2019. https://doi.org/10.48550/arXiv.1705.08415
55. Hoffmann T., Peel L., Lambiotte R., Jones N.S. Community detection in networks without observing edges. *Sci. Adv*. 2020;6(4):eaav1478. https://doi.org/10.1126/sciadv.aav1478
56. Bashuev Ya., Grigorjev V. Social nets deanonimization methods. *Vestnik RGGU. Seriya Dokumentovedenie i arkhivovedenie. Informatika. Zashchita informatsii i informatsionnaya bezopasnost' = RGGU BULLETIN. Series: Records Management and Archival Studies. Computer Science. Data Protection and Information Security*. 2016;3(5):125–146 (in Russ.). Available from URL: https://www.rsuh.ru/upload/main/vestnik/pmorv/Vestnik_daizi3(5)-16.pdf#page=125
57. Wondracek G., Holz T., Kirda E., Kruegel C. A practical attack to de-anonymize social network users. *Technical Report TR-iSecLab-0110-001*. 2013. Available from URL: https://anonymous-proxy-servers.net/paper/sonda-tr.pdf
58. Simon B., Gulyás G., Imre S. Analysis of grasshopper, a novel social network de-anonymization algotithm. *Periodica Polytechnica: Electrical Engineering and Computer Science.* 2014;58(4):161–173. https://doi.org/10.3311/PPee.7878

59. Peng W., Li F., Zou X., Wu J. A two-stage deanonymization attack against anonymized social networks. *IEEE Trans. Comp*. 2014;63(2):290–303. https://doi.org/10.1109/TC.2012.202
60. Khvatova T., Zaltsman A., Zhukov D. Information processes in social networks: Percolation and stochastic dynamics. In: *CEUR Workshop*. *Proceedings 2nd International Scientific Conference "Convergent Cognitive Information Technologies*." 2017;1–2064:277–288.
61. Zhukov D., Khvatova T., Zaltsman A. Stochastic dynamics of influence expansion in social networks and managing users' transitions from one state to another. In: *Proceedings of the 11th European Conference on Information Systems Management* (*ECISM*). 2017. P. 322–329. Available from URL: http://www.scopus.com/inward/record.url?eid=2-s2.0-85039839600&partnerID=MN8TOARS

About the authors

Julia P. Perova, Senior Lecturer, Department of Telecommunications, Institute of Radio Electronics and Informatics, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: perova_yu@mirea.ru. Scopus Author ID 57431908700, https://orcid.org/0000-0003-4028-2842

Vitaly P. Grigoriev, Cand. Sci. (Eng.), Associate Professor, Head of the Department of Information Warfare, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: grigorev@mirea.ru. RSCI SPIN-code 4088-0403

Dmitry O. Zhukov, Dr. Sci. (Eng.), Professor, Department of Information Warfare, Institute for Cybersecurity and Digital Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: zhukov_do@mirea.ru. Scopus Author ID 57189660218, RSCI SPIN-code 1798-8891

Об авторах

Перова Юлия Петровна, старший преподаватель кафедры телекоммуникаций Института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: perova_yu@mirea.ru. Scopus Author ID 57431908700, https://orcid.org/0000-0003-4028-2842

Григорьев Виталий Робертович, к.т.н., доцент, заведующий кафедрой «Информационное противоборство» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: grigorev@mirea.ru. SPIN-код РИНЦ 4088-0403.

Жуков Дмитрий Олегович, д.т.н., профессор, профессор кафедры «Информационное противоборство» Института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: zhukov_do@mirea.ru. Scopus Author ID 57189660218, SPIN-код РИНЦ 1798-8891.

Translated from Russian into English by Evgenii I. Shklovskii
Edited for English language and spelling by Thomas A. Beavitt

Micro- and nanoelectronics. Condensed matter physics

Микро- и наноэлектроника. Физика конденсированного состояния

UDC 535.015
https://doi.org/10.32362/2500-316X-2023-11-2-50-57

RESEARCH ARTICLE

Improving the efficiency of an optical-to-terahertz converter using sapphire fibers

Nikolay V. Zenchenko [1, 2, 3, @], Denis V. Lavrukhin [1, 2], Igor A. Glinskiy [1, 2, 3], Dmitry S. Ponomarev [1, 2]

[1] *V.G. Mokerov Institute of Ultra High Frequency Semiconductor Electronics, Russian Academy of Sciences, Moscow, 117105 Russia*
[2] *Prokhorov General Physics Institute, Russian Academy of Sciences, Moscow, 119991 Russia*
[3] *MIREA – Russian Technological University, Moscow, 119454 Russia*
[@] *Corresponding author, e-mail: zenchenko.nikolay@yandex.ru*

Abstract

Objectives. The study aims to improve the efficiency of a large-area photoconductive terahertz (THz) emitter based on an optical-to-terahertz converter (OTC) having a radiating area of 0.3 × 0.3 mm^2 for generating high-power THz radiation by using an array of close-packed profiled sapphire fibers having a diameter in the range of 100–300 μm as focusing optics.

Methods. As a photoconductive substrate, we used a semi-infinite LT-GaAs layer (low-temperature grown GaAs; GaAs layer grown by molecular beam epitaxy at a low growth temperature). Additional Si_3N_4 and Al_2O_3 layers are intended for reducing leakage currents in the OTC and reducing the reflection of the laser pump pulse from the air/semiconductor interface (Fresnel losses), respectively, at a gap width of 10 μm. For forming the antenna electrodes and feed strips, the Ti/Au metal system was used. The simulation was carried out by the finite element method in the *COMSOL Multiphysics* environment.

Results. The use of a profiled sapphire fiber whose diameter has been optimized with respect to the gap parameters to significantly increase the concentration of charge carriers in the immediate vicinity of the electrodes of an OTC is demonstrated. The integrated efficiency of a large-area photoconductive THz emitter was determined taking into account the microstrip topology of the array with a characteristic size of feed strips proportional to the gap width in the OTC and with the upper (masking) metal layer. The maximum localization of the electromagnetic field in close proximity to the edges of electrodes at the "fiber–semiconductor" interface is achieved with a profiled sapphire fiber diameter of 220 μm.

Conclusions. By optimizing the diameter of the sapphire fiber, the possibility of improving the localization of incident electromagnetic waves in close proximity to the edges of the OTC electrodes by ~40 times compared to the case without fiber, as well as increasing the overall efficiency of a large-area emitter by up to ~7–10 times, was demonstrated.

Keywords: pulsed terahertz spectroscopy, emitters and detectors of THz radiation, subwavelength radiation, terahertz optical elements and systems, optical-to-terahertz conversion, metalens

• **Submitted:** 20.05.2022 • **Revised:** 07.10.2022 • **Accepted:** 27.01.2023

For citation: Zenchenko N.V., Lavrukhin D.V., Glinskiy I.A., Ponomarev D.S. Improving the efficiency of an optical-to-terahertz converter using sapphire fibers. *Russ. Technol. J.* 2023;11(2):50–57. https://doi.org/10.32362/2500-316X-2023-11-2-50-57

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Повышение эффективности оптико-терагерцового преобразователя за счет профилированных сапфировых волокон

Н.В. Зенченко [1, 2, 3, @], **Д.В. Лаврухин** [1, 2], **И.А. Глинский** [1, 2, 3], **Д.С. Пономарев** [1, 2]

[1] *Институт сверхвысокочастотной полупроводниковой электроники им. В.Г. Мокерова Российской академии наук, Москва, 117105 Россия*
[2] *Институт общей физики им. А.М. Прохорова Российской академии наук, Москва, 119991 Россия*
[3] *МИРЭА – Российский технологический университет, Москва, 119454 Россия*
[@] *Автор для переписки, e-mail: zenchenko.nikolay@yandex.ru*

Резюме

Цели. Цель работы – повышение эффективности фотопроводящего ТГц-излучателя большой площади на основе оптико-терагерцового преобразователя (ОТП) (излучающая область составляет 0.3 × 0.3 мм^2) для генерации мощного ТГц-излучения с помощью применения в качестве фокусирующей оптики массива плотноупакованных профилированных сапфировых волокон диаметром в диапазоне 100–300 мкм.

Методы. В качестве фотопроводящей подложки использовался полубесконечный слой LT-GaAs (LT, low-temperature grown GaAs – слой GaAs, выращиваемый методом молекулярно-лучевой эпитаксии при пониженной температуре роста). Далее следуют слои Si_3N_4 и Al_2O_3 для снижения токов утечки в ОТП и уменьшения отражения импульса лазерной накачки от границы «воздух/полупроводник» (френелевские потери) соответственно. Ширина зазора составляет 10 мкм, система металлов Ti/Au используется для формирования электродов антенны и подводящих полосков. Моделирование проводилось методом конечных элементов в среде *COMSOL Multiphysics*.

Результаты. Продемонстрирована способность профилированного сапфирового волокна после оптимизации диаметра относительно параметров зазора значительно увеличить концентрацию носителей заряда в непосредственной близости к электродам ОТП. Определена интегральная эффективность фотопроводящего ТГц-излучателя большой площади с учетом микрополосковой топологии массива с характерным размером подводящих полосков, пропорциональным ширине зазора в ОТП, и с верхним (маскирующим) металлическим слоем. Максимальная локализация электромагнитного поля в непосредственной близости к краям электродов на интерфейсе «волокно/полупроводник» достигается при диаметре профилированного сапфирового волокна, равном 220 мкм.

Выводы. Путем оптимизации диаметра сапфирового волокна продемонстрирована возможность увеличения в ~40 раз локализации падающих электромагнитных волн в непосредственной близости к краям электродов ОТП по сравнению со случаем без волокна, а также повышение до ~7–10 раз общей эффективности излучателя большой площади.

Ключевые слова: терагерцовая импульсная спектроскопия, источники и детекторы ТГц-излучения, субволновая фокусировка излучения, терагерцовые оптические элементы и системы, оптико-терагерцовая конверсия, металинза

• **Поступила:** 20.05.2022 • **Доработана:** 07.10.2022 • **Принята к опубликованию:** 27.01.2023

Для цитирования: Зенченко Н.В., Лаврухин Д.В., Глинский И.А., Пономарев Д.С. Повышение эффективности оптико-терагерцового преобразователя за счет профилированных сапфировых волокон. *Russ. Technol. J.* 2023;11(2):50–57. https://doi.org/10.32362/2500-316X-2023-11-2-50-57

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Today, optical-to-terahertz converters (OTCs) are widely used in terahertz (THz) spectroscopy systems for generating and detecting broadband THz radiation [1]. Due to the flexibility of their manufacturing technology (the possibility of variations in the topology and geometry of antenna electrodes, as well as the choice of photoconductive semiconductor material) OTCs are of considerable interest for creating single-channel and multi-channel detection systems for imaging objects in the THz range [2, 3].

However, the efficiency of OTC emitters is limited by the fact that only a small fraction of the laser pump pulse energy is converted into THz electromagnetic oscillations [4, 5]. One approach for increasing efficiency is to structuring electrode edges by forming periodic metallic (plasmonic) nanostructures in the antenna gap [4–7]. An alternative approach to plasmonic is based on dielectric structures, which are not subject to ohmic losses and overheating with Joule heat generation, allowing the laser pulse to be focused to form local caustics [8–10]. By localizing optical radiation, the efficiency of pump energy transfer to the photoconductive layer can be significantly increased (about 7-fold), leading to improved THz radiation generation efficiency by increasing photocurrent density [6]. The paper describes a means by which this effect can be achieved using lenses based on profiled sapphire fibers (PSF) with diameters of 100–300 μm. Such fibers allow a significant amount of energy to be focused along the entire electrode surfaces of the photoconductive THz emitter [7].

Previously, the formation of subwavelength local caustics (regions of maximum concentration of charge carriers) at the interface with semiconductor have been described [10]. Due to high refractive index of sapphire in a wide range of the electromagnetic spectrum [11], a significant optical contrast at the "fiber–semiconductor" interface can be created, thus allowing localizing photo-excited charge carriers fundamentally near OTC electrodes (at optimum fiber diameter). Localization (focusing) results in increased efficiency of the pump energy transfer into photoconductor and improved THz generation power due to the increase in the photocurrent density [12].

SIMULATION

In the present paper, a large-area photoconductive THz OTC-based radiator concept (with an emitting region is 0.3 × 0.3 mm^2) for generating powerful THz radiation is proposed, in which an array of densely packed PSFs having diameters in the range of 100–300 μm, manufactured by the Bauman Moscow State Technical University, Russia, is used as the focusing optics. The OTC model and the cross section of the structure is shown in Fig. 1; here, d is the PSF diameter, g is the gap size between electrodes, and a is diameter of the pulsed laser pump beam.

The OTC was created at the V.G. Mokerov Institute of Ultra High Frequency Semiconductor Electronics of the Russian Academy of Sciences (IUHFSE RAS, Russia). The OTC structure comprised a sequence of semiconductor and dielectric layers. The LT-GaAs (LT, low-temperature grown GaAs is the GaAs layer grown by molecular beam epitaxy at reduced growth temperature) was used as photoconductive substrate. The next Si_3N_4 and Al_2O_3 layers were intended to reduce leakage currents in OTC and reduce the Fresnel reflection of laser pumping from the air/semiconductor interface, respectively. The gap width was 10 μm. Gold was used to form antenna electrodes and feeder strips. All technological procedures (in particular,

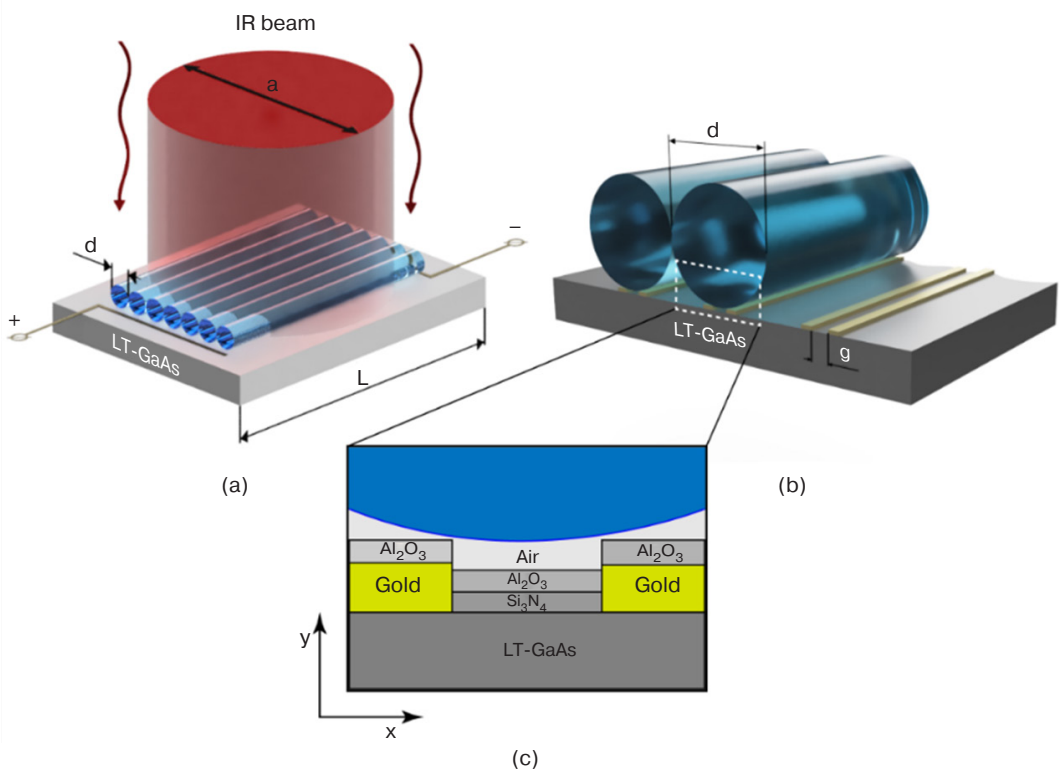

Fig. 1. OTC model (a); enlarged image of the OTC element with focusing optics based on PSF (b); cross-section (c)

deposition of Si_3N_4, Al_2O_3, and gold) were performed at IUHFSE RAS. Further, the resulting substrate was used to create OTC.

The electromagnetic calculation was carried out using the finite element method in the *COMSOL Multiphysics*[1] software environment. The sizes of the finite-element mesh were varied from λ/8 for the gap region to λ/4 for other regions (λ is the wavelength of the laser pumping pulse; in calculations, λ = 780 and 1560 nm). It should be noted that the obtained electromagnetic field distributions for both wavelengths are almost identical due to the optical properties (in particular, the refractive index) of PSF samples differing only to the second decimal place. The electric field propagation vector of the laser pump pulse is oriented along the normal to the OTC surface. The parameter (x/g), where x is the lateral coordinate, was used to make the solution dimensionless and scale simulation results for different PSF diameters (d) and gap sizes (g). The results of simulating the spatial distribution of the electromagnetic wave (EMW) electric field square for three different values of d/g = 14, 18, and 22 are shown in Fig. 2.

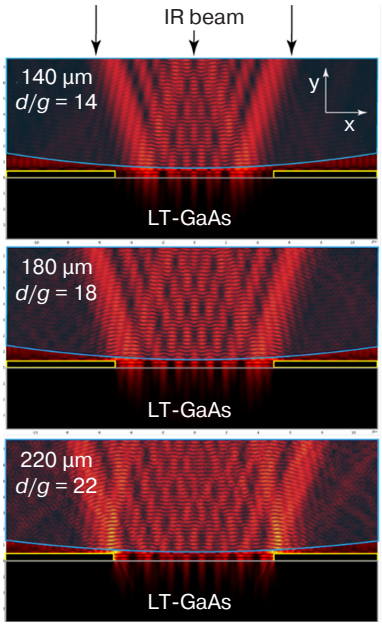

Fig. 2. EMW distribution for different values of parameter d/g

The EMW distribution patterns clearly show subwavelength caustics (characteristic regions of maximum localization of the laser pumping field in the semiconductor) formed near OTC electrodes. The size of the caustics is observed to increase with increasing ratio (d/g) to reach its maximum (in other words, the

[1] https://www.comsol.ru/. Accessed February 01, 2022 (in Russ.).

maximum localization of the laser pump pulse energy) at $d/g = 22$.

The intensity of the electromagnetic field in the OTC gap region is integrated in the following way to qualitatively estimate the number of photoexcited charge carriers that can reach OTC electrodes before their recombination in the semiconductor:

$$I = \int_{-g/2}^{g/2} |E^2(x)| \exp\left(-\left(|x - g/2|\right)/r_d\right) dx,$$

where r_d characterizes the drift length for charge carriers.

Typical values r_d = 100, 300, and 500 nm are selected based on the characteristic values of saturated velocities and carrier lifetime in LT-GaAs [13, 14]. Then, the coefficient of the EMW intensity increase (K) is introduced to quantify the degree of localization of the laser pump pulse in the OTC gap; this is determined by the ratio between integrals I_s and I_0, where I_s and I_0 are calculated both for the case of OTC with PSF and without it, respectively. The simulation results are shown in Fig. 3.

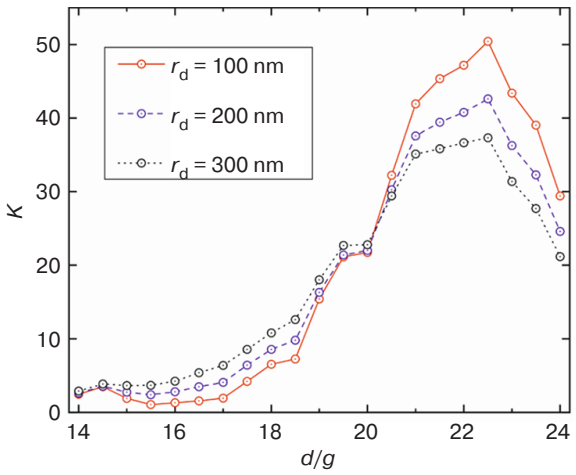

Fig. 3. Coefficient of the EMW intensity increase in the gap

It can be seen that all three curves (corresponding to different r_d values) retain their shape, thus confirming the calculation correctness, while the coefficient K value monotonically increases with increasing parameter d/g reaching its maximum of ~40 at $d/g = 22$. It should be noted that this value characterizes the case when the subwavelength caustics are located strictly at the edges of OTC electrodes, thus allowing a larger number of photoexcited charge carriers to contribute to THz radiation generation. The latter, in turn, results in the increasing photocurrent generated by OTC and potentially increases the efficiency of the optical-THz conversion.

Based on the calculation results obtained, the integral efficiency K of the large-area photoconductive THz emitter with allowance for the microstrip array topology with the characteristic size of feed strips proportional to the gap width in OTC and with an upper (masking) metal layer can be estimated as:

$$K \sim I_s(a/d)/I_0(a/4g) = 4K(d/g)/(d/g),$$

where parameter a characterizes the typical spot diameter of laser pump pulse (1.0–1.5 mm), while digit 4 corresponds to the period of microstrip array structure consisting of two photoconductive gaps and two widths of feed strips.

Possible approaches for the optimization of the design of the large-area THz emitter using PSF should also be noted. As shown in Fig. 1a, the laser beam covers $n = a/d$ of fibers. It would be logical to reduce both the PSF diameter and the gap size for increasing the power of THz radiation generation. However, the decrease of the gap is equivalent to a sharp increase in the electric field strength due to decreasing distance between two adjacent metal strips. This significantly increases the probability of electrical breakdown in LT-GaAs (especially, in photoconductors for IR laser radiation pumping—with relatively small band gap width, for example, InGaAs). In other words, "lower" boundary values for d and g in practice would be: g ~ 3–5 μm, d ~ 100 μm. In addition, any reduction in the gap width imposes additional requirements on the accuracy of PSF alignment with the surface of the OTC sample. According to our estimates, the number of adjoining radiating elements with microstrip topology for the large-area THz emitter should not exceed 10. In this case, the combination of OTC + PSF is arranged so as to achieve maximum efficiency.

CONCLUSIONS

The paper proposes using an array of lenses made of sapphire fiber to increase the efficiency of the large-area photoconductive THz emitter. Using numerical simulation, each lens is shown to provide spatial redistribution of the density of photoexcited charge carriers in the gap between electrodes of single antenna. By optimizing the diameter of the sapphire fiber, the possibility of increasing the localization of incident EMWs in close proximity to the edges of OTC electrodes by ~40 times compared to the case without fiber, as well as increasing to ~7–10 times the overall efficiency of the large-area radiator is demonstrated.

Since the incident laser beam has the diameter of 1.0–1.5 mm in practice, the number of pairs of strip lines on the large-area THz emitter crystal is about 5 at d = 220 μm and g = 10 μm. Therefore, 10 pairs of strip lines with 5 μm gap (d = 110 μm) may be used for increasing the performance of a large-area THz emitter.

ACKNOWLEDGMENTS

The work was supported by the RTU MIREA grant (Innovations in the Implementation of Priority Areas for the Development of Science and Technology). The design on an OTC was developed within the Russian Scientific Foundation grant No. 18-79-10195 while the simulations were supported by the Fund for Assistance to Small Innovative Enterprises, grant No. 16298GU/2021.

Authors' contributions

N.V. Zenchenko—numerical simulation of OTC.

D.V. Lavrukhin—methodology for calculating the enhancement coefficient for the intensity of the EMW in the OTC.

I.A. Glinskiy—numerical simulation of OTC.

D.S. Ponomarev—project management.

