НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВО–ТЕХНІЧНИЙ ЖУРНАЛ

№ 4 2022

ЗАСНОВАНО У ЛИПНІ 2001 р.

АДРЕСА РЕДАКЦІЇ:

03056, м. Київ,
просп. Перемоги, 37, корп. 35,
ННК «ІПСА» КПІ ім. Ігоря Сікорського
Тел.: **204-81-44**; факс: **204-81-44**
E-mail: journal.iasa@gmail.com
http://journal.iasa.kpi.ua

У номері:

NATIONAL ACADEMY OF SCIENCES OF UKRAINE
EDUCATIONAL AND SCIENTIFIC COMPLEX
«INSTITUTE FOR APPLIED SYSTEM ANALYSIS»
OF THE NATIONAL TECHNICAL UNIVERSITY OF UKRAINE
«IGOR SIKORSKY KYIV POLYTECHNIC INSTITUTE»

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC AND TECHNICAL JOURNAL

№ 4 2022

IT IS FOUNDED IN JULY 2001

THE EDITION ADDRESS:

03056, Kyiv,
av. Peremogy, 37, building 35,
Institute for Applied System Analysis
at the Igor Sikorsky Kyiv Polytechnic Institute
Phone: **204-81-44**; Fax: **204-81-44**
E-mail: journal.iasa@gmail.com
http://journal.iasa.kpi.ua

In the issue:

Шановні читачі!

Навчально-науковий комплекс «Інститут прикладного системного аналізу» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» видає міжнародний науково-технічний журнал

«СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ».

Журнал публікує праці теоретичного та прикладного характеру в широкому спектрі проблем, що стосуються системних досліджень та інформаційних технологій.

Провідні тематичні розділи журналу:

Теоретичні та прикладні проблеми і методи системного аналізу; теоретичні та прикладні проблеми інформатики; автоматизовані системи управління; прогресивні інформаційні технології, високопродуктивні комп'ютерні системи; проблеми прийняття рішень і управління в економічних, технічних, екологічних і соціальних системах; теоретичні та прикладні проблеми інтелектуальних систем підтримання прийняття рішень; проблемно і функціонально орієнтовані комп'ютерні системи та мережі; методи оптимізації, оптимальне управління і теорія ігор; математичні методи, моделі, проблеми і технології дослідження складних систем; методи аналізу та управління системами в умовах ризику і невизначеності; евристичні методи та алгоритми в системному аналізі та управлінні; нові методи в системному аналізі, інформатиці та теорії прийняття рішень; науково-методичні проблеми в освіті.

Головний редактор журналу — ректор Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», академік НАН України Михайло Захарович Згуровський.

Журнал «Системні дослідження та інформаційні технології» включено до переліку фахових видань ВАК України.

Журнал «Системні дослідження та інформаційні технології» входить до таких наукометричних баз даних: Scopus, EBSCO, Google Scholar, DOAJ, Index Copernicus, реферативна база даних «Україніка наукова», український реферативний журнал «Джерело», наукова періодика України.

Статті публікуються українською та англійською мовами.

Журнал рекомендовано передплатити. **Наш індекс 23918.** Якщо ви не встигли передплатити журнал, його можна придбати безпосередньо в редакції за адресою: 03056, м. Київ, просп. Перемоги, 37, корп. 35.

Завідувачка редакції **С.М. Шевченко**
Редакторка **Р.М. Шульженко**
Молодша редакторка **Л.О. Тарин**
Комп'ютерна верстка, дизайн **А.А. Патіохи**
Свідоцтво про реєстрацію КВ № 23234–13074 ПР від 22.03.2018 р.

Підписано до друку 27.12.2022. Формат 70×108 1/16. Папір офс. Гарнітура Times.
Спосіб друку – цифровий. Ум. друк. арк. 14,411. Обл.-вид. арк. 28,56. Наклад 100 пр. Зам. № 11/04

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Свідоцтво про державну реєстрацію: ДК № 5354 від 25.05.2017 р.
просп. Перемоги, 37, м. Київ, 03056.
ФОП Пилипенко Н.М., вул. Мічуріна, б. 2/7, м. Київ, 01014.
Виписка з Єдиного державного реєстру № 2 070 000 0000 0214697 від 17.05.2019 р.,
тел.. (044) 361 78 68.

Dear Readers!

Educational and Scientific Complex «Institute for Applied System Analysis» of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute» is published of the international scientific and technical journal

«SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES».

The Journal is printing works of a theoretical and applied character on a wide spectrum of problems, connected with system researches and information technologies.

The main thematic sections of the Journal are the following:

Theoretical and applied problems and methods of system analysis; theoretical and applied problems of computer science; automated control systems; progressive information technologies, high-efficiency computer systems; decision making and control in economic, technical, ecological and social systems; theoretical and applied problems of intellectual systems for decision making support; problem- and function-oriented computer systems and networks; methods of optimization, optimum control and theory of games; mathematical methods, models, problems and technologies for complex systems research; methods of system analysis and control in conditions of risk and uncertainty; heuristic methods and algorithms in system analysis and control; new methods in system analysis, computer science and theory of decision making; scientific and methodical problems in education.

The editor-in-chief of the Journal is rector of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», academician of the NASU Michael Zaharovich Zgurovsky.

The articles to be published in the Journal in Ukrainian and English languages are accepted. Information printed in the Journal is included in the Catalogue of periodicals of Ukraine.

СИСТЕМНІ ДОСЛІДЖЕННЯ ТА ІНФОМАЦІЙНІ ТЕХНОЛОГІЇ

4 • 2022

ЗМІСТ

SYSTEM RESEARCH AND INFORMATION TECHNOLOGIES

4 • 2022

CONTENT

UDC 62-50
DOI: 10.20535/SRIT.2308-8893.2022.4.01

1D CNN MODEL FOR ECG DIAGNOSIS BASED ON SEVERAL CLASSIFIERS

MAHMOUD M. BASSIOUNI, ISLAM HEGAZY, NOUHAD RIZK, EL-SAYED A. EL-DAHSHAN, ABDELBADEEH M. SALEM

Abstract. One of the main reasons for human death is diseases caused by the heart. Detecting heart diseases in the early stage can stop heart failure or any damage related to the heart muscle. One of the main signals that can be beneficial in the diagnosis of diseases of the heart is the electrocardiogram (ECG). This paper concentrates on the diagnosis of four types of ECG records such as myocardial infarction (MYC), normal (N), variances in the ST-segment (ST), and supraventricular arrhythmia (SV). The methodology captures the data from six main datasets, and then the ECG records are filtered using a pre-processing chain. Afterward, a proposed 1D CNN model is applied to extract features from the ECG records. Then, two different classifiers are applied to test the extracted features' performance and obtain a robust diagnosis accuracy. The two classifiers are the softmax and random forest (RF) classifiers. An experiment is applied to diagnose the four types of ECG records. Finally, the highest performance was achieved using the RF classifier, reaching an accuracy of 98.3%. The comparison with other related works showed that the proposed methodology could be applied as a medical application for the early detection of heart diseases.

Keywords: Electrocardiogram (ECG), Continuous wavelet transform (CWT), 1D convolutional neural network (CNN) model.

INTRODUCTION

Heart diseases are one of the main reasons for death worldwide and they are sometimes called cardiovascular diseases (CVD). Various people suffer and die from heart diseases annually based on recent research and survey studies. In 2022 [1], it is estimated that about 17.9 million people died from CVD, and this represents about 32% of the global death, and about 85% of these people have died from heart attack and stroke. Moreover, CVD was responsible for 38% of all premature deaths (under the age of 70) due to non-communicable diseases. About 3 quarters of the deaths caused by CVD occur in the low-and middle-income countries. Arrhythmia is one of the salient groups of CVDs. They represent the abnormal electrical conduction or impulse origin in the heart. Most of the arrhythmias are non-life-threatening, while some of them can cause many cardiovascular complications and sudden death. The early diagnosis of arrhythmia can assist in preventing sudden death and help in treating many different cardiovascular diseases. Physicians, experts, and doctors detect arrhythmias based on electrocardiograms (ECG) signals. The ECG measures the variations in the electrical potential in one cycle of the heartbeat. A single ECG signal consists of a group of peaks defined by P, Q, R, S and T. Moreover, various types of arrhythmia do not appear in a short time and may require a large amount of ECG heartbeats. As a result, a diagnosis method automated should be investigated for the identification of different ECG records and this is the main focus of the proposed methodology.

Several methodologies based on machine learning have been built for extracting features and classifying ECG records. On one hand, extracting features from ECG signals is essential before the classification process because it provides a great impact on the results of the classification. P-QRS-T segment and RR interval were used in almost every research [2]. In addition to this, there are other conventional features extracted from the ECG based on morphological features, wavelet transform features, higher-order statistics, random projection, and wavelet packet entropy. These methodologies require providing a hand-crafted feature before applying any conventional classifier. There are several disadvantages in these processes of feature extraction which are depthless, large time-consuming and they lack any implicit knowledge. On the other hand, several numbers of classifiers were applied such as a k-nearest neighbor, artificial neural network, support vector machine, random forest, and Gaussian mixture models. When these conventional features are fit to these conventional classifiers they suffer from overfitting obstacles. The main causes for overfitting are as follows: (i) noise and unclean data used for training (ii) high variance and complexity of the model (iii) size of the training set is not enough (iv) learning from a small dataset. Deep learning (DL) is preserved to be part of machine learning. It is known by the word "deep" because the network structure consists of many hidden layers [3]. The main concept in DL is that the low-level features are integrated to obtain high-level features. In DL no hand-craft features are obtained and implicit knowledge can be learned easily. DL has also been used in some of the ECG studies, and it showed excellent classification results in diagnosis. Several DL structures were used such as recurrent neural network, stacked de-noising auto-encoder, deep neural network, convolutional neural networks, and restricted Boltzmann machine. Finally, based on the advantages of the DL the proposed methodology used the merits of the DL and delivered the following contributions.

The contributions stemming from this paper are two-main folds:

1. The proposed DL is used to diagnose four main ECG records based on balanced datasets of records.

2. Development of a proposed 1D CNN model for the diagnosis of several ECG diseases.

The manuscript is summarized based on different sections. Related works are presented in section 2, while the proposed methodology in terms of capturing ECG records, filtering the ECG signals, extracting features, and classifying the ECG records is presented in section 3. Moreover, the main results and the discussion are illustrated in sections 4 and 5 respectively. Finally, section 6 manifests the conclusion and the future directions.

RELATED WORK

Various approaches are applied for the diagnosis of ECG signals. These approaches depend on machine learning and deep learning methodologies and techniques. Some methods deal with the ECG signals in the form of 1D signal and other methods convert the 1D ECG signals to 2D images using several techniques such as fast Fourier transform and wavelet transforms. Moreover, numerous studies applied 1D CNN models for extracting feature from the ECG records and heartbeats. A study presented by L.A. Abdullah et al. [4] for the diagnosis of ECG signals. The proposed model is based on a 1D CNN model for learning features, and the results are fed to a long short term memory (LSTM). The 1D CNN model consists of 4 (1D) convolutional and 2 fully connected layers, while the LSTM model consists of 2 LSTM and 2 fully connected layers. Two main datasets were used in their study which are MIT-BIH arrhythmia and PTB diagnostic datasets. The CNN-LSTM model has achieved an accuracy of 98.1% and 98.66% in the diagnosis of myocardial infraction (MYC) and other arrhythmia respectively.

Another study presented by E. Butun. et al. [5] for detecting various heart diseases using ECG signals. The methodology is based on 1D version of capsule networks (CapsNet). The 1D CapsNet model consist of several layers based on convolutional and fully connected layers. The model starts with 1 input and 2 (1D) convolutional layers. Then, the model consists of 1D convolutional, 1 reshape, and 1 squash layers. Afterwards, the output of the squashing is input to an ECG caps. The ECG caps consists of a masking layer and 3 fully connected layers. Finally, the model ends with CapsNet for the ECG diagnosis. The model classifies normal and coronary artery diseases (CAD) using 5-fold cross validation achieving an accuracy of 99.44% and 98.62% for 2*s* and 5*s* ECG segments respectively. In addition to this, a study presented by X. Hau et al. [6] for the diagnosis of several ECG diseases. The methodology proposed is based on preprocessing, data augmentation, and data segmentation using R-R-R strategy. The data were selected from the MIT-BIH arrhythmia, and the number of ECG heartbeats selected are normal, left bundle block beat (LBBB), right bundle block beat (RBBB), premature ventricular contraction, and the paced beat. Then, the features are extracted using a proposed 1D CNN model. The model consists of 3 convolutional, 3 pooling, 1 fully connected layer, and 1 classification layer. It was tested on 5 classes of ECG heart beats, and the results using accuracy, area under the curve (AUC), sensitivity, and F1-score performance measurements have achieved 0.9924, 0.9994, 0.99, and 0.99 respectively.

Moreover, a study provided by G. Petmeza et al. [7] for the diagnosis of ECG diseases. The methodology proposed relies on Butterworth filter for ECG signal de-noising. In addition to that, an improved version of cross-entropy loss to solve the problem of unbalanced data. Then, the *R* peaks and the beats of the ECG signals are separated before extracting features. Also, the features are extracted from a hybrid model depending on 1D CNN layers and LSTM layers. The model consists of 3 (1D) convolutional, 3 max pooling, 1 LSTM, and 1 dense layers. Ten-fold cross validation is applied to denoise normal, atrial fibrillation (AFID), atrial flutter (AFL), and AV junctional rhythm (J). The data was obtained from the MIT-BIH atrial fibrillation database, and the model achieved a sensitivity and specificity of 97.87% and 99.29% respectively. Finally, it can be seen from the previous works that various types of 1D CNN models were proposed, and the results achieved or resulted from them was robust in performance. Therefore, it was recommended to develop a 1D CNN model for diagnosis of ECG records.

METHODOLOGY

The methodology consists of four main stages which are obtaining ECG data, filtering ECG signals, extracting various ECG features, and classifying ECG records. In the data acquisition phase, six main data sets online are downloaded that consist of four different ECG heartbeats. In the second stage, filtering or denoising is performed on the ECG heartbeats as the ECG signal consists of three main common noises which are line drifting, power interference, and noise based on a high frequency.

These distortions are removed using wavelets and a set of filters. The next stage is to pass the filtered ECG records to a proposed 1D CNN model for feature extraction. Finally, in the classification phase, two different classifiers are employed for ECG diagnosis which are Softmax and Random Forest (RF) as shown in Fig. 1.

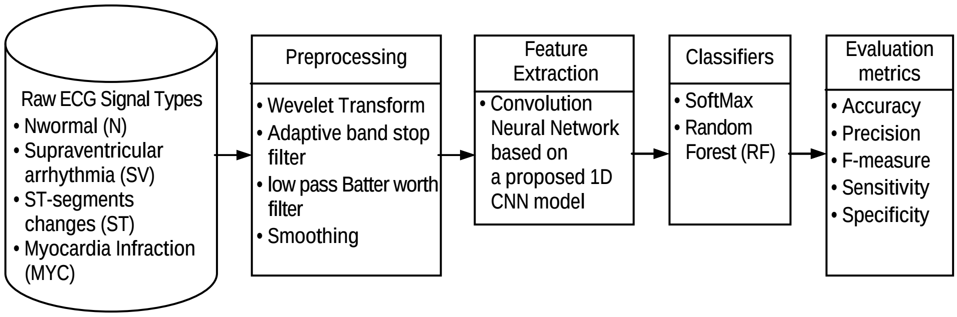


Fig. 1. Proposed Overall Methodology

Data Acquisition

This stage is one of the most important stages in the methodology proposed. There exist two concepts for capturing the ECG signals. The first concept is the application of a medical device for capturing the ECG heartbeats at different leads, whereas the second way is to download an available ECG signal online. In this methodology, the second way is employed, and the ECG signals are captured from six datasets which are Normal Sinus Rhythm Database (nsrdb) [8], Normal Sinus Rhythm RR Interval Database (nsr2db) [9], MIT-BIH Supraventricular arrhythmia database (svdb) [10], MIT-BIH ST change database (stdb) [11], Long Term ST database (ltsdb) [12] and PTB diagnostic ECG [13], where the number of ECG records in the former datasets are 18, 54, 84, 28, 86, and 549 respectively. Four main different ECG heartbeats were chosen from these datasets.