REFERENCES

1. Yachmenev A.E., Pushkarev S.S., Reznik R.R., Khabibullin R.A., Ponomarev D.S. Arsenides-and related III-V materials-based multilayered structures for terahertz applications: Various designs and growth technology. *Prog. Cryst. Growth Charact. Mater*. 2020;66(2):100485. https://doi.org/10.1016/j.pcrysgrow.2020.100485
2. Gueroboukha H., Nallappan K., Skorobogatiy M. Toward real-time terahertz imaging. *Adv. Opt. Photonics*. 2018;10(4):843–938. https://doi.org/10.1364/AOP.10.000843
3. Henri R., Nallappan K., Ponomarev D.S., Guerboukha H., Lavrukhin D.V., Yachmenev A.E., Khabibullin R.A., Skorobogatiy M. Fabrication and characterization of an 8 × 8 terahertz photoconductive antenna array for spatially resolved time domain spectroscopy and imaging applications. *IEEE Access*. 2021;9:117691–117702. https://doi.org/10.1109/ACCESS.2021.3106227
4. Yachmenev A.E., Lavrukhin D.V., Glinskiy I.A., Zenchenko N.V., Goncharov Y.G., Spektor I.E., Khabibullin R.A., Otsuji T., Ponomarev D.S. Metallic and dielectric metasurfaces in photoconductive terahertz devices: a review. *Optical Engineering*. 2019;59(6):061608 (19 p.). https://doi.org/10.1117/1.OE.59.6.061608
5. Yardimci N.T., Jarrahi M. Nanostructure-enhanced photoconductive terahertz emission and detection. *Small*. 2018;14(44):1802437. https://doi.org/10.1002/smll.201802437
6. Lepeshov S., Gorodetsky A., Krasnok A., Rafailov E., Belov P. Enhancement of terahertz photoconductive antenna operation by optical nanoantennas. *Laser & Photonics Reviews*. 2017;11(1):1600199. https://doi.org/10.1002/lpor.201600199
7. Castro-Camus E., Alfaro M. Photoconductive devices for terahertz pulsed spectroscopy: a review. *Photon. Res.* 2016;4(3):A36–A42. https://doi.org/10.1364/PRJ.4.000A36
8. Glinskiy I.A., Zenchenko N.V., Ponomarev D.S. All-dielectric metalens based on a single colloidal particle for photoconductive optical-to-terahertz switches. *Russ. Technol. J.* 2020;8(6):78–86 (in Russ.). https://doi.org/10.32362/2500-316X-2020-8-6-78-86
9. Zenchenko N.V., Lavrukhin D.V., Goncharov Yu.G., Yakovlev E.V., Zaytsev K.I., Ponomarev D.S. Subdiffractional local caustics in THz antennas with metasurfaces. In: *The 10th International Scientific and Practical Conference on the physics and technology of nanoheterostructural microwave electronics, MOKEROV READINGS*. Moscow: NIYaU MIFI; 2020. P. 107–108 (in Russ.). Available from URL: http://www.mokerov.ru/%d1%81%d0%b1%d0%be%d1%80%d0%bd%d0%b8%d0%ba-%d1%82%d1%80%d1%83%d0%b4%d0%be%d0%b2-2020/
10. Zenchenko N.V., Lavrukhin D.V., Goncharov Yu.G., Frolov T.V., Katyba G.M., Khabibullin R.A., Zaytsev K.I., Ponomarev D.S. Focusing elements based on sapphire fibers aimed at the enhancement of terahertz radiation generation. In: *The 12th International Scientific and Practical Conference on the physics and technology of nanoheterostructural microwave electronics, MOKEROV READINGS*. Moscow: NIYaU MIFI; 2021. P. 101–102 (in Russ.). Available from URL: http://www.mokerov.ru/%d1%81%d0%b1%d0%be%d1%80%d0%bd%d0%b8%d0%ba-%d1%82%d1%80%d1%83%d0%b4%d0%be%d0%b2-4/
11. Katyba G., Zaytsev K., Dolganova I., Shikunova I., Chernomyrdin N., Yurchenko S., Komandin G., Reshetov I., Nesvizhevsky V., Kurlov V. Sapphire shaped crystals for waveguiding, sensing and exposure applications. *Prog. Cryst. Growth Charact. Mater.* 2018;64(4):133–151. https://doi.org/10.1016/j.pcrysgrow.2018.10.002

12. Lai W., Abdulmunem O.M., Pino P., Pelaz B., Parak W.J., Zhang Q., Zhang H. Enhanced terahertz radiation generation of photoconductive antennas based on manganese ferrite nanoparticles. *Sci. Rep*. 2017;7:46261. https://doi.org/10.1038/srep46261
13. Roux J.-F., Coutaz J.-L., Krotkus A. Time-resolved reflectivity characterization of polycrystalline low-temperature-grown GaAs. *Appl. Phys. Lett.* 1999;74(17):2462. https://doi.org/10.1063/1.123881
14. Liliental-Weber Z., Cheng H.J., Gupta S., Whitaker J., Nichols K., Smith F.W. Structure and carrier lifetime in LT-GaAs. *J. Electron. Mater*. 1993;22(12):1465–1469. https://doi.org/10.1007/BF02650000

About the authors

Nikolay V. Zenchenko, Researcher, V.G. Mokerov Institute of Ultra High Frequency Semiconductor Electronics, Russian Academy of Sciences (7/5, Nagorny pr., Moscow, 117105 Russia); Researcher, Prokhorov General Physics Institute, Russian Academy of Sciences (38, Vavilova ul., Moscow, 119991 Russia); Senior Lecturer, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: zenchenko.nikolay@yandex.ru. Scopus Author ID 56891470400, ResearcherID K-2233-2015, RSCI SPIN-code 7667-6535, https://orcid.org/0000-0002-7932-1821

Denis V. Lavrukhin, Researcher, V.G. Mokerov Institute of Ultra High Frequency Semiconductor Electronics, Russian Academy of Sciences (7/5, Nagorny pr., Moscow, 117105 Russia); Researcher, Prokhorov General Physics Institute, Russian Academy of Sciences (38, Vavilova ul., Moscow, 119991 Russia). E-mail: denis_lavruhin@mail.ru. Scopus Author ID 55794617500, ResearcherID K-2107-2014, RSCI SPIN-code 4006-0978, https://orcid.org/0000-0002-8594-7855

Igor A. Glinskiy, Researcher, V.G. Mokerov Institute of Ultra High Frequency Semiconductor Electronics, Russian Academy of Sciences (7/5, Nagorny pr., Moscow, 117105 Russia); Researcher, Prokhorov General Physics Institute, Russian Academy of Sciences (38, Vavilova ul., Moscow, 119991 Russia); Senior Lecturer, Department of Nanoelectronics, Institute for Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: glinskiy.igor@yandex.ru. Scopus Author ID 57190616854, ResearcherID I-4334-2015, RSCI SPIN-code 6254-0273, https://orcid.org/0000-0002-0477-608X

Dmitry S. Ponomarev, Cand. Sci. (Phys.-Math.), Deputy Director, Leading Researcher, V.G. Mokerov Institute of Ultra High Frequency Semiconductor Electronics, Russian Academy of Sciences (7/5, Nagorny pr., Moscow, 117105 Russia); Leading Researcher, Prokhorov General Physics Institute, Russian Academy of Sciences (38, Vavilova ul., Moscow, 119991 Russia). E-mail: ponomarev_dmitr@mail.ru. Scopus Author ID 37124831400, ResearcherID K-1632-2014, RSCI SPIN-code 6099-9599, https://orcid.org/0000-0002-9567-8927

Об авторах

Зенченко Николай Владимирович, научный сотрудник, Институт сверхвысокочастотной полупроводниковой электроники им. В.Г. Мокерова Российской академии наук (117105, Россия, Москва, Нагорный проезд, д. 7, стр. 5); научный сотрудник, Институт общей физики им. А.М. Прохорова Российской академии наук (119991, Россия, Москва, ул. Вавилова, д. 38); старший преподаватель кафедры наноэлектроники Института перспективных технологий и индустриального программирования ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: zenchenko.nikolay@yandex.ru. Scopus Author ID 56891470400, ResearcherID K-2233-2015, SPIN-код РИНЦ 7667-6535, https://orcid.org/0000-0002-7932-1821

Лаврухин Денис Владимирович, научный сотрудник, Институт сверхвысокочастотной полупроводниковой электроники им. В.Г. Мокерова Российской академии наук (117105, Россия, Москва, Нагорный проезд, д. 7, стр. 5); научный сотрудник, Институт общей физики им. А.М. Прохорова Российской академии наук (119991, Россия, Москва, ул. Вавилова, д. 38). E-mail: denis_lavruhin@mail.ru. Scopus Author ID 55794617500, ResearcherID K-2107-2014, SPIN-код РИНЦ 4006-0978, https://orcid.org/0000-0002-8594-7855

Глинский Игорь Андреевич, научный сотрудник, Институт сверхвысокочастотной полупроводниковой электроники им. В.Г. Мокерова Российской академии наук (117105, Россия, Москва, Нагорный проезд, д. 7, стр. 5); научный сотрудник, Институт общей физики им. А.М. Прохорова Российской академии наук (119991, Россия, Москва, ул. Вавилова, д. 38); старший преподаватель кафедры наноэлектроники Института перспективных технологий и индустриального программирования ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: glinskiy.igor@yandex.ru. Scopus Author ID 57190616854, ResearcherID I-4334-2015, SPIN-код РИНЦ 6254-0273, https://orcid.org/0000-0002-0477-608X

Пономарев Дмитрий Сергеевич, к.ф.-м.н., доцент, заместитель директора по научной работе, ведущий научный сотрудник, Институт сверхвысокочастотной полупроводниковой электроники им. В.Г. Мокерова Российской академии наук (117105, Россия, Москва, Нагорный проезд, д. 7, стр. 5); ведущий научный сотрудник, Институт общей физики им. А.М. Прохорова Российской академии наук (119991, Москва, ул. Вавилова, д. 38). E-mail: ponomarev_dmitr@mail.ru. Scopus Author ID 37124831400, ResearcherID K-1632-2014, SPIN-код РИНЦ 6099-9599, https://orcid.org/0000-0002-9567-8927

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 615.471:616.12
https://doi.org/10.32362/2500-316X-2023-11-2-58-71

RESEARCH ARTICLE

Methodological features of the analysis of the fractal dimension of the heart rate

Margarita O. Bykova [@], Vyacheslav A. Balandin

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] Corresponding author, e-mail: margaritabyckova@yandex.ru

Abstract

Objectives. The aim of the present work is to determine the fractal dimension parameter calculated for a sequence of R–R intervals in order to identify the boundaries of its change for healthy and sick patients, as well as the possibility of its use as an additional factor in the detection of cardiac pathology.

Methods. In order to determine the fractal dimension parameter, the Hurst-, Barrow-, minimum coverage area-, and Higuchi methods are used. For assessing the stationarity of a number of electrocardiography (ECG) intervals, a standard method is used to compare arithmetic averages and variances from samples of the total data array of ECG intervals. To identify differences in fractal dimensions of healthy and sick patients, this parameter was ranked. Using the Kolmogorov–Smirnov two-sample criterion, the difference between the distribution laws in the samples for healthy and sick patients is shown.

Results. Among the considered methods for calculating the fractal dimension, the Higuchi method demonstrates the smallest data spread between healthy patients. By ranking the calculated fractional dimension values, it was possible to identify the difference between this parameter for healthy and sick patients. The difference in the distribution of fractal dimension of healthy and sick patients is shown to be statistically significant for the coverage and Higuchi methods. At the same time, when using the traditional Hurst method, there is no reason to reject the null hypothesis that two groups of patients belong to the same general population.

Conclusions. Based on the obtained data, the difference between the fractal dimension indicators of the duration of R–R intervals of healthy and sick patients is shown to be statistically significant when using the Higuchi method. The fractal dimensions of healthy and sick patients can be effectively distinguished by ranking samples. The results of the research substantiate prospects for further studies aimed at using fractal characteristics of the heart rhythm to identify abnormalities of the latter, which can serve as an additional factor in determining heart pathologies.

Keywords: fractal, fractal dimension, coronary heart disease, chronic heart failure, Higuchi method, minimum coverage area method, Hurst method

• **Submitted:** 30.05.2022 • **Revised:** 14.09.2022 • **Accepted:** 25.01.2023

For citation: Bykova M.O., Balandin V.A. Methodological features of the analysis of the fractal dimension of the heart rate. *Russ. Technol. J*. 2023;11(2):58–71. https://doi.org/10.32362/2500-316X-2023-11-2-58-71

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Методические особенности анализа фрактальной размерности сердечного ритма

М.О. Быкова [@], **В.А. Баландин**

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: margaritabyckova@yandex.ru*

Резюме

Цель. Целью работы было определение параметра фрактальной размерности, рассчитанного для последовательности длительностей R-R интервалов, выявление границы его изменения для здоровых и больных пациентов, а также возможности его использования в качестве дополнительного фактора при выявлении патологии сердечной деятельности.

Методы. Для определения параметра фрактальной размерности используются методики Херста, Барроу, минимальной площади покрытия и Хигучи. При оценке стационарности ряда кардиоинтервалов применяется стандартный метод сравнения средних арифметических и дисперсий по выборкам общего массива данных кардиоинтервалов. Для выявления различий фрактальных размерностей здоровых и больных пациентов выполнено ранжирование данного параметра. С помощью использования двухвыборочного критерия Колмогорова – Смирнова показано различие законов распределения в выборках для здоровых и больных пациентов.

Результаты. Показано, что из рассмотренных методов расчета фрактальной размерности наименьший разброс данных между здоровыми пациентами демонстрирует метод Хигучи. Выполнено ранжирование рассчитанных значений фрактальной размерности, позволившее выявить различие данного параметра для здоровых и больных пациентов. Показано, что различие в распределении фрактальной размерности здоровых и больных пациентов является статистически значимым для методов покрытия и Хигучи. В то же время при использовании традиционного метода Херста нет основания отвергать нулевую гипотезу о принадлежности двух групп пациентов одной генеральной совокупности.

Выводы. На основании полученных данных было показано, что статистически значимое различие между показателями фрактальной размерности длительностей R-R интервалов здоровых и больных пациентов имеет место при применении метода Хигучи. Установлено, что ранжирование выборок позволяет эффективно различать фрактальные размерности здоровых и больных пациентов. Результаты работы показывают перспективность дальнейших исследований, направленных на использование фрактальных характеристик кардиоритма для выявления нарушений последнего, что может служить дополнительным фактором при определении патологии деятельности сердца.

Ключевые слова: фрактал, фрактальная размерность, ишемическая болезнь сердца, хроническая сердечная недостаточность, метод Хигучи, метод минимальной площади покрытия, метод Херста

• **Поступила:** 30.05.2022 • **Доработана:** 14.09.2022 • **Принята к опубликованию:** 25.01.2023

Для цитирования: Быкова М.О., Баландин В.А. Методические особенности анализа фрактальной размерности сердечного ритма. *Russ. Technol. J.* 2023;11(2):58–71. https://doi.org/10.32362/2500-316X-2023-11-2-58-71

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Heart rate variability (HRV) has become one of the effective methods for assessing the neural regulation of the heart, analyzing the interaction between sympathetic and vagal fluctuations, as well as examining their effect on heart rate. Heart rate fluctuations exhibit various linear, non-linear, periodic and irregular oscillation patterns.

Thanks to an intensive study of heart rate variability carried out over several decades, it is now possible to correlate changes in the functioning of the cardiovascular system with the presence of pathologies. In Russia, fundamental research in this area carried out at the scientific institution founded by R.M. Baevsky[1], as well as other scientists [1–4], has made it possible to give a physiological interpretation of heart rate variability through the analysis of information about the state and functioning of systems that regulate heart rhythm.

The conducted studies led to the conclusion that the assessment of the overall heart rate variability helps to carry out early diagnosis of disorders of the cardiovascular system. There are a number of different approaches to the analysis of the heartbeat process. In particular, the analysis of variability based on the study of the statistical parameters of rhythmograms is successfully used. The spectral analysis of rhythmograms based on the fast Fourier transform and subsequent analysis of the spectral density distribution over the frequency range has also become widespread. As a result of the performed studies, it has been shown, for example, that the high-frequency region (0.15–0.4 Hz) is a marker of wandering modulation, while the low-frequency region (0.04–0.15 Hz) mainly reflects sympathetic tone and baroreflex activity. In common with a number of other methods, those based on frequency and time measurements are based on the assumption that HRV signals cannot reflect and quantify the dynamic structure of the signal due to their linear character. A number of methods have also been proposed to evaluate nonlinear properties, including fractal dimension, Lyapunov exponents, correlation dimension, approximate entropy, and downtrend analysis of fluctuations [5–9]. All these methods define the properties of HRV as a non-linear process that reacts to external disturbances in a nonlinear manner. Over the past 2–3 decades, attempts have been made to describe such nonlinear systems from the standpoint of deterministic chaos [10–12], which, in contrast to the everyday understanding of disorder as an absolutely random process, refers to processes characterized by limited randomness such as the heart rate. As noted in [13], a certain norm in terms of randomness is necessary for the normal functioning of an organism. Any significant deviation from the norm, both in the direction of greater order and in the direction of greater randomness, may indicate a disease of the body.

In this paper, we consider fractal dimension analysis, comprising one of the methods for nonlinear study of heart rhythm. Fractal dimension analysis is based on a coefficient that describes fractal structures or sets based on a quantitative assessment of their complexity.[2] This parameter is determined for the sequence of durations of R–R intervals of healthy patients and patients with chronic heart failure (CHF) and coronary heart disease (CHD).

METHODS FOR DETERMINING THE FRACTAL DIMENSION *D*

Benoit Mandelbrot [14] gives the following definition of fractals: “A fractal is a structure consisting of parts that are in some sense similar to the whole.” The main characteristic of self-similar structures, which determines the degree of space indentation, is the fractal dimension D. There are a number of different ways to define it. One of the most popular is the Hurst method [15, 16], based on the ratio of the range of the accumulated deviation to the standard deviation. Another name for this technique is R/S-analysis.

Its essence is expressed by the following formula:

$$\frac{R}{S} = (aN)^H,$$

where a is some constant for a particular process; N is the amount of data; H is the Hurst exponent; S is the standard deviation of the series; R is the range of the accumulated deviation, i.e., the difference between the maximum and minimum values of the accumulated deviation from the average value of series Z in the interval $[1; u]$. In turn, u belongs to the interval from 1 to N.

Figure 1 shows changes in some value X, its accumulated deviation Z and the average value X_{avu} in the interval $[1; u]$:

[1] Roman Markovich Baevsky, Dr. Sci. (Med.), Professor, Honored Scientist of the Russian Federation, Academician at the International Academy of Astronautics, Academician at the International Academy of Informatization. He is one of the founders of aerospace cardiology, space telemetry, and the concept of prenosological diagnostics, having personally carried out the development of a medical control system for Yu.A. Gagarin. In Baevsky’s scientific school, three fundamental directions can be distinguished: ballistocardiography and cardiography; heart rate variability; space medicine and prenosological diagnostics.

[2] Fractal dimension. https://en.wikipedia.org/wiki/Fractal_dimension. Accessed April 23, 2022 (in Russ.).

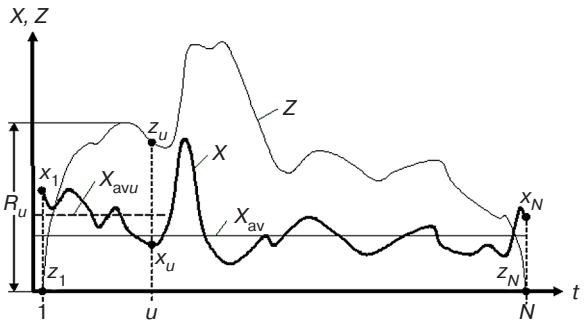

Fig. 1. Changes in the X value (thick line), its accumulated deviation Z (thin line), and the average value of X_{av} (R_u is the range of the accumulated deviation; X_{avu} is the average value of X in the interval; x_1 is the initial value of X; z_1 is the initial value of the accumulated deviation Z; z_u is the value of the accumulated deviation Z taken at the point u; x_u is the value of X, taken at the point u; x_N is the final value of X; z_N is the final value of the accumulated deviation Z)[3]

The fractal dimension is related to the Hurst exponent by the relation:

$$D = 2 - H$$

Another approach for determining the fractal dimension is the Barrow method[4], which consists in finding the average variance of increments W, calculated by the following formula:

$$W(\Delta N) = \frac{1}{N - \Delta N} \sum_{i=1}^{N-\Delta N} \left(x_{i+\Delta N} - x_i \right)^2,$$

where $1 \le \Delta N \le N - 1$.

The dependence $W = f(\Delta N)$ is described by the following equation:

$$W = (a\Delta N)^B,$$

where a is some constant value for the given series of data, B is the Barrow exponent.

The fractal dimension is defined as follows:

$$D = 2 - B.$$

The method of minimum coverage area has also become widespread. It is used, in particular, in the field of economics, as well as having application in the analysis of meteorological series [17]. In this case, after determining the fractal dimension by dividing the data volume N into δ parts, the sum of the amplitude variations for each of the obtained parts is calculated. Then δ changes; following several repetitions of the algorithm, a graph is plotted on a logarithmic scale of the dependence of V on δ. The resulting set of points is approximated by a straight line, after which the slope k is calculated using the least squares method. The process of calculating the coverage area for various values of δ is illustrated by Fig. 2.

In this case, the fractal dimension is found by the formula:

$$D = k + 1.$$

In some studies related to the fractal analysis of biological processes, the Higuchi algorithm was used to

[3] Kobenko V.Yu. *Fractals in science and technology. Guidelines for performing laboratory work in the Microsoft Excel application*. Omsk: OmGTU; 2005. P. 6 (in Russ.).

[4] Ibid. P. 13–15.

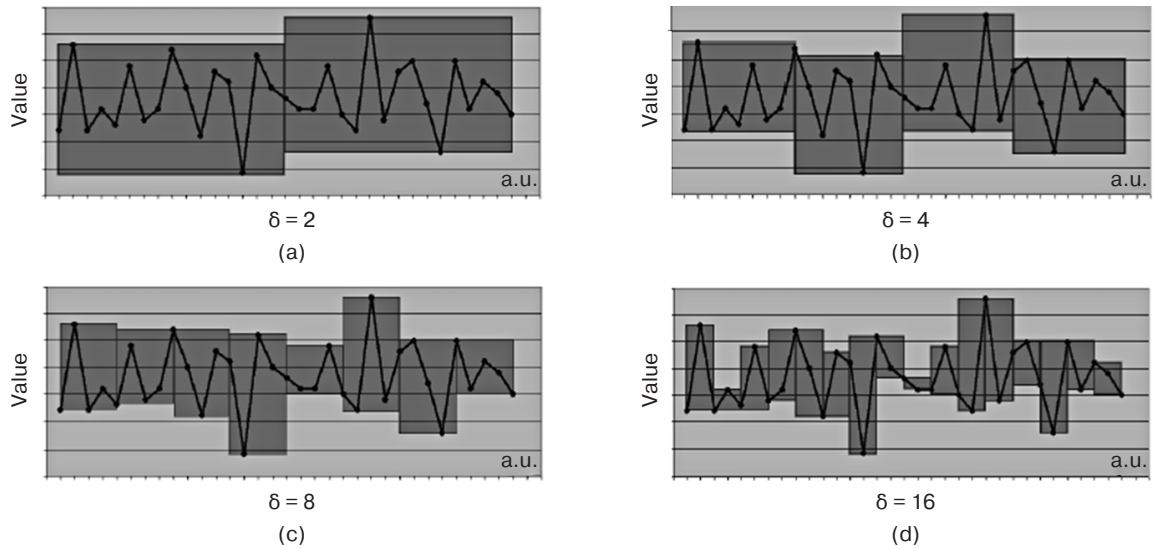

Fig. 2. Calculation of the coverage area for various values of δ [17]

estimate the fractal dimension [15, 18]. This technique will also be used in the present work.

To obtain the fractal dimension D, Higuchi examined a finite set of observations obtained at a regular interval:

$$X(1), X(2), X(3), ..., X(N).$$

A new series, X_k^m, is compiled from these time series as follows:

$$X_k^m : X(m),\ X(m+k), X(m+2k), ..., X\left(m+\left[\frac{N-m}{k}\right]k\right),$$

where m = 1, 2, …, k; k and m are integers; m and k are the serial number of the sample and the size of the interval, respectively.

Within the framework of this method, the length of the curve associated with each time series X_k^m is determined as follows:

$$L_m(k) = \\ = \frac{1}{k}\left(\sum_{i=1}^{\left[\frac{N-m}{k}\right]}(X(m+ik)-X(m+(i-1)k))\right)\left(\frac{N-1}{\left[\frac{N-m}{k}\right]k}\right),$$

where $\frac{N-1}{\left[\frac{N-m}{k}\right]k}$ is the normalizing factor.

The average value $\langle L(k)\rangle$ of the lengths associated with the time series determines the fractal dimension D. At the same time, the following relation takes place:

$$<L(k)> \propto k^{-D}.$$

FRACTIONAL DIMENSION OF HEART RATE VARIABILITY IN HEALTHY AND ILL PATIENTS

To calculate the fractal dimension of HRV in healthy patients, records of R–R intervals from the database "RR interval time series from healthy subjects" were used. For sick patients corresponding data were taken from the databases "Congestive Heart Failure RR Interval Database" and "St. Petersburg INCART 12-lead Arrhythmia Database." The bases are presented in the Physionet[5] open resource of medical signals. The durations of R–R intervals were obtained using the Show RR intervals as text provided by this database, which automatically determines this parameter at a given time interval. At the beginning of the study, a time interval of 450 R–R values was considered, which corresponded to approximately a 5-minute recording. All further calculations were carried out in Microsoft Excel. Algorithms for determining the fractal dimension by the Hurst, Barrow, and coverage methods are described in detail in [15–18]. The procedure for constructing a fractal plane by the Higuchi method consisted of the following steps:

1. Compilation of time series, $k \in [2; 10]$.
2. Calculation of the length of the curve $L_m(k)$ of each series.
3. Determining the average value $\langle L(k)\rangle$ of the lengths associated with the time series for each set of observations.
4. Plotting the dependence of $\langle L(k)\rangle$ on k on a logarithmic scale and determining the value of the fractal dimension D by the least squares method.

The fractal planes determined by the four methods used for one healthy patient are shown in Fig. 3.

The coordinates of points for fractal planes of the R/S method, the Barrow method, the area of least coverage method, and the Higuchi method are highlighted in blue, red, green, and purple, respectively. The black dotted line indicates the approximating trend line, whose slope is used to calculate the value of the fractal dimension. As follows from the graphs, the values of the fractal dimension determined by the last two methods have a significantly lower determination error due to the smaller scatter of the data to be approximated. Using the methods described above, fractal dimensions were found for ten healthy patients, as well as patients with diagnoses of chronic heart failure (CHF) and coronary heart disease (CHD). The obtained values are presented in Table 1.