Pre-processing

ECG signals are always nested in it some distortions and noises that are produced from variant origins. The main three types of noises concentrated in the ECG signals are the drifting in the baseline of ECG signal, interference in the power line, high noise frequency in the main components of the signal, and in some cases, a combination between these types of noises can be found. As a result, a preprocessing chain of filters and wavelets is developed to eliminate these noises by saving the main information of the signal. The pre-processing chain should be summarized in three main tasks which are correcting the drifting in the ECG signal, reducing the interference, selecting low-frequency components, and enhancing the overall signal [14]. The chain contains four main stages based on wavelet drift correction, adaptive band stop filter, low pass filter, and smoothing. The baseline drift is removed by applying wavelet decomposition with db8 and a decomposition level equal to 9.

Then, the powerline interference is removed using an adaptive band stop filter with a stopband frequency corner W_s equals to 50Hz. In addition to this, high frequency located in the ECG signals is removed using a low pass Butterworth filter with a passband frequency corner and a stopband frequency corner equal to 40Hz and 60Hz respectively. The values for the passband ripple and stopband ripple attenuation are 0.1dB and 30dB respectively [15]. Finally, a smoothing filter based on Savitzky–Golay (SG) is applied to remove the remaining noise with a smoothing value equal to 5.

Feature Extraction

Convolutional Neural Networks based on 1D-CNN Model. The datasets faced in this study does not have any previous information or knowledge about the features that can be extracted from them. Also, it may be very difficult to extract robust features using traditional machine learning or feature engineering techniques. It is recommended to learn information or features automatically using deep learning. Therefore, the CNN model is developed to obtained robust features from the ECG records [16].

The main core of the CNN model is the convolutional layers because these layers work by applying a convolution operation between the local and filter regions of the input. It is also known that CNN models are designed for two-dimensional data that appear in most cases in the form of images. The proposed 1D CNN model depends mainly on convolutional layers at the beginning of the model and at the middle of the model. The convolutional layers at the beginning extracts low level features from the ECG signals that can appear in the form of sudden variances, while the convolutional layers in the middle extracts high level and more abstract features related to the ECG signals.

To use the CNN model that relies on the convolutional layers for 1D signals, the convolutional layers must be redesigned to match the input. The proposed CNN model consists of an input layer, 3 convolutional layers, 3 ReLU layers, 3 batch normalization layers, 3 max pooling layers, and ending with 3 fully connected layers [17]. The structure of the proposed 1D CNN is shown in Fig. 2, and the details (Filter size, Stride, Padding) of each layers in the proposed 1D CNN are shown in table 1.

Table 1. The Whole Parameters of the proposed 1D CNN Model

Layer No	Layers Name	Activations	Learnables	Parameters
1	Input Layer	1 · 65536 · 1	–	Input Size = [1 65536 1] Normalization = "zero center"
2	Convolutional	1 · 4 · 32	W = [1 × 23 × 1 × 32] B = [1 × 1 × 32]	Filter Size = [1 32] No. Filters = 32 Stride = [3 3] Padding = [0 0 0 0]
3	Activation	1 · 4 · 32	–	Function = "ReLU"
4	Batch Normalization	1 · 4 · 32		Offset = 1 · 1 · 16 Scale = 1 · 1 · 16

Continued table 1

Layer No	Layers Name	Activations	Learnables	Parameters
5	Max Pooling	1 · 1 · 32	–	Pool Size = [1 2] Stride = [1 1] Padding = [0 0 0 0]
6	Convolutional	1 · 1 · 32	W = [1 □ 23 · 1·· 32] B = [1 · 1 · 32]	Filter Size = [1 32] No. Filters = 32 Stride = [2 2] Padding = [0 0 0 0]
7	Activation	1 · 4 · 32	–	Function = "ReLU"
8	Batch Normalization	1 · 4 · 32		Offset = 1 · 1 · 16 Scale = 1 · 1 · 16
9	Max Pooling	1 × 1 × 32	–	Pool Size = [1 2] Stride = [1 1] Padding = [0 0 0 0]
10	Convolutional	1 · 1 · 16	W = [1 · 23 · 1 · 32] B = [1 · 1 · 32]	Filter Size = [1 16] No. Filters = 16 Stride = [1 1] Padding = [0 0 0 0]
11	Activation	1 · 4 · 16	–	Function = "ReLU"
12	Batch Normalization	1 · 4 · 16		Offset = 1 · 1 · 16 Scale = 1 · 1 · 16
13	Max Pooling	1 · 1 · 16	–	Pool Size = [1 2] Stride = [1 1] Padding = [0 0 0 0]
14	Fully Connected	1 · 1 · 100	W = [100 · 16] B = [100 · 1]	Output Size = 100
15	Fully Connected	1 · 1 · 4	W = [100 · 4] B = [100 · 1]	Output Size = 4

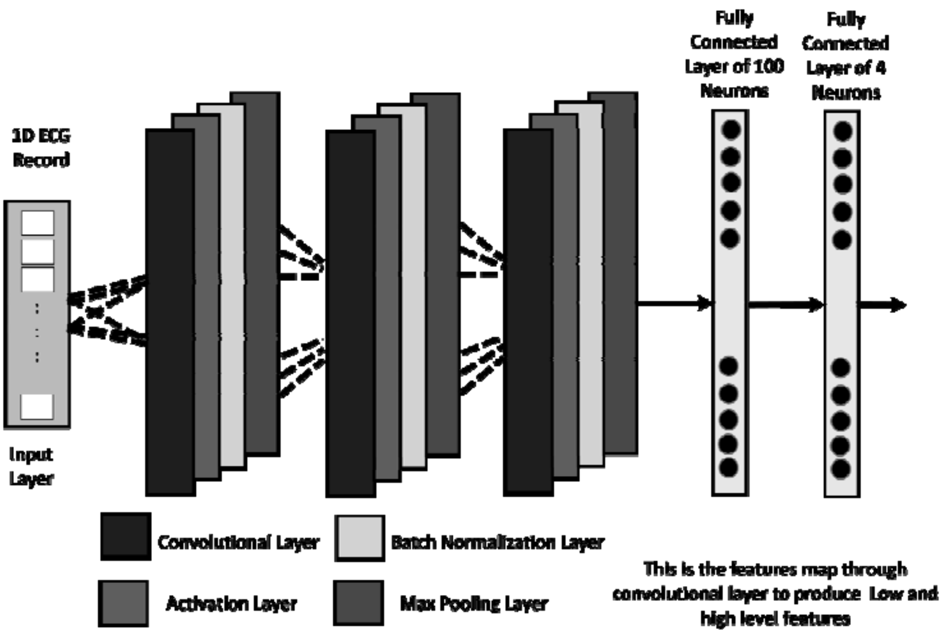


Fig. 2. The proposed 1D CNN model for ECG records Diagnosis

Classification

This step is the final step of the proposed methodology in which the result of the diagnosis will be determined with average accuracy. Two main classifiers are determined to examine the performance of the methodology, and these classifiers are Softmax, Random forest (RF), and XGBoost classifier.

Softmax Classifier. Softmax is known as a multinomial logistic regression and it is well accepted in statistical mathematics as it is applied to classify a categorical class placement. Softmax gives a more intuitive classification output and probabilistic interpretation [18]. For instance, let us assume that 4 classes are presented, the Softmax classifier will have 4 main nodes generated and defined by P_i , where $i = 1,2,3,4$. These probabilities depend on a discrete target function, and these probabilities are input to the Softmax classifier in the form of the following equation:

$$S_i = \sum_k Y_k t_{ki} ,$$

where y is the activation produced from the nodes found in the last layers, and the t is the weight that joins the last layers of nodes to the last layers in the DL model [64].

The probability of the S_i will be defined using the following equation:

$$P_i = \frac{\exp(S_i)}{\sum_j^2 \exp(S_i)} .$$

The predicated class i will be obtained by the following equation:

$$i = \operatorname{argmax}(P_i) .$$

Random Forest (RF) Classifier. Random Forests (RF) are one of the powerful ensemble learning methods. RF was developed to overcome the drawbacks of decision trees (DT). The major disadvantage of the DT is the high variance. In order words, it is not natural that a small variance in the training data can lead to a major change in the structure of a decision tree. This makes the decision trees as a classifier largely unstable in comparison to other decision predictors. Also, if an error happens in a node that is near the root, it propagates to the leaves of the tree. This leads to different and worse classification results. Therefore, the classifier of the random forest is invented by Breiman [19]. RF is built based on the combination of various decision trees. It integrates the output obtained from each separate decision tree to generate the final result. In addition to that, RF relies on uncorrelated decision trees. In other words, if similar decision trees are used in the forest, then the overall result will not vary so much and it will be equivalent to the result of a single decision tree. To achieve the concept of uncorrelated decision trees in RF features randomness and bootstrapping are applied. Random forests work considering a learning set known by $L = ((X_1, Y_1), \ldots$ $\ldots, (X_i, Y_i))$ designed with i vector. Where X is a set of features and samples and the Y is the set of labels. In the classification problems, RF maps X to Y and new input features are recognized by each tree of the forest. Then, each tree produces a specific classification result and the decision forest selectsthe classification based on the most votes obtained over all the trees in the forest.

The training of the RF is achieved relying on the result obtained from each decision tree. The training data is distributed randomly based on drawing N examples with a special kind of replacement in which the N is considered the original size of the training data. The learning method produces a classifier obtained from various trials and then the classifiers are gathered together to form the final classifier. In the classification stage, each classifier starts to record a vote for the class to which it belongs and the feature is drawn to the class with the highest votes.

EXPERIMENTAL RESULTS

The experimental result was reached using the proposed model based on two main classifiers which are SoftMax, and RF. The deep learning model was implemented using MATLAB software. An experiment is applied based on the proposed methodology for the diagnosis of four different ECG records. The whole experiment is performed on a computer with Intel (R) Core i7-8565U CPU of 1.99 GHz, 12 GB memory, and NVIDIA graphical card with GM 310M. The total number of records selected from 6 datasets for the four types of ECG heartbeats is 294 records. These records are collected as follows: 72 normal records (NSR) from the first two datasets (18 from nsrdb) and (54 from nsr2db), 74 supraventricular arrhythmias (SV) records from the third dataset (svdb), 74 records representing ST-segment changes from the fourth and the fifth datasets (28 from stdb) and (46 from ltsdb), and finally 74 myocardial infractions (MYC) records from the sixth dataset. The experiment was based on dividing the whole ECG records into three different parts training, validation, and test. This division made 177 records used for training and 57 records for validation and 60 for the test. The parameters of the training are adjusted properly to achieve the highest training performance and the lowest loss error.

Training Parameters Setting

The parameters of the 1D CNN model applied for the ECG diagnosis are determined in the Table 2. There exist various hyper-parameters that can be set before the training process. The selected parameters are the optimizer, mini-batch size, maximum epochs, and total number of iterations, regularization factor, and the validation frequency.

Table 2. Parameters adjusted for the proposed 1D CNN Model

t				
Optimizer	**Mini Batch Size**	**Maximum Epochs**	**Number of iterations**	**Validation Accuracy (%)**
Stochastic gradient descent Momentum (Sgdm)	8	100	2100	94.73
	16	100	1100	92.98
	32	100	500	89.47
	35	100	500	94.73
Adaptive Moment estimation (adam)	8	100	2100	92.98
	16	100	1100	94.73
	32	100	500	94.73
	35	100	500	91.22
Root mean square propagation (RMSprop)	8	100	2100	98.24
	16	100	1100	89.47
	32	100	500	91.22
	35	100	500	89.47

The optimal parameters selected for the 1D CNN model for the diagnosis of the ECG records are determined experimentally. The optimizers used for training the 1D CNN model are the stochastic gradient descent with momentum (SGDM), adaptive moment estimation (adam), and root mean square propagation (RMSprop). The mini-batch size parameter is applied with different values such as 8, 16, 32, and 35 on the three optimizers. The maximum epochs are 100 and the iterations vary relying on the size of the data and the mini-batch size values.

A validation data is input during the training process with a validation frequency equal to 30, and an L2 regularization factor is defined with a value equal to $1 \cdot 10^{-4}$. It can be seen that when the optimizer is set to rmsprop, and the mini-batch size is 8 the performance of the validation data has the highest accuracy. Therefore, the test data are passed to the model with the highest validation accuracy. In the training stage, the accuracy and the loss curves are obtained for each of the pre-trained models which is the 1D CNN model. Fig. 3 shows the highest performance achieved based on the validation data. The blue curve presents the training curve, whereas, the black dashed curve presents the validation accuracy curve during the training phase.

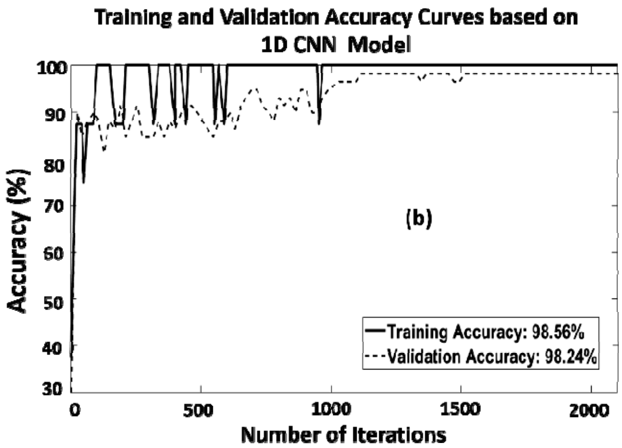


Fig. 3. Training and Vcalidation Accuracy curves performance of proposed 1D CNN Model

Classification Parameters

It is important to determine the hyper-parameters required before training the 1D CNN model. It is also essential to determine the parameters used on each classifier after applying 1D CNN model for feature extraction. As mentioned before, two main classifiers are used to determine the diagnosis performance of the 1D CNN features.

Table 3. Classifiers applied in the methodology and its optimal parameters

Classifiers	Optimal Parameters
Softmax	Loss function : "Cross Entropy (CE)"
Random Forest	Number of trees = 100 Max depth of each tree = 0 (zero indicates unlimited) Number of features = (log2(no.of.predictors)+1)

The first classifier is the Softmax classifier and its main parameter is the loss function which is defined by the cross-entropy. The next classifier is the random forest and it has a set of parameters such as the number of features extracted, number of trees, and the maximum depth of each tree. Finally, the last classifier applied is the XGBoost classifier and it also has several parameters. These parameters are chosen depending on the kind of classification that XGBoost will perform. In the case of multi-class classification (as in this study) the booster and the evaluation matrix must be defined by gbtree and mlogloss respectively. The rest of the XGBoost parameters are used based on their default values in the library of XGBoost. Table 3 shows the main parameters' values for the classifiers used after applying the 1D CNN model.

Classification Results

The features obtained from the fully connected layer of the 1D CNN model are forward for the two classifiers. The classifiers start to operate on the test data to ensure the performance of the validation accuracy obtained during the training. Table.4 shows various statistical performance measurements such as true positive rate (TPR), precision, false-positive rate (FPR), recall, receiver operating characteristic (ROC), Mathew's correlation coefficient (MCC), and precision-recall characteristic (PRC) value [22].

Table. 4. Classifiers performance using 1D CNN model based on different statistical measurements

Classifiers Performance	Performance Measurements (%)								
	TPR	FPR	Precision	Recall	F-Measure	Accuracy	MCC	ROC Value	PRC Value
Softmax	0.967	0.011	0.971	0.967	0.967	0.967	0.957	0.992	0.980
RF	0.983	0.006	0.984	0.983	0.983	98.33	0.978	0.997	0.993

These measurements are calculated for each classifier on the test data. In addition to this, the confusion matrix is manifested to determine the overall diagnosis performance on the classifiers. The confusion matrix is a figure or a table that is needed to describe the diagnosis performance of the tested data. It is a heat map in which the true value must be known. It gives the chance to visualize the performance of the two applied classifiers on the 1D CNN deep learning model as shown in Fig. 4 (*a* and *b*). It can be manifested that the RF classifier has the highest accuracy performance over other classifiers.

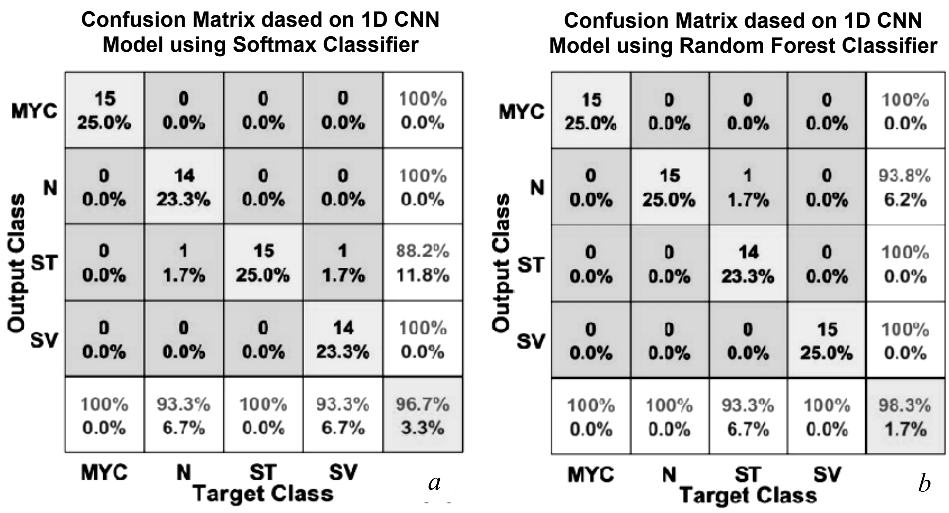


Fig. 4. Confusion matrices of Softmax (*a*) and RF Classifiers on the features obtained from 1D CNN Model (*b*)

DISCUSSION

The paper proposed a methodology for the diagnosis of four main types of ECG heartbeats. The methodology consists of four main phases, and these phases are obtaining ECG data, filtering ECG signals, extracting various ECG features, and classifying ECG records. In the phase of gathering ECG data, the ECG records are obtained from six various online ECG datasets. In the phase of signal filtration, the records are filtered using wavelets and a set of filters based on band stop, low pass, and smoothing filter to eliminate the main common noises in the ECG signals. In the phase of extracting features, the most discernment feature was obtained from a proposed 1D CNN model. Finally, the classification is applied based on two main classifiers and these classifiers are Softmax, and RF classifier. To overcome the chances of overfitting in the proposed model, a regularization factor is defined to shrink the learned estimates to zero. In other words, this regularization can tune the loss function by providing a penalty term to the optimizer of the 1D CNN model, and this will encourage smaller weights avoiding excessive changes of the coefficient. In addition to this, the number of ECG records in each of the four ECG classes are nearly equal leading to a balanced number of records in each category, dropping the probability of overfitting. Moreover, the ECG records are filtered using a pre-processing chain for reducing common noises that can cause overfitting during the training. Finally, the training accuracy obtained from the model with the highest accuracy validation is 99.5% and the value of the highest test accuracy is 98.3%, the slight difference between the training and the test accuracies shows that the model appropriately fits.

For comparison with others, several algorithms applied different methodologies for ECG diagnosis as shown in Table 5. S. Yu and M. Lee. [23], the authors approached an accuracy of 96.38% with the bispectrum feature set and SVM as the classifier, and when the authors added the genetic feature selector to the bispectrum and the SVM was used for classification, the accuracy increased to 98.10%. The number of records used was 54 and 29 from each of the normal sinus rhythm (NSR) and cognitive heart failure (CHF) data sets, respectively. K.H. Boon et al. [24] applied a diagnosis method to differentiate between normal and abnormal based on PAF. The features were produced from 106 ECG data collected from 53 ECG recordings. The SVM classifies based on 5 mins heart rate variability (HRV) segment and its distance from the PAF event. If it is at least 45 min distant from the event, the recording is called normal, but if the HVR segment goes before the event the recording is called abnormal. The accuracy achieved was 87.7%. Based on the improvement in the deep learning models in the diagnosis of the ECG heartbeats. H.B. Bae et al. [25] tried to classify normal NSR and abnormal ECG records such as AF, and ventricular fibrillation (VF) and they also focused on balancing the number of records used. The classification was based on Gamma distribution using probability output networks (CPON), and it proved that the performance was higher than KNN, SVM, aiming at an accuracy of 97.33%. R.R. Janghel et al. [26] aimed at building automated classification of regular and irregular ECG heartbeats. They applied their system on 47 records and the best results were achieved by using the decision tree, obtaining an accuracy of 88.2%.

Table 5. Proposed DL model compared to other previous work for ECG diagnosis

Authors	Records	Methodology	Classes	Databases	Performance
S.N. Yu et al. [23] 2012	54 R from NSR + 29 R from CHF	Features: Bisecpectrum + genetic feature set Classifier: SVM-	2	MIT-BIH NSR and CHF	Bisecptrum + SVM = 96.38% Bispecturm + genetic feature set + SVM = 98.10%
K.H. Boon et al. [24] 2018	106 data from 53 R pairs	Features: Time domain, spectral, Bispectrum, nonlinear dynamics features Classifiers: SVM	2	Atrial Fibrillation prediction (AFPDB) Database	ACC = 87.7%
H.B. Bae et al. [25] 2019	NSR: 15 R VF: 15 R AF: 15 R	R–R interval + (CPON)	3	MIT-BIH (NSRDB), (VFDB), (AFDB)	ACC = 97.33%
R.R. Janghel et al. [26] 2020	47 R 40% of the 47 R records are patients	Naïve Bayes SVM Ada-boost RF, Decision Tree, and KNN	2	MIT-BIH arrhythmia database	ACC of the Decision Tree = 88.2%
The proposed Method	294 R 177 R for train 117 R for validation and test	Proposed 1D CNN Model	4	6 main datasets	Softmax ACC =96.7% RF ACC = 98.3%

The proposed methodology worked on 294 recordings obtained from 4 different ECG heartbeats. The features are obtained from a 1D CNN model. Two main classifiers were applied to reach 96.7% using Softmax and 98.3% using RF classifier. The advantages of the proposed model are illustrated in three main points. The first point is the removal of the three common noises related to the ECG signals using a well-defined pre-processing chain. The second point is obtaining robust features from the 1D CNN model. The last point is the superiority of the XB-boost in the classification because it is highly flexible, can be paralleled, supports generalization, and is faster than gradient boosting.

CONCLUSIONS AND FUTURE WORK

In this study, a methodology is presented for the diagnosis of the four different types of ECG heartbeats based on a proposed 1D CNN model. The proposed methodology produces better results makes it adaptable for the diagnosis of different ECG records. The data were collected from 6 public available datasets. The ECG records were filtered to drifting in the ECG signals, powerline interference, and the high noise frequencies. The filtering chain is based on wavelets and a set of filters. Then, the ECG records are passed to a 1D CNN model for feature extraction. Finally, the classification is based on Softmax and RF, classifiers achieving an accuracy of 96.6% and 98.3%, respectively. ECG signals have future directions that can contribute and provide assistance in the field of medical informatics. There is a need for a real-time diagnosis application that can verify various types of heart diseases. In addition to this, it was discovered recently that the ECG signals can diagnose COVID patients based on the ECG image reports. It is recommended to develop diagnosis systems that can identify COVID patients from normal and various abnormal heartbeats. It is also suggested to use stratified k-fold cross-validation in future experiments to provide more information about the methodology performance. It is also advised to select the hyper-parameters based on various methods such as grid or random search or various metaheuristic techniques to reach the optimal values on the parameters for the proposed model. Finally, the XB-Boost classifier can be replaced with a sparse representation classifier as it is considered a powerful technique for pixel-wise classification of images [27].

REFERENCES

1. *World Health Organization, Cardiovascular Diseases* (*CVDs*). 2022. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs317/en/
2. S.K. Berkaya, A.K. Uysal, E.S. Gunal, S. Ergin, S. Gunal, and M.B. Gulmezoglu, "A survey on ECG analysis," *Biomed Signal Process Control*, 43, pp. 216–235, 2018.
3. O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, and U.R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Biomed. Comput. Meth. Prog. Bio*., 161, pp. 1–13, 2018.
4. L.A. Abdullah and M.S. Al-Ani, "CNN-LSTM based model for ECG arrhythmias and myocardial infarction classification," *Adv. Sci. Technol. Eng. Syst*., 5(5), pp. 601–606, 2020.
5. E. Butun, O. Yildirim, M. Talo, R.S. Tan, and U.R. Acharya, "1D-CADCapsNet: One dimensional deep capsule networks for coronary artery disease detection using ECG signals," *Physica Medica*, 70, pp. 39–48, 2020.
6. X. Hua et al., "A novel method for ECG signal classification via one-dimensional convolutional neural network," *Multimedia Systems*, pp. 1–13, 2020.
7. G. Petmezas et al., "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets," *Biomedical Signal Processing and Control*, 63, 102194, 2021.
8. *MIT-BIH Normal Sinus Rhythm*. [Online]. Available: https://www.physionet.org/ content/nsrdb/1.0.0/ last accessed 2-10-2021.
9. *Normal Sinus Rhythm RR Interval*. [Online]. Available: https://physionet.org/ content/nsr2db/1.0.0/ last accessed 2-10-2021.
10. *MIT-BIH Supraventricular Arrhythmia*. [Online]. Available: https://physionet.org/ content/svdb/1.0.0/ last accessed 2-10-2021.
11. *MIT-BIH ST Change Database*. [Online]. Available: https://physionet.org/content/ stdb/1.0.0/last accessed 2-10-2021.
12. *Long Term ST Database*. [Online]. Available: https://physionet.org/content/l tstdb/1.0.0/ last accessed 2-10-2021.
13. *PTB Diagnostic ECG Database*. [Online]. Available: https://www.physionet.org/ content/ptbdb/1.0.0/ last accessed 2-10-2021.
14. M.M. Bassiouni, E.S.A. El-Dahshan, W. Khalefa, and A.M. Salem, "Intelligent hybrid approaches for human ECG signals identification," *Signal Image Video Process*, 12(5), pp. 941–949, 2018.
15. M. Bassiouni, W. Khaleefa, E.A. El-Dahshan, and A.B.M Salem, "A machine learning technique for person identification using ECG signals," *Int. J. Appl. Phys.*, 1, pp. 37–41, 2016.
16. M.M. Bassiouni, I. Hegazy, N. Rizk, S.A. El-Dahshan, and A.M. Salem, "Combination of ECG and PPG Signals for Healthcare Applications: A Survey", *Advances in Modelling and Analysis*, 64(1-4), pp. 63–70, 2021.
17. M.M. Bassiouni, I. Hegazy, N. Rizk, E.S.A. El-Dahshan, and A.M. Salem, "Automated Detection of COVID-19 Using Deep Learning Approaches with Paper-Based ECG Reports," *Circuits, Systems, and Signal Processing*, 41, pp. 1–43,2022.
18. A.F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint*, arXiv:1803.08375, 2018.
19. A.T. Azar, H.I. Elshazly, A.E. Hassanien, and A.M. Elkorany, "A random forest classifier for lymph diseases," *Comput. Biol. Med*., 113(2), pp. 465–473, 2014.
20. S.Y. El-Bakry, E.S. El-Dahshan, and M.Y. El-Bakry, "Total cross section prediction of the collisions of positrons and electrons with alkali atoms using Gradient Tree Boosting," *Indian J. Phys*., 85(9), pp. 1405–1415, 2011.

21. H.T. Weldegebriel, H. Liu, A.U. Haq, E. Bugingo, and D. Zhang, "A new hybrid convolutional neural network and eXtreme gradient boosting classifier for recognizing handwritten Ethiopian characters," *IEEE Access*, 8, pp.17804–17818, 2019.
22. S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability," *Soft Comput*., 13(10), 959, 2009.
23. S.N. Yu and M.Y. Lee, "Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability," *Comput. Biol. Med*., 42(8), pp. 816–825, 2012.
24. K.H. Boon, M. Khalil-Hani, and M.B. Malarvili, "Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm," *Comput. Biol. Med*., 153, pp. 171–184, 2018.
25. H.B. Bae, M.S. Park, R.M. Kil, and H.Y. Youn, "Classifying heart conditions based on class probability output networks," *Neurocomputing*, 360, pp. 198–208, 2019.
26. R.R. Janghel and S.K. Pandey, "A Classification of ECG Arrhythmia Analysis Based on Performance Factors Using Machine Learning Approach," in *Computational Network Application Tools for Performance Management*. Springer, Singapore, 2020, pp. 65–74.
27. M.M. Bassiouni, I. Hegazy, N. Rizk, E.S.A. El-Dahshan, and A.M. Salem, "Deep learning approach based on transfer learning with different classifiers for ECG diagnosis," *International Journal of Intelligent Computing and Information Sciences*, 22(2), pp.44–62, 2022.

Received 23.09.2022

INFORMATION ON THE ARTICLE

Mahmoud M. Bassiouni, ORCID: 0000-0002-8617-8867, Egyptian E-Learning University (EELU), Egypt, e-mail: mbassiouni@eelu.edu.eg

Islam Hegazy, ORCID: 0000-0002-1572-463X, Ain Shams University, Egypt, e-mail: islheg@cis.asu.edu.eg

Nouhad Rizk, ORCID: 0000-0001-9277-9741, Houston University, USA, e-mail: njrizk@uh.ed

El-Sayed A. El-Dahshan, ORCID: 0000-0002-1221-0262, Ain Shams University, Egypt, e-mail: seldahshan@eelu.edu.eg

Abdelbadeeh M. Salem, ORCID: 0000-0003-0268-6539, Ain Shams University, Egypt, e-mail: absalam@cis.asu.edu.eg

1D МОДЕЛЬ CNN ДЛЯ ДІАГНОСТИКИ ЕКГ НА КІЛЬКОХ КЛАСИФІКАТОРАХ / М.М. Басіуні, І. Хегазі, Н. Різк, Е.С.А. Ел-Дашан, А.М. Салем

Анотація. Однією з основних причин смерті людини є захворювання серця. Виявлення серцевих захворювань на ранній стадії може запобігти серцевій недостатності або будь-якому пошкодженню серцевого м'яза. Одним з основних сигналів, які можуть бути корисними в діагностиці захворювань серця, є електрокардіограма (ЕКГ). Розглянуто діагностику чотирьох типів записів ЕКГ, таких як інфаркт міокарда (MYC), норма (N), відхилення сегмента ST (ST) і надшлуночкова аритмія (SV). Методологія збирає дані з шести основних наборів даних, а потім записи ЕКГ фільтруються за допомогою ланцюжка попереднього оброблення. Після цього запропонована модель 1D CNN використовується для вилучення ознак із записів ЕКГ. Потім застосовуються два різні класифікатори, щоб перевірити ефективність виділених ознак і отримати надійну точність діагностики. Два класифікатори – це softmax і класифікатор випадкового лісу (RF). Застосовується експеримент для діагностики чотирьох типів записів ЕКГ. Зрештою найвищої продуктивності досягнуто за допомогою радіочастотного класифікатора з точністю 98,3%. Порівняння з іншими суміжними роботами показало, що запропоновану методику можна застосовувати для раннього виявлення захворювань серця.

Ключові слова: електрокардіограма (ECG), безперервне вейвлет-перетворення (CWT), одновимірна модель згорткової нейронної мережі (CNN).

УДК 330.4
DOI: 10.20535/SRIT.2308-8893.2022.4.02

АНАЛІЗ ТА ПРОГНОЗУВАННЯ РІВНЯ СТАЛОГО РОЗВИТКУ В ЄВРОПЕЙСЬКОМУ КОНТЕКСТІ

І.О. ПИШНОГРАЄВ, І.О. ТКАЧЕНКО

Анотація. Висвітлено результати проведеного дослідження із прогнозування рівня сталого розвитку в європейському контексті. На підставі аналізу наукових здобутків вітчизняних та зарубіжних науковців визначено, що наявні методології мають ряд проблем, зумовлених використанням великої кількості показників, що унеможливлює швидке приблизне оцінювання нового об'єкта чи періоду. З огляду на це, дослідження спрямовано на побудову моделі розрахунку рівня сталого розвитку на основі обмеженого набору відкритих даних, що значно полегшить процес як його оцінювання, так і прогнозування. Базою дослідження є дані Світового центру даних з геоінформатики та сталого розвитку і проекту «Sustainable development index». Моделювання та аналіз виконано у застосунках MS Excel і RStudio. Отримані результати демонструють, що прогнозувати рівень сталого розвитку можна на основі моделі апроксимації, використовуючи обмежений набір індикаторів розвитку територій, що призведе до втрати мінімальної кількості інформації.

Ключові слова: сталий розвиток, європейський контекст, модель апроксимації, індикатори розвитку територій, прогнозування.