Since the technique proposed by Barrow uses a similar approach to the traditional Hurst method, it will not be considered further. The fractal dimension D obtained by the Higuchi method shows a smaller spread of data between healthy patients. In this regard, the results obtained by this method will be considered in more detail.

The fact that most of the values of the fractal dimension fall within the interval from 1.5 to 2 is an indicator of the antipersistence of the series. This concept indicates a more frequent change in the direction of the vector of the system development than would be the case with a random sequence. Approximation of the fractal dimension parameter to D = 2, which indicates an increase in the variability of the series, is typical for a situation where the studied sequence tends to completely fill the fractal plane. Values approaching 1.5 indicate that the process tends to complete randomness, i.e., white noise. The fact that the values of the fractal dimension fall into the interval (1.5–2.0) indicates the ergodicity of the process, a special property of certain dynamic systems, which consists in the fact that during development any state with rare exceptions have a certain probability to pass near each other state of the system[6].

[5] https://physionet.org/. Accessed February 14, 2022.

[6] Ergodicity. https://en.wikipedia.org/wiki/Ergodicity. Accessed April 17, 2022 (in Russ.).

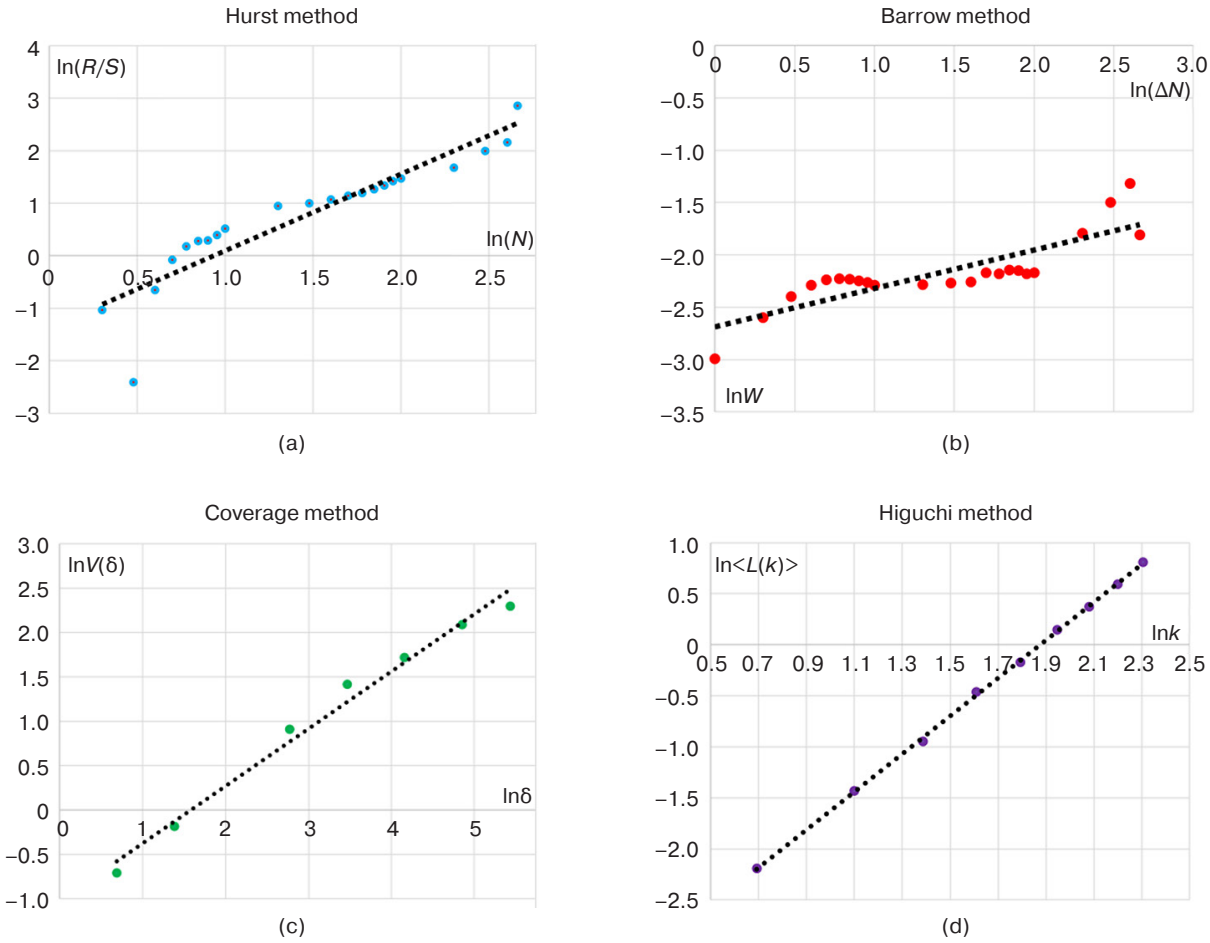

Fig. 3. Fractal plane of a healthy patient:
(a) Hurst method, (b) Barrow method, (c) area of least coverage method, (d) Higuchi method

Table 1. Fractal dimensions *D* of ten healthy patients and patients with CHF and IHD, calculated by various methods

Patient No.	Fractal dimension of healthy patients	Fractal dimension of sick patients (IHD)	Fractal dimension of sick patients (CHF)	Method
1	1.896	1.499	1.754	Hurst method
	1.791	1.502	1.753	Barrow method
	1.612	1.601	1.572	Coverage method
	1.638	1.695	1.685	Higuchi method
2	1.937	1.933	1.787	Hurst method
	1.794	1.886	1.943	Barrow method
	1.722	1.595	1.606	Coverage method
	1.757	1.822	1.981	Higuchi method
3	1.944	1.815	1.432	Hurst method
	1.845	1.767	1.871	Barrow method
	1.665	1.746	1.619	Coverage method
	1.702	1.910	1.805	Higuchi method
4	1.944	1.449	1.518	Hurst method
	1.845	1.790	1.944	Barrow method
	1.665	1.494	1.627	Coverage method
	1.702	1.654	1.797	Higuchi method

Table 1. Continued

Patient No.	Fractal dimension of healthy patients	Fractal dimension of sick patients (IHD)	Fractal dimension of sick patients (CHF)	Method
5	1.529	1.783	1.770	Hurst method
	1.463	1.275	1.994	Barrow method
	1.608	1.598	1.821	Coverage method
	1.630	1.665	1.982	Higuchi method
6	1.536	1.685	1.709	Hurst method
	1.633	1.579	1.994	Barrow method
	1.646	1.528	1.683	Coverage method
	1.767	1.538	2.000	Higuchi method
7	1.459	1.979	1.964	Hurst method
	1.615	1.722	1.863	Barrow method
	1.627	1.814	1.766	Coverage method
	1.751	2.000	1.928	Higuchi method
8	1.627	1.895	1.909	Hurst method
	1.721	1.912	1.776	Barrow method
	1.663	1.738	1.703	Coverage method
	1.708	1.956	1.803	Higuchi method
9	1.665	1.739	1.735	Hurst method
	1.988	1.747	1.699	Barrow method
	1.698	1.573	1.591	Coverage method
	1.772	1.608	1.560	Higuchi method
10	1.559	1.598	1.702	Hurst method
	1.787	1.665	1.949	Barrow method
	1.605	1.648	1.600	Coverage method
	1.675	1.692	1.989	Higuchi method

At the same time, in certain time intervals the statistical characteristics coincide. The alternation of such intervals is due to the presence of a latent periodicity of the process, which is a characteristic, in particular, for the heart rhythm.

DETERMINATION OF SAMPLE SIZE

Sample size is an important feature of any empirical study that aims to draw inferences about a population of parameters from sample observations. The chance of detecting statistically significant differences depends on the sample size and the magnitude of the true difference between the compared indicators [19].

The values of the Higuchi fractal dimension obtained for a healthy patient with an increase in the sample size from 200 to 1500 R–R intervals are presented in Table 2.

Based on the data of Table 2, the fractal dimensions D calculated by the Higuchi method were determined on a sample size equal to 1000 values of the R–R interval.

It should be noted that the fractal dimension calculated by the coating method does not show any dependence on the recording duration for both sick and healthy patients. The D values calculated by this method for samples of 450 and 1000 values are presented in Table 3.

Table 2. Comparison of fractal dimensions for a healthy patient obtained by the Higuchi method for samples of 200, 450, 900, 1000, 1200, and 1500 values of the R–R interval

Fractal dimension calculated by the Higuchi method for different sizes of input data						
Sample size	200	450	900	1000	1200	1500
Healthy patient	1.673	1.702	1.710	1.709	1.707	1.714

Table 3. Comparison of fractal dimensions of 3 patients obtained by the coating method for samples of 450 and 1000 values of R–R intervals

Fractal dimension by the coverage method at interval value		Patient
450	1000	
1.72	1.75	Healthy
1.74	1.72	IHD
1.57	1.59	CHF

Based on the data obtained, it follows that the fractal dimension D calculated by the coverage method is practically independent of the sample size. Therefore, for this method, the sample size is not critical in the considered range of values. Taking into account the obtained results presented in Table 2, we will use a sample with an input data volume of 1000 values of R–R intervals in future studies.

STATIONARITY ESTIMATE

While fractals are closely related to the concept of dynamic chaos, reference is also occasionally made to the concept of deterministic chaos. One of the conditions for the existence of such chaos is non-linearity. These concepts are described in detail in [20]. In this section, the issues of stationarity of a sample of 1000 R–R intervals are considered.

As is known, stationarity is the invariability of the characteristics of a random process over time: the average value and variance of a stationary process remain constant regardless of time, and the autocorrelation function depends only on the difference between the time points at which it is determined. Table 4 shows the average values determined for successive samples consisting of 100 durations of R–R intervals. As can be seen from the presented data, the average values change by 11–25%; meanwhile, the dispersion value also varies significantly from ~50% for healthy patients to 300% for CHF patients. Thus, a number of values of the durations of R–R intervals are not stationary, whether in the presence of pathologies or in their absence.

This fact is an additional confirmation of the expediency of using the concepts of deterministic chaos to describe the array of R–R interval durations.

Using all the previously mentioned algorithms and conclusions, we compiled a table containing the fractal dimensions D for the array of 1000 R–R intervals. The fractal dimensions were calculated using the Hurst-, Minimum Area Coverage- and Higuchi methods.

It should be noted that the used database of patients with CHF contains information on R–R intervals for more than 20 patients. Therefore, the sample size for this pathology was increased to 20 patients. The results obtained are presented in Table 5.

It can be seen that the average values of the fractal dimension for all three methods are not sufficient to

Table 4. Average values and variances for samples of 100 intervals taken from an array of 1000 R–R intervals

Parameter	1–100	100–200	200–300	300–400	400–500	500–600	600–700	700–800	800–900	900–1000	Patient
Average value, s	0.754	0.785	0.779	0.754	0.725	0.639	0.610	0.638	0.660	0.734	CHF
Variance, $s^2 \cdot 10^3$	1.568	3.241	9.201	3.334	1.221	2.159	3.101	7.528	1.116	5.211	
Average value, s	0.799	0.807	0.801	0.755	0.722	0.750	0.742	0.748	0.783	0.753	Healthy
Variance, $s^2 \cdot 10^3$	1.857	1.744	2.400	2.006	2.465	2.608	1.868	3.088	2.253	2.180	

Table 5. Fractal dimensions calculated by three different methods for the array of input data equal to 1000 R–R intervals

Patient No.	Fractal dimension calculated with the Higuchi method			Fractal dimension calculated with the coverage method			Fractal dimension calculated with the Hurst method		
	CHF	Healthy	ISH	CHF	Healthy	ISH	CHF	Healthy	ISH
1	1.621	1.720	1.755	1.587	1.650	1.645	1.728	1.863	1.629
2	1.966	1.783	1.674	1.650	1.746	1.576	2.000	1.979	1.902
3	1.834	1.709	1.897	1.651	1.717	1.719	1.679	1.948	2.000
4	1.832	1.717	1.650	1.665	1.717	1.577	1.519	1.948	1.722
5	1.983	1.703	1.659	1.816	1.648	1.598	2.000	1.718	1.713

Table 5. Continued

Patient No.	Fractal dimension calculated with the Higuchi method			Fractal dimension calculated with the coverage method			Fractal dimension calculated with the Hurst method		
	CHF	Healthy	ISH	CHF	Healthy	ISH	CHF	Healthy	ISH
6	2.028	1.714	1.573	1.693	1.646	1.547	2.000	1.824	1.674
7	1.848	1.752	2.008	1.704	1.676	1.873	1.946	1.704	2.000
8	1.756	1.746	1.903	1.733	1.711	1.720	2.000	1.707	1.816
9	1.488	1.758	1.659	1.641	1.676	1.588	1.884	1.704	1.709
10	1.965	1.679	1.747	1.637	1.630	1.709	1.865	1.691	1.705
11	1.811	–	–	1.710	–	–	1.939	–	–
12	1.783	–	–	1.619	–	–	1.430	–	–
13	1.949	–	–	1.677	–	–	1.786	–	–
14	1.818	–	–	1.668	–	–	1.969	–	–
15	1.879	–	–	1.542	–	–	2.000	–	–
16	1.739	–	–	1.687	–	–	1.669	–	–
17	1.579	–	–	1.632	–	–	1.703	–	–
18	1.739	–	–	1.414	–	–	1.883	–	–
19	1.510	–	–	1.552	–	–	1.507	–	–
20	1.513	–	–	1.616	–	–	1.801	–	–
Average value ± interval of validity	1.782 ± 0.077	1.728 ± 0.022	1.752 ± 0.100	1.645 ± 0.039	1.682 ± 0.028	1.655 ± 0.072	1.866 ± 0.121	1.808 ± 0.084	1.827 ± 0.151

reliably identify the presence of pathology. It should be also noted that the average values of the fractal dimension *D* of healthy patients, determined by the coverage and Higuchi methods, correspond to the data of [21], while the results by the Hurst method give overestimated values. Since the tabular presentation of data is quite difficult for perception and analysis, the data in Table 5 were subjected to additional processing in this work.

RANKING AND STATISTICAL PROCESSING OF THE RESULTS

The ranking method used in the present work consists in arranging objects or phenomena in descending or ascending order of a certain feature inherent in each of them. In this case, the ranking is performed in ascending order of the fractal dimension *D*. Thus, it is necessary to determine the number of ranks and the range of *D* values to which each rank will correspond.

In this work, 10 ranks were used. Despite the fact that the fractal dimension is determined in the range of values from 1 to 2, the vast majority of *D* values is concentrated in the range from 1.5 to 2. This area is divided into 10 ranks through 0.05 with the boundaries of each rank of ±0.025.

After analyzing the data presented in Table 5, it was noticed that most of the values of the fractal dimensions *D* of healthy patients fell into the range of values from 1.675 to 1.825, corresponding to ranks 4 and 5. The distribution by ranks of the fractal dimension of healthy and sick patients is presented in Table 6.

As can be seen from the table, 90% of the dimension values for healthy patients calculated by the Higuchi method have a rank of 4–5. For CHF patients, only 3 out of 20 people have this rank (15%), while for IHD patients this value is 30% (3 out of 10 people). For the dimensions determined by the coverage and Hurst methods, the picture of the distribution of ranks between healthy and sick patients becomes more blurred.

To assess the statistical significance of the division into ranks of healthy and sick patients, the Kolmogorov–Smirnov criterion was used to test the homogeneity of the distribution of two samples [22]. This criterion is based on a comparison of empirical distribution functions that are determined for two samples. The calculated values of the criterion, which are compared with the table values at a significance level of 0.01, are presented in Table 7.

The first line shows the tabular value of the Kolmogorov–Smirnov criterion for 10 healthy and

Table 6. Ranking of the results of Table 5

Patient No.	Ranking for the Higuchi method			Ranking for the coverage method			Ranking for the Hurst method		
	CHF	Healthy	IHD	CHF	Healthy	IHD	CHF	Healthy	IHD
1	2	4	5	2	4	3	4	7	2
2	9	6	4	3	5	2	10	10	8
3	7	4	8	3	4	4	4	9	10
4	7	4	3	3	4	2	0	8	4
5	10	4	3	6	3	2	10	4	4
6	10	4	2	4	3	1	10	6	3
7	7	5	10	4	4	7	9	4	10
8	5	5	8	5	4	4	10	4	6
9	0	5	3	3	4	2	8	4	4
10	9	4	5	3	3	4	6	4	4
11	6			4			9		
12	6			2			0		
13	9			4			6		
14	6			3			9		
15	7			1			10		
16	5			4			3		
17	2			3			4		
18	5			0			8		
19	0			1			0		
20	0			2			6		

20 patients with CHF, the second line is for 10 healthy patients and 10 with coronary artery disease.

The data in Table 7 indicate that the difference in the distribution of the fractal dimension between healthy and sick patients is statistically significant ($D_{\max\,calc} > D_{\max\,tab}$) for the Higuchi method. At the same time, when using the Hurst and coverage methods, there is no reason to reject the null hypothesis concerning the absence of differences in the distribution of results for the two groups of patients, i.e., two samples may belong to the same general population. We also note that the previously reached conclusion, i.e., that the fractal dimension calculated by the Higuchi method is apparently the most preferable in detecting the pathology of the heart, is confirmed by the results of the present study.

Figure 4 shows a diagram of the process of implementing the method of dividing patients into groups according to the presence and absence of pathology based on the value of the fractal dimension of the heart rate.

The diagram in Fig. 4 reflects some of the processing steps. The input data of R–R intervals were obtained on the basis of already analyzed Holter records of healthy and sick patients presented in the Physionet open resource. Holter monitoring data were recorded for

Table 7. Comparison of calculated and tabulated values of the two-sample Kolmogorov–Smirnov D_{max} criterion for healthy and sick patients

Tabular value of criterion D_{max} (for 30 values)	0.290
Tabular value of criterion D_{max} (for 20 values)	0.352
Calculated D_{max} value of Higuchi ranking (CHD and healthy)	0.500
Calculated D_{max} value of Higuchi ranking (IHD and healthy)	0.400
Calculated D_{max} value of coverage ranking (CHF and healthy)	0.350
Calculated D_{max} value of coverage ranking (IHD and healthy)	0.500
Calculated D_{max} value of Hurst ranking (CHF and healthy)	0.200
Calculated D_{max} value of Hurst ranking (IHD and healthy)	0.200

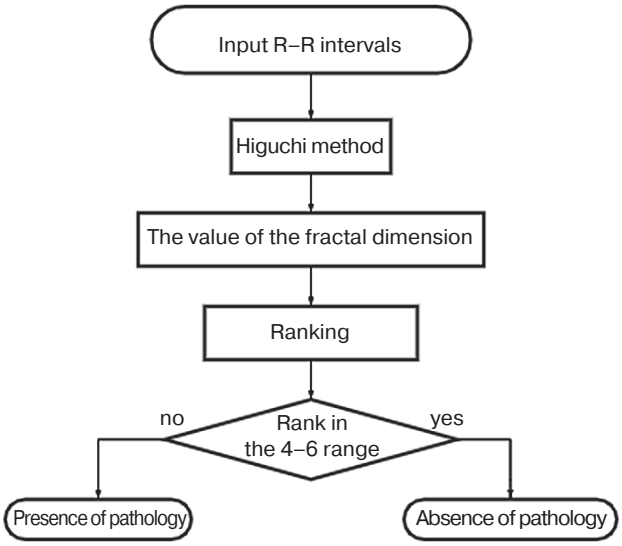

Fig. 4. Algorithm for the process of dividing patients into groups according to the presence and absence of pathology based on the value of the fractal dimension of the heart rate

24 h using DMS300-7 and DMS300-3A digital three-channel recorders, as well as Galix recorders using 3M electrodes. The Galix recorders had programmable read sampling rates of 512 and 1024 Hz and write sampling rates of 128 Hz. The DMS recorders had a sampling rate of 1024 Hz per channel for electrocardiography (ECG) analysis with an averaged signal, a read sampling rate of 512 Hz, and a recording sampling rate of 128 Hz otherwise. Signals were analyzed using Galix software and CardioScan 10.0, 11.0 software[7] for DMS recorders. The error in determining the R–R interval was approximately 8 ms (2 times more than the error in determining the R-wave). After automatic detection and classification using the *Holter* software[8], cardiac events in the records were then checked and corrected by two cardiologists. The recordings were then analyzed beat by beat to identify and correct as many R-waves as possible. Thus, the number and duration of artifacts in the signal was reduced.

The input R–R intervals then need to be processed using the Higuchi method presented above.

To determine the rank of the obtained value D, it is necessary to use Table 8, which presents the rank numbers, the ranges of their values and the average values of each.

Table 8. Ordinal numbers of ranks, their ranges and average values

Rank No.	Range of D values	Average value of D
0	1.500–1.525	1.5125
1	1.525–1.575	1.55
2	1.575–1.625	1.6
3	1.625–1.675	1.65
4	1.675–1.725	1.7
5	1.725–1.775	1.75
6	1.775–1.825	1.8
7	1.825–1.875	1.85
8	1.875–1.925	1.9
9	1.925–1.975	1.95
10	1.975–2.000	1.9875

If the rank value falls within the interval 4–6, then we can conclude that the input data of the R–R intervals belong to a healthy patient, otherwise there is reason to believe that the patient has pathology.

CONCLUSIONS

Mathematical approaches based on a study of the fractality of the non-stationarity or irregularity (in the geometric sense) of processes represent one of the

[7] https://vdd-pro.ru/ru/usb-kardiograf/programmnoe-obespechenie/. Accessed February 14, 2022 (in Russ.).

[8] https://dms-at.ru/products/programs/programmnoe-obespechenie-kholter/. Accessed February 14, 2022 (in Russ.).

possible approaches for their assessment. If considered from the point of view of amplitude changes in the magnitude of the electric potential over time, the graphic record of the ECG also represents a curve that has a strongly irregular shape. If we look at the rhythmogram in time, then we can come to the conclusion that in some way it has the properties of a fractal.

In the present work, we have considered several methods for obtaining the fractal dimension of a sequence of cardio interval durations, namely, the Hurst, Barrow, minimum coverage area and Higuchi methods.

Based on the data obtained, it is shown that the difference between the fractal dimensions of the durations of the R–R intervals of healthy and sick patients is statistically significant at a significance level of 0.01 when using the Higuchi method. The ranking of samples has demonstrated the possibility of effectively distinguishing between the fractal dimensions of healthy and sick patients. The results of the work demonstrate the promise of further research aimed at using the fractal characteristics of the heart rhythm to identify its failure, which can serve as an additional factor in determining the pathology of the heart.

Authors' contributions

M.O. Bykova—collection and analysis of information for the article, calculations and analysis of fractal dimension obtained by various methods.

V.A. Balandin—conceptual idea, discussion and analysis of the obtained results.

REFERENCES

1. Baevskii P.M. Cybernetic analysis of heart rate control processes. In: *Aktual'nye problemy fiziologii i patologii krovoobrashcheniya* (*Actual Problems of Physiology and Pathology of Blood Circulation*). Moscow: Meditsina; 1976. P. 161–175 (in Russ.).
2. Baevskii P.M., Berseneva A.P. *Otsenka adaptatsionnykh vozmozhnostei organizma i risk razvitiya zabolevanii* (*Assessment of the Adaptive Capabilities of the Body and the Risk of Developing Diseases*). Moscow: Meditsina; 1997. P. 265. (in Russ.).
3. Baevskii P.M. *Prognozirovanie sostoyanii na grani normy i patologii* (*Prediction of Conditions on the Verge of Norm and Pathology*). Moscow: Meditsina; 1979. 205 p. (in Russ.).
4. Baevskii P.M., Ivanov G.G. Cardiac rhythm variability: the theoretical aspects and the opportunities of clinical application (lecture). *Ul'trazvukovaya i funktsional'naya diagnostika = Ultrasound and Functional Diagnostics*. 2001;3:106–127 (in Russ.).
5. Antonov V.I., Zagainov A.I., Vu van Quang. Dynamic fractal analysis of heart rate variability. *Nauchno-tekhnicheskie vedomosti Sankt-Peterburgskogo gosudarstvennogo politekhnicheskogo universiteta. Informatika. Telekommunikatsii. Upravlenie = St. Petersburg Polytechnical University Journal. Computer Science. Telecommunication and Control Systems.* 2012;1(140):88–94 (in Russ.).
6. Fedorov V.A., Mizirin A.V., Khramtsov P.I., Agafonova N.A. Fractal analysis of heart rhythm. *Voprosy sovremennoi pediatrii = Current Pediatrics*. 2006;5(1):596 (in Russ.).
7. Yatsyk V.Z., Paramzin V.B., Bolotin A.E., Vorotova M.S. Fractal analysis of variability of heart rhythm among female biathletes with a different training level. *Fizicheskaya kul'tura, sport – nauka i praktika = Physical Education, Sports – Science and Practice*. 2018;4:95–102 (in Russ.).
8. Sen J., McGill D. Fractal analysis of heart rate variability as a predictor of mortality: A systematic review and meta-analysis. *Chaos*: *An Interdisciplinary Journal of Nonlinear Science*. 2018;28(7):072101. https://doi.org/10.1063/1.5038818

СПИСОК ЛИТЕРАТУРЫ

1. Баевский Р.М. Кибернетический анализ процессов управления сердечным ритмом. В сб.: *Актуальные проблемы физиологии и патологии кровообращения*. М.: Медицина; 1976. С. 161–175.
2. Баевский Р.М., Берсенева А.П. *Оценка адаптационных возможностей организма и риск развития заболеваний*. М.: Медицина; 1997. С. 265.
3. Баевский Р.М. *Прогнозирование состояний на грани нормы и патологии*. М.: Медицина; 1979. 205 с.
4. Баевский Р.М., Иванов Г.Г. Вариабельность сердечного ритма: теоретические аспекты и возможности клинического применения. *Ультразвуковая и функциональная диагностика*. 2001;3:106–127.
5. Антонов В.И., Загайнов А.И., Ву ван Куанг. Динамический фрактальный анализ вариабельности сердечного ритма. *Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление.* 2012;1(140):88–94.
6. Федоров В.А., Мизирин А.В., Храмцов П.И., Агафонова Н.А. Фрактальный анализ ритма сердца. *Вопросы современной педиатрии*. 2006;5(1):596.
7. Яцык В.З., Парамзин В.Б., Болотин А.Э., Воротова М.С. Фрактальный анализ вариабельности сердечного ритма у биатлонисток с разным уровнем тренированности. *Физическая культура, спорт – наука и практика.* 2018;4:95–102.
8. Sen J., McGill D. Fractal analysis of heart rate variability as a predictor of mortality: A systematic review and meta-analysis. *Chaos*: *An Interdisciplinary Journal of Nonlinear Science*. 2018;28(7):072101. https://doi.org/10.1063/1.5038818
9. Рахимов Н.Г., Олимзода Н.Х., Мурадов А.М., Мурадов А.А., Хусаинова М.Б. Некоторые показатели фрактального анализа вариабельности сердечного ритма, как предикторы тяжелой преэклампсии и эклампсии. *Вестник последипломного образования в сфере здравоохранения*. 2018;1:70–75.