ВСТУП

Останні десятиріччя XX ст. досить гостро поставили перед людством проблему виживання та подальшого існування. Ця проблема детермінована складним поєднанням багатьох чинників, найвагомішими серед яких є значне вичерпання природних ресурсів, екологічна криза, несприятлива демографічна ситуація, голод і злидні в багатьох регіонах світу, безліч конфліктів у суспільстві, постійні війни з використанням засобів масового знищення людей, постійна небезпека міжнародного тероризму. У зв'язку з цим постала необхідність нової політики та стратегії, яка б дозволила вирішити ці глобальні проблеми сучасності, запобігти погіршенню якості навколишнього середовища, забезпечити не тільки теперішнє, але і майбутнє суспільство ресурсами, потрібними для задоволення його потреб [1]. Відповіддю на зазначені виклики стало розроблення концепції сталого розвитку, спрямованої на встановлення балансу між задоволенням сучасних потреб людства і захистом інтересів майбутніх поколінь.

З огляду на те, що сталий розвиток є керованим процесом, а основою його керованості є системний підхід та сучасні інформаційні технології, що уможливлюють моделювання різних варіантів розвитку та прогнозування їх результатів з метою обрання найбільш оптимального, розроблення дієвих моделей прогнозування сталого розвитку окремих територій набуває дедалі більшої актуальності, що сприятиме сталому розвитку загалом.

Питання щодо визначення сталого розвитку та оцінювання його рівня для країн світу досліджувалося багатьма вітчизняними та іноземними вченими. Так, різні тлумачення поняття «сталий розвиток» подано в наукових працях М. Згуровського [2], Б. Буркинського, В. Степанова, С. Харічкова [3], Р. Нуртдінова, А. Нуртдінова [4], Л. Скутару [5], О. Ханової, С. Скібіної [6] та А. Цвикилевич [7]. Проблемі оцінювання рівня сталого розвитку свої наукові доробки присвятили такі вчені: Джерфі Сакс, Крістіан Кроль, Гійом Лафортюн, Грейсон Фуллер та Фінн Вьольм [8, 9], Джейсон Хікель [10], а також фахівці Світового центру даних [11] .

МЕТОЛОДОГІЇ ВИЗНАЧЕННЯ РІВНЯ СТАЛОГО РОЗВИТКУ КРАЇН

Натепер ідея сталого розвитку є безальтернативною системою принципів і способів розроблення стратегій розвитку суспільства. Вона відіграє узагальнювальну роль, оскільки поєднує в собі основні положення більшості теорій суспільного розвитку, визначає цінності, пріоритети та стратегії розвитку сучасної цивілізації. Ідея сталого розвитку стала сьогодні парадигмою, особливим стилем наукового мислення та узагальнення буття, у межах якої розвивається суспільство. Вона є системою теоретичних, методологічних та аксіологічних установ, що поділяються багатьма науковцями та враховуються у вирішенні завдань суспільного розвитку [12].

Еволюція концепції сталого розвитку охоплює значний період часу та може бути поділена на три етапи: початковий (1968–1972), політичний (1972–1992) та етап соціально-економічних проблем (1992–дотепер) [13]. За весь цей час відбувалися виокремлення, розуміння та дослідження глобальних проблем та викликів, що загрожують теперішньому та майбутнім поколінням.

Визначення сталого розвитку розрізняються різною спрямованістю акцентів на найбільш пріоритетних проблемах: характері економічного зростання, якості життя, захисті навколишнього середовища, взаєминах розвинених країн і тих, що розвиваються, необхідності управління світовими процесами, раціональному використанні ресурсів, збалансованості сфер суспільного розвитку та ін [14]. Сталий розвиток визначено багатьма способами, але найчастіше цитується визначення з праці «Наше спільне майбутнє», також відомого як звіт Брундтланда: сталий розвиток – це розвиток, який відповідає потребам сьогодення без шкоди для майбутніх поколінь задовольняти свої власні потреби.

Ідея екологічно стійкого економічного зростання не нова. Багато культур протягом історії людства визнавали необхідність гармонії між навколишнім середовищем, суспільством та економікою. «Екологічно стійке економічне зростання» є синонімом поширеної концепції сталого розвитку. Його метою є досягнення балансу (гармонії) між екологічною, економічною та соціально-політичною стійкістю.

Мета сталого розвитку – збалансувати наші економічні, екологічні та соціальні потреби, забезпечуючи процвітання для сьогодення та майбутніх поколінь. Стійкий розвиток складається з довгострокового комплексного підходу до розвитку та досягнення здорової спільноти шляхом спільного вирішення економічних, екологічних та соціальних проблем, уникаючи надмірного споживання основних природних ресурсів [15].

Досягнення поставленої мети сталого розвитку неможливе без чіткого плану дій. Саме тому Організацією Об'єднаних Націй (ООН) у 2015 р. було прийнято Цілі сталого розвитку (ЦСР), також відомі як Глобальні цілі, як універсальний заклик до дій щодо припинення бідності, захисту планети та забезпечення миру та процвітання людства до 2030 р. [16].

Відстеження результату з досягнення кожної із сімнадцяти ЦСР відбувається різними способами. Наприклад, науковці Джерфі Сакс, Крістіан Кроль, Гійом Лафортюн, Грейсон Фуллер та Фінн Вьольм у працях [8, 9] аналізують рівень досягнення цілей, загрози та проблеми, які виникали в окремих країнах, а також можливі шляхи їх подолання. Так, в останньому звіті подано діаграму прогресу досягнення кожної ЦСР у відсотковому вимірі у 2021 р. порівняно з 2015 р. (рис. 1).

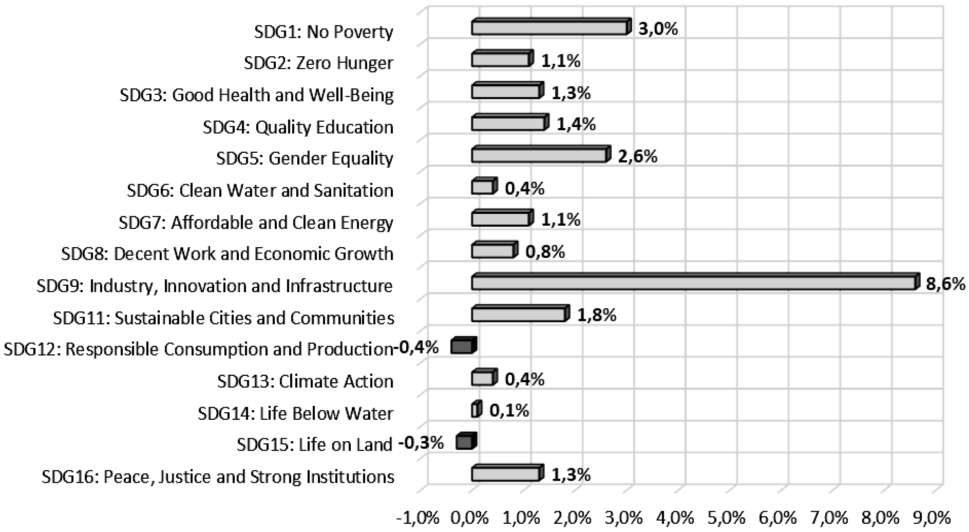


Рис. 1. Прогрес досягнення ЦСР у 2021 р. порівняно з 2015 р.
Джерело: побудовано авторами на основі [9].

Також для кожної країни вони визначають загальну оцінку досягнення ЦСР, надаючи однакову вагу кожній цілі. У результаті дана оцінка відображає позицію країни між найгіршим можливим результатом (0) і найкращим або цільовим результатом (100). На рис. 2 зображено розподіл країн відповідно до досягнення зазначених цілей сталого розвитку.

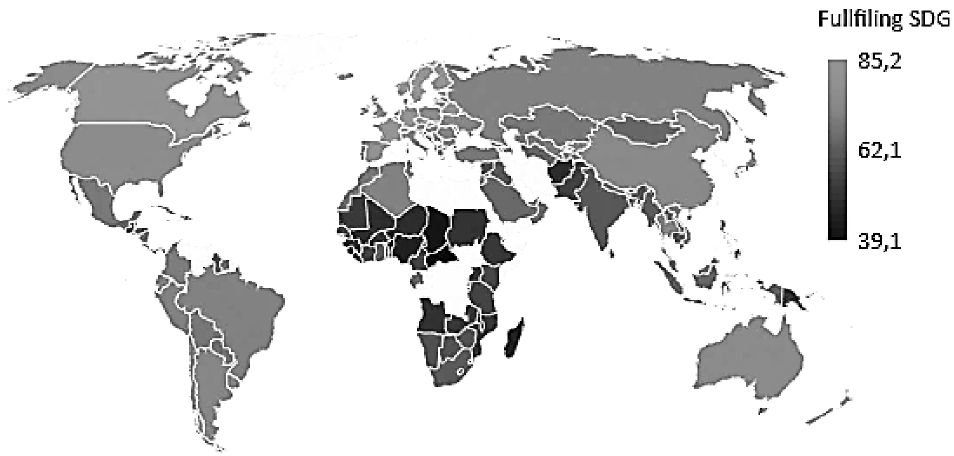


Рис. 2. Карта розподілу загальної оцінки досягнення ЦСР за країнами світу

Цей звіт про сталий розвиток надає всебічну оцінку того, наскільки країни близькі до досягнення цілей ЦСР на основі найновіших даних, доступних для всіх 193 держав-членів ООН. Цьогорічний звіт містить 91 глобальний індикатор, а також 30 додаткових показників для країн ОЕСР.

Аналізуючи звіт 2021 р., можна зробити висновок, що три скандинавські країни очолюють оцінку досягнення ЦСР (Фінляндія, Швеція та Данія). Усі країни в топ-20, крім Хорватії, є країнами організації економічного співробітництва та розвитку (ОЕСР), що можна простежити за табл. 1.

Таблиця 1. Топ-20 країн за оцінкою досягнення ЦСР у 2021 р.

Rank	Country	Score	Rank	Country	Score
1	Finland	85,9	11	Netherlands	81,6
2	Sweden	85,6	12	Czech Republic	81,4
3	Denmark	84,9	13	Ireland	81,0
4	Germany	82,5	14	Croatia	80,4
5	Belgium	82,2	15	Poland	80,2
6	Austria	82,1	16	Switzerland	80,1
7	Norway	82,0	17	United Kingdom	80,0
8	France	81,7	18	Japan	79,8
9	Slovenia	81,6	19	Slovak Republic	79,6
10	Estonia	81,6	20	Spain	79,5

Джерело: складено авторами на основі [*9*].

Проте навіть країни-члени ОЕСР стикаються зі значними проблемами у досягненні кількох ЦСР. Кожна країна ОЕСР характеризується червоною позначкою (головні проблеми залишаються) принаймні за однією ЦСР. Виходячи з наявних (до пандемії) траєкторій, прогрес у багатьох країнах з високим рівнем доходу був недостатнім у сферах сталого споживання та виробництва, кліматичних заходів та захисту біорізноманіття (ЦСР 12–15).

Країни з низьким рівнем доходу, як правило, мають нижчі показники індексу. Частково це зумовлено характером ЦСР, які значною мірою зосереджені на подоланні бідності та забезпеченні доступу всього населення до основних послуг та інфраструктури (ЦСР 1–9). Як правило, бідніші країни не мають належної інфраструктури та механізмів для вирішення ключових екологічних проблем, які вирішуються в ЦСР 12–15. Проте до початку пандемії більшість країн із низьким рівнем доходу – за винятком тих, у яких тривають збройні конфлікти чи громадянська війна – досягали прогресу в напрямі припинення крайньої бідності та забезпечення доступу до основних послуг та інфраструктури, особливо у зв’язку із ЦСР 3 (міцне здоров’я) та ЦСР 8 (гідна праця та економічне зростання). Однак у багатьох випадках пандемія COVID-19 призвела до зворотного прогресу в досягненні ЦСР.

Іншим варіантом відстеження прогресу досягнення ЦСР є модифікований індекс людського розвитку, який розроблений антропологом і доктором Джейсоном Хікелем [10]. Даний індекс сталого розвитку (SDI) вимірює екологічну ефективність людського розвитку, визнаючи, що розвиток має бути досягнуто в межах планетарних кордонів.

Розрахунок SDI визначається показником «людського розвитку» країни, який отриманий шляхом аналізу статистичних даних щодо очікуваної тривалості життя, стану здоров'я, освіти і доходів. Отриманий результат ділиться на два ключові показники: викиди CO_2 та екологічний матеріальний слід, які розраховані як на душу населення, так і в розрахунку перевищення природних меж Землі. Країни з високим рівнем людського розвитку та меншим негативним впливом на навколишнє середовище мають високу оцінку, відповідно низькою оцінкою відзначаються країни з невисокою тривалістю життя та низьким рівнем грамотності. Але особливим фактором, що впливає на низьку оцінку країн, є потужний негативний вплив на навколишнє середовище. У табл. 2 наведено ТОП-20 країн за розрахунком індексу сталого розвитку станом на 2019 р.

Таблиця 2. ТОП-20 країн за оцінкою індексу сталого розвитку на основі даних 2019 р.

Rank	Country	Score	Rank	Country	Score
1	Costa Rica	0,853	11	Algeria	0,803
2	Sri Lanka	0,843	12	Dominican Republic	0,802
3	Georgia	0,839	13	Colombia	0,801
4	Armenia	0,827	14	Azerbaijan	0,796
5	Albania	0,826	15	Fiji	0,788
6	Kerala (India)	0,825	16	Tunisia	0,786
7	Panama	0,821	17	Ecuador	0,783
8	Peru	0,818	18	Mexico	0,780
9	Cuba	0,814	19	Argentina	0,777
10	Moldova	0,808	20	Bolivia	0,773

Джерело: складено авторами на основі [17].

Карту, яка відображає індекс сталого розвитку для кожної країни станом на 2019 р., показано на рис. 3.

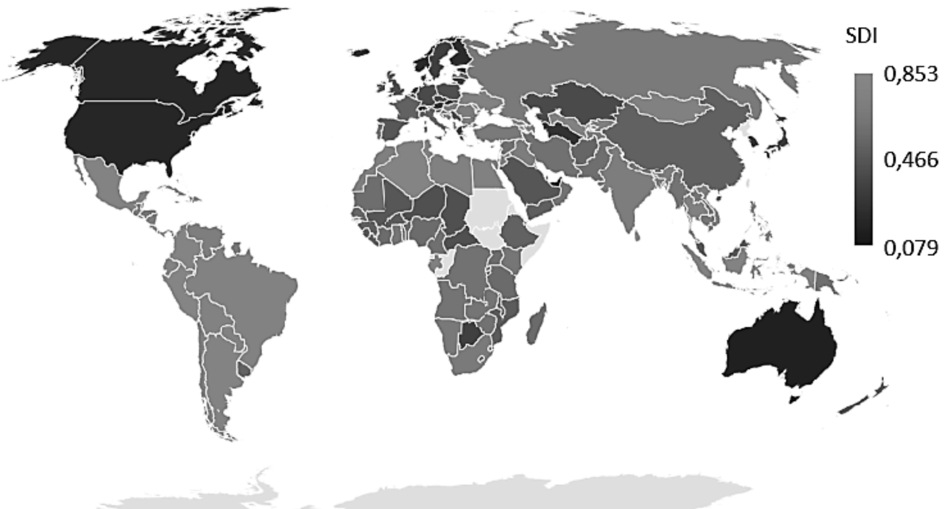


Рис. 3. Карта розподілу SDI за країнами світу
Джерело: побудовано авторами на основі [*17*]

Аналізуючи наявні дані, можна стверджувати, що деякі країни мають достатньо високі значення індексу, але вони не перетинають межу значення 0,9. Це свідчить про те, що ідеального показника за всіма складовими SDI не досягає жодна країна.