9. Rakhimov N.G., Olimzoda N.Kh., Muradov A.M., Muradov A.A., Khusainova M.B. Some indicators of fractal analysis heart rate variability as predictors of the development of severe preeclampsia and eclampsia. *Vestnik poslediplomnogo obrazovaniya v sfere zdravookhraneniya = Herald of Postgraduate Education Health Sphere*. 2018;1:70−75 (in Russ.).
10. Goldberger E. L. Goldberger A.L., Rigney D.R., West B.J. Chaos and fractals in human physiology. *V mire nauki* (Russian version) = *Sci. American.* 1990;4:25−32 (in Russ.).
11. Shuster G. *Determinirovannyi khaos: Vvedenie* (*Deterministic chaos: An Introduction*). Moscow: Mir; 1988. 248 p. (in Russ.).
12. Lorenz E. Deterministic non-periodic motion. In: *Strange attractors: collection of articles*. Moscow: Fizmatlit; 1981. P. 88−116 (in Russ.).
13. Klimontovich Yu.L. *Turbulentnoe dvizhenie i struktura khaosa: Novyi podkhod k statisticheskoi teorii otkrytykh system* (*Turbulent motion and the structure of chaos: A new approach to the statistical theory of open systems*). Moscow: URSS; 2021. 326 p (in Russ.).
14. Mandelbrot B. *Fractal geometry of nature*. Moscow: Institut komp'yuternykh issledovanii; 2002. 656 p. (in Russ.).
15. Cervantes-De la Torre F., González-Trejo J.I., Real-Ramírez C.A., Hoyos-Reyes L.F. Fractal dimension algorithms and their application to time series associated with natural phenomena. *J. Phys.: Conf. Ser*. 2013;475(1):012002. http://doi.org/10.1088/1742-6596/475/1/012002
16. Antipov O.I., Nagornaya M.Yu. Bioenergetic signal Hearst index. *Infokommunikatsionnye tekhnologii = Infocommunication Technologies*. 2011;9(1):75−77 (in Russ.).
17. Aptukov V.N., Mitin V.Yu. Fractal analysis of meteorological series based on the minimum covering method. *Geograficheskii vestnik = Geographical Bulletin*. 2019;2(49):67–79 (in Russ.). https://doi.org/10.17072/2079-7877-2019-2-67-79
18. Gladun K.V. Higuchi fractal dimension as a method for assessing response to sound stimuli in patients with diffuse axonal brain injury. *Sovremennye tekhnologii v meditsine = Modern Technologies in Medicine*. 2020;12(4):63–71 (in Russ.). https://doi.org/10.17691/stm2020.12.4.08
19. Koichubekov B.K., Sorokina M.A., Mkhitaryan K.E. Sample size determination in planning of scientific research. *Mezhdunarodnyi zhurnal prikladnykh i fundamental'nykh issledovanii = International Journal of Applied and Fundamental Research*. 2014;4:71−74 (in Russ.).
20. Anishchenko V.S. Deterministic chaos. *Sorosovskii obrazovatel'nyi zhurnal = Soros Educational Journal*. 1997;6:70–76 (in Russ.).
21. Meganur R., Zadidul K., Maksudul H., Jarin S. Successive RR interval analysis of PVC with sinus rhythm using fractal dimension, Poincaré plot and sample entropy method. *Int. J. Image, Graphics and Signal Processing* (*IJIGSP*). 2013;5(2):17−24. https://doi.org/10.5815/ijigsp.2013.02.03
10. Голдбергер Э.Л., Ригни Д.Р., Уэст Б.Дж. Хаос и фракталы в физиологии человека. *В мире науки*. 1990;4:25−32.
11. Шустер Г. *Детерминированный хаос: Введение.* М.: Мир; 1988. 248 с.
12. Лоренц Э. Детерминированное непериодическое движение. *Странные аттракторы: сборник статей*. М.: Физматлит; 1981. С. 88−116.
13. Климонтович Ю.Л. *Турбулентное движение и структура хаоса: Новый подход к статистической теории открытых систем.* М.: URSS; 2021. 326 с. ISBN 978-5-9710-8442-6
14. Мандельброт Б. *Фрактальная геометрия природы.* М.: Институт компьютерных исследований; 2002. 656 с.
15. Cervantes-De la Torre F., González-Trejo J.I., Real-Ramírez C.A., Hoyos-Reyes L.F. Fractal dimension algorithms and their application to time series associated with natural phenomena. *J. Phys.: Conf. Ser*. 2013;475(1):012002. http://doi.org/10.1088/1742-6596/475/1/012002
16. Антипов О.И., Нагорная М.Ю. Показатель Херста биоэнергетических сигналов. *Инфокоммуникационные технологии*. 2011;9(1):75−77.
17. Аптуков В.Н., Митин В.Ю. Фрактальный анализ метеорологических рядов с помощью метода минимального покрытия. *Географический вестник.* 2019;2(49):67−79. https://doi.org/10.17072/2079-7877-2019-2-67-79
18. Гладун К.В. Фрактальная размерность Хигучи как метод оценки реакции на звуковые стимулы у пациентов с диффузным аксональным повреждением головного мозга. *Современные технологии в медицине*. 2020;12(4):63–71. https://doi.org/10.17691/stm2020.12.4.08
19. Койчубеков Б.К., Сорокина М.А., Мхитарян К.Э. Определение размера выборки при планировании научного исследования. *Международный журнал прикладных и фундаментальных исследований*. 2014;4:71−74.
20. Анищенко В.С. Детерминированный хаос. *Соросовский образовательный журнал*. 1997;6:70–76.
21. Meganur R., Zadidul K., Maksudul H., Jarin S. Successive RR interval analysis of PVC with sinus rhythm using fractal dimension, Poincaré plot and sample entropy method. *Int. J. Image, Graphics and Signal Processing* (*IJIGSP*). 2013;5(2):17−24. https://doi.org/10.5815/ijigsp.2013.02.03
22. Орлов А.И. Система моделей и методов проверки однородности двух независимых выборок. *Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета*. 2020;157:145−169. https://doi.org/10.21515/1990-4665-157-012

22. Orlov A.I. System of models and methods of testing the homogeneity of two independent samples. *Politematicheskii setevoi elektronnyi nauchnyi zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta = Polythematic online Electronic Scientific Journal of Kuban State Agrarian University.* 2020;157:145–169 (in Russ.). https://doi.org/10.21515/1990-4665-157-012

About the authors

Margarita O. Bykova, Student, Department of Biocybernetic Systems and Technologies, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: margaritabyckova@yandex.ru. https://orcid.org/0000-0001-5398-3184

Vyacheslav A. Balandin, Cand. Sci. (Phys.–Math.), Assistant Professor, Department of Biocybernetic Systems and Technologies, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: admiral49@mail.ru. Scopus Author ID 7003691025, RSCI SPIN-code 1288-9918

Об авторах

Быкова Маргарита Олеговна, магистрант кафедры биокибернетических систем и технологий Института искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: margaritabyckova@yandex.ru. https://orcid.org/0000-0001-5398-3184

Баландин Вячеслав Алексеевич, к.ф.-м.н., доцент кафедры биокибернетических систем и технологий Института искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: admiral49@mail.ru. Scopus Author ID 7003691025, SPIN-код РИНЦ 1288-9918

Translated from Russian into English by Evgenii I. Shklovskii
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 519.857
https://doi.org/10.32362/2500-316X-2023-11-2-72-83

RESEARCH ARTICLE

Optimization of spline parameters in approximation of multivalued functions

Dmitry A. Karpov,
Valery I. Struchenkov @

MIREA – Russian Technological University, Moscow, 119454 Russia
@ *Corresponding author, e-mail: str1942@mail.ru*

Abstract

Objectives. Methods for spline approximation of a sequence of points in a plane are increasingly used in various disciplines. A spline is defined as a single-valued function consisting of a known number of repeating elements, of which the most widely used are polynomials. When designing the routes of linear structures, it is necessary to consider a problem with an unknown number of elements. An algorithm implemented for solving this problem when designing a longitudinal profile was published earlier. Here, since the spline elements comprise circular arcs conjugated by line segments, the spline is a single-valued function. However, when designing a route plan, the spline is generally a multivalued function. Therefore, the previously developed algorithm is unsuitable for solving this problem, even if the same spline elements are used. The aim of this work is to generalize the obtained results to the case of approximation of multivalued functions while considering various features involved in designing the routes of linear structures. The first stage of this work consisted in determining the number of elements of the approximating spline using dynamic programming. In the present paper, the next stage of solving this problem is carried out.

Methods. The spline parameters were optimized using a new mathematical model in the form of a modified Lagrange function and a special nonlinear programming algorithm. In this case, it is possible to analytically calculate the derivatives of the objective function with respect to the spline parameters in the absence of its analytical expression.

Results. A mathematical model and algorithm were developed to optimize the parameters of a spline as a multivalued function consisting of circular arcs conjugated by line segments. The initial approximation is the spline obtained at the first stage.

Conclusions. The previously proposed two-stage spline approximation scheme for an unknown number of spline elements is also suitable for approximating multivalued functions given by a sequence of points in a plane, in particular, for designing a plan of routes for linear structures.

Keywords: route, plan and longitudinal profile, spline, nonlinear programming, objective function, constraints

• **Submitted:** 02.03.2022 • **Revised:** 01.06.2022 • **Accepted:** 26.01.2023

For citation: Karpov D.A., Struchenkov V.I. Optimization of spline parameters in approximation of multivalued functions. *Russ. Technol. J*. 2023;11(2):72–83. https://doi.org/10.32362/2500-316X-2023-11-2-72-83

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Оптимизация параметров сплайна при аппроксимации многозначных функций

Д.А. Карпов,
В.И. Струченков [@]

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: str1942@mail.ru*

Резюме

Цели. Методы сплайн-аппроксимации последовательности точек на плоскости получают все более широкое применение в различных областях. Сплайн рассматривается как однозначная функция с известным числом повторяющихся элементов. Наиболее широкое применение получили полиномиальные сплайны. Применительно к проектированию трасс линейных сооружений приходится рассматривать задачу с неизвестным числом элементов. Алгоритм решения задачи применительно к проектированию продольного профиля реализован и опубликован ранее. В этой задаче элементами сплайна являются дуги окружностей, сопрягаемые отрезками прямых, и сплайн представляет собой однозначную функцию. Однако при проектировании плана трассы в общем случае сплайн является многозначной функцией. Поэтому разработанный ранее алгоритм не пригоден для решения этой задачи, даже в случае использования тех же элементов сплайна. Цель настоящей статьи – обобщение полученных результатов на случай аппроксимации многозначных функций с учетом особенностей проектирования трасс линейных сооружений. На первом этапе работы было определено число элементов аппроксимирующего сплайна с помощью динамического программирования. В статье рассматривается следующий этап решения задачи.

Методы. Для оптимизации параметров сплайна используется новая математическая модель в виде модифицированной функции Лагранжа и специальный алгоритм нелинейного программирования. При этом удается вычислять аналитически производные целевой функции по параметрам сплайна при отсутствии ее аналитического выражения через эти параметры.

Результаты. Разработаны математическая модель и алгоритм оптимизации параметров сплайна (как многозначной функции), состоящего из дуг окружностей, сопрягаемых отрезками прямых. Начальным приближением является сплайн, полученный на первом этапе.

Выводы. Двухэтапная схема сплайн-аппроксимации при неизвестном числе элементов сплайна, предложенная ранее, пригодна и для аппроксимации многозначных функций, заданных последовательностью точек на плоскости, в частности для проектирования плана трасс линейных сооружений.

Ключевые слова: трасса, план и продольный профиль, сплайн, нелинейное программирование, целевая функция, ограничения

• **Поступила:** 02.03.2022 • **Доработана:** 01.06.2022 • **Принята к опубликованию:** 26.01.2023

Для цитирования: Карпов Д.А., Струченков В.И. Оптимизация параметров сплайна при аппроксимации многозначных функций. *Russ. Technol. J.* 2023;11(2):72–83. https://doi.org/10.32362/2500-316X-2023-11-2-72-83

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

The previously proposed [1] method for approximating multivalued functions defined discretely by a special type of spline uses a two-stage scheme for solving the problem. At the first stage, the number of elements of the spline along with a calculation of the approximate spline parameter values is determined using dynamic programming method. At the second stage, the parameters are optimized by nonlinear programming using the spline obtained at the first stage as an initial approximation. The first stage was performed in our previous work [2]. The present article, which is a continuation of that work, considers the second stage.

A spline represents a chain of repeating "circular arc + + line segment" elements. At this stage, the starting point, both the direction of the tangent at this point, as well as the lengths of all the arcs and their conjugating line segments, are known. Despite the fact that the desired spline is a multivalued function, this allows continuous optimization methods to be applied—in particular, methods of nonlinear programming of the gradient type.

Optimization of the parameters of the spline obtained at the first stage is necessary not only due to the insufficient accuracy of the solution of the problem at the first stage, which is due to the discreteness of the search, but also because of the impossibility of strictly imposing the constraints on fixed points at the first stage, i.e., the starting points, which are not displaced during the approximation.

As is common practice in dynamic programming, accuracy can be improved by repeating the calculations at smaller search steps. In this problem, this is particularly important because, at a known number of elements, the amount of computation is sharply reduced, which enables one to solve the problem at reduced discretes with an increase in their number in a reasonable time on public computers.

The problem is considered as applied to designing a plan for the routes of linear structures. For some of them, e.g., for a trench for laying pipelines of various purposes, the spline of the considered type is final. When designing horizontal road alignment, straight lines and circles should be conjugated by clothoids to ensure continuity not only of the tangent, but also of the curvature. It was shown [2] that, if the clothoids are short, their addition to the resulting spline with circles leads to insignificant displacements. However, for the general case, it is necessary to implement a step-by-step spline approximation scheme with repeating elements "straight line + clothoid + circle + clothoid." The solution of this problem will be the subject of further research.

As shown previously [2], this approach differs significantly from the method of selecting elements in interactive mode as accepted in design practice, from various semiautomatic methods for searching for curve boundaries based on curvature graphs and angular diagrams, as well as from a novel heuristic method for searching curve boundaries [3] with subsequent application of genetic algorithms [4–12]. In contrast, the use of adequate mathematical models and mathematically correct algorithms seems to be more promising.

1. PROBLEM STATEMENT AND ITS FORMALIZATION

The problem is to find a spline of a given type that satisfies all the constraints and best approximates a given sequence of points in a plane [2, Fig. 1].

The starting point A and the direction of the tangent to the desired spline at this point are set and remain unchanged during the search for the spline. The quality of the approximation is estimated by the sum of the squared deviations h_i of the given points of the spline.

It is necessary to find

$$\min F(\mathbf{h}) = \sum_1^n h_j^2. \quad (1)$$

Here, $\mathbf{h}(h_1, h_2, \ldots, h_n)$ is the vector of unknowns, while n is their number. Instead of a simple sum, a weighted sum of squares can be given.

Deviations h_j are calculated differently in design practice in different countries and in corresponding studies carried out by various researchers. Typically, the deviation of a point from a spline is calculated along the normal to the spline [2]. In Russia, however, it is customary to calculate the deviation along the normal to the original polyline [2], i.e., toward the center of the circle connecting three adjacent points. If three points lie on the same straight line, then the deviation is calculated along the normal to this straight line.

Since the noted difference in calculation methods does not affect the search for the number of spline elements, the simplest method was adopted at the first stage, i.e., calculation along the normal to the desired spline. At the second stage, when optimizing the spline parameters, we use precomputed normals to the original polyline, i.e., fixed directions that do not need to be recalculated in an iterative process. These are the same normals that contain the points that determine the "states of the system" in dynamic programming [2, Fig. 2].

The starting point of the first curve may not coincide with the starting point A; therefore, the length L_1 of the initial line is considered unknown, and, unlike the first stage, the spline elements are considered in the order "straight line + circle". If the number of such repeating elements is m, then the system of constraints has the form

$$L_j^{\text{line}} \geq L_{\min}^{\text{line}}, \quad (2)$$

$$L_j^{\text{curve}} \geq L_{\min}^{\text{curve}}, \quad (3)$$

$$R_{\min} \leq \left|R_j\right| \leq R_{\max}, \; j = \overline{1,m}. \quad (4)$$

Here, as at the first stage [2], L_j^{line} and L_j^{curve} are the lengths of the straight line and the curve in the jth element, respectively, while R_j are the radii of the circles, whose signs are known. This makes it possible to avoid taking the absolute value in constraint (4) and obtain a linear constraint on each R_j in the form of a two-sided inequality:

$$R_{\min} \leq R_j \leq R_{\max}, \text{ if } R_j > 0, \quad (5)$$

$$-R_{\max} \leq R_j \leq -R_{\min}, \text{ if } R_j < 0. \quad (6)$$

The end point of the spline is fixed, but its length is unlimited. If the final direction is also fixed, then constraints are imposed not only on h_n, but also on h_{n-1}. In addition, constraints can also be imposed on the displacements of individual points in the form of both inequalities, including double ones,

$$h_{\min} \leq h_m \leq h_{\max}, \quad (7)$$

and equalities,

$$h_m = h_0. \quad (8)$$

These are the same fixed points whose presence cannot be taken into account in dynamic programming.

As a result, we obtain a nonlinear programming problem with the objective function $F(\mathbf{h})$ under constraints (2), (3), and (5)–(8), some of which may be absent.

2. FEATURES OF THE PROBLEM

Constraints (2), (3), (5), and (6) are not expressed in terms of unknown displacements h_i, but if all the lengths and radii are known, then all the h_j can be calculated. Further, all the lengths and radii are considered as the main variables, and all the h_i are regarded as intermediate variables, which depend on the main ones. Analytical expressions of these dependences are unknown and will not be determined. There is also no analytical expression of the objective function $F(\mathbf{h})$ in terms of the main variables. As a result, we obtain a nonlinear programming problem under a simple system of constraints (2), (3) and (5), (6) on the main variables, under several constraints (7), (8) on intermediate variables, and with the objective function expressed in terms of intermediate variables.

Nonlinear programming algorithms, with all their diversity[1] [13–26], reduce to an iterative process with the following steps:

1) construct an admissible initial approximation;
2) determine the direction of descent from the next iteration point, in particular, from the starting point;
3) check the conditions for terminating the account. If they are not met, then go to the next item, otherwise, end the calculations;
4) find the step in the found direction from the condition that the constraints are satisfied and the minimum point in the direction is reached;
5) go to a new point, and then go to step 2.

In order to solve our problem, we need to repeatedly calculate intermediate variables (normal displacements) as the main variables are changed. To do this, the intersection points of two straight lines and a straight line with a circle have to be found (Fig. 1).

The shifts of the initial points to the design position are considered positive if they are directed along with the outward normal.

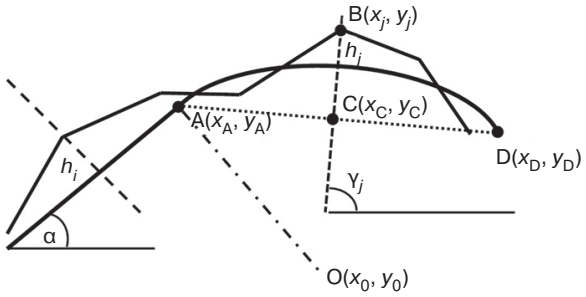

Fig. 1. Calculation of normal displacements

Let x_A and y_A be the coordinates of the beginning of the arc of the circle (point A in Fig. 1) and α is the angle of the tangent at this point with the OX axis. Then the coordinates of the center of the circle are written as $x_0 = x_A - R\sin\alpha$ and $y_0 = y_A + R\cos\alpha$. Here and henceforth, the radius is positive when moving along the curve counterclockwise. The point C of intersection with the normal can be both outside and inside the arc of the circle. Without loss of generality, for the point of intersection of the normal with the circle, one can write $x_C = x_j + h_j\cos\gamma_j$ and $y_C = y_j + h_j\sin\gamma_j$. Here and henceforth, γ_j is the angle of the jth normal with the OX axis.

From the condition that the point C belongs to the circle, one can obtain a quadratic equation for h_j, the solution of which gives a formula for h_j:

$$h_j = (x_A - R\sin\alpha - x_j)\cos\gamma_j + (y_A + R\cos\alpha - y_j)\sin\gamma_j \pm \pm\sqrt{(R^2 - [(x_A - R\sin\alpha - x_j)\sin\gamma_j - (y_A + R\cos\alpha - y_j)\cos\gamma_j]^2)}. \quad (9)$$

[1] Panteleev A.V., Letova T.A. *Optimization methods*: A handbook. Moscow: Logos; 2011. 424 p. ISBN 978-5-98704-540-4 (in Russ.).

To select a point on the circle that is closest to the analyzed point, a minus sign is placed in front of the root if the expression in front of the root in formula (9) is positive, and *vice versa*. The deviations h_j are calculated sequentially from the starting point to the end point. In this case, for angles of rotation of the arcs of circles that are greater than π, there are features of determining whether or not the point of intersection of the circle and the normal (point C in Fig. 1) falls inside the arc. For example, in the general case, it cannot be assumed that, if chord AC < AD (Fig. 1), then point C will fall inside the arc AD (Fig. 2).

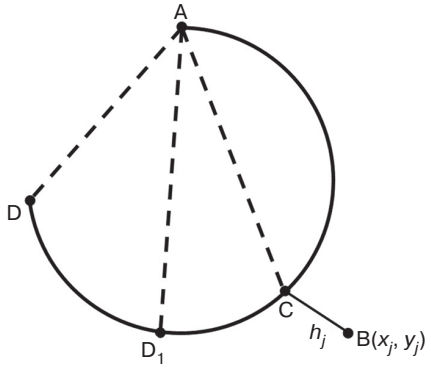

Fig. 2. Determination of whether or not the normal and the arc intersect

In the example in Fig. 2, AC > AD, but AC < AD_1, and the rule of determining the position of a point on an arc by comparing the lengths of the chords does not work.

The position of the point C with respect to the chord AD (Fig. 1) is determined by the sign of *d*.

$$d = (y_C - y_A)(x_D - x_A) - (x_C - x_A)(y_D - y_A).$$

If $d > 0$, then the point C is to the left of the direction of AD, while if $d < 0$, it is to the right.

This can be easily verified by passing to a coordinate system centered at the point A and directing the *OX* axis along AD. Hence, we obtain the rule: if $Rd > 0$, then the point C is outside the arc AD; otherwise, it is inside.

Another feature of the problem being solved is that the admissible domain is unlimited due to the one-sidedness of because inequalities (2) and (3). However, this circumstance is not significant in this case because the search for a step in the direction at each iteration can be limited to a maximum increase of 1 m in the radii and in the lengths of line segments and circular arcs.

A more important complicating feature is the already noted absence of an analytical expression for the objective function in terms of the main variables. On the other hand, a significant simplification consists in an extremely simple form of constraints on the main variables, owing to which the advisability of changing the set of active constraints on the main variables is easy to check by considering the sign of the corresponding components of the gradient at each iteration when searching for the direction of descent.

3. CALCULATION OF THE DERIVATIVES OF THE OBJECTIVE FUNCTION WITH RESPECT TO THE MAIN VARIABLES

Here, we are talking about an attempt to analytically calculate the gradient of the objective function with respect to the main variables without having its analytical expression in terms of these variables by recalculating the derivatives. It turned out that, in the context of our problem, such a recalculation is quite possible.

Let us assume that the set of the main variables forms vector **x**. Then the formula for recalculating the derivatives has the form

$$\frac{\partial F}{\partial x_i} = \sum_1^n \frac{\partial F}{\partial h_j} \cdot \frac{\partial h_j}{\partial x_i}, \; i = \overline{1,n}, \qquad (10)$$

where $\frac{\partial F}{\partial h_j} = 2h_j$ on the strength of expression (1).

This comes down to a calculation of the derivatives of the displacements with respect to the normals in terms of the main variables. Let us show how this can be done in our case. To do this while omitting the subscripts and keeping the notation *R* for the radius of an arbitrary circular arc, we denote the length of an arbitrary line segment by *l* and the length of the circular arc by *L*.

Let us start with the length of the line segment and assign it increment δl without changing all other lengths and radii. Obviously, the desired $\frac{\partial h_j}{\partial l} = 0$ for all normals that are closer to the start of the spline than the end of the line segment being varied.

If we find the displacement δh_j of the point of intersection of the *j*th normal with the spline along this normal caused by change δl at unchanged values of all the other variables, then, by passing to the limit in $\delta hj/\delta l$ as $\delta l \to 0$, we obtain the desired derivative without having an analytical expression for the function $h_j(l)$.

An increase in the length of the line segment by δl at unchanged values of all the other variables results in a shift of the entire remaining part of the spline in the direction of this straight line by δl. This is the simplest variation of the spline. If the point of intersection of the spline and the *j*th normal lies on the straight line (Fig. 3), then

$$\frac{\partial h_j}{\partial l} = \frac{\sin(\alpha - \beta)}{\sin(\gamma_j - \beta)}, \qquad (11)$$

where β is the angle of this line (spline element AB in Fig. 3) with the *OX* axis; α is the angle of the line being varied with the *OX* axis (establishing the direction of the displacement); γ_j is the angle of the normal (C_0C_1 in Fig. 3) with the *OX* axis.

In Fig. 3, point C is the initial position of the point of intersection of the normal and the spline, which corresponds to the intermediate variable h_j. When shifting in the direction determined by the angle α by δl, AB transforms into A_1B_1, point C transforms into C_2, while C_1 becomes the point of intersection of the normal with the spline. The displacement h_j gets an increment of $h_j = CC_1$.

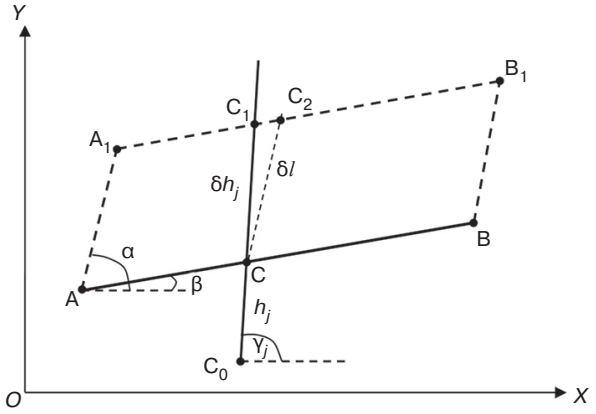

Fig. 3. Calculation of partial derivatives with changing length of the line segment

When applied to the triangle C_1CC_2, formula (11) follows from the law of sines. In this formula, $\sin(\gamma_j - \beta) \neq 0$ because the normal to the initial route at the point C and the normal to the spline, i.e., to line AB are close to each other; i.e., $\gamma_j - \beta \approx \pi/2$. At $\alpha = \beta$, the direction of the displacement coincides with the direction of the straight line; therefore, $\delta h_j = 0$. Formula (11) is also valid at $\alpha > \pi$.

If the point of intersection of the spline and the normal lies not in a straight line, but a circular arc, then β is the angle between the *OX* axis and the tangent to the circular arc at the point of its intersection with the normal, and formula (11) remains unchanged.

Let us consider the effect of the increment of the length of the circular arc on δL at unchanged values of all the other variables. In this case, the spline element from the beginning to *L* inclusive is not changed; in the remaining part, there is a shift by δL in the direction making angle α with the *OX* axis and a rotation by angle δα of the entire next section of the route plan. Here, α is the angle of the tangent at the end of the circular arc with the *OX* axis, while δα is its increment when the arc length is changed by δL. The center of rotation is at the end of the arc being varied.

The effect of the shift is taken into account by formula (11).

Let us now consider the rotation of the element of the route plan by the angle δα under the action of the elongation δL. Since $\delta\alpha = \delta L/R$, it is sufficient to calculate $\dfrac{\partial h_j}{\partial \alpha}$.

To calculate $\dfrac{\partial h_j}{\partial \alpha}$, it is necessary to calculate the radius of rotation *S* from the coordinates of the end of the arc (center of rotation: point A in Fig. 4) and the point of intersection of the spline element (the line segment or the tangent to the circle) with the normal (point C in Fig. 4).

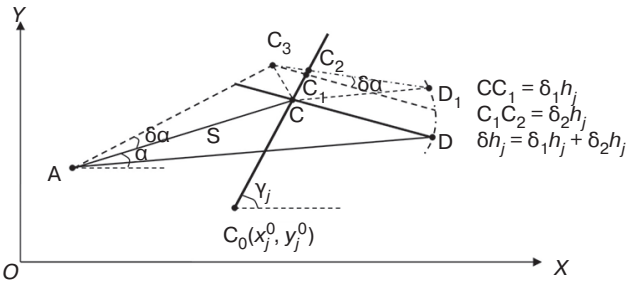

Fig. 4. Calculation of partial derivatives by rotation

The position of the straight line CD after rotation can be obtained in two ways. First, by rotating AC by the angle δα, point C_3 is found. Then, by rotating AD by the angle δα, point D_1 is determined. The intersection of the line C_3D_1 with the normal gives the sought-for $\delta h_j = CC_2$. However, CC_2 cannot be analytically expressed in terms of the known angles and coordinates.

One can make a parallel shift of CD in the direction of CC_3 (to obtain point C_1 at the intersection with the normal) followed by a rotation with the center at C_3 by the angle δα. Thereby, the same points D_1 and then C_2 are obtained.

Let us represent increment δh_j as the sum $\delta h_j = \delta_1 h_j + \delta_2 h_j$. The increment $\delta_1 h_j = CC_1$ arises by rotating point C about point A (which transforms to point C_3) followed by a parallel displacement in the direction of CC_3. Since we want to calculate partial derivatives, the lengths of the chord, arc and tangent are of the same order at small angles of rotation; therefore, we take $CC_3 = S\delta\alpha$. Here, CC_3, CC_1, and C_1C_3 are of the order δα, while the increment $\delta_2 h_j = C_1C_2$, which is caused by the rotation about the point C_3 at a first-order radius by the angle δα, has a higher order of smallness than $\delta_1 h_j$. Therefore, the point C_2 is not needed at all.

The increment $\delta_1 h_j$ is calculated from expression (11) as above by taking into account the shift by $S\delta\alpha$ in the direction along the normal to AC, which makes the angle $\alpha + \pi/2$ with the *OX* axis.

According to formula (11),

$$\delta_1 h_j = \frac{S\delta\alpha \sin(\pi/2 + \alpha - \beta)}{\sin(\gamma_j - \beta)},$$

where β is the angle of the displaced straight line CD with the *OX* axis. Hence, $\frac{\delta_1 h_j}{\delta\alpha} = \frac{S\cos(\alpha - \beta)}{\sin(\gamma_j - \beta)}$.

Let x_A and y_A be the coordinates of the center of rotation, while x_C and y_C are the points of intersection with the normal. Then the derivative can be expressed as

$$\frac{\delta_1 h_j}{\delta\alpha} = \frac{(x_C - x_A)\cos\beta + (y_C - y_A)\sin\beta}{\sin(\gamma_j - \beta)}. \quad (12)$$

Here, as above, β is the angle between the *OX* axis and the tangent to the spline at the point of intersection with the *j*th normal, and γ_j is the angle of this normal with the *OX* axis.

The expression $S\cos(\alpha - \beta)$ is replaced by $(x_C - x_A)\cos\beta + (y_C - y_A)\sin\beta$.

Taking into account tangential shift (11) and rotation (12), which reduces to a shift by $S\delta\alpha$ for the sought-for derivative of the displacement h_j along the length of the circular arc L_i^{curve} (here, $\delta\alpha = \delta L_i^{\text{curve}} / R_i$, where R_i is the radius of the circular curve being varied), we obtain the formula

$$\frac{\partial h_j}{\partial L_i^{\text{curve}}} = \frac{\sin(\alpha - \beta) + \frac{(x_C - x_A)\cos\beta + (y_C - y_A)\sin\beta}{R_i}}{\sin(\gamma_j - \beta)}. \quad (13)$$

Formulas (11)–(13) can also be applied if the normal intersects the circular arc rather than the straight line. In this case, β is the angle between the *OX* axis and the tangent to the circle at the point of intersection.

Let us turn to the calculation of the partial derivatives of the intermediate variables with respect to the radii.

In Fig. 5, AC is the initial position of the arc, while AC_1 is the position of this arc at a changed value of the radius and constant values of the starting point $A(x_A, y_A)$, the angle α of the tangent with the *OX* axis, and the length *L* of the entire arc AC. Instead of point B in the normal, we obtain point B_1. Although the displacement along the normal is $\partial h_j = BB_1$, the new position of the point B is not B_1 (Figs. 3 and 5) because the point B leaves the normal.

Knowing the coordinates of the points $A(x_A, y_A)$ and $B(x_B, y_B)$, the angles of the tangent at these points with the *OX* axis (α and β, respectively), the length *L* of the arc AB, and the angle γ of the normal with the *OX* axis, one can calculate the derivative of the displacement along the normal $\partial h/\partial R$ (the subscripts of the normal and the curve are omitted because the point B is an arbitrary point of an arbitrary arc):

$$x_B - x_A = R(\sin\beta - \sin\alpha).$$

Here, x_A is constant, and x_B and β depend on *R*. Hence, it follows that

$$\frac{\partial x_B}{\partial R} = \sin\beta - \sin\alpha + R\cos\beta \cdot \frac{\partial\beta}{\partial R}.$$

Hereinafter, $\beta - \alpha = \frac{L}{R}$. *L* and α are fixed, while $\frac{\partial\beta}{\partial R} = -\frac{L}{R^2}$.

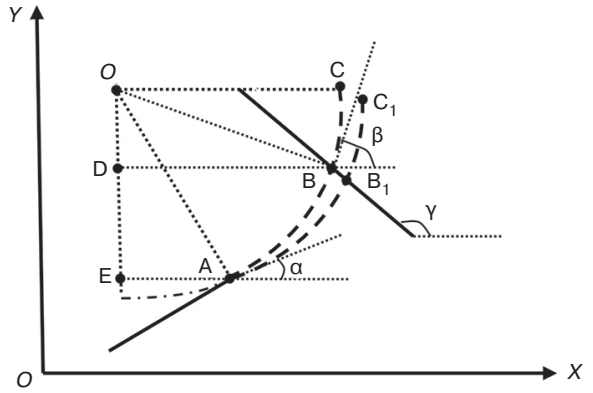

Fig. 5. Calculation of the derivatives of the displacements inside the arc with changing radius

Finally, we obtain

$$\frac{\partial x_B}{\partial R} = \sin\beta - \sin\alpha - (\beta - \alpha)\cos\beta. \quad (14)$$

Similarly, we obtain

$$\frac{\partial y_B}{\partial R} = \cos\alpha - \cos\beta - (\beta - \alpha)\sin\beta. \quad (15)$$

The increment δ*R* of the radius gives the linear part of the increment of the coordinates of the point B:

$$\delta x_B = \frac{\partial x_B}{\partial R}\delta R \text{ and } \delta y_B = \frac{\partial y_B}{\partial R}\delta R. \quad (16)$$

The shift δh_j along the normal is obtained as a result of the shift of the tangent at point B along the *OX* axis by δx_B and then along the *OY* axis by δy_B. In the former case, the direction of the shift in formula (11) is given by the angle α = 0, and in the latter case, α = π/2.

As a result, for points inside the curve, the linear part of the normal displacement is represented by the formula

$$\delta h_j = \frac{\delta y_B \cos\beta - \delta x_B \sin\beta}{\sin(\gamma - \beta)}.$$

Using expressions (14)–(16), the derivative is expressed as

$$\frac{\delta h_j}{\delta R} = \frac{(\cos\alpha - \cos\beta - (\beta - \alpha)\sin\beta)\cos\beta}{\sin(\gamma - \beta)} - \\ - \frac{(\sin\beta - \sin\alpha - (\beta - \alpha)\cos\beta)\sin\beta}{\sin(\gamma - \beta)} = \\ = \frac{\cos(\beta - \alpha) - 1}{\sin(\gamma - \beta)}. \quad (17)$$

Application of formulas (14) and (15) to the end point of the curve C gives the linear part of the increments of its coordinates:

$$\delta x_C = \left[\sin\beta - \sin\alpha - \frac{L}{R}\cos\beta\right]\delta R,$$
$$\delta y_C = \left[\cos\alpha - \cos\beta - \frac{L}{R}\sin\beta\right]\delta R.$$

Here and henceforth, α and β are the angles between the *OX* axis and the tangents at the initial and final points of the arc, respectively.

All the subsequent points of the spline are allocated the same increments (shift in the same direction). Therefore, for the linear part of the displacement along the normal for the point of intersection with the spline, we obtain

$$\delta_1 h_j = \frac{\delta y_C \cos\beta_t - \delta x_C \sin\beta_t}{\sin(\gamma_j - \beta_t)}\delta R. \quad (18)$$

Here and henceforth, β_t is the angle of the straight line or the tangent to the circle with the *OX* axis at the point of intersection of the *j*th normal, and γ_j is the angle of the normal with the *OX* axis.

Formula (18) after simplifications takes the form

$$\delta_1 h_j = \\ = \frac{\cos(\beta_t - \alpha) - \cos(\beta_t - \beta) + (\beta - \alpha)\sin(\beta_t - \beta)}{\sin(\gamma_j - \beta_t)}\delta R. \quad (19)$$

This is only a consequence of the shift with changing radius. It is also necessary to take into account the rotation of the tangent at the end of the arc (point C in Fig. 5) with changing radius; as a result, the rotation of all the next points of the spline centered at the end of the arc (point C in Fig. 5) by the angle $\delta\varphi = -\frac{L}{R^2}\delta R$, where *L* is the length of the arc AC.

As the rotation is taken into account with changing length of the curve (12), so the linear part $\delta_2 h_j$ of the displacement is taken into account by the shift by $S\delta\varphi$ along the normal to the straight line (or the tangent to the circle) at the point D of intersection with the normal. Here, *S* = CD is the radius of rotation.

According to (12), $\frac{\delta_2 h_j}{\delta\varphi} = \frac{S\cos(\varphi - \beta_t)}{\sin(\gamma_j - \beta_t)}$. Here, φ, β_t, and γ_j are the angles between the *OX* axis and the straight line CD, the straight line (or the tangent), and the intersected normal, respectively.

As a result, we obtain

$$\frac{\delta_2 h_j}{\delta\varphi} = \frac{(x_D - x_C)\cos\beta_t + (y_D - y_C)\sin\beta_t}{\sin(\gamma_j - \beta_t)}.$$

and, then,

$$\delta_2 h_j = -\frac{(x_D - x_C)\cos\beta_t + (y_D - y_C)\sin\beta_t}{\sin(\gamma_j - \beta_t)R^2} L\delta R. \quad (20)$$

Using expressions (19) and (20), the derivative of the displacement along the normal with respect to the radius is written as

$$\frac{\partial h_j}{\partial R} = \\ = \frac{\cos(\beta_t - \alpha) - \cos(\beta_t - \beta) + (\beta - \alpha)\sin(\beta_t - \beta)}{\sin(\gamma_j - \beta_t)} - \\ - \frac{(x_D - x_C)\cos\beta_t + (y_D - y_C)\sin\beta_t}{\sin(\gamma_j - \beta_t)R^2} L. \quad (21)$$

Formulas (13), (17), and (21) can be applied to any normal as well as to all spline elements preceding it. This means that it is possible to calculate the gradient of objective function (1) without having its analytical expression in terms of the main variables.

4. CONSTRUCTION AND USE OF THE MODIFIED LAGRANGE FUNCTION

Let us consider the problem of imposing constraints (7) and (8) on the intermediate variables.

Even though we do not have the expressions for the are nonlinear constraints (7) and (8) on the main variables, the penalty method can be used if the derivatives of the intermediate variables can be calculated with respect to the main variables [19, 24]. In this case, one can add a term, called a penalty

function, to the objective function, with this penalty function determining a penalty for violation of constraints. In other words, instead of the original objective function, a modified Lagrange function is constructed, which takes into account both equality and nonequality constraints.

There are several versions of this method, which differ in the form of the penalty function and in methods for changing its parameters [18, 19, 23, 24].

When solving practical problems, good results were obtained using Powell's method, in which the minimum of the original objective function $F(\mathbf{x})$ under constraints $c_j(\mathbf{x}) \le 0,\ j = \overline{1,m}$, is searched for using the function

$$\Phi(\mathbf{x},\boldsymbol{\sigma},\boldsymbol{\theta}) = F(\mathbf{x}) + 1/2\sum_1^m \sigma_j (c_j(\mathbf{x}) + \theta_j)_+^2.$$

Here, $\mathbf{x}$ is the vector of unknowns, and $F(\mathbf{x})$ is the original objective function. In the context of the present problem, the components of the vector $\mathbf{x}$ are the lengths of the spline elements, while the radii (main variables) and $c_j(\mathbf{x})$ are implicit functions of $\mathbf{x}$; however, since their dependence on the intermediate variables $\mathbf{h}$ is explicit ((7) and (8)), we can calculate their partial derivatives. Constraints (7) and (8) can always be represented as one-sided inequalities.

The setting by the user vectors σ and θ, which have m components each (according to the number of constraints), represent a set of parameters of the penalty function, with two parameters per constraint. The plus sign means that the sum includes only the terms for which $c_j(\mathbf{x}) + \theta_j > 0$. Here, $\theta_j > 0$ is the overconstraint in the jth constraint; i.e., a penalty is imposed on not only an actual violation at $c_j(\mathbf{x}) > 0$, but also at $c_j(\mathbf{x}) > -\theta_j$.

If there are equalities in the system of constraints, then the terms corresponding to them are always present in the sum. If $\theta_j = 0$ and $\sigma_j = k_n$ (k_n are set by the user), $j = \overline{1,m}$, the penalty function is simpler; however, its second derivatives with respect to x_i are discontinuous at the boundary of the admissible domain. These discontinuities increase with greater k_n, which have to be increased in each new iterative minimization cycle in order to reduce the residuals of the constraints. It is a different matter when σ_j are constant and only θ_j are varied. In this case, the surfaces of the discontinuities of the second derivatives are far from the minimum points determined when solving problems in each optimization cycle [19].

The initial values of the parameters $\theta_j > 0$ and $\sigma_j > 0$ should be selected based on the meaning and importance of the corresponding constraints and the residuals of the constraints at the initial approximation point. In our case, the solution was started with $\theta_j = 0.1$ and $\sigma_j = 1$ for all j.

Then, after solving the minimum problem $\Phi(\mathbf{x}, \boldsymbol{\sigma}, \boldsymbol{\theta})$ under simple constraints (2), (3), (5), and (6) on the main variables, constraints (7) and (8) were checked. If there were violations, then the parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\theta}$ were changed according to the following rule: if there was an overconstraint in the jth constraint, i.e., if $c_j(\mathbf{x}) > -\theta_j$, then the new value $\theta^1_j = 0$; otherwise, $\theta^1_j = \theta_j + c_j(\mathbf{x})$. Such a substitution was carried out in all the constraints. To recalculate σ_j, another rule was applied: if, as a result of solving the problem, the residual of the jth constraint decreased rapidly, then σ_j was not changed; however, if the residual decreased slowly, then σ_j was increased. The following constants were used: if the residual decreased by a factor of less than 4, then the corresponding σ_j was multiplied by 10, and θ_j was divided by 2.

After the parameters were recalculated, the process was repeated; i.e., the next outer iteration was done. The calculation was terminated in the following cases:

1. A solution with acceptable residuals was obtained. In this case, one could make one more outer iteration for control to make sure that the solution remains virtually unchanged.
2. After a specified limit of outer iterations was exhausted, no solution was obtained. At the same time, there was every reason to doubt the compatibility of the system of constraints—and, consequently, the existence of a solution to the original problem. Such situations arose when specifying fixed points through which it was impossible to pass at the given minimum length values of elements and radii.

5. MAIN PROVISIONS OF THE METHOD FOR SOLVING THE PROBLEM

The initial approximation of the sought-for spline, which was obtained using dynamic programming, is used to calculate the parameters of the spline optimization problem. To do this, the following steps are performed:

1. Outer normals to the original polyline are constructed successively at the given points, and their angles with the OX axis are memorized.
2. The points of intersection of the normals with the spline are determined and memorized. At large angles of rotation, one normal can intersect the spline at two points. In this case, the point closest to the given point is selected.
3. The values of all intermediate variables are calculated.
4. At each point of intersection of the normals with the spline element (straight line or circle), the angle with the OX axis of the straight line or tangent to the circle is calculated and memorized.

The results of the calculations are sufficient to calculate the gradient of the modified Lagrange function.

Due to the simple form of the constraints on the main variables, various gradient methods can be used, including the simple coordinate-wise descent

method [18, 19]. For example, when using the gradient projection method, after setting the gradient components corresponding to the variables that take limit values (the so-called active set) to zero, the standard algorithm [19] is applied. It has been experimentally established that this method does not guarantee obtaining exact solutions if penalty functions are used.

The application of optimally efficient second-order methods [19] requires the inversion of the matrix of second derivatives (Hessian matrix), which, in our case, cannot be calculated. Therefore, we used the variable metric method, namely, the so-called Davidon–Fletcher–Powell (DFP) optimization. In this method, during the course of the descent, increasingly accurate approximations of the matrices $\mathbf{H}_i$ to the inverse Hessian matrix $\mathbf{G}^{-1}$ are carried out using the gradients of the objective function at already passed points of iteration [23].

Thus [23, 24], if $\mathbf{x}_i$ be the iteration point, $\mathbf{g}_i$ be the gradient, $\mathbf{p}_i$ be the direction of descent at the ith iteration, $\mathbf{z}_i = \mathbf{x}_{i+1} - \mathbf{x}_i$, and $\mathbf{y}_i = \mathbf{g}_{i+1} - \mathbf{g}_i$, then $\mathbf{H}_0 = \mathbf{E}$ and $\mathbf{p}_i = -\mathbf{H}_i\mathbf{g}_i$. Under no constraints, we have

$$\mathbf{H}_{i+1} = \mathbf{H}_i + \mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} / (\mathbf{z}_i, \mathbf{y}_i) - \mathbf{H}_i\mathbf{y}_i\mathbf{y}_i^{\mathrm{T}}\mathbf{H}_i / (\mathbf{H}_i\mathbf{y}_i, \mathbf{y}_i^{\mathrm{T}}). \quad (22)$$

This formula is applicable to unconstrained problems. However, in our case, constraints (2), (3), (5), and (6) on the main variables remain. If the initial approximation contains the limiting lengths or radii, then the identity matrix $\mathbf{E}$ should not be used to begin with. Instead, a projection matrix should be used, which in our cause is simply constructed as follows: in $\mathbf{E}$, 1 is replaced by 0 in the rows the numbers of which coincide with the numbers of variables that have taken limit values (active set).

When changing the set of active constraints, the matrix $\mathbf{H}_i$ should be modified [24] before calculating the direction of descent. This happens both when a constraint is included in the active set and when a constraint is excluded from the active set. The simple form of the constraints allowed the corresponding formulas [24] to be significantly simplified using the noted simple method of constructing a projection matrix.

Since DFP optimization works well (in the sense of approaching the inverse Hessian matrix) for points close to the extremum [19, 23, 24], a combination of methods was used: the simple gradient projection method (ensures a descent into the "ravine") with subsequent DFP optimization.

6. MAIN RESULTS AND OBJECTIVES OF FURTHER RESEARCH

The main result of this work is a solution to the problem of optimizing a sequence of points in a plane by a spline that is not a single-valued function. This solution is obtained not by heuristic techniques, but by mathematically correct methods (dynamic and nonlinear programming). However, due to the variety of nonlinear programming methods used, we cannot claim that the method used in the calculations is the most efficient. Of particular interest is the use of ravine algorithms [23]. It was stated [23] that the ravine conjugate gradient method, in which $\mathbf{p}_0 = -\mathbf{g}_0$, $\mathbf{p}_{i+1} = -\mathbf{g}_{i+1} + b_i\mathbf{p}_i$, and $b_i = (\mathbf{g}_{i+1} - \mathbf{g}_i, \mathbf{g}_i)/|\mathbf{p}_i, \mathbf{g}_i|$, has advantages for inaccurate one-dimensional minimization and for ravine bends.

In the context of our problem, instead of the gradient, its projection should be used; here, when the active set is changed, one should start with updating, i.e., from the step along the antigradient projection.

There are more complex algorithms than DFP, e.g., the Broyden–Fletcher–Goldfarb–Shanno method [23]. In this method, the term $\mathbf{v}_i\mathbf{v}_i^{\mathrm{T}}$ is added to formula (22) for calculating the matrix $\mathbf{H}$, where the vector $\mathbf{v}_i = (\mathbf{y}_i, \mathbf{H}_i)^{1/2}[\mathbf{z}_i/(\mathbf{z}_i, \mathbf{y}_i) - \mathbf{H}_i\mathbf{y}_i/(\mathbf{y}_i, \mathbf{H}_i\mathbf{y}_i)]$.