Методологія кількісної оцінки сталого розвитку країн, яку досліджували фахівці Світового центру даних (СЦД) з геоінформатики та сталого розвитку, ґрунтується на двох базових складових: компонентах якості (Cql) та безпеки (Csl) життя [11]. *Компонента якості життя* — інтегральна оцінка, яка враховує сумісно усі три виміри сталого розвитку, і, тим самим, відображає взаємозв'язок між трьома нероздільними сферами розвитку суспільства: економічною, екологічною та соціальною. Її структура включає 72 індикатори. Ступінь гармонізації сталого розвитку відображає баланс між його економічним, екологічним та соціально-інституціональним вимірами. *Компонента безпеки життя* описує сукупний вплив глобальних загроз на сталий розвиток різних країни та їх груп. На основі аналізу звітів міжнародних організацій СЦД виділяє такі загрози людства: глобальне зниження енергетичної безпеки, порушення балансу між біологічною можливістю землі та потребами людства, стрімке поширення інфекційних хвороб, наростання корупції, глобальне потепління та інші. Індекс вразливості країни до впливу сукупності глобальних загроз відображає ступінь наближення цієї країни одночасно до всіх загроз у просторі, який визначається нормою Мінковського.

Розподіл країн відповідно до значення індексу сталого розвитку показано на рис. 4. Рейтинг країн за індексом сталого розвитку, розрахованого за методикою СЦД, описує табл. 3.

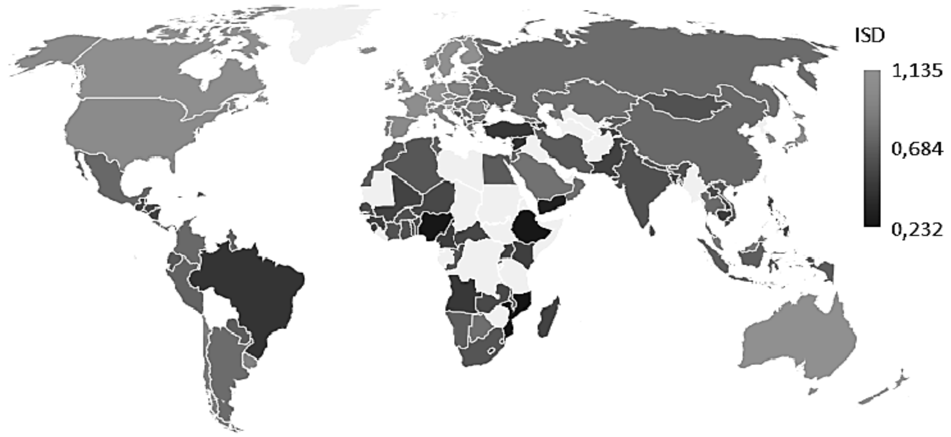


Рис. 4. Карта розподілу країн за значенням ISD
Джерело: побудовано авторами на основі [11].

Т а б л и ц я 3 . Топ-20 країн за оцінкою індексу сталого розвитку (методологія СЦД) на основі даних 2020 р.

Rank	Country	Score	Rank	Country	Score
1	Switzerland	1,135	11	Iceland	1,091
2	Germany	1,130	12	Austria	1,090
3	Denmark	1,128	13	United Kingdom	1,088
4	Australia	1,114	14	Luxembourg	1,081
5	Sweden	1,112	15	Ireland	1,073
6	Canada	1,108	16	Belgium	1,068

Продовження табл. 3

Rank	Country	Score	Rank	Country	Score
7	Japan	1,098	17	France	1,061
8	Norway	1,095	18	Netherlands	1,060
9	New Zealand	1,094	19	Singapore	1,047
10	Finland	1,093	20	Czech Republic	1,026

Джерело: складено авторами на основі [*11*].

Аналіз табл. 1 – 3 показав, що найвищі показники, розраховані за методологією проекту «Sustainable development index», мають зовсім інші країни, аніж за двома попередніми. Це пояснюються різним змістом, що вкладається в поняття «сталий розвиток». Також можна стверджувати, що за своєю сутністю та складовою оцінка досягнення ЦСР та індекс сталого розвитку, запропонований науковцями СЦД, подібні та системно підходять до оцінки всіх вимірів сталого розвитку. З урахуванням додаткового впливу сукупності глобальних загроз на рівень сталого розвитку в роботі розглянемо індекс сталого розвитку, розрахований за методологією [11, 17].

ПОБУДОВА АПРОКСИМУЮЧОЇ МОДЕЛІ ОЦІНЮВАННЯ РІВНЯ СТАЛОГО РОЗВИТКУ

Прогнозування рівня сталого розвитку країн Європи включатиме два етапи: прогнозування складових (метод Хольта) та прогнозування індексу сталого розвитку (за існуючою методологією та на основі регресійної моделі).

У межах дослідження прогнозування рівня сталого розвитку відбувалося для країн Європи, для яких характерною є розвинена та перехідна економіка. Усього було розглянуто 39 країн, які виокремлені в зазначені групи за типом економіки на основі офіційної класифікації ООН [18]. Для порівняння отриманих результатів також розраховано прогноз за існуючою методологією для вибірки з 3 країн (Україна, Фінляндія та Норвегія). Проведення дослідження було можливим завдяки використанню даних СЦД (за період 2005–2020 рр.) [1] та проекту SUSTAINABLE DEVELOPMENT INDEX (за період 1990–2019 рр.) [17]. Як допоміжні використовувалися дані про прогноз чисельності населення планети [19]. Прогнозування рівня сталого розвитку країн Європи розраховано на 5 періодів (2020–2024 рр.). При цьому прогноз рівня сталого розвитку за методологією [11] визначений шляхом апроксимації регресійною моделлю, що дозволило розглянути не всі індикатори розвитку, а лише кілька основних, а прогнозні значення індексу сталого розвитку за [17] обчислювалися за формулами методології. Моделювання та аналіз виконано у застосунках MS Excel та RStudio.

Побудова багатофакторної регресійної моделі передбачає вибір таких показників, залежність між якими могла б найбільш точно описувати рівень сталого розвитку. Для цього проведено кореляційний аналіз зв’язку між індексом сталого розвитку та багатьма показниками трьох вимірів людського життя (економічного, соціального та екологічного). У результаті відібрано 9 показників, для яких значення коефіцієнта кореляції більше за 0,7 (наявний сильний зв'язок між змінними). Інформацію про ці показники та коефіцієнти наведено в табл. 4.

Т а б л и ц я 4 . Обрані показники для моделювання рівня сталого розвитку

Номер показника	Indicators	Correlation
1	CP (corruption perception)	0,884
2	EF (economic freedom)	0,768
3	EPI (environmental performance index)	0,902
4	GPE (GDP per person employed)	0,822
5	IDI (ICT development index)	0,840
6	ITU (internet users)	0,830
7	LEX (life expectancy)	0,748
8	SRSC1 (research and development expenditure)	0,750
9	SRSC2 (researchers in R&D)	0,847

Джерело: складено авторами

Також прийнято рішення про поповнення переліку показників індексом яскравості нічних вогнів, оскільки проведене дослідження [20] показало наявність тісного зв'язку між цим індексом та показниками соціально-економічної складової сталого розвитку. Саме тому використання зазначених даних є доцільним для побудови багатофакторної регресійної моделі.

Таким чином, за допомогою програмної реалізації мовою R у застосунку RStudio на основі зазначених десяти показників отримано регресійну модель:

$$\mathrm{ISD} \sim \mathrm{CP} + \mathrm{EF} + \mathrm{EPI} + \mathrm{IDI} + \mathrm{GPE} + \mathrm{ITU} + \mathrm{LEX} + \mathrm{SRSC1} + \mathrm{SRSC2} + \mathrm{LIGHTS} - 1$$

Інформацію про модель зазначено в табл. 5.

Т а б л и ц я 5 . Звіт моделювання рівня сталого розвитку (модель 1)

Показник	Estimate	Std.Error	t value	Pr(>|t|)
CP	1,885e-03	2,215e-04	8,507	<2e-16(***)
EF	4,417e-03	3,213e-04	13,748	<2e-16(***)
EPI	6,093e-03	3,193e-04	19,084	<2e-16(***)
IDI	-7,676e-03	2,448e-03	-3,135	0,00181(**)
GPE	1,270e-07	6,737e-08	1,886	0,05988(.)
ITU	-1,829e-04	2,177e-04	-0,840	0,40113
LEX	1,501e-03	2,806e-04	5,351	1,30e-07(***)
SRSC1	2,671e-02	4,264e-03	6,264	7,76e-10(***)
SRSC2	-4,523e-06	2,078e-06	-2,177	0,02993(*)
LIGHTS	6,930e-04	1,123e-04	6,168	1,37e-9(***)
Signif. codes: 0 «***» 0,001 «**» 0,01 «*» 0,05 «.» 0.1 «»				
Multiple R-squared: 0,9987 **Adjusted R-squared**: 0,9986 **F-statistic**: 4,017e+04 on 10 and 533 DF **p-value**: <2,2e-16				

Джерело: складено авторами.

Проаналізувавши звіт першої моделі, можна дійти висновку, що показник кількості користувачів інтернету (ITU) не значущий, а отже, його можна вилучити з моделі. Найменш значущий показник ВВП, тому спробуємо його значення змінити відповідно до масштабу інших наявних даних. Для цього введемо в модель логарифм значень цього показника. Маємо модель такого вигляду:

$$ISD \sim CP + EF + EPI + IDI + LN(GPE) + LEX + SRSC1 + SRSC2 + LIGHTS - 1 .$$

Інформацію про результати моделювання наведено в табл. 6.

Таблиця 6. Звіт моделювання рівня сталого розвитку (модель 2)

Показник	**Estimate**	**Std.Error**	**t value**	**Pr(>|t|)**
CP	2,092e-03	1,922e-04	10,884	<2e-16(***)
EF	3,517e-03	3,337e-04	10,541	<2e-16(***)
EPI	6,043e-03	2,976e-04	20,304	<2e-16(***)
IDI	-7,129e-03	1,604e-03	-4,445	1,07e-05(***)
LN_GPE	2,926e-02	5,298e-03	5,522	5,23e-08(***)
LEX	-1,977e-03	6,807e-04	-2,904	0,00384(**)
SRSC1	2,783e-02	3,961e-03	7,028	6,42e-12(***)
SRSC2	-5,524e-06	1,995e-06	-2,769	0,00582(**)
LIGHTS	6,275e-04	1,100e-04	5,707	1,91e-08(***)
Signif. codes: 0 «***» 0,001 «**» 0,01 «*» 0,05 «.» 0.1 «»				
Multiple R-squared: 0,9987 **Adjusted R-squared**: 0,9987 **F-statistic**: 4,69e+04 on 10 and 534 DF **p-value**: <2,2e-16				

Джерело: складено авторами.

Наразі видно, що всі зазначені в моделі показники значущі, коефіцієнт детермінації дуже близький до 1 і становить 0,9987. Отже, відповідно до побудованої нами моделі зміна рівня сталого розвитку на 99,87% пояснюється зміною дев'яти показників різних вимірів життя людства. Саме тому ця модель цілком придатна для прогнозування.

АНАЛІЗ ПРОГНОЗІВ АПРОКСИМУЮЧИХ МОДЕЛЕЙ ОЦІНЮВАННЯ РІВНЯ СТАЛОГО РОЗВИТКУ

Наступним кроком обчислюємо рівень сталого розвитку країн Європи, використовуючи прогнозовані значення складових та побудовану регресійну модель. Результати прогнозування для всіх країн наведено в табл. 7.

Аналіз отриманих результатів є важливим етапом дослідження. Рівень сталого розвитку окремих країн (Україна, Норвегія та Фінляндія), а також прогнозовані значення до 2024 р. Зліва відображено результати моделювання на основі методології СЦД шляхом апроксимації (ISD), а справа – методології проекту SUSTAINABLE DEVELOPMENT INDEX (SDI).

Таблиця 7. Результати прогнозування рівня сталого розвитку на основі багатофакторної регресійної моделі

Country	2019	2020	2021	2022	2023	2024
Albania	0,737	0,761	0,762	0,763	0,764	0,765
Austria	1,091	1,121	1,122	1,122	1,122	1,123
Belarus	0,782	–	–	–	–	–
Belgium	1,058	1,069	1,068	1,067	1,066	1,065
Bosnia and Herzogovina	0,698	0,731	0,732	0,733	0,735	0,736
Bulgaria	0,865	0,842	0,841	0,841	0,840	0,840
Croatia	0,883	0,880	0,884	0,888	0,892	0,895
Cyprus	0,916	0,920	0,919	0,918	0,917	0,916
Czech Republic	1,023	1,005	1,005	1,005	1,004	1,004
Denmark	1,124	1,161	1,161	1,161	1,160	1,160
Estonia	1,017	1,023	1,021	1,020	1,019	1,018
Finland	1,113	1,146	1,147	1,148	1,149	1,150
France	1,040	1,051	1,050	1,050	1,049	1,048
Germany	1,127	1,114	1,113	1,112	1,110	1,109
Greece	0,882	0,891	0,891	0,890	0,890	0,890
Hungary	0,891	0,894	0,893	0,892	0,891	0,889
Iceland	1,082	1,073	1,072	1,070	1,068	1,067
Ireland	1,081	1,089	1,089	1,089	1,088	1,088
Italy	0,985	0,955	0,954	0,953	0,952	0,952
Kazakhstan	0,742	0,742	0,742	0,742	0,742	0,742
Latvia	0,902	0,939	0,938	0,938	0,937	0,937
Lithuania	0,960	0,961	0,961	0,961	0,960	0,960
Luxembourg	1,074	1,135	1,134	1,133	1,132	1,131
Macedonia	0,788	0,839	0,840	0,841	0,841	0,842
Moldova	0,682	0,672	0,671	0,670	0,669	0,668
Montenegro	0,782	0,752	0,753	0,754	0,755	0,756
Netherlands	1,114	1,102	1,101	1,100	1,100	1,099
Norway	1,098	1,115	1,116	1,117	1,119	1,120
Poland	0,947	0,925	0,925	0,925	0,925	0,925
Portugal	0,988	0,938	0,938	0,938	0,937	0,937
Romania	0,862	0,903	0,904	0,904	0,904	0,904
Serbia	0,795	0,803	0,804	0,806	0,807	0,809
Slovakia	0,925	0,918	0,918	0,918	0,917	0,917
Slovenia	0,998	0,985	0,984	0,984	0,984	0,984
Spain	1,000	0,986	0,986	0,985	0,985	0,985
Sweden	1,105	1,148	1,148	1,148	1,148	1,148
Switzerland	1,125	1,187	1,187	1,186	1,186	1,186
Ukraine	0,694	0,694	0,692	0,691	0,690	0,689
United Kingdom	1,092	1,129	1,129	1,129	1,128	1,128

Джерело: складено авторами.

Як видно з рис. 5, для України загальний рівень сталого розвитку за обома індексами досить схожий — 0,7–0,8. Для Норвегії та Фінляндії, які є розвиненими країнами, значення індексів дещо різняться, що пояснюється різною спрямованістю розуміння сталого розвитку науковцями. SDI більшою мірою має екологічне спрямування у той час, як ISD враховує усі три виміри (соціальний, економічний та екологічний) у рівних пропорціях.

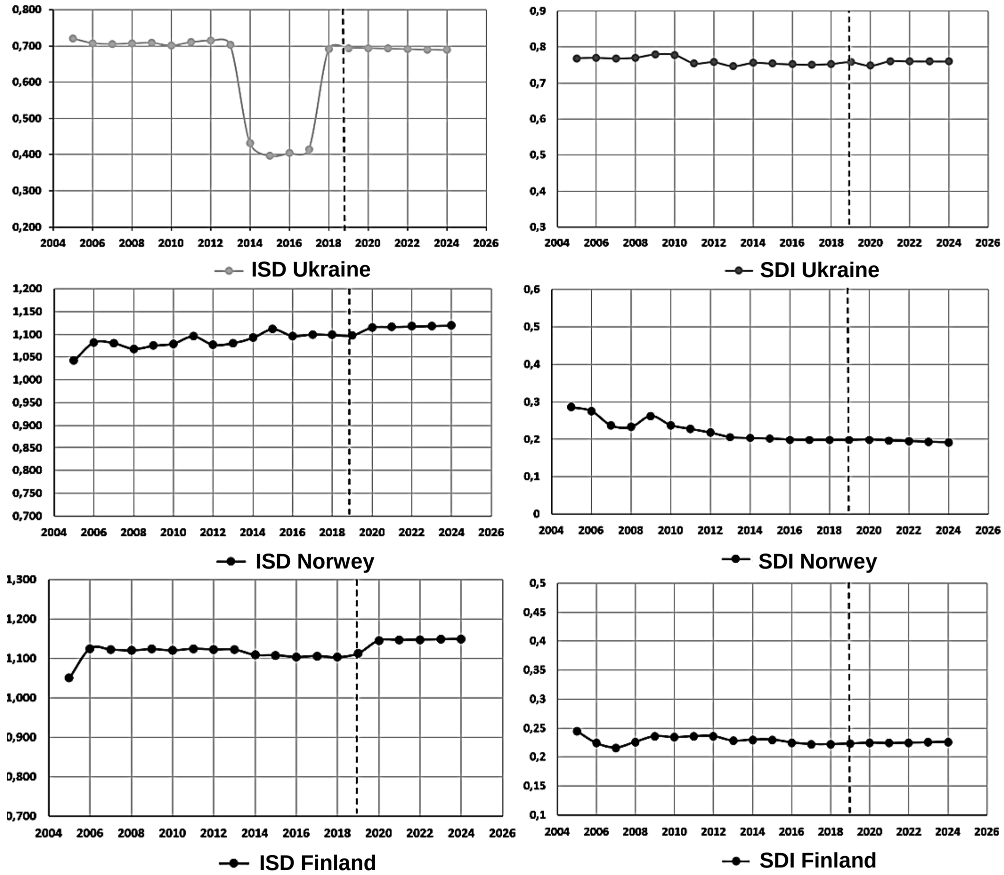


Рис. 5 Результати прогнозування рівня сталого розвитку для України, Норвегії та Фінляндії (з використанням апроксимації та існуючої методики)

Джерело: побудовано авторами.