The efficiency of using more complex methods can only be determined experimentally.

It should be noted that gradient methods give a local minimum of the objective function. Therefore, to obtain an initial approximation in our problem, it is especially important to use dynamic programming (possibly with repetition of calculations with decreasing search increments), since dynamic programming gives a global minimum, excluding the discreteness effect.

To combat local minima, descents from different points were used. Following completion of the optimization process, the obtained solution is checked. The process starts anew with the obtained solution used as the initial approximation. All the coefficients of the modified Lagrange function take initial values. At the first iterations, the sum of their squares takes a smaller value due to the violation of constraints on the intermediate variables. However, the coefficients of the modified Lagrange function then change, resulting in the disappearance of the constraint violations. As a result, in experimental calculations, virtually the same solution was obtained. While this is not a guarantee of reaching the global minimum, it offers a real opportunity to recede from a local minimum point. To select the most efficient method of nonlinear programming for solving the problem under consideration, additional experimental studies are needed.

The main objective of further research is to generalize the results obtained for a spline with circles to the more complex problem of approximation by a spline with clothoids. First of all, it is necessary to obtain formulas for calculating derivatives in the absence of an analytical expression of the objective function in terms of the parameters of the clothoid in addition to the formulas for lines and arcs of circles that were presented in this article.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Karpov D.A., Struchenkov V.I. Two-stage spline-approximation in linear structure routing. *Russ. Technol. J.* 2021;9(5):45−56 (in Russ.). https://doi.org/10.32362/2500-316X-2021-9-5-45-56
2. Karpov D.A., Struchenkov V.I. Spline approximation of multivalued function in leaner structures routing. *Russ. Technol. J*. 2022;10(4):65–74 (in Russ.). https://doi.org/10.32362/2500-316X-2022-10-4-65-74
3. Li W., Pu H., Schonfeld P., et al. A method for automatically recreating the horizontal alignment geometry of existing railways. *Comput.-Aided Civ. Inf*. 2019;34(1):71–94. https://doi.org/10.1111/mice.12392
4. Jha M.K., McCall C., Schonfeld P. Using GIS, genetic algorithms, and visualization in highway development. *Comput.-Aided Civ. Inf*. 2001;16(6):399–414. https://doi.org/10.1111/0885-9507.00242
5. Jha M.K., Schonfeld P. A highway alignment optimization model using geographic information systems. *Transportation Research Part A: Policy and Practice*. 2004;38(6):455–481. https://doi.org/10.1016/j.tra.2004.04.001
6. Jong J.C., Jha M.K., Schonfeld P. Preliminary highway design with genetic algorithms and geographic information systems. *Comput.-Aided Civ. Inf*. 2000;15(4):261–271. https://doi.org/10.1111/0885-9507.00190
7. Kang M.W., Schonfeld P., Yang N. Prescreening and repairing in a genetic algorithm for highway alignment optimization. *Comput.-Aided Civ. Inf*. 2009;24(2):109–119. https://doi.org/10.1111/j.1467-8667.2008.00574.x
8. Pushak Y., Hare W., Lucet Y. Multiple-path selection for new highway alignments using discrete algorithms. *Eur. J. Oper. Res.* 2016;248(2):415–27. https://doi.org/10.1016/j.ejor.2015.07.039
9. Sarma K.C., Adeli H. Bilevel parallel genetic algorithms for optimization of large steel structures. *Comput.-Aided Civ. Inf*. 2001;16(5):295–304. https://doi.org/10.1111/0885-9507.00234
10. Shafahi Y., Bagherian M. A customized particle swarm method to solve highway alignment optimization problem. *Comput.-Aided Civ. Inf*. 2013;28(1):52–67. https://doi.org/10.1111/j.1467-8667.2012.00769.x
11. Bosurgi G., D'Andrea A. A polynomial parametric curve (PPC-curve) for the design of horizontal geometry of highways. *Comput.-Aided Civ. Inf*. 2012;27(4):303–312. https://doi.org/10.1111/j.1467-8667.2011.00750.x
12. Cerf R. The quasispecies regime for the simple genetic algorithm with roulette wheel selection. *Adv. Appl. Probability*. 2017;49(3):903–926. https://doi.org/10.1017/apr.2017.26
13. Borodakii Yu.V., Zagrebaev A.M., Kritsyna N.A., Kulyabichev Yu.P., Shumilov Yu.Yu. *Nelineinoe programmirovanie v sovremennykh zadachakh optimizatsii* (*Nonlinear Programming in Modern Optimization Problem*). Moscow: NIYAU MEPhI; 2008. 244 p. (in Russ.).
14. Bazaraa M., Sherali Y., Shetty C. *Nonlinear programming. Theory and algorithms*. 3rd ed. Hoboken, NJ: Wiley; 2006. 872 p. ISBN 978-0-471-48600-8

СПИСОК ЛИТЕРАТУРЫ

1. Карпов Д.А., Струченков В.И. Двухэтапная сплайн-аппроксимация в компьютерном проектировании трасс линейных сооружений. *Russ. Technol. J.* 2021;9(5):45–56. https://doi.org/10.32362/2500-316X-2021-9-5-45-56
2. Карпов Д.А., Струченков В.И. Сплайн-аппроксимация многозначных функций в проектировании трасс линейных сооружений. *Russ. Technol. J.* 2022;10(4):65–74. https://doi.org/10.32362/2500-316X-2022-10-4-65-74
3. Li W., Pu H., Schonfeld P., et al. A method for automatically recreating the horizontal alignment geometry of existing railways. *Comput.-Aided Civ. Inf*. 2019;34(1):71–94. https://doi.org/10.1111/mice.12392
4. Jha M.K., McCall C., Schonfeld P. Using GIS, genetic algorithms, and visualization in highway development. *Comput.-Aided Civ. Inf*. 2001;16(6):399–414. https://doi.org/10.1111/0885-9507.00242
5. Jha M.K., Schonfeld P. A highway alignment optimization model using geographic information systems. *Transportation Research Part A: Policy and Practice*. 2004;38(6):455–481. https://doi.org/10.1016/j.tra.2004.04.001
6. Jong J.C., Jha M.K., Schonfeld P. Preliminary highway design with genetic algorithms and geographic information systems. *Comput.-Aided Civ. Inf*. 2000;15(4):261–271. https://doi.org/10.1111/0885-9507.00190
7. Kang M.W., Schonfeld P., Yang N. Prescreening and repairing in a genetic algorithm for highway alignment optimization. *Comput.-Aided Civ. Inf*. 2009;24(2):109–119. https://doi.org/10.1111/j.1467-8667.2008.00574.x
8. Pushak Y., Hare W., Lucet Y. Multiple-path selection for new highway alignments using discrete algorithms. *Eur. J. Oper. Res.* 2016;248(2):415–27. https://doi.org/10.1016/j.ejor.2015.07.039
9. Sarma K.C., Adeli H. Bilevel parallel genetic algorithms for optimization of large steel structures. *Comput.-Aided Civ. Inf*. 2001;16(5):295–304. https://doi.org/10.1111/0885-9507.00234
10. Shafahi Y., Bagherian M. A customized particle swarm method to solve highway alignment optimization problem. *Comput.-Aided Civ. Inf*. 2013;28(1):52–67. https://doi.org/10.1111/j.1467-8667.2012.00769.x
11. Bosurgi G., D'Andrea A. A polynomial parametric curve (PPC-curve) for the design of horizontal geometry of highways. *Comput.-Aided Civ. Inf*. 2012;27(4):303–312. https://doi.org/10.1111/j.1467-8667.2011.00750.x
12. Cerf R. The quasispecies regime for the simple genetic algorithm with roulette wheel selection. *Adv. Appl. Probability*. 2017;49(3):903–926. https://doi.org/10.1017/apr.2017.26
13. Бородакий Ю.В., Загребаев А.М., Крицына Н.А., Кулябичев Ю.П., Шумилов Ю.Ю. *Нелинейное программирование в современных задачах оптимизации*. М.: НИЯУ МИФИ; 2011. 244 с. ISBN 987-5-7262-1451-1
14. Bazaraa M., Sherali Y., Shetty C. *Nonlinear programming. Theory and algorithms*. 3rd ed. Hoboken, NJ: Wiley; 2006. 872 p. ISBN 978-0-471-48600-8
15. Betts J.T. *Practical methods for optimal control using nonlinear programming*. Ser. Advances in Design and Control. Philadelphia: SIAM; 2001. 190 p.

15. Betts J.T. *Practical methods for optimal control using nonlinear programming*. Ser. Advances in Design and Control. Philadelphia: SIAM; 2001. 190 p.
16. Lee J., Leyffer S. *Mixed integer nonlinear programming*. NY: Springer; 2011. 707 p. https://doi.org/10.1007/978-1-4614-1927-3
17. Sun W., Yuan Y.-X. *Optimization theory and methods. Nonlinear programming*. NY: Springer; 2006. 688 p. https://doi.org/10.1007/b106451
18. Struchenkov V.I. *Metody optimizatsii trass v SAPR lineinykh sooruzhenii* (*Methods for route optimization in CAD of linear structures*). Moscow: SOLON-Press; 2014. 271 p. (in Russ.). ISBN 978-5-91359-139-5
19. Gill F., Myurrei U., Rait M. *Prakticheskaya optimizatsiya* (*Practical Optimization*): transl. from Engl. Moscow: Mir; 1985. 509 p. (in Russ.).
[Gill Ph.E., Murray W., Wright M.H. *Practical Optimization*. London: Academic Press; 1981. 402 p.]
20. Audet C., Hare W. *Derivative-free and blackbox optimization.* Springer Series in Operations Research and Financial Engineering. Springer International Publishing; 2017. 302 p. https://doi.org/10.1007/978-3-319-68913-5
21. Kokhenderfer M.D., Uiler T.A. *Algoritmy optimizatsii* (*Algorithms for Optimization*). Moscow: Vil'yams; 2020. 528 p. (in Russ.).
[Kochenderfer M.D., Wheeler T.A. *Algorithms for Optimization*. London; The MIT Press; 2019. 520 p.]
22. Chernorutskii I.G. *Metody optimizatsii. Komp'yuternye tekhnologii* (*Methods of optimization. Computer technologies*). St. Petersburg: BHV-Petersburg; 2011. 329 p. (in Russ.).
23. Larichev O.I., Gorvits G.G. *Metody poiska lokal'nykh ekstremumov ovrazhnykh funktsii* (*Methods for Finding Local Extrema of Ravine Functions*). Moscow: Nauka; 1990. 96 p. (in Russ.).
24. Gill F., Myurrei U. *Chislennye metody uslovnoi optimizatsii* (*Numerical methods of conditional optimizatio*n): transl. from Engl. Moscow: Mir; 1977. 296 p. (in Russ.).
[Gill Ph.E., Murray W. *Numerical methods for constrained optimization*. London: Academic Press; 1974. 283 p.]

16. Lee J., Leyffer S. *Mixed integer nonlinear programming*. NY: Springer; 2011. 707 p. https://doi.org/10.1007/978-1-4614-1927-3
17. Sun W., Yuan Y.-X. *Optimization theory and methods. Nonlinear programming*. NY: Springer; 2006. 688 p. https://doi.org/10.1007/b106451
18. Струченков В.И. *Методы оптимизации трасс в САПР линейных сооружений*. М.: СОЛОН-Пресс; 2015. 271 с. ISBN 978-5-91359-139-5
19. Гилл Ф., Мюррей У., Райт М. *Практическая оптимизация*: пер. с англ. М.: Мир; 1985. 509 с.
20. Audet C., Hare W. *Derivative-free and blackbox optimization.* Springer Series in Operations Research and Financial Engineering. Springer; 2017. 302 p. https://doi.org/10.1007/978-3-319-68913-5
21. Кохендерфер М.Д., Уилер Т.А. *Алгоритмы оптимизации*. М.: Вильямс; 2020. 528 с.
22. Черноруцкий И.Г. *Методы оптимизации. Компьютерные технологии*. СПб.: БХВ-Петербург; 2011. 329 с.
23. Ларичев О.И., Горвиц Г.Г. *Методы поиска локальных экстремумов овражных функций*. М.: Наука; 1990. 96 с.
24. Гилл Ф., Мюррей У. *Численные методы условной оптимизации*: пер. с англ. М.: Мир; 1977. 296 с.

About the authors

Dmitry A. Karpov, Cand. Sci. (Eng.), Head of the General Informatics Department, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: karpov@mirea.ru. RSCI SPIN-code 2619-7100, https://orcid.org/0000-0003-3734-7182

Valery I. Struchenkov, Dr. Sci. (Eng.), Professor, General Informatics Department, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: str1942@mail.ru. RSCI SPIN-code 4581-4698, https://orcid.org/0000-0002-9801-745

Об авторах

Карпов Дмитрий Анатольевич, к.т.н., заведующий кафедрой общей информатики Института искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: karpov@mirea.ru. SPIN-код РИНЦ 2619-7100, https://orcid.org/0000-0003-3734-7182

Струченков Валерий Иванович, д.т.н., профессор, профессор кафедры общей информатики Института искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: str1942@mail.ru. SPIN-код РИНЦ 4581-4698, https://orcid.org/0000-0002-9801-7454

Translated from Russian into English by Vladislav V. Glyanchenko
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 519.224.22
https://doi.org/10.32362/2500-316X-2023-11-2-84-91

RESEARCH ARTICLE

Extremum in the problem of paired comparisons

Igor S. Pulkin [@],

Andrey V. Tatarintsev

MIREA – Russian Technological University, Moscow, 119454 Russia
[@] *Corresponding author, e-mail: pulkin@mirea.ru*

Abstract

Objectives. An analysis of the problem of evaluating alternatives based on the results of expert paired comparisons is presented. The importance and relevance of this task is due to its numerous applications in a variety of fields, whether in the technical and natural sciences or in the humanities, ranging from construction to politics. In such contexts, the problem frequently arises concerning how to calculate an objective ratings vector based on expert evaluations. In terms of a mathematical formulation, the problem of finding the vector of objective ratings can be reduced to approximating the matrices of paired comparisons by consistent matrices.

Methods. Analytical analysis and higher algebra methods are used. For some special cases, the results of numerical calculations are given.

Results. The theorem stating that there is always a unique and consistent matrix that optimally approximates a given inversely symmetric matrix in a log-Euclidean metric is proven. In addition, derived formulas for calculating such a consistent matrix are presented. For small dimensions, examples are considered that allow the results obtained according to the derived formula to be compared with results for other known methods of finding a consistent matrix, i.e., for calculating the eigenvector and minimizing the discrepancy in the log-Chebyshev metric. It is proven that all these methods lead to the same result in dimension 3, while in dimension 4 all results are already different.

Conclusions. The results obtained in the paper allow us to calculate the vector of objective ratings based on expert evaluation data. This method can be used in strategic planning in cases where conclusions and recommendations are possible only on the basis of expert evaluations.

Keywords: expert estimates, paired comparisons, inversely symmetric matrix, consistent matrix, metric, discrepancy minimization

• **Submitted:** 08.11.2021 • **Revised:** 29.11.2022 • **Accepted:** 22.01.2023

For citation: Pulkin I.S., Tatarintsev A.V. Extremum in the problem of paired comparisons. *Russ. Technol. J.* 2023;11(2):84–91. https://doi.org/10.32362/2500-316X-2023-11-2-84-91

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Экстремум в задаче о парных сравнениях

И.С. Пулькин [@],
А.В. Татаринцев

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] *Автор для переписки, e-mail: pulkin@mirea.ru*

Резюме

Цели. Рассмотрена задача оценки альтернатив на основе результатов экспертных парных сравнений. Важность и актуальность этой задачи обусловлены ее многочисленными применениями в самых разных областях – как в технических и естественных, так и в гуманитарных, от строительства до политики. Ставится задача вычисления вектора объективных рейтингов на основе экспертных оценок. В математической формулировке задача нахождения вектора объективных рейтингов сводится к аппроксимации матриц парных сравнений согласованными матрицами.

Методы. Используются аналитические методы анализа и высшей алгебры. Для некоторых частных случаев приведены результаты численных расчетов.

Результаты. В работе доказана теорема, утверждающая, что согласованная матрица, наилучшим образом аппроксимирующая заданную обратно-симметрическую матрицу в лог-евклидовой метрике, всегда существует и единственна. Кроме того, выведены формулы для вычисления такой согласованной матрицы. Для малых размерностей рассматриваются примеры, позволяющие сравнить результаты, полученные по выведенной формуле, с результатами для других известных способов нахождения согласованной матрицы – для вычисления собственного вектора и для минимизации невязки в лог-чебышевской метрике. Доказано, что в размерности 3 все эти способы приводят к одному и тому же результату, а уже в размерности 4 все результаты различны.

Выводы. Полученные в статье результаты позволяют вычислять вектор объективных рейтингов по данным экспертной оценки. Этот метод может быть использован в стратегическом планировании в тех случаях, когда выводы и рекомендации возможны только на основании экспертных суждений.

Ключевые слова: экспертные оценки, парные сравнения, обратно-симметрическая матрица, согласованная матрица, метрика, минимизация невязки

• **Поступила:** 08.11.2021 • **Доработана:** 29.11.2022 • **Принята к опубликованию:** 22.01.2023

Для цитирования: Пулькин И.С., Татаринцев А.В. Экстремум в задаче о парных сравнениях. *Russ. Technol. J.* 2023;11(2):84–91. https://doi.org/10.32362/2500-316X-2023-11-2-84-91

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Since a person responsible for making decisions does not always have complete information, decisions can be taken on the basis of criteria that are not always objective. In cases where a large number of factors must be taken into account, an error may lead to disastrous outcomes. Such situations include, for example, strategic planning issues, particularly in construction, as well as medicine, politics, economics, and many other areas of human activity.

The science according to which a strategy is selected under conditions of incomplete information, as well as providing a rationale for such choices, is commonly referred to as decision-making theory. Such studies attract close attention of experts in various fields [1, 2]. Many aspects are discussed in the books of Thomas L. Saaty [3, 4], one of the founders of this theory.

The continued growth of research in this direction can confidently predicted due to the possibility of applying decision theory methods to machine learning. In fact, the use of inconsistent and inaccurate expert

evaluations which nevertheless allow necessary information to be obtained when based on large datasets are broadly similar to situations that typically arise, for example, when training neural networks or constructing an ensemble of decision trees that comprises a random forest [5].

Numerous recent studies discuss the technique of comparing heterogeneous assets as applied to various problems from the field of information technology, in particular, to select storage formats for big data for various computing complexes, both local and distributed. For example, studies [6–11] are devoted to this topic.

A situation commonly arises when neither priori distributions nor prior statistics are available but a forecast or decision must be made on the basis of earlier forecasts and expert recommendations. Thus, the task of processing expert evaluations should be given a mathematical formulation.

Let us deal with expert evaluations. For example, an expert compares an apple (A), orange (O), and banana (B). He compares the fruits in pairs. Suppose the following opinions are expressed:

- A banana is three times better than an apple;
- An orange is five times better than an apple;
- An orange is twice as good as a banana.

Based on these evaluations, the following comparison matrix wherein the first column and the first row correspond to the apple, the second to banana, and the third to orange can be constructed:

$$\mathbf{W} = \begin{pmatrix} 1 & 3 & 5 \\ 1/3 & 1 & 2 \\ 1/5 & 1/2 & 1 \end{pmatrix}.$$

At the intersection of the *i*th column and the *j*th row, there is a number equal to the ratio of the values of the *i*th and the *j*th fruit. For such a positive-definite symmetric matrix, all of whose elements are strictly positive, the following relation is satisfied:

$$a_{ij} = a_{ji}^{-1}.$$

However, the matrix is inconsistent. If O is 5 times better than A and O is 3 times better than B, then it would be appropriate to assume that B should be 5/3 times better than A, not 2 times better.

This situation arises commonly when carrying out expert evaluations. Moreover, there are also non-transitive evaluations, e.g., when A is better than B, B is better than C, but C is better than A. This occurs, for example, in tournaments when A beats B, B beats C, and C beats A.

For making an objective evaluation, we assume that there are objective ratings w_1, w_2, and w_3, for the evaluated items, and that the expert evaluations are the same ratings distorted by random errors. The problem then arises as to how to reconstruct these ratings based on the given expert evaluations.

If ratings w_i are found, then the comparison matrix elements may be written as follows:

$$x_{ij} = \frac{w_i}{w_j}.$$

We shall denote this matrix, whose rank is 1, by $\mathbf{W}_0$. Such matrix is called a consistent matrix.

Thus, the task of processing expert evaluations is reduced to that of finding the consistent matrix $\mathbf{W}_0$ approximating the inversely symmetric matrix $\mathbf{W}$ in the best way. Here, it turns out that the answer changes significantly depending on the metric which the difference between these matrices is calculated in.

The study by N.K. Krivulin and his students [12] proposes to calculate this difference in a log-Chebyshev metric. In particular, it is noted there that this results in the problem of processing expert evaluations becoming the problem from the field of so-called idempotent or tropical mathematics [13], new direction in modern mathematics that is rapidly developing. However, it is also noted there that this metric in high dimensions results in non-uniqueness of the solution.

The studies by Saaty [3, 4] propose to consider the correspondingly normalized eigenvector of matrix A corresponding to its maximal eigenvalue as the required rank vector. The well-known Perron–Frobenius theorem [14] states that any positive matrix (consisting only of positive numbers) has a single maximal modulo eigenvalue; the multiplicity of such a strictly positive matrix is equal to 1. However, the metric according to which the obtained solution is optimal is not specified in those studies.

Thus, the present work is aimed at finding the optimal solution in the log-Euclidean metric.

DERIVING OPTIMAL EVALUATION

We shall consider the comparison matrix of arbitrary dimension ($n \times n$). The discrepancy of the original comparison matrix $\mathbf{W} = (a_{ij})$ and its matched counterpart $\mathbf{W}_0 = (x_{ij})$ in the log-Euclidean metric under consideration may be calculated in the following way:

$$\Phi = \sum_{i,j=1}^{n} \log^2 \left(\frac{x_{ij}}{a_{ij}} \right).$$

If we consider that the consistent matrix elements are expressed through the components of the matrix eigenvector in the form of $x_{ij} = w_i/w_j$, then the function Φ depends on n variables, as follows:

$$\Phi(w_1, ..., w_n) = \sum_{i,j=1}^{n} \log^2\left(\frac{w_i}{w_j a_{ij}}\right).$$

The conditions of equality to zero of the derivative of the residual function on the kth component of the eigenvector w_k can give the following system of equations:

$$n \log w_k - \sum_{\beta=1}^{n} \log w_\beta = \sum_{\beta=1}^{n} \log a_{k\beta}.$$

The solution to this system of equations is the following:

$$w_k = N \prod_{\beta=1}^{n} (a_{k\beta})^{1/n},$$

where N is the arbitrary normalization factor.

It is considered that the product of all elements of the original inversely symmetric comparison matrix is equal to one as follows:

$$\prod_{\alpha,\beta=1}^{n} a_{\alpha\beta} = 1.$$

Thus, the following statement is proved.

Theorem. Consider an inversely symmetric matrix (a_{ij}). Then the components of the consistent matrix (x_{ij}) minimizing the residual function for the log-Euclidean metric have the following form:

$$x_{ij} = \frac{w_i}{w_j} = \prod_{\beta=1}^{n} (a_{i\beta} \cdot a_{\beta j})^{1/n}.$$

In other words, the matrix element x_{ij} is equal to the product of the geometric mean of the ith row and the jth column of the original matrix.

The inversely symmetric matrices of small dimensions may be considered as an example.

Let $\mathbf{W} = (a_{ij})$ be the three-dimensional matrix of expert comparisons:

$$\mathbf{W} = \begin{pmatrix} 1 & a & b \\ 1/a & 1 & c \\ 1/b & 1/c & 1 \end{pmatrix}.$$

The matrix elements are positive $a_{ij} > 0$ and inversely symmetric $a_{ij} = a_{ji}^{-1}$. Note that a matrix is called consistent if the condition $c = b/a$ or $ac/b = 1$ is satisfied for its elements a, b, and $c > 0$. The similar parameter is called "tropical radius" and notated $R=(ac/b)^{1/3}$ in [12]; this notation is also used in the paper.

It is found in [15] that the eigenvalues of the comparison matrix for $n = 3$ are the following:

$$\lambda_1 = 1 + \left(R + \frac{1}{R}\right);$$

$$\lambda_2 = \lambda_3 = 1 - \frac{1}{2}\left(R + \frac{1}{R}\right) \pm \frac{i\sqrt{3}}{2}\left(R - \frac{1}{R}\right).$$

One of the roots of the characteristic equation is real, while the other two are complex-conjugate. The real root has the largest value in absolute value. For the consistent matrix, $R = 1$ and the eigenvalues are $\lambda_1 = 3$; $\lambda_2 = \lambda_3 = 0$. The largest nonzero eigenvalue coincides with the dimension of the comparison matrix in general.

The eigenvector of the original comparison matrix for the first eigenvalue is also easy to find. It has the following form (in normalizing $w_1 = 1$):

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 1 \\ R/a \\ 1/bR \end{pmatrix}.$$

In this case, the elements of the consistent matrix may be found as follows: $\mathbf{W}_0 = (x_{ij}) = w_i/w_j$, and hence:

$$\mathbf{W}_0 = \begin{pmatrix} 1 & a/R & bR \\ R/a & 1 & c/R \\ 1/bR & R/c & 1 \end{pmatrix}.$$

The found eigenvector of the original matrix is also the eigenvector of the consistent matrix. It corresponds to the eigenvalue of this matrix equal to the dimension $\lambda = 3$.

For the following three-dimensional comparison matrix and its corresponding consistent matrix

$$\mathbf{W} = (a_{ij}) = \begin{pmatrix} 1 & a & b \\ 1/a & 1 & c \\ 1/b & 1/c & 1 \end{pmatrix};$$

$$\mathbf{W}_0 = (x_{ij}) = \begin{pmatrix} 1 & x & y \\ 1/x & 1 & y/x \\ 1/y & x/y & 1 \end{pmatrix}$$

the problem in the log-Euclidean metric is reduced to finding the minimum of the residual function $\Phi(x, y)$ of two variables included in the following consistent matrix:

$$\Phi(x,y)=\sum_{i,j=1}^{3}(\log a_{ij}-\log x_{ij})^2,$$

which may be written in the following form with provision for the explicit form of matrices:

$$\Phi(x,y)=2\log^2\left(\frac{x}{a}\right)+2\log^2\left(\frac{y}{b}\right)+2\log^2\left(\frac{y}{cx}\right).$$

The function extremum (minimum) is reached at $x = a/R$, $y = bR$. The minimum value of the function $\min \Phi = 6\log^2 R$ depends on the inconsistency of the original matrix of expert judgments.

The results for the three-dimensional case can be also obtained in another way.

As before, the inversely symmetric matrix in the three-dimensional case is following:

$$\mathbf{W}=\begin{pmatrix}1 & a & b\\ 1/a & 1 & c\\ 1/b & 1/c & 1\end{pmatrix}.$$

When logarithmic, it becomes cosymmetric:

$$\mathbf{H}=\begin{pmatrix}0 & u & v\\ -u & 0 & w\\ -v & w & 0\end{pmatrix}.$$

Here, $u = \log a$, $v = \log b$, and $w = \log c$.

The consistent matrix has the following form:

$$\mathbf{W}_0=\begin{pmatrix}w_1/w_1 & w_2/w_1 & w_3/w_1\\ w_1/w_2 & w_2/w_2 & w_3/w_2\\ w_1/w_3 & w_2/w_3 & w_3/w_3\end{pmatrix}.$$

After logarithmizing, it has the following form:

$$\mathbf{L}=\begin{pmatrix}0 & y_1 & y_2\\ -y_1 & 0 & y_3\\ -y_2 & -y_3 & 0\end{pmatrix}.$$

Here, $y_1 = \log w_2 - \log w_1$; $y_2 = \log w_3 - \log w_1$; $y_3 = \log w_3 - \log w_2$.

In addition, the condition $y_1 - y_2 + y_3 = 0$ is satisfied. Thus, the problem is reduced to finding the point **Q** on the plane $y_1 - y_2 + y_3 = 0$ being the closest to the given point $\mathbf{P}(u, v, w) \in R^3$. The solution to this problem depends on the metric.

For the Euclidean metric, draw a line perpendicular to the plane through point **P** and find the intersection point:

$$(u + t) - (v - t) + (w + t) = 0.$$

We obtain:

$$t=-\frac{1}{3}(u-v+w).$$

Hence:

$$y_1=u+t=\frac{2}{3}u+\frac{1}{3}v-\frac{1}{3}w,\quad y_2=v-t=\frac{1}{3}u+\frac{2}{3}v+\frac{1}{3}w,$$
$$y_3=w+t=-\frac{1}{3}u+\frac{1}{3}v+\frac{2}{3}w.$$

Let us assume without loss of generality that $w_1 = 1$. Then

$$w_2=e^{y_1}=a^{2/3}b^{1/3}c^{-1/3},$$
$$w_3=e^{y_2}=a^{1/3}b^{2/3}c^{1/3}.$$

These are the elements of the first row of the matrix. The first column contains their inverse elements. The first column is equal to

$$\mathbf{V}=\begin{pmatrix}1\\ a^{-2/3}b^{-1/3}c^{1/3}\\ a^{-1/3}b^{-2/3}c^{-1/3}\end{pmatrix},$$

which coincides with the earlier obtained result

$$\mathbf{V}=\begin{pmatrix}1\\ R/a\\ 1/bR\end{pmatrix}.$$

This column (as well as the other two) is an eigenvector of the original matrix **A**. This can be easily checked by direct calculation. However, unfortunately, this calculation method is not generalized to higher dimensions.

We shall also consider other metrics. It seems most natural to consider the most common log-Manhattan and log-Chebyshev metrics.

It is easy to demonstrate that the solution is not unique in the log-Manhattan metric even in dimension 3. Indeed, in geometrical terms, the solution to the problem of the minimum distance from a point to a plane is reduced to constructing a sphere centered at this point and touching this plane. However, in Manhattan metrics, the "sphere" is an octahedron, one of whose facets lies just on the plane $y_1 - y_2 + y_3 = 0$. All points of this facet are solutions. In addition, the solution in Euclidean metric belongs to the same facet, i.e., it is one of the solutions in Manhattan metric.

In the Chebyshev metric in dimension 3, the solution is singular and coincides with the solution in the Euclidean metric. Indeed, the "ball" in this metric is actually a cube. The vector **PQ** has coordinates $(t, -t, t)$

and is half of the diagonal of the cube, so the cube also touches the plane at a single point, point **Q**.

In [12], another approach is used to minimize the discrepancy in the eigenvector in the log-Chebyshev metric in dimension 3.

Thus, all the described ways of computing the matrix consistent in dimension 3—computing the eigenvector and computing the vector by minimizing the discrepancy—lead to the same result (although this solution is not the only one in the log-Manhattan metric).

In dimension 4, these solutions are already different.

In [12], the numerical example with such a matrix is considered for dimension 4:

$$\mathbf{D} = \begin{pmatrix} 1 & 2 & 4 & 1 \\ 1/2 & 1 & 1/2 & 1/3 \\ 1/4 & 2 & 1 & 2 \\ 1 & 3 & 1/2 & 1 \end{pmatrix}.$$

The above paper proves that any vector belonging to the segment **AB** is optimal in the log-Chebyshev metric, where

$$\mathbf{A} = \begin{pmatrix} 1 \\ 1/4 \\ 1/2 \\ 1/2 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} 1 \\ 1/3 \\ 1/2 \\ 1/2 \end{pmatrix}.$$

Thus, in the log-Chebyshev metric in dimension 4, the solution is not unique in general.

The results of the method of calculating the eigenvector are as follows: the eigenvalue is $\lambda_{max} = 4.5056$, while the eigenvector in normalization when its first coordinate is equal to 1 may be written as follows:

$$\mathbf{V} = \begin{pmatrix} 1.0000 \\ 0.2837 \\ 0.5818 \\ 0.6110 \end{pmatrix}.$$

Calculating the solution giving the minimum discrepancy in the log-Euclidean metric, in accordance with the theorem proved above, results in the following:

$$\mathbf{V} = \begin{pmatrix} 1.0000 \\ 0.3195 \\ 0.5946 \\ 0.6580 \end{pmatrix}.$$

Thus, already in dimension 4, the methods of calculating the rating vector described above lead to different results.

CONCLUSIONS

The theorem proved in the paper can be used to process expert opinions by reducing them to the form of a ranking list. It is shown to give the best evaluation in the log-Euclidean metric. Examples demonstrate that this evaluation in high dimensions may not coincide with those obtained by other methods. Thus, the selection of the desired method should be related to the specifics of the problem under consideration.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Korobov V.B. *Teoriya i praktika ekspertnykh metodov* (*Theory and Practice of Expert Methods*). Moscow: INFRA-M; 2019. 279 p. (in Russ.). ISBN 978-5-16-015053-6. https://doi.org/10.12737/monography_5caee0067f1835.43206494
2. Andreichikov A.V., Andreichikova O.N. *Analiz, sintez, planirovanie reshenii v ekonomike* (*Analysis, Synthesis, Planning of Decisions in the Economy*). Moscow: Finansy i statistika; 2004. 467 p. (in Russ.). ISBN 5-279-02901-7
3. Saaty T. *Prinyatie reshenii. Metod analiza ierarkhii* (*Decision Making. Hierarchy Analysis Method*). Moscow: Radio i svyaz'; 1993. 341 p. (in Russ.). ISBN 5-256-00443-3

СПИСОК ЛИТЕРАТУРЫ

1. Коробов В.Б. *Теория и практика экспертных методов*. М.: ИНФРА-М; 2019. 279 с. ISBN 978-5-16-015053-6. https://doi.org/10.12737/monography_5caee0067f1835.43206494
2. Андрейчиков А.В., Андрейчикова О.Н. *Анализ, синтез, планирование решений в экономике*. М.: Финансы и статистика; 2004. 467 с. ISBN 5-279-02901-7
3. Саати Т. *Принятие решений. Метод анализа иерархий*. М.: Радио и связь; 1993. 314 с. ISBN 5-256-00443-3
4. Саати Т. *Принятие решений при зависимостях и обратных связях: аналитические сети*. М.: URSS; 2010. 357 с. ISBN 978-5-397-01622-3

4. Saaty T. *Prinyatie reshenii pri zavisimostyakh i obratnykh svyazyakh: analiticheskie seti* (*Decision Making with Dependencies and Feedbacks: Analytical Networks*). Moscow: URSS; 2010. 357 p. (in Russ.). ISBN 978-5-397-01622-3
5. Breiman L. Random forests. *Machine Learning*. 2001;45(1): 5–32. https://doi.org/10.1023/A:1010933404324
6. Belov V., Tatarintsev A., Nikulchev E. Comparative characteristics of big data storage formats. *J. Phys.: Conf. Ser.* 2021;1727(1):012005. http://doi.org/10.1088/1742-6596/1727/1/012005
7. Belov V., Tatarintsev A., Nikulchev E. Choosing a data storage format in the Apache Hadoop system based on experimental evaluation using Apache Spark. *Symmetry.* 2021;13(2):195. https://doi.org/10.3390/sym13020195
8. Moro Visconti R., Morea D. Big data for the sustainability of healthcare project financing. *Sustainability*. 2019;11(13):3748. https://doi.org/10.3390/su11133748
9. Gusev A., Ilin D., Nikulchev E. The dataset of the experimental evaluation of software components for application design selection directed by the artificial bee colony algorithm. *Data*. 2020;5(3):59. https://doi.org/10.3390/data5030059
10. Munir R.F., Abelló A., Romero O., Thiele M., Lehner W. A cost-based storage format selector for materialized results in big data frameworks. *Distrib. Parallel Databases*. 2020;38(3):335–364. https://doi.org/10.1007/s10619-019-07271-0
11. Gusev A., Ilin D., Kolyasnikov P., Nikulchev E. Effective selection of software components based on experimental evaluations of quality of operation. *Eng. Lett*. 2020;28(2):420–427.
12. Krivulin N.K., Ageev V.A., Gladkikh I.V. Application of methods of tropical optimization for evaluating alternatives based on pairwise comparisons. *Vestnik Sankt-Peterburgskogo universiteta*. *Prikladnaya matematika. Informatika. Protsessy upravleniya = Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes.* 2017;13(1):27–41. https://doi.org/10.21638/11701/spbu10.2017.103
13. Litvinov G.L. The Maslov dequantization, idempotent and tropical mathematics: a briff introduction. *Zapiski nauchnykh seminarov Sankt-Peterburgskogo otdeleniya matematicheskogo instituta im. V.A. Steklova RAN* (*Zapiski Nauchnykh Seminarov POMI*). 2005;326(13):145–182 (in Russ.).
14. Gantmakher F R. *Teoriya matrits* (*Matrix Theory*). Moscow: Fizmatlit; 2004. 560 p. (in Russ.). ISBN 5-9221-0524-8
15. Evseeva O.A., Pulkin I.S., Tatarintsev A.V. On the problem of processing expert judgments. In: *Innovatsionnye tekhnologii v elektronike i priborostroenii:* sbornik trudov konferentsii (*Innovative technologies in electronics and instrumentation:* collection of conference proceedings). Moscow: MIREA; 2021. V. 1. P. 355–359 (in Russ.).

5. Breiman L. Random forests. *Machine Learning*. 2001;45(1): 5–32. https://doi.org/10.1023/A:1010933404324
6. Belov V., Tatarintsev A., Nikulchev E. Comparative characteristics of big data storage formats. *J. Phys.: Conf. Ser.* 2021;1727(1):012005. http://doi.org/10.1088/1742-6596/1727/1/012005
7. Belov V., Tatarintsev A., Nikulchev E. Choosing a data storage format in the Apache Hadoop system based on experimental evaluation using Apache Spark. *Symmetry.* 2021;13(2):195. https://doi.org/10.3390/sym13020195
8. Moro Visconti R., Morea D. Big data for the sustainability of healthcare project financing. *Sustainability*. 2019;11(13):3748. https://doi.org/10.3390/su11133748
9. Gusev A., Ilin D., Nikulchev E. The dataset of the experimental evaluation of software components for application design selection directed by the artificial bee colony algorithm. *Data*. 2020;5(3):59. https://doi.org/10.3390/data5030059
10. Munir R.F., Abelló A., Romero O., Thiele M., Lehner W. A cost-based storage format selector for materialized results in big data frameworks. *Distrib. Parallel Databases*. 2020;38(3):335–364. https://doi.org/10.1007/s10619-019-07271-0
11. Gusev A., Ilin D., Kolyasnikov P., Nikulchev E. Effective selection of software components based on experimental evaluations of quality of operation. *Eng. Lett*. 2020;28(2):420–427.
12. Кривулин Н.К., Агеев В.А., Гладких И.В. Применение методов тропической оптимизации для оценки альтернатив на основе парных сравнений. *Вестник СПбГУ. Прикладная математика. Информатика. Процессы управления*. 2017;13(1):27–41. https://doi.org/10.21638/11701/spbu10.2017.103
13. Литвинов Г.Л. Деквантование Маслова, идемпотентная и тропическая математика: краткое введение. *Записки научных семинаров Санкт-Петербургского отделения математического института им. В.А. Стеклова РАН* (*Записки научных семинаров ПОМИ*). 2005;326(13):145–182.
14. Гантмахер Ф Р. *Теория матриц*. М.: Физматлит; 2004. 560 с. ISBN 5-9221-0524-8
15. Евсеева О.А., Пулькин И.С., Татаринцев А.В. О задаче обработки экспертных суждений. *Инновационные технологии в электронике и приборостроении*: сборник трудов конференции. М.: РТУ МИРЭА; 2021. Т. 1. С. 355–359.

About the authors

Igor S. Pulkin, Cand. Sci. (Phys.-Math.), Associate Professor, Higher Mathematics Department, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: pulkin@mirea.ru. RSCI SPIN-code 3381-669, https://orcid.org/0000-0002-5907-2151

Andrey V. Tatarintsev, Cand. Sci. (Phys.-Math.), Associate Professor, Department of Higher Mathematics and Programming, Institute of Advanced Technologies and Industrial Programming, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). Scopus Author ID 57221996001, 7004076246, https://orcid.org/0000-0003-2969-8740

Об авторах

Пулькин Игорь Сергеевич, к.ф.-м.н., доцент кафедры высшей математики Института искусственного интеллекта, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: pulkin@mirea.ru. SPIN-код РИНЦ 3381-669, https://orcid.org/0000-0002-5907-2151

Татаринцев Андрей Владимирович, к.ф.-м.н., доцент кафедры высшей математики и программирования Института перспективных технологий и индустриального программирования, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). Scopus Author ID 57221996001, 7004076246, https://orcid.org/0000-0003-2969-8740

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

Mathematical modeling

Математическое моделирование

UDC 534-16
https://doi.org/10.32362/2500-316X-2023-11-2-92-99

RESEARCH ARTICLE

Identification of a longitudinal notch of a rod by natural vibration frequencies

Ilnur M. Utyashev [1, 2], Alfir F. Fatkhelislamov [3, @]

[1] *Mavlyutov Institute of Mechanics, Ufa Federal Research Center, Russian Academy of Sciences, Ufa, 450054 Russia*
[2] *Bashkir State Agrarian University, Ufa, 450001 Russia*
[3] *Ufa University of Science and Technology, Ufa, 450076 Russia*
[@] *Corresponding author, e-mail: alfir93@mail.ru*

Abstract

Objectives. To study the direct and inverse problem of vibrations of a rectangular rod having a longitudinal notch, to analyze regularities of the behavior of natural frequencies and natural forms of longitudinal vibrations when changing the location and size of the notch, and to develop a method for uniquely identifying the parameters of the longitudinal notch using the natural frequencies of longitudinal vibrations of the rod.

Methods. The rod with a longitudinal notch is modeled as two rods, where the first one does not have a notch, while the second one does. For connection, conjugation conditions are used, in which longitudinal vibrations and deformations are equated. The solution of the inverse problem is based on the construction of a frequency equation under the assumption that the desired parameters are included in the equation. Substituting natural frequencies into this equation, the nonlinear system with respect to unknown parameters is derived. The solution of the latter is the desired notch parameters.

Results. Tables of eigenfrequencies and graphs of eigenforms are given for different notch parameters. The results for different boundary conditions are obtained and analyzed. A method for identifying notch parameters by a finite number of eigenfrequencies is presented. The inverse problem is shown to have two solutions, which are symmetrical about the center of the rod. The unambiguous solution requires eigenfrequencies of the same problem with different boundary conditions at the right end. By adding additional conditions at the ends of the rod, the inverse problem can be solved with new boundary conditions to construct the exact solution and develop an algorithm for checking the uniqueness of the solution.

Conclusions. The developed method can be used to solve the problem of identification of geometric parameters of various parts and structures modeled by rods.

Keywords: longitudinal vibrations, natural frequency, eigenform, rod, identification problem, direct problem, inverse problem, Sturm–Liouville problem, boundary conditions

• **Submitted:** 08.11.2021 • **Revised:** 29.11.2022 • **Accepted:** 22.01.2023

For citation: Utyashev I.M., Fatkhelislamov A.F. Identification of a longitudinal notch of a rod by natural vibration frequencies. *Russ. Technol. J*. 2023;11(2):92–99. https://doi.org/10.32362/2500-316X-2023-11-2-92-99

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

НАУЧНАЯ СТАТЬЯ

Идентификация продольного надреза стержня по собственным частотам колебаний

И.М. Утяшев [1, 2],
А.Ф. Фатхелисламов [3, @]

[1] *Институт механики им. Р.Р. Мавлютова Уфимского федерального исследовательского центра Российской академии наук, Уфа, 450054 Россия*
[2] *Башкирский государственный аграрный университет, Уфа, 450001 Россия*
[3] *Уфимский университет науки и технологий, Уфа, 450076 Россия*
[@] *Автор для переписки, e-mail: alfir93@mail.ru*

Резюме

Цели. Цели работы: рассмотреть прямую и обратную задачу о колебании прямоугольного стержня с продольным надрезом; исследовать закономерности поведения собственных частот и собственных форм продольных колебаний при изменении места и размера надреза; разработать метод, позволяющий однозначно идентифицировать параметры продольного надреза с помощью собственных частот продольных колебаний стержня.

Методы. Стержень с продольным надрезом моделируется как два стержня, где первый не имеет надреза, а второй – имеет. Для соединения используются условия сопряжения, в которых приравниваются продольные колебания и деформации. Решение обратной задачи основано на построении частотного уравнения в предположении, что искомые параметры входят в уравнение. При подстановке собственных частот в это уравнение получим нелинейную систему относительно неизвестных параметров. Решение последнего есть искомые параметры надреза.

Результаты. Приведены таблицы собственных частот и графики собственных форм для разных параметров надреза. Получены и проанализированы результаты для различных краевых условий. Представлен метод идентификации параметров надреза по конечному числу собственных частот. Показано, что обратная задача имеет два решения, симметричных относительно центра стержня. Для однозначного решения требуются собственные частоты той же задачи с другими граничными условиями на правом конце. Добавление дополнительных условий на концах стержня позволило решить обратную задачу с новыми краевыми условиями, дающими возможность построить точное решение и разработать алгоритм проверки однозначности решения.

Выводы. Разработанный метод позволяет решить задачу идентификации геометрических параметров различных деталей и конструкций, моделируемых стержнями.

Ключевые слова: продольные колебания, собственная частота, собственная форма, стержень, задача идентификации, прямая задача, обратная задача, задача Штурма – Лиувилля, граничные условия

• **Поступила:** 24.07.2022 • **Доработана:** 06.09.2022 • **Принята к опубликованию:** 22.01.2023

Для цитирования: Утяшев И.М., Фатхелисламов А.Ф. Идентификация продольного надреза стержня по собственным частотам колебаний. *Russ. Technol. J.* 2023;11(2):92–99. https://doi.org/10.32362/2500-316X-2023-11-2-92-99

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

INTRODUCTION

Vibrations are one of the most common forms of motion. The study of vibrations is of great practical importance both in terms of utilizing their positive properties in engineering and technology, as well as avoiding undesirable effects of vibrations by limiting their level. Important practical problems of structural dynamics can be solved on the basis of the profound study of different kinds of vibrations only [1]. Studying natural frequencies of vibrations is of great scientific and applied interest in many engineering problems.

Many studies are devoted to the identification of cracks [2–14]. Starting from studies [3–5], transverse open cracks are typically modeled according to spring conjugation conditions. In the contemporary literature, other conjugation conditions for modeling transverse cracks are also proposed [2, 6, and 7]. However, longitudinal cracks cannot be modeled by conjugation conditions. The study [2] proposes to solve the rod identification problem on the basis of changes in the moments of inertia around the axes and cross-sectional areas at the notch. Another study [8] gives the simplest model of longitudinal vibrations of the rod with nascent transverse cracks, determining natural frequencies of vibrations along with coordinates and dimensions of cracks by experimental values of natural frequencies. In [9], a rod with a rigidly fixed left end is considered; here, the fixation on the right end can be either free, or elastic or rigid. The first three natural frequencies for different cross-section profiles are given. The closest problem in terms of the formulation is given in [10], where the evolution of characteristics of natural longitudinal vibrations and natural shapes of a circular rod when increasing its cross-section defect is considered. The study [14] deals with the method of solving inverse problems of defectoscopy for rods performing longitudinal vibrations. Using numerical modeling, the use of several low frequencies for satisfactorily determining defect properties is demonstrated. In [15], experimental data is compared with different theoretical models for describing the longitudinal vibrations of a rod. In the present paper, the result is obtained for the case when the longitudinal notch does not run along the entire length of a rectangular rod. The results of the study can find application in the acoustic diagnostics of various rods, such as I-beams, rails, frame bridges, etc.

The results show that natural frequencies of longitudinal vibrations can be used to find the start point of the longitudinal notch of the rod along with its depth and width. The dependence of the problem output data on the input data can be determined by analyzing the graphs of the vibration natural forms.

DIRECT PROBLEM

We consider a homogeneous isotropic rectangular rod of length $L = 1$, density ρ and cross-sectional area F (Fig. 1). The boundary conditions are as follows: the rod is embedded on the left end and free on the right. The cross section of the rod has height H and width B. A longitudinal notch of depth h and width b is located from point x_c to the right end.

In order to determine the influence of the size and location of the notch start point on these frequencies, it is necessary to determine the natural frequencies of longitudinal vibrations of the rod. For clarity of the solution, natural forms of vibrations are constructed.

The equation of longitudinal vibrations may be described by the following equation [11, p. 189]:

$$EF\frac{d^2U(x,t)}{dx^2}+\rho F\frac{d^2U(x,t)}{dt^2}=0, \qquad (1)$$

where $U = U(x, t)$ is the longitudinal displacement; EF is the bending stiffness of the rod; ρ is the rod density; F is the cross-sectional area of the rod.

The solution of Eq. (1) is sought in the form of $U(x, t) = y(x)\cos\omega t$. Then (1) may be reduced to the following equation:

$$y''+\lambda^2 y=0, \qquad (2)$$

where spectral parameter $\lambda^2=\frac{\rho F\omega^2}{E}$. Since the rod to the left and right of point x_c has a different cross-sectional shape, the equations of longitudinal vibrations to the left and right of point x_c may be written in the following form:

$$y_-''+\lambda^2 y_-=0,\quad y_+''+\lambda^2 y_+=0, \qquad (3)$$

where y_-, y_+ are longitudinal displacements to the left and right of the point x_c.