Щодо прогнозних значень, то загалом вони відповідають установленій в останні роки тенденції. Так, наприклад, для Норвегії у 2020–2024 рр. значення ISD буде незначно зростати та оцінюватиметься на рівні 1,1–1,12, а SDI – зменшуватись з 0,2 до 0,19. Для Фінляндії характерним є помітне збільшення ISD у 2020 р. з 1,1 до 1,15 і надалі спостерігається стабільне значення – приблизно 1,15. Щодо SDI, то значення є стабільним на рівні 0,22–0,23 увесь період прогнозування.

Для України прогнозні значення рівня сталого розвитку ISD відповідають загальній тенденції 2005–2012 рр. і 2018–2019 рр. Значне зниження цього показника у 2013–2017 рр. спричинено воєнними діями на сході країни, але вже у 2018 р. показник сталого розвитку повернувся до звичного рівня, саме тому значення на період 2020–2024 рр. цілком описують загальну тенденцію. SDI для України у прогнозований період має стабільне значення 0,76, незважаючи на зниження показника до 0,75 у 2020 р. Звісно, прогноз для України не може враховувати воєнне вторгнення агресора на територію країни, що розпочалось 24 лютого 2022 р. Це в будь-якому випадку призведе до надвеликого зниження рівня сталого розвитку.

ВИСНОВКИ

У ході виконання дослідження розглянуто методології оцінювання рівня сталого розвитку країн, визначено їх основні відмінності.

З використанням кореляційно-регресійного аналізу та аналізу часових рядів побудовано апроксимуючу модель для визначення рівня сталого розвитку, що складається з 9 параметрів і забезпечує необхідний рівень достовірності. На основі запропонованої моделі виконано розрахунки прогнозних значень індексу сталого розвитку для 39 країн.

Проведений аналіз дає підстави стверджувати, що визначати рівень сталого розвитку можна на основі моделі апроксимації, використовуючи набагато меншу кількість індикаторів розвитку територій, що не призведе до значної втрати інформації.

ЛІТЕРАТУРА

1. O.O. Muzhylko, "The main origins of the idea of sustainable development and the path of Ukraine's transition to the principles of sustainable development," (in ukr.), *Derzhavne upravlinnya: udoskonalennya ta rozvytok*, no. 7, 2011. Available: http://www.dy.nayka.com.ua/?op=1&z=300
2. Michael Zgurovsky, "Impact of the information society on sustainable development: global and regional aspects," *Data Science Journal*, vol. 6, Supplement, pp. 137–145, 11 March 2007. Available: https://www.jstage.jst.go.jp/article/dsj/6/0/6_0_S137/_pdf
3. V.B. Burkinskij, V.N. Stepanov, and S.K. Harichkov, *Economic and environmental bases of regional nature management and development*, (in rus.). Odessa : Feniks, 2005, 575 p.
4. R.M. Nurtdinov and A.R. Nurtdinov, "From the theory of economic growth to the concept of sustainable development: issues of rethinking," (in rus.), *Vestnik Kazanskogo tehnologicheskogo universiteta,* 2012. Available: https://cyberleninka.ru/ article/n/ot-teorii-ekonomicheskogo-rosta-k-kontseptsii-ustoychivogo-razvitiya-voprosy-pereosmysleniya
5. L. Scutaru, "Economic Development Versus Sustainable Development," *Ecoforum*, vol 2, no. 1, 2013. Available: http://www.ecoforumjournal.ro/index.php/ eco/article/view/19
6. O.V. Khanova and S.O. Skibina, "Sustainable development of the EU countries: evaluation methodology and indicators," (in ukr.), *Problemy ekonomiky*, no. 3, pp. 20–32, 2017.
7. A.V. Tsvikilevich, *Improvement of management of municipal education development*, (in rus.). M.: Akademija estestvoznanija, 2006, 138 p.
8. "Sustainable Development Report 2019," *Sustainable Development Report*. Available: https://sdgindex.org/reports/sustainable-development-report-2019/
9. "Sustainable Development Report 2021," *Sustainable Development Report*. Available: https://www.sustainabledevelopment.report/reports/europe-sustainable-development-report-2021/
10. Jason Hickel, "The sustainable development index: Measuring the ecological efficiency of human development in the Anthropocene," *Ecological Economics*, vol. 167, 106331, January 2020. Available: https://www.sciencedirect.com/ science/article/abs/ pii/S0921800919303386
11. *Sustainable Development Analysis: Global and Regional Contexts. Part. 1. Global Analysis of Quality and Security of Life;* International Council for Science (ISC) and others; Scientific Supervisor of the Project M. Zgurovsky. K.: Igor Sikorsky KPI, 2019, 328 p. Available: http://wdc.org.ua/sites/default/files/SD2019-P1-FULL-EN.pdf
12. I.M. Koreneva, "Peculiarities and genesis of the idea of sustainable development as a methodology of modern education," (in ukr.), *Visnyk Cherkaskoho universytetu. Seriya*

"Pedahohichni nauk", issue 5, pp. 95–102, 2018. Available: https://ped-ejournal.cdu.edu.ua/article/view/2512/2655
13. K.Ju. Belousov, "The current stage of the evolution of the sustainable development concept and forming the paradigm of corporate sustainability," (in rus.), *Problemy sovremennoj jekonomiki*, 1 (45), 2013. Available: http://www.m-economy.ru/art.php?nArtId=4424
14. V.Ye. Khaustova and Sh.A. Omarov, "The concept of sustainable development as a paradigm for the development of the society," (in rus.), *Problemi ekonomiki*, 1 (35), pp. 265–273, 2018. Available: https://www.problecon.com/export_pdf/problems-of-economy-2018-1_0-pages-265_273.pdf
15. Puja Mondal, "What is the Importance of Sustainable Development?", *Your article library the next generation library.* Available: https://www.yourarticlelibrary.com/ environment/what-is-the-importance-of-sustainable-development/9910
16. *The SDGS in action. United nations Development programme*. Available: https://www.undp.org/sustainable-development-goals
17. *Sustainable Development index*. Available: https://www.sustainabledevelopmentindex.org/
18. *World Economic Situation and Prospects: statistical annex*. United Nations New York, 2020. Available: https://drive.google.com/drive/u/0/folders/1GlKLeIOeXaGMxgpLNUd3WO9XcUv-EGsd
19. *World population pyramid from 1950 to 2100*. Available: https://www.populationpyramid.net/ru/%D0%BC%D0%B8%D1%80-%D0%B7%D0%B5%D0%BC%D0%BB%D1%8F/2024/
20. I.O. Pyshnohrayev, I.O. Tkachenko, and S.V. Hapon, "Potential of night-time lights data to critical analysis of countries' socio-economic indicators," (in ukr.), *Ekonomichnyj visnyk NTUU «KPI»*, no. 19, pp. 54–62, 2021. Available: https://doi.org/10.20535/2307-5651.19.2021.240492

Received 11.07.2022

INFORMATION ON THE ARTICLE

Ivan O. Pyshnograiev, ORCID: 0000-0002-3346-8318, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: pyshnograiev@gmail.com.

Ivanna O. Tkachenko, ORCID: 0000-0002-2684-0126, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: kpivanna.tkachenko@gmail.com.

ANALYSIS AND FORECASTING THE LEVEL OF THE SUSTAINABLE DEVELOPMENT IN THE EUROPEAN CONTEXT / I.O. Pyshnograiev, I.O. Tkachenko

Abstract. The article highlights the results of the conducted research on forecasting the level of sustainable development in the European context. Based on the analysis of the scientific achievements of domestic and foreign scientists, it was determined that the existing methodologies had several problems associated with using a large number of indicators, which made it impossible to estimate a new object or period quickly. While considering this fact, the research aimed at constructing a model for calculating the level of sustainable development based on a limited set of open data, which would significantly facilitate the process of its assessment and forecasting. The basis of the research is data from the World Data Center for Geoinformatics and Sustainable Development and the "Sustainable development index" project. Modeling and analysis were carried out in MS Excel and RStudio applications. The obtained results demonstrate that it is possible to predict the level of sustainable development based on the approximation model using a limited set of territorial development indicators, which will lead to the loss of a minimal amount of information.

Keywords: sustainable development, European context, approximation model, territorial development indicators, forecasting.

UDC 519.925.51
DOI: 10.20535/SRIT.2308-8893.2022.4.03

COMPARATIVE ANALYSIS OF MODIFIED SEMI-SUPERVISED LEARNING ALGORITHMS ON A SMALL AMOUNT OF LABELED DATA

L.M. LYUBCHYK, K.S. YAMKOVYI

Abstract. The paper is devoted to improving semi-supervised clustering methods and comparing their accuracy and robustness. The proposed approach is based on expanding a clustering algorithm for using an available set of labels by replacing the distance function. Using the distance function considers not only spatial data but also available labels. Moreover, the proposed distance function could be adopted for working with ordinal variables as labels. An extended approach is also considered, based on a combination of unsupervised k-medoids methods, modified for using only labeled data during the medoids calculation step, supervised method of k nearest neighbor, and unsupervised k-means. The learning algorithm uses information about the nearest points and classes' centers of mass. The results demonstrate that even a small amount of labeled data allows us to use semi-supervised learning, and proposed modifications improve accuracy and algorithm performance, which was found during experiments.

Keywords: center of mass, clustering, distance function, medoids, nearest neighbor, semi-supervised learning.

INTRODUCTION

A large amount of data was produced recently, and nowadays humanity has the opportunity to store and process all this data. In all spheres of life, people try to use various data for optimizing business and life-improving using AI and data mining.

There are several approaches to data processing and analysis problems within the framework of machine learning (ML) paradigms. One of them is unsupervised learning when one tries to detect inner structure or patterns without human supervision. The most efficient approach in ML is supervised learning when we have some data with labels and try to learn a model function on data points as pairs of feature vectors and suitable labels. In many cases, there is no opportunity to label all data from different cases, causes are too complex and expensive experiments, data streaming with large frequency or just high cost of data labeling. Therefore, in this case, a satisfactory compromise is semi-supervised learning [1, 2], when we use datasets with a small amount of label that allows learning better its inner structure, which is illustrated by (Fig. 1).

Semi-supervised learning includes a variety of different approaches and can be used for any popular data analysis problems, such as clustering, anomaly detection, latent variables models, and many overs. In this paper, the object of the study is the process of the data points classifications, namely, identifying to which of a set of categories a new observation belongs to using a training set of data containing observations whose category membership is known for a piece of data. The purpose is to develop an improved combined semi-supervised method using already existing supervised and unsupervised algorithms and compare their accuracy and robustness.

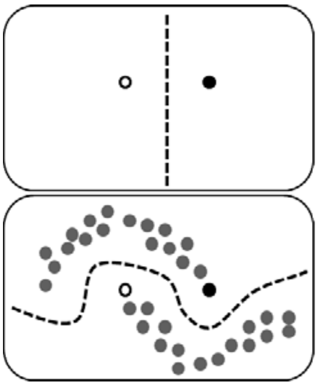


Fig. 1. Example of unlabeled data in semi-supervised learning (adapted from [3])

PROBLEM STATEMENT

Given a set of l labeled examples $\{\langle x_1, y_1\rangle, \ldots, \langle x_l, y_l\rangle\}$, where x_i – feature vector of i-th example and y_i – its label (class), $y_1, y_2, \ldots, y_l \in Y$, and a set of u unlabeled data $\{x_{l+1}, \ldots, x_{l+u}\}$ $x_1, x_2, \ldots, x_{l+u} \in X$. The goal is to determine some function using given sets that will give correct mapping of points from X to Y: $f(x_j) = y_j$ for any point from X.

REVIEW OF LITERATURE

The semi-supervised learning approach described in the literature is not so widely investigated as unsupervised or supervised, especially algorithms implementation. In [2] presented an overview of semi-supervised approaches that describe assumptions of semi-supervised learning especially: smoothness, low-density, and manifold.

In particular, the semi-supervised approach demonstrates high efficiency in solving clustering problems. The idea of the corresponding improvement of clustering algorithm was described in the review [4]. Majority of these methods are modifications of the popular *k*-means clustering method. As the base method chosen for improvement within the semi-supervised paradigm, the unsupervised *k*-medoids approach also known as PAM (Partitioning Around Medoid) algorithm, proposed in [5]. A medoid is a point in the cluster, whose average dissimilarities with all the other points in the cluster is minimum. *k*-medoid is a partitioning technique of clustering, which clusters the data set of n objects into *k* clusters, with the number *k* of clusters assumed known *a priori*. Both the *k*-means and *k*-medoids algorithms are partitional, which breaks the dataset up into groups, and both attempt to minimize the distance between points labeled to be in a cluster, and a point designated as the center of that cluster. In contrast to the *k*-means algorithm, *k*-medoids choose data points as centers and can be used with arbitrary distances, while in *k*-means the center of a cluster is the average between the points in the cluster (Fig. 2). Consequently, *k*-medoids are more robust to noise and outliers as compared to *k*-means.

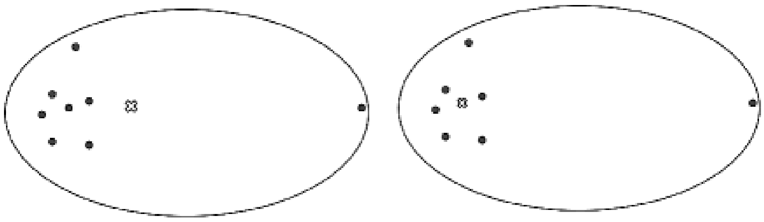


Fig. 2. Mean and medoid difference (adapted from [6]): *a* —mean; *b* — medoids

Another clustering method refined within the implementation of semi-supervised paradigm is DBSCAN – Density-Based Spatial Clustering of Applications with Noise proposed in [7]. The idea is to find core samples of high density and expand clusters from them. Such an approach is suitable for data that contains clusters of similar density. Based on a set of points, DBSCAN groups together points that are close to each other based on distance measurement, wherein it also marks as outliers the points that are in low-density regions.

A widespread clustering algorithm is also agglomerative clustering, which is the typical type of hierarchical clustering used to group objects in clusters based on their distance to each other. The algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects. The result is a tree-based representation of the objects – dendrogram (Fig. 3) [8].

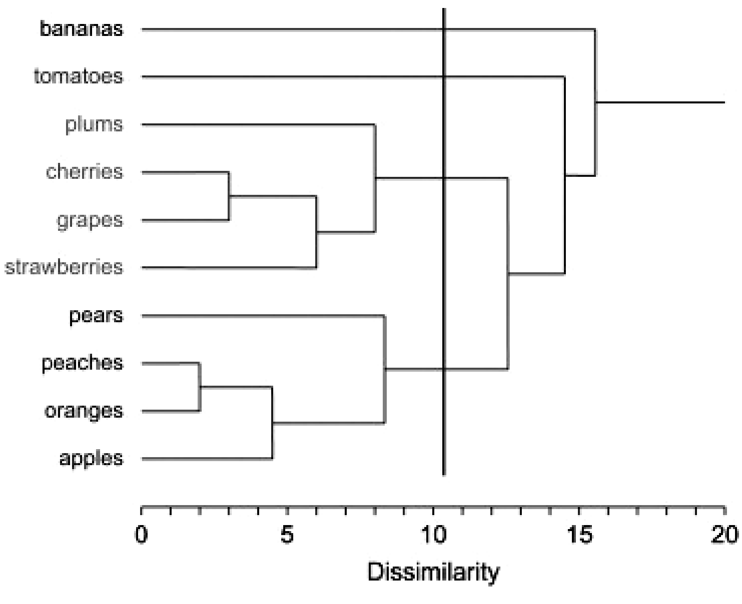


Fig. 3. Dendrogram of hierarchical clustering (adapted from [9])

The supervised approach for clustering problem is described in [10]. The nearest neighbor decision rule assigns to an unclassified sample point the classification of the nearest of a set of previously classified points. Thus, for any number of categories, the probability of error of the nearest neighbor rule is bounded above by twice the Bayes probability of error. In this sense, it may be said that half the classification information in an infinite sample set is contained in the nearest neighbor.

MATERIALS AND METHODS

Distance function extension

As was shown above the major clustering methods form clusters based only on distance function. So we made the assumption that feature space can be extended by additional dimensions with information about available labels. We develop multiple distance functions that take to account that label dimension. The proposed approach allows concentrating attention on distance function creation and the use of already implemented and optimized clustering algorithms.

As a base distance metric, we use Euclidean distance. If there is additional data from the label space, it is advisable to use this information. An example is a naive solution - to reduce the distance between points if they have the same labels and increase in the opposite case:

$$D_{labeled}(p,q) = (1 + W * S(p,q)) * d(p,q), \quad (1)$$

where W — weight coefficient, $W \in [0,1]$; $S(p,q) = \{-1,$ if $label_p = label_q$ 1, if $label_p \neq label_q$ and 0, otherwise $d((p,q) =$ euclidean distance :

$$\sqrt{\sum_{k=1}^{n} (p_k - q_k)^2}\ .$$

In (1) weight coefficient W is used for tuning influence of labels: 0 — has no influence, ignoring label information; 1 — the distance between points with the same label equal to zero.

The suggestions concerning distance function not only decrease the distance between points with the same labels but also increase if points have different labels and improve robustness in cases with noised data and close clusters.

In real cases, data often occurs with labels as ordinal variables wherein the labels should be number type (0, 1, 2…). In this case, we can also use the distance between labels, because rank 1 is closer to rank 2 than rank 3 (for example, “cat” is closer to “dog” than to “fish”) [11]. However, it required additional data analysis before clustering.

Considering the idea above, one can expand the (1) with the distance between labesls:

$$D_{labeled}(p,q) = (1 + K * S(p,q) * * \left|label_p - label_q\right|) * d(p,q)\ .$$

The methods described above are intuitively understood and easy to implement, but have one con:

– labeled and unlabeled data have the same influence on cluster formation, while the labeled point should have more influence;

– only points with labels are considered and do not take neighborhood points without labels, but in most cases, the neighborhood has the same class.

Semi-supervised *k*-medoids algorithm

We will propose some improved techniques that can resolve these issues and use the *k*-medoid approach as a base idea. However, unlike *k*-medoids the proposed algorithm first calculates medoids using only labeled data and next processes unlabeled classes – assign labels of nearest medoid. This approach is described by Algorithm 1.

This algorithm has the following pros:

– reduced processing time, because required only multiple iteration throw points unlike standard *k*-medoid;

– more robustness to wrong assigned labels, because the algorithm gives higher weights to labeled data in the medoids calculation step.

Algorithm 1. Modified *k*-medoids algorithm

Input:

X — feature matrix $n*m$, n — number of objects, m — number of features

y — labels vector of length n, $y[i] = -1$ if no label data for i-th object

Output:

y_predicted – vector of length n with object labels

```
k ← number of clusters, e.g. number of unique labels in y
X_l ←labeled point from X
X_u ←unlabeled point from X
select k random points out of the X_l as the medoids
associate each data point to the closest medoid
while the cost of the configuration decreases:
    for each medoid m, and for each non-medoid data point o from X_l:
        Consider the swap of m and o, and compute the cost change
        If the cost change is the current best, remember this m and o combination
    associate each point from X_u with the nearest medoid
    for each point o in X:
        fill y_predicted with assigned medoid of point o
    return y_predicted
```

Semi-supervised *k*-nearest neighbors algorithm

Another proposed approach uses the idea of *k*-nearest neighbors and the *k*-mean algorithm, because for classifying we use both information about the nearest points and classes centers of mass. As a distance metric was used Euclidean distance but any metric could be used.

Classes' centers do not recalculate after each assignment, because experiments show that it does not bring benefits but takes more computation time.

Algorithm 2 implements the proposed approach.

Algorithm 2. Object clustering using *k*-NN based approach

Input:

X – feature matrix $n*m$, n – number of objects, m – number of features

y – labels vector of length n, $y[i] = -1$ if no label data for i-th object

```
    K – number of nearest points
    C – the weight of the nearest class center
Output:
    y_predicted – vector of length n with object labels
y_predicted ← empty list of length n
unlabeled_idxs ← list of indexes where y = -1
labeled_idxs ←list of indexes where y > -1
center_coordinates ← list of center coordinates for each class, calculated using
available labels
random shuffle unlabeld_idxs
for i in unlabeld_idxs do
    distances_i ← distances from i-th object to each object with indexes in la-
    beled_idxs
    argsort distances_i
    nearest_idxs ← indexes of first K elements from distances_i
    classes_dist_i ← distance from i-th object to each classes' center
    neares_class_idx ← index of nearest class to i-th object
    cls_counts ← list, where j-th element denote numbers of points belonging to j-th
    class among nearest_idxs
    cls_counts[neares_class_idx] ← cls_counts[neares_class_idx] + C // add addi-
    tional value for class with nearest center
    label ← argmax(cls_counts)
    y_predicted[i] ← label
end for
for i in labeled_idxs do
    y_predicted[i] ← y[i]
end for
return y_predicted
```

So, the method described above allows:

– consider information about the nearest point, because in most cases point has the same label as its neighbors;

– combine a different kind of information;

– tune the weight of different sources using input parameters.

EXPERIMENTS

For experiments purpose was generated synthetic multiple datasets using sklearn library. Each dataset contains 250 points in 2D space. Available only 10% of labels as default. In addition, datasets have multiple clusters with different distributions and shapes (Fig. 4).

We will compare different approaches to find the average accuracy score on all these datasets for each approach with different combinations of base clustering methods and distance functions. Table included combinations that have improved compared to the base clustering method.

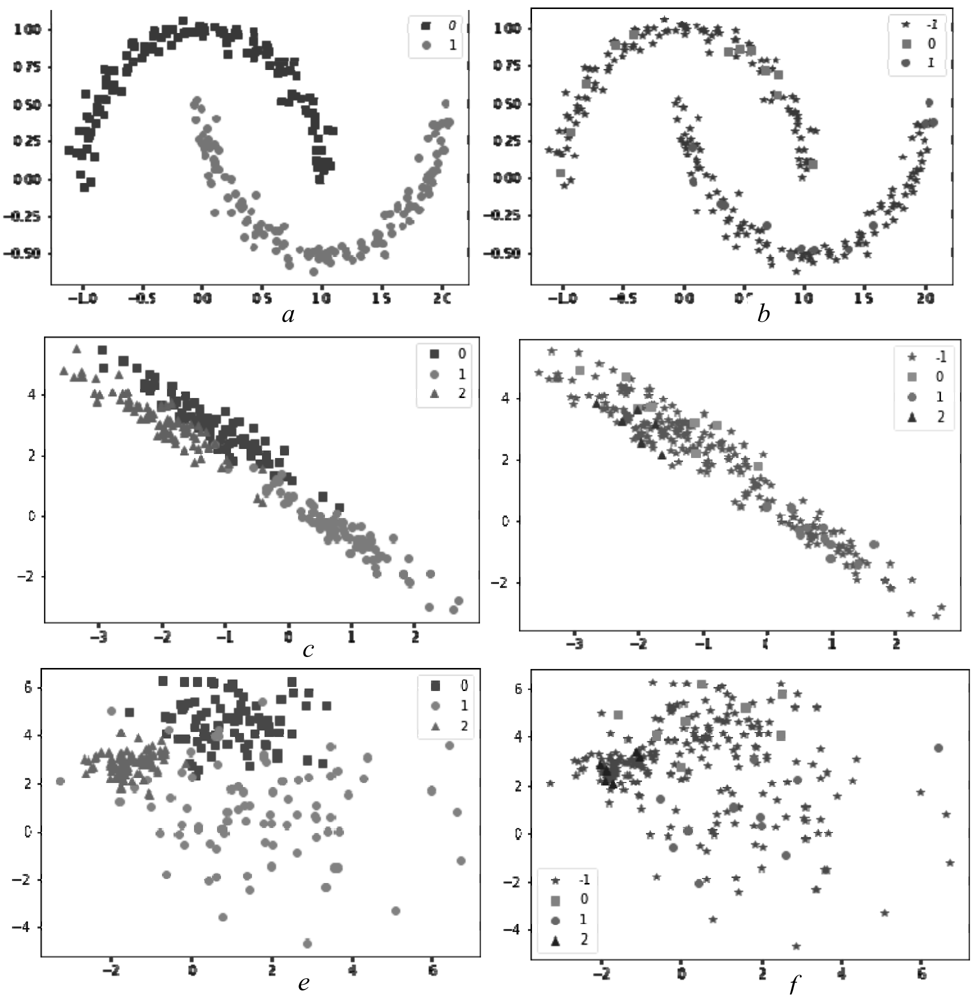


Fig. 4. Datasets visualization. The legend shows classes' labels, -1 – unlabeled point; *a, b* – Moons dataset, 2 classes, with non-convex and separable shapes; *c, d* – Aniso dataset, 3 classes, convex shape with same class variation, not separable; *e, f* – Varied dataset, 3 classes with a convex shape and different class variation, also not separable

Accuracy comparison

Method name	**Dataset name**			**Avg accuracy**
	Moons	Aniso	Varied	
Agglomerative + custom distance with ordinal variables ($W = 0.8$)	1.000	0.824	0.888	0.904
DBSCAN + custom distance ($W = 1.0$)	1.000	0.488	0.360	0.616
K-Medoids based	0.86	0.864	0.904	0.876
k-NN based ($N = 5$, $C = 2$)	0.904	0.900	0.912	0.905

The results shown in Table show that the best-unsupervised method is *k*-medoid and the *k*-NN based algorithm has higher average accuracy.

Fig. 5 illustrates the difference between unsupervised and semi-supervised methods, which is especially pronounced for non-convex data localization areas and for clusters with the same variation and located nearest to each other.

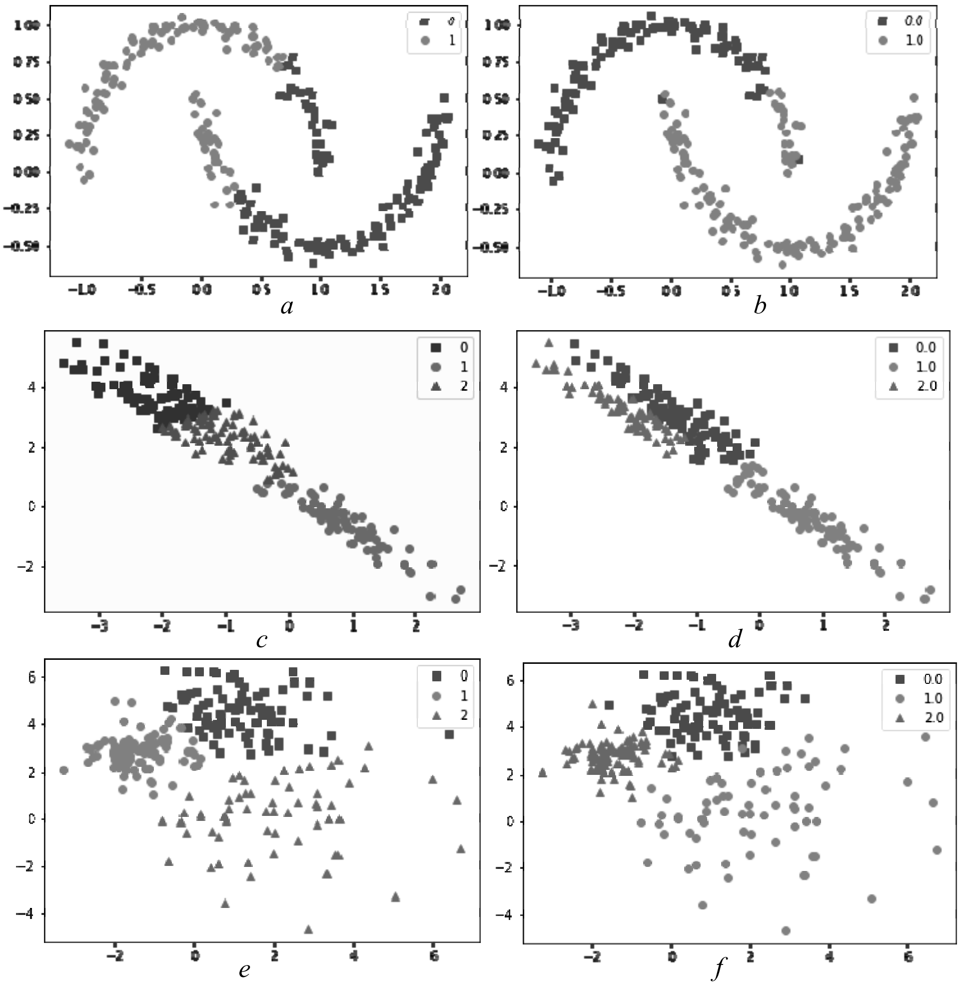


Fig. 5. Predicted labels visualization. *a*, *c*, *e* — unsupervised *k*-medoids, *b*, *d*, *f* — semi-supervised *k*-NN based method

Another required feature of a semi-supervised algorithm is quality versus a umber of labels dependency: more labels – higher quality and vice versa. However, Fig. 6 shows that clustering methods with custom distance functions do not have this feature. Therefore, this approach can be easy and fast, because it requires implementation only of the distance function. However, on the other hand, it is necessary to develop and tune the distance function for each case with a different number of available labels.

DISCUSSIONS

In Fig. 6 we can see that with the percentage of available labels increasing the accuracy of *k*-NN based and *k*-medoids based algorithms increased too. In addition, these algorithms have high accuracy according to Table. At that time, DBSCAN and Agglomerative methods did not respond to increasing labels. It means that we need to develop and tune the distance function for each case with a different number of available labels.

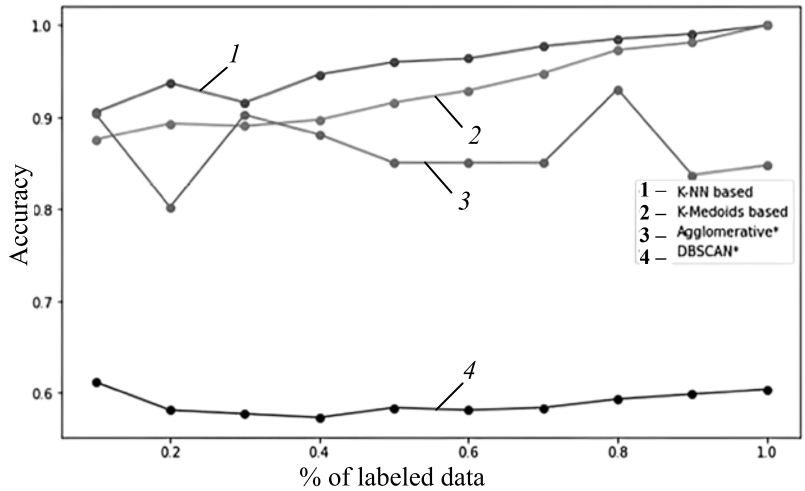


Fig. 6. Accuracy versus the quantity of labeled data comparison plot

CONCLUSIONS

In this study, we had shown that even small amounts of labeled data allow the use of semi-supervised learning and improve accuracy. At that, semi-supervised learning can improve algorithm performance too. Multiple approaches to semi-supervised learning were proposed, one of them is using a distance metric that considers available label information.

Further development of this work was a modification of other methods of classification and clustering and a deeper study of the influence of the distance function on the accuracy of clustering.