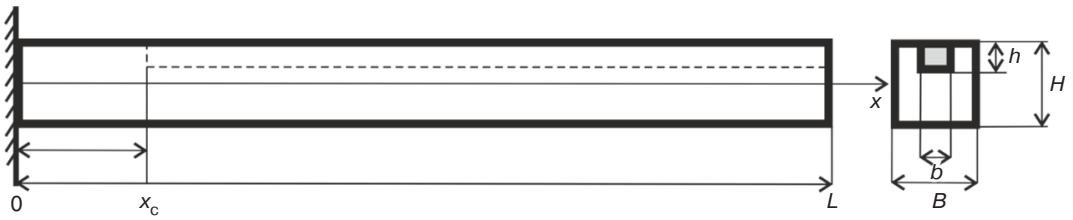

Fig. 1. Rod with a notch

The conjugation condition at point x_c for the rod sections may be written as follows:

$$EF_- \frac{dy_-}{dx} = EF_+ \frac{dy_+}{dx}. \quad (4)$$

We shall denote the ratio of areas $\frac{F_+}{F_-}$ by P:

$$F_+ = BH - bh, \quad F_- = BH, \quad P = \frac{F_+}{F_-}. \quad (5)$$

The conjugation conditions using P are written as follows:

$$y_-(x_c) = y_+(x_c), \quad y'_-(x_c) = y'_+(x_c)P. \quad (6)$$

Since the rod is embedded at the left end and free at the right end, the boundary conditions may be written in the following form:

$$y_-(0) = 0, \quad y'_+(1) = 0. \quad (7)$$

The general solution of equations (3) may be written as follows:

$$y_- = C_1^- \cos\lambda x + C_2^- \frac{\sin\lambda x}{\lambda},$$

$$y_+ = C_1^+ \cos\lambda x + C_2^+ \frac{\sin\lambda x}{\lambda}. \quad (8)$$

Substituting Eqs. (8) into (6) and (7), the following equation may be written:

$$y_-(0) = C_1^- 1 + C_2^+ 0 = C_1^- = 0. \quad (9)$$

From Eq. (9), $C_1^- = 0$.

$$y'_+(1) = -\lambda C_1^+ \sin\lambda + C_2^+ \cos\lambda = 0. \quad (10)$$

$$C_1^- + C_2^- \frac{\sin\lambda x_c}{\lambda} - C_1^+ \cos\lambda x_c - C_2^+ \frac{\sin\lambda x_c}{\lambda} = 0. \quad (11)$$

$$-\lambda C_1^- + C_2^- \cos\lambda x_c - P\left(-\lambda C_1^+ \sin\lambda x_c + C_2^+ \cos\lambda x_c\right) = 0. \quad (12)$$

The system of Eqs. (10)–(12) for finding constants C_1^+, C_2^+, C_2^- has a nonzero solution if and only if the determinant of the system is zero, as follows:

$$D = \begin{vmatrix} 0 & -\lambda\sin\lambda & \cos\lambda \\ \frac{\sin\lambda x_c}{\lambda} & -\cos\lambda x_c & -\frac{\sin\lambda x_c}{\lambda} \\ \cos\lambda x_c & P\lambda\sin\lambda x_c & -P\cos\lambda x_c \end{vmatrix} = 0. \quad (13)$$

The result is the following equation for finding the eigenvalues (natural frequencies):

$$D = -\lambda\sin\lambda\cos\lambda x_c \sin\lambda x_c P + \cos\lambda\sin^2\lambda x_c + \\ + \sin\lambda\cos\lambda x_c \sin\lambda x_c + \cos^2\lambda x_c \cos\lambda = 0. \quad (14)$$

Two kinds of problems—direct and inverse—can be solved on the basis of this equation.

The solution of the direct problem with different initial data allows the dependence of natural frequencies of vibrations on the rod parameters to be analyzed to derive main conclusions on the problem's solution. Therefore, the next stage of the work consists in analyzing the results of the direct problem.

The measurement values are dimensioned for the convenience of calculations.

Initial data: $H = B = 0.1$; $h = b = 0.01$.

It is necessary to find natural frequencies of longitudinal vibrations.

Consider the dependence of the first five natural frequencies on the position of the notch start point of the rod.

In values x_c equidistant from the middle (1, 2, 5, and 7 in Table 1), the same values of longitudinal vibration frequencies can be seen. Hence, the solution of the inverse problem is dual, i.e., there are two solutions symmetric to the middle of the rod. The solution duality can be visualized by plotting the first three eigenforms of the function (Figs. 2–4).

Table 1. Natural frequencies of longitudinal vibrations when changing the notch location

No.	Position x_c	λ_1	λ_2	λ_3	λ_4	λ_5
1	0.1	1.57389	4.72052	7.86408	11.00378	14.14031
2	0.25	1.57791	4.71956	7.84681	10.98846	14.14428
3	0.3	1.57895	4.71552	7.84388	10.99868	14.14536
4	0.5	1.58090	4.70229	7.86408	10.98547	14.14727
5	0.75	1.57791	4.71956	7.84681	10.98846	14.14428
6	0.8	1.57670	4.72201	7.85398	10.98595	14.13126
7	0.9	1.57389	4.72052	7.86408	11.00378	14.14031

Note: λ_n is natural frequencies of longitudinal vibrations of the rod found numerically in *Maple* software[1].

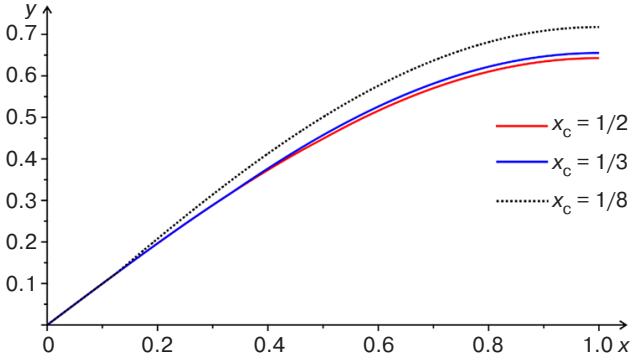

Fig. 2. First eigenform

[1] https://www.maplesoft.com/. Accessed January 01, 2023.

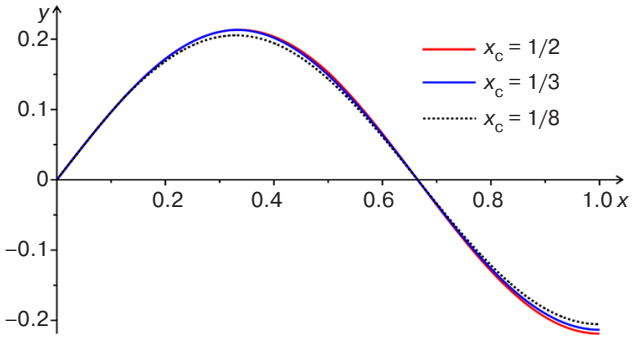

Fig. 3. Second eigenform

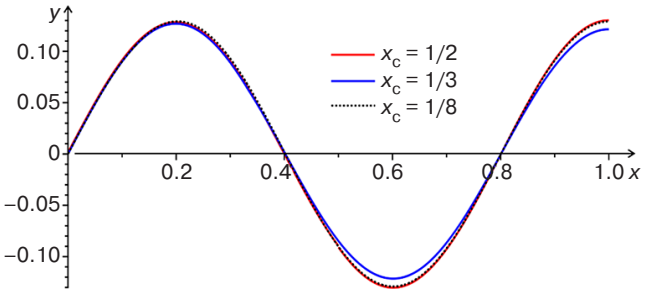

Fig. 4. Third eigenform

The graphs of eigenforms (Figs. 2–4) show that the problem of identifying the notch start point is unstable and strongly depends on the error of the input data.

We shall consider changes in the natural frequency from height h and width b of the rod notch.

Tables 2 and 3 show that a change in the notch size results in a more significant alteration to the natural frequencies of vibrations than a change in the notch location.

INVERSE PROBLEM

We shall consider the inverse problem, where it becomes necessary to find the notch start point not located in the middle of the rod by a finite number of eigenfrequencies.

Let eigenvalues $\lambda_1 = 1.57952$, $\lambda_2 = 4.71238$, and $\lambda_3 = 7.84524$, rod length $L = 1$, width and height $H = 0.1$, $B = 0.1$, and notch width and depth $b = 0.01$, $h = 0.01$, respectively, be known. It is necessary to find the start point of the notch x_c.

Substituting the ratio of areas (5) and known values into the frequency Eq. (14), the following may be written:

$$D = -\lambda_n \sin\lambda_n \cos\lambda_n x_c \sin\lambda_n x_c P + \\ + \cos\lambda_n \sin^2 \lambda_n x_c + \\ + \sin\lambda_n \cos\lambda_n x_c \sin\lambda_n x_c + \\ + \cos^2 \lambda_n x_c \cos\lambda_n = 0. \qquad (15)$$

Equation (15) gives two valid solutions (Fig. 5), symmetric about the middle of the segment L:

$$x_{c1} = 0.25, x_{c2} = 0.75.$$

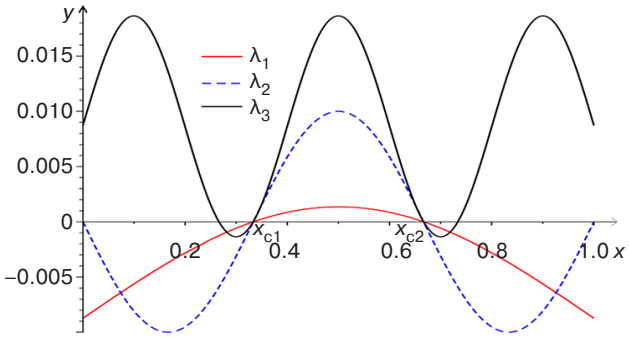

Fig. 5. The inverse problem solution

It follows from the solution of the inverse problem that an unambiguous solution would not be obtained by identifying the rod by means of longitudinal vibrations only. The exact solution requires adding supplementary

Table 2. Change in the frequency of natural vibrations from the notch height increase

No.	Height h	λ_1	λ_2	λ_3	λ_4	λ_5
1	0.01	1.57582	4.70736	7.85901	10.9905	14.14219
2	0.02	1.58090	4.70229	7.86408	10.9854	14.14727
3	0.04	1.59120	4.69198	7.87439	10.9752	14.15758
4	0.05	1.59644	4.68675	7.87963	10.9699	14.16281
5	0.06	1.60173	4.68146	7.88491	10.9646	14.16810
6	0.08	1.61247	4.67071	7.89566	10.9539	14.17884
7	0.09	1.61793	4.66525	7.90112	10.9484	14.18430

Table 3. Change in the frequency of longitudinal vibrations from the position and height of the rod notch

No.	Position x_c	Height h	λ_1	λ_2	λ_3	λ_4	λ_5
1	0.9	0.09	1.58485	4.74968	7.90108	11.03455	14.15227
2	0.1	0.01	1.57234	4.71644	7.85901	10.99965	14.13873
3	0.5	0.05	1.59644	4.68675	7.87963	10.96993	14.16281

conditions, e.g., changing the boundary condition at one of its ends.

We shall add an elastic element to the right end of the rod. Accordingly, the Robin condition appears.

The boundary conditions (Robin conditions) for a rod of unit length ($L = 1$) may be written in the following way:

$$y'_x(1) - Ky(1) = 0, \qquad (16)$$

where K is the stiffness of the spring at the end.

Substituting the general solution of (8) into (16), the following may be written:

$$M = \begin{vmatrix} 0 & K\cos\lambda_n - \lambda_n \sin\lambda_n & \cos\lambda_n + \dfrac{\sin\lambda_n}{\lambda_n} K \\ \dfrac{\sin\lambda_n x_c}{\lambda_n} & -\cos\lambda_n x_c & -\dfrac{\sin\lambda_n x_c}{\lambda_n} \\ \cos\lambda_n x_c & P\lambda\sin\lambda_n x_c & -P\cos\lambda_n x_c \end{vmatrix} = 0.$$

Find the determinant

$$\begin{aligned} D = \frac{1}{\lambda_n}\Big(&K\cos\lambda_n \cos(\lambda_n x_c)\sin(\lambda_n x_c)P + \\ &+K\sin\lambda_n \sin(\lambda_n x_c)^2 P + \\ &+\cos\lambda_n \sin(\lambda_n x_c)^2 P\lambda_n - \\ &-\sin\lambda_n \cos\left(\lambda_n x_c\right)\sin\left(\lambda_n x_c\right)P\lambda_n - \\ &-K\cos\left(\lambda_n x_c\right)\sin\left(\lambda_n x_c\right) + \\ &+K\sin\left(\lambda_n\right)\cos(\lambda_n x_c)^2 + \\ &+\lambda_n \cos(\lambda_n x_c)^2 \cos\lambda_n + \\ &+\sin(\lambda_n)\cos(\lambda_n x_c)\sin(\lambda_n x_c)\lambda_n\Big) = 0. \end{aligned} \qquad (17)$$

The solution (17) gives the eigenvalues of λ_n.

We shall solve the inverse problem and plot graphical solutions of the equation.

Let the first three natural frequencies $\lambda_1 = 2.2972263$, $\lambda_2 = 5.0846367$, and $\lambda_3 = 8.0885773$, cross-section area ratio $P = 0.98$, and spring stiffness factor $K = 2$ be known. It is necessary to find the notch location x_c. We shall substitute by order natural frequencies of longitudinal vibrations in (17). The solution of the equation is shown in Fig. 6.

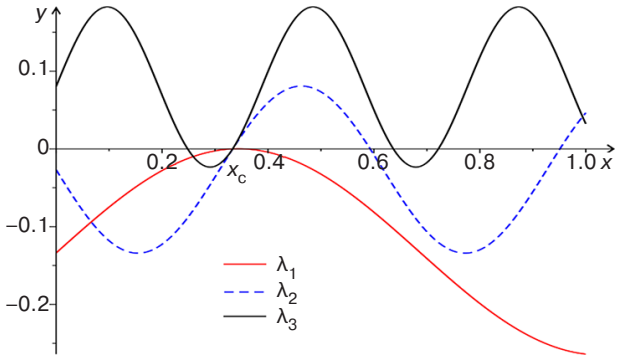

Fig. 6. Solutions to the problem with elastic anchoring

It is seen from Fig. 6 that the problem has an exact solution at the point $x_c = 0.25$ under the elastic anchoring condition. It follows that, in order to solve the inverse problem unambiguously, it will be necessary to apply elastic anchoring at one of the ends of the rod.

CONCLUSIONS

With boundary conditions (7), the solution of the inverse problem of identifying the longitudinal notch by natural frequencies of longitudinal vibrations is unstable since natural shapes are close to each other at the points symmetric from the middle of the notch; moreover, in order to solve the problem unambiguously, it is necessary to replace the fixing one of the rod ends by an elastic condition. The plotted graphs and given tables of the direct problem solution have shown the dependence of natural frequencies on the initial data. An example of the inverse problem solution is given showing that two natural frequencies obtained for different boundary conditions are sufficient for unambiguous determination of the start of the notch. This method may be applied to solving problems on identifying geometric parameters of parts modeled by the rod.

Authors' contribution. All authors equally contributed to the research work.

REFERENCES

1. Shakirzyanov R.A., Shakirzyanov F.R. *Dinamika i ustoichivost'sooruzhenii:* uchebnoe posobie (*Dynamics and stability of structures:* textbook). Moscow: IPR Media; 2022. 119 p. (in Russ.). ISBN 978-5-4497-1379-7
2. Akhtyamov A.M., Il'gamov M.A. Flexural model for a notched beam: Direct and inverse problems. *J. Appl. Mech. Tech. Phy.* 2013;54(1):132–141. https://doi.org/10.1134/S0021894413010161

СПИСОК ЛИТЕРАТУРЫ

1. Шакирзянов Р.А., Шакирзянов Ф.Р. *Динамика и устойчивость сооружений:* учебное пособие. М.: Ай Пи Ар Медиа; 2022. 119 с. ISBN 978-5-4497-1379-7
2. Ахтямов А.М., Ильгамов М.А. Модель изгиба балки с надрезом: прямая и обратная задачи. *Прикладная механика и техническая физика*. 2013;54(1):152–162.
3. Rice J.R., Levy N. The part through surface crack in an elastic plate. *J. Appl. Mech.* 1972;39(1):185–194. https://doi.org/10.1115/1.3422609

[Original Russian Text: Akhtyamov A.M., Il'gamov M.A. Flexural model for a notched beam: Direct and inverse problems. *Prikladnaya Mekhanika i Tekhnicheskaya Fizika*. 2013;54(1):132–141 (in Russ.).]

3. Rice J.R., Levy N. The part through surface crack in an elastic plate. *J. Appl. Mech.* 1972;39(1):185–194. https://doi.org/10.1115/1.3422609
4. Freund L.B., Herrmann G. Dynamic fracture of a beam or plate in plane bending. *J. Appl. Mech.* 1976;43(1): 112–116. https://doi.org/10.1115/1.3423760
5. Narkis Y. Identification of crack location in vibrating simply-supported beames. *J. Sound Vib*. 1994;172(4): 549–558. https://doi.org/10.1006/jsvi.1994.1195
6. Vatul'yan A.O., Osipov A.V. Transverse vibrations of beam with localized heterogeneities. *Vestnik Donskogo gosudarstvennogo tekhnicheskogo universiteta = Vestnik of Don State Technical University* (*Advanced Engineering Research*). 2012;12(8):34–40 (in Russ.).
7. Il'gamov M.A., Khakimov A.G. Diagnosis of damage of a cantilever beam with a notch. *Russ. J. Nondestruct. Test*. 2009;45(6):430–435. https://doi.org/10.1134/S1061830909060072
   [Original Russian Text: Ilgamov M.A., Khakimov A.G. Diagnosis of damage of a cantilever beam with a notch. *Defektoskopiya*. 2009;45(6):83–89 (in Russ.).]
8. Il'gamov M.A. Longitudinal vibrations of a bar with incipient transverse cracks. *Mech. Solids*. 2017;52(1): 18–24. https://doi.org/10.3103/S0025654417010034
   [Original Russian Text: Il'gamov M.A. Longitudinal vibrations of a bar with incipient transverse cracks. *Izvestiya Akademii Nauk. Mekhanika Tverdogo Tela*. 2017;1:23–31 (in Russ.).]
9. Utyashev I.M. Longitudinal oscillation of a rod with a variable cross section *Mnogofaznye sistemy = Multiphase Systems.* 2019;14(2):138–141 (in Russ.). https://doi.org/10.21662/mfs2019.2.019
10. Akulenko L.D., Baidulov V.G., Georgievskii D.V., Nesterov S.V. Evolution of natural frequencies of longitudinal vibrations of a bar as its cross-section defect increases. *Mech. Solids.* 2017;52(6):708–714. https://doi.org/10.3103/S0025654417060103
    [Original Russian Text: Akulenko L.D., Baidulov V.G., Georgievskii D.V., Nesterov S.V. Evolution of natural frequencies of longitudinal vibrations of a bar as its cross-section defect increases. *Izvestiya Akademii Nauk. Mekhanika Tverdogo Tela*. 2017;6:136–144 (in Russ.).]
11. Bolotin V.V. (Ed.). *Vibratsii v tekhnike*: Spravochnik v 6 t. T. 1. *Kolebaniya lineinykh system* (*Vibrations in Technology:* in 6 v. V. 1. *Oscillations of Linear Systems*). Moscow: Mashinostroenie; 1978. 352 p. (in Russ.).
12. Narkis Y. Identification of crack location in vibrating simply-supported beames. *J. Sound Vib.* 1994;172(2): 549–558. https://doi.org/10.1006/jsvi.1994.1195
13. Akhtyamov A.M., Fatkhelislamov A.F. Identification of location of a cut of the rod on natural frequency. *Doklady Bashkirskogo Universiteta* 2017;2(2):204–208 (in Russ.)
14. Akulenko L.D., Gavrikov A.A., Nesterov S.V. Identification of cross-section defects of the rod by using eigenfrequencies and features of the shape of longitudinal oscillations. *Mech. Solids*. 2019;54(8): 1208–1215. https://doi.org/10.3103/S0025654419080119

4. Freund L.B., Herrmann G. Dynamic fracture of a beam or plate in plane bending. *J. Appl. Mech.* 1976;43(1): 112–116. https://doi.org/10.1115/1.3423760
5. Narkis Y. Identification of crack location in vibrating simply-supported beames. *J. Sound Vib*. 1994;172(4): 549–558. https://doi.org/10.1006/jsvi.1994.1195
6. Ватульян А.О., Осипов А.В. Поперечные колебания балки с локализованными неоднородностями. *Вестник Донского государственного технического университета*. 2012;12(8):34–40.
7. Ильгамов М.А., Хакимов А.Г. Диагностика повреждений консольной балки с надрезом. *Дефектоскопия*. 2009;6:83–89.
8. Ильгамов М.А. Продольные колебания стержня с зарождающимися поперечными трещинами. *Известия Российской академии наук. Механика твердого тела*. 2017;1:23–31.
9. Утяшев И.М. Продольные колебания стержня с переменным сечением. *Многофазные системы*. 2019;14(2):138–141. https://doi.org/10.21662/mfs2019.2.019
10. Акуленко Л.Д., Байдулов В.Г., Георгиевский Д.В., Нестеров С.В. Эволюция собственных частот продольных колебаний стержня при увеличении дефекта поперечного сечения. *Известия Российской академии наук. Механика твердого тела*. 2017;6:136–144.
11. Болотин В.В. (ред.). *Вибрации в технике:* справочник в 6 т. Т. 1. *Колебания линейных систем.* М.: Машиностроение; 1978. 352 с.
12. Narkis Y. Identification of crack location in vibrating simply-supported beames. *J. Sound Vib.* 1994;172(2): 549–558. https://doi.org/10.1006/jsvi.1994.1195
13. Ахтямов А.М., Фатхелисламов А.Ф. Идентификация местоположения надреза стержня по собственной частоте. *Доклады Башкирского университета.* 2017;2(2):204-208.
14. Акуленко Л.Д., Гавриков А.А., Нестеров С.В. Идентификация дефектов поперечного сечения стержня по собственным частотам и особенностям формы продольных колебаний. *Известия Российской академии наук. Механика твердого тела.* 2019;6:98–107. https://doi.org/10.1134/S0572329919060023
15. Попов А.Л., Садовский С.А. О соответствии теоретических моделей продольных колебаний стержня экспериментальным данным. *Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия.* 2021;8(2):270–281. https://doi.org/10.21638/spbu01.2021.207

[Original Russian Text: Akulenko L.D., Gavrikov A.A., Nesterov S.V. Identification of cross-section defects of the rod by using eigenfrequencies and features of the shape of longitudinal oscillations. *Izvestiya Akademii Nauk. Mekhanika Tverdogo Tela*. 2019;6:98–107 (in Russ.).]

15. Popov A.L., Sadovsky S.A. Correspondence of theoretical models of longitudinal rod vibrations to experimental data. *Vestnik St. Peterb. Univ. Math*. 2021;54(2):162–170. https://doi.org/10.1134/S1063454121020114
[Popov A.L., Sadovsky S.A. Correspondence of theoretical models of longitudinal rod vibrations to experimental data. *Vestnik Sankt-Peterburgskogo Universiteta. Matematika. Mekhanika. Astronomiya*. 2021;54(2):270–281 (in Russ.). https://doi.org/10.21638/spbu01.2021.207]

About the authors

Ilnur M. Utyashev, Cand. Sci. (Phys.-Math.), Researcher, Mavlyutov Institute of Mechanics, Ufa Federal Research Center, Russian Academy of Sciences (71, Oktyabrya pr., Ufa, Republic of Bashkortostan, 450054 Russia); Assistant Professor, Department of Mathematics, Bashkir State Agrarian University (34, 50-Letiya Oktyabrya ul., Ufa, Republic of Bashkortostan, 450001 Russia). E-mail: utyashevim@mail.ru. Scopus Author ID 56966700200, ResearcherID J-1064-2018, RSCI SPIN-code 7856-5351, https://orcid.org/0000-0002-2342-0492

Alfir F. Fatkhelislamov, Senior Lecturer, Department of Information Security Management, Ufa University of Science and Technology (32, Zaki Validi ul., Ufa, Republic of Bashkortostan, 450076 Russia). E-mail: alfir93@mail.ru. RSCI SPIN-code 5470-4819

Об авторах

Утяшев Ильнур Мирзович, к.ф.-м.н., научный сотрудник, Институт механики им. Р.Р. Мавлютова Уфимского федерального исследовательского центра Российской академии наук (450054, Россия, Республика Башкортостан, Уфа, пр. Октября, д. 71); доцент кафедры математики, Башкирский государственный аграрный университет (450001, Россия, Республика Башкортостан, Уфа, ул. 50-летия Октября, д. 34). E-mail: utyashevim@mail.ru. Scopus Author ID 56966700200, ResearcherID J-1064-2018, SPIN-код РИНЦ 7856-5351, https://orcid.org/0000-0002-2342-0492

Фатхелисламов Альфир Фирдависович, старший преподаватель кафедры управления информационной безопасностью, Уфимский университет науки и технологий (450076, Россия, Республика Башкортостан, Уфа, ул. Заки Валиди, д. 32). E-mail: alfir93@mail.ru. SPIN-код РИНЦ 5470-4819

Translated from Russian into English by Kirill V. Nazarov
Edited for English language and spelling by Thomas A. Beavitt

MIREA – Russian Technological University.
78, Vernadskogo pr., Moscow, 119454 Russian Federation.
Publication date March 31, 2023.
Not for sale.

МИРЭА – Российский технологический университет.
119454, РФ, г. Москва, пр-т Вернадского, д. 78.
Дата опубликования 31.03.2023 г.
Не для продажи.


МИРЭА
Российский
технологический
университет