REFERENCES

1. L. Lyubchyk, A. Galuza, and G. Grinberg, "Semi-supervised Learning to Rank with Nonlinear Preference Model," *Recent Developments in Fuzzy Logic and Fuzzy Sets Studies in Fuzziness and Soft Computing*, pp. 81–103, 2020.
2. J.E.V. Engelen and H.H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2019.
3. Wikipedia contributors, "Semi-supervised learning", in *Wikipedia, The Free Encyclopedia*. [Online]. Available: https://en.wikipedia.org/wiki/Semi-supervised_ learning
4. E. Bair, "Semi-supervised clustering methods," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 5, pp. 349–361, 2013.
5. A.S. Hadi, L. Kaufman, and P.J. Rousseeuw, "Finding Groups in Data: An Introduction to Cluster Analysis," *Technometrics*, vol. 34, no. 1, pp. 111, 1992.
6. X. Jin and J. Han, "*K*-Medoids Clustering," in *Encyclopedia of Machine Learning*. Boston, MA: Springer, 2011
7. M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second InternationalConference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, USA*, E. Simoudis, J. Han, and U.M. Fayyad, Eds. AAAI Press, 1996, pp. 226–231. [Online]. Available: http://www.aaai.org/Library/KDD/1996/kdd96-037.php
8. Daniel Müllner, *Modern hierarchical, agglomerative clustering algorithms*. [Online]. Available: https://arxiv.org/pdf/1109.2378.pdf

9. T. Tullis and A. Bill, *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Elsevier/Morgan Kaufmann, 2013
10. T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
11. L. Lyubchyk, G. Grinberg, and K. Yamkovyi, "Integral Indicator for Complex System Building Based on Semi-Supervised Learning," *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, 2018.

Received 09.12.2021

INFORMATION ON THE ARTICLE

Leonid M. Lyubchyk, National Technical University «Kharkiv Polytechnic Institute», Ukraine, e-mail: Leonid.Liubchyk@kpi.edu.ua

Klym S. Yamkovyi, National Technical University «Kharkiv Polytechnic Institute», Ukraine, e-mail: klym.yamkovyi@cs.khpi.edu.ua

ПОРІВНЯЛЬНИЙ АНАЛІЗ МОДИФІКОВАНИХ АЛГОРИТМІВ НАВЧАННЯ З ЧАСТКОВИМ ЗАЛУЧЕННЯМ УЧИТЕЛЯ НА МАЛІЙ КІЛЬКОСТІ РОЗМІЧЕНИХ ДАНИХ / Л.М. Любчик, К.С. Ямковий

Анотація. Присвячено вдосконаленню методів кластеризації з частковим підкріпленням, а також порівнянню їх точності та стійкості. Запропонований підхід заснований на розширенні алгоритмів кластеризації шляхом використання доступного набору міток класів за допомогою заміни функції відстані, при цьому за використання запропонованої функції відстані враховуються не тільки просторові дані, але й мітки. Більше того, запропонована функція відстані може бути адаптована для роботи з порядковими змінними як мітки. Також запропоновано підхід, заснований на методі навчання без вчителя *k*-медоїдів, модифікований для використання лише розмічених даних на етапі обчислення медоїдів кластерів, комбінацію методу навчання з учителем *k* найближчих сусідів та без вчителя – *k*-середніх. При цьому алгоритм навчання використовує інформацію як про найближчі точки, так і про центри мас класів. Отримані результати демонструють, що навіть невеликий обсяг помічених даних дає змогу використовувати навчання з частковим підкріпленням, а запропоновані модифікації забезпечують підвищення точності і стійкості алгоритму, що продемонстровано під час експериментів.

Ключові слова: центр мас, кластеризація, функція відстані, найближчий сусід, навчання з частковим залученням вчителя, медоід.

UDC 629.7.05
DOI: 10.20535/SRIT.2308-8893.2022.4.04

METHODOLOGICAL ASPECTS OF OPERATIVE CONTROL SYSTEM INTELLECTUALIZATION FOR DYNAMIC OBJECTS

S.V. MELNYKOV, P.M. MALEZHYK, A.S. GASANOV, P.I. BIDYUK

Abstract. The problems of intelligent control system organization are considered: determining the number of intellectualization levels, the sequence of actions required for analysis of the control process, adding to the control system new elements providing for enhancement degree of its intellectualization, special features of its structural organization, estimating the possibilities of intellectualization, providing examples of practical intellectualization. The primary purpose of the study is to determine the purposeful organization of intelligent control systems as well as the necessity and usefulness of systemic consideration that takes into consideration the following: requirements of the problem statement, characteristics of the environment, means for acquiring and processing necessary information, working control mechanisms, functional characteristics and experience of user-operator. As a result of the analysis performed, characteristic levels of the intellectual development of a system were determined, the stages of performing intellectualization of a control system were proposed, and the effectiveness of proposed solutions for practical problems was shown.

Keywords: control system, intellectualization, organization of control systems, intelligent control, man-machine systems.

INTRODUCTION

The urgency of the problem of control system intellectualization is determined by the modern requirements to development and implementation new control systems for dynamic processes and systems that could use large volumes of informational data and are capable to perform intellectual processing of the data to reach high quality control in real time. Dynamic modern development of the computational means and particularly the means for parallel data processing like graphical and tensor processors as well as new intellectual data processing methods create favorable conditions for reaching high quality control. Together with the possibilities mentioned the requirements regarding data optimization are preserved aiming to determining the confidence intervals and characteristic numbers.

The complex approach to taking into consideration the purposeful functioning of a system, to determining dimensionality of complex systems characteristics and control environment, to description of functioning machine operator provides creating conditions in the direction mentioned.

The trend to automatizing of all sides of human activities (transportation, industrial production, agriculture, economy, home activities, social sphere etc.) require development of high-tech systems with intellectual functions, capable to interact with distributed data systems and use remote services.

The paper contains methodological aspects and examples of organizing intellectual control systems for complex dynamic objects and processes. It is also proposed the methodology for organizing control systems with functional structures that implement some separate characteristics of intellect.

Intellectual technology, intellectual control, intellectualization – all these notions are mostly accepted intuitively. When transition occurs from some purposeful functioning of a system to the process that exhibits intellectual features, then different subjective opinions come to being that gradually transform into the processes of search for high quality solutions and more complicated tasks.

The considerations proposed in the study are linked to emergence and development of natural evolutional intellect but the possibilities are preserved for hybrid combinations with other approaches. The intellectual level of a control system is linked to effectiveness of its functioning but there is no one-to-one correspondence. The general estimate of effectiveness of the system created and the measure of its intellectualization is rather relative and can be determined via investigating of its functioning using various methods in natural conditions or via computer simulations. The qualitative estimation of control system intellectuality supposes availability of some definite structure as well as adaptation of the structure and model parameters, memorizing control actions and their results, the possibilities for constructing models of environment and the system under study etc. The quantitative estimation of intellectuality can be expressed by the number of micro-situations in which within limited time can be found solution for stated problem. Time is a critical factor that influences general effectiveness of practical functioning of a system being studied. The effectiveness of a system is often considered as necessary initial but not sufficient condition of intellectuality. When strict conditions on the decision making time exist the effectiveness of general functioning of a control system that has lower intellectuality level may become higher. Certainly here we have to higher numerical effectiveness criterion.

PROBLEM STATEMENT

The use of conforming to the laws of evolutionary development biological systems for constructing engineering systems is carried out during all history of mankind. Development and improvement of intellectual functions of biological system stands in the same position, i.e. this is not exception. Here it is proposed to use basic natural conforming to the laws intellectual developments for creating technical systems though we don't remove the possibilities of using effective engineering solutions in hybrid systems. Development of such engineering systems requires paying special attention to selecting and processing only necessary information for solving specific goal problems and further usage of the data received.

One of key moments of carrying out analysis while solving complex practical problems and touching upon determining characteristics related to confidence interval and control of dynamic objects is taking into consideration and coordination of used environment characteristics and the controlled object with the goal requirements and possibilities of receiving information from gauges that is limited in its precision in noisy conditions as well as taking into consideration discrete time intervals for making control decisions. As an example, the problem is considered to control flying object in such a way to lead it into given limited space from the space of initial states. Taking into consideration of such requirements provides the possibility for carrying out micro-situational analysis and find operative control, to determine the areas of reachability for the goal values of variables and parameters for various control algorithms, and make conclusions regarding effectiveness of the usage for the algorithms applied.

Today there exist a large number of studies directed to description of intellectual features of biological systems and their implementation in technical systems [1, 2]. Of particular popularity and success today has the use of analogs for neural nets [3]. Practical implementation of intellectual systems is performed in the most diverse spheres of human activities; it is especially topical and substantiated problem for the unmanned transportation systems [4]. The problem of practical implementation of the intellectual systems is mainly related to the high dimension of input data, availability of noisy components, uncertainties in internal and external parameters of control processes as well as in absence of effective basic control solutions and structures that could be modernized. The systematic considerations of forming intellectual features of biological organisms helps to overcome the difficulties mentioned in the practice of creating and implementing intellectual technical systems.

Using quite generalized considerations it is possible to analyze the process of forming some system with control structures and control object in chosen goal environment as gradual refinement of the system organization.

The system that performs definite goal task is functioning in a specific selected environment and basically presents a part of the environment, and it can serve as external sphere for some other systems. The structural division of the sphere and system, of an object and control system can be quite conditional, that could be used for logical understandable description of the processes related to functioning of the system.

Selection of the goal environment is performed with taking into consideration the specificity of the problem being solved. Out of the world where everything is happening the part of environment is selected in which the specific task is implemented. Then the dimensionality of substantial characteristics and states of the goal environment is determined, selection of the controlled object, control system, input and output data, conforming to the laws functioning of a man in the process of search the goal solutions. Then the goal situations are described and determining of the sets of goal micro-situations that create situations, and it is determined the goal functioning in the situations and micro-situations.

The search of decisions necessary for reaching the goals determined in selected situations and micro-situations can be characterized as determining the function of data for output variables at the inputs of the goal (under investigation) system (for example, control system or decision support system). Taking into consideration the evolutional development of biological systems determine the basic directions of their improvement and further development in the sense of enhancement of the level of intellectual possibilities without strict division on availability or absence of intellect. In the problems that are solved by simple biological organisms the similarity of intellectual features can be highlighted by the specificity of subsystems receiving and processing external information (intellectual sensors). Further enhancement of the intellectuality level is linked to constructing of environment models and models of possible behavior of an object (being studied) in this environment with estimation of final results.

When intellectual control system is synthesized the attention should be paid to the urgent methodological questions related to selection and processing input information. Here also the aspects of constructing distributed control structures and databases as well as the questions of decision forecasting in the nonlinear non-stationary environments should be taken into consideration.

THE CONCEPTUAL IDEAS REGARDING ORGANIZATION OF INTELLECTUAL CONTROL SYSTEMS

When intellectual control system is synthesized it is necessary to take into consideration a series of general structural and functional norms. The norm touching upon adequate reflection of features of living beings and a man in the structure and functions of modules and systems being developed should be necessarily considered. The basic structural and functional similarity is reflected through development of logic-and-dynamic models of control processes. Extended functional similarity is reflected through constructing the models of adaptive and optimization stochastic transforms of control processes.

Consider the idea of compact discrete representation of the recognition processes and object control. Such representation supposes selection of substantial variables, forming of input and output spaces of features, the measure of discretization, and optimal division for specific control environment.

Now consider the idea of finding compact transforms that provide for optimal solution and correspond to representations of man-operator and experts. In particular, here it is necessary to perform processing of several data streams with their mutual comparison in functional space, and perform the goal transforms with several levels of decision making based only on the short time forecasts, but also estimate development of situation on the long-term time intervals.

The intellectual integrated control represents multicomponent process of searching for solutions that supposes forming of a series of goal images for control object and control environment, their further refinement, transforming and consolidation in the generalized goal space.

The control process is related to discrete or situational ones. The intervals of discretization or situations are determined by the interactions of control object and control subject as well as influence of environment. Here the most substantial influence on the discretization of control process makes man-operator and correspondingly his characteristics and practical experience. The next substantial reason for control process discretization is availability of noisy components in the control system and environment. Another factor influencing discrete characteristics of the control process is related to bad quality of data characterizing initial and current characteristics of controlled object and control environment. A specific feature of this representation is also taking into account of interaction between man-operator and controlled object even if the control is performed in fully automatic mode. Here automatic control is constructed with taking into consideration of man-operator representation about possible control process and his experience of work in conditions of preserving monitoring functions and correction of possible control decisions. Thus, the decisions that are performed (implemented) are directly or indirectly related to representation of man-operator and his conformity of laws regarding information transforming. Here it is of particular importance for development of separate optimal systems have the goals set by man-operator and their actual practical implementation.

The process of system intellectualization is considered as gradual enhancement of its possibilities regarding effective goal stated functioning on the extended set of environment states. We can state that when the control system reaches some goal using its actions more effectively than without such activity, then there exists some level of its intellectuality.

As an example of possible representation of the development levels for intellectual characteristics of automatic and automatized control systems the following levels are given:

1. The intellectualization level at which preliminary data processing is performed in a single control channel. For example, it can be the case of using specialized goal sensor or the system based upon directive control.

2. The intellectualization level at which several information channels are working and extra preliminary data processing is performed together with selection of substantial data necessary for reaching the goal. It is also performed identification and taking into account of possible uncertainties.

3. The intellectualization level at which development of information transformers is performed that take into account stable conformity to the laws of man-operator activities like logical inference or stereotype behavior.

4. The level of intellectualization at which adaptation is performed to the specificity of accepting and search for solution of the specific operator, the generation takes place and expansion of possibilities for selecting alternative information transforms.

5. The level of intellectualization at which during the control processes is taken into account the set of data- and knowledge bases about control object and environment; the adequacy is observed about system functioning according to basic notions and criteria with the man-operator what can be considered as initial level of image-like control.

6. The level of intellectualization at which exists adequacy of control system functioning according to the image representation of situations by the man-operator and decision forming by the technical part of the system in the definite specific control area.

7. The intellectualization level at which there exists adequacy of system functioning according to the basic information data (and outside of separate specific control area). Here is performed in concordance with multiple situations interaction of a man-operator with the engineering part of the control system.

8. The level of intellectualization at which the systems under study exhibit the characteristics of creative development and refinement.

ORGANIZATION OF GOAL REACHING SYSTEMS

The conformity to the laws for organizing goal reaching systems is expressed in structural and functional characteristics of highly organized living beings [5]. Analysis of these characteristics allows for determining the following basic conformities to the law.

Availability of some time and space continuum for the control system existence in the form of a part of selected environment.

Availability of definite levels of external and internal activities, and freedom of the system under study.

Availability of signals of definite levels for control and interaction of the system with environment and control of internal states.

The number of distinctive signal levels allows for selection of a limited number of possible states in the system and states of the system in the environment.

Availability of initial structural organization that determines initial choice of possible system reactions to external and internal influences.

Availability in the initial structure of motivation basis (for example, for satisfying of the life necessities as well as role necessities and self-development) that is expressed in the form of definite behavior or program of actions.

Availability in the initial structure of plastic elements that allow for recording stimuli signals, emergence of initial reactions and results of system interaction with the environment.

Availability of hierarchical structural system organization that determines interactions of its basic structures, motivation basis, plastic elements and the elements that are responsible for actualization of system reaction.

An example of the sequence of organization and search for the solutions in goal set man-machine systems:

- General formulation of the goal settings.
- Selection of the goal space characteristics and control object.
- Activities analysis of a man-operator under realization of various general and specific goal settings.
- Analysis of selected environment in the space of goal set functioning of a system – goal set reduction of the control environment.
- Analysis of the control object in the space of goal-set functioning of a system – the goal set dimension reduction of control object.
- Creation of a basic structure of a control system in realization of goal settings.
- Distribution of information and control system resources directed to effective and reliable realization of goal settings.
- Creating of possibilities for database and knowledge base expansion, adaptation, learning and repeated learning while realizing of the goal settings.
- Multilevel integration of resources and structures for generating decisions.
- Reaching integrated functioning of technical part of a control system and man-operator.
- Introducing into integrated goal system the features of creative developments.

ORGANIZING DISTRIBUTED REMOTE CONTROL SYSTEMS

Development of effective distributed remote control systems includes the following stages [6].

The problem statement formulation for remote control and determining requirements to the functional characteristics of control object and control system itself.

Formulation of specific temporal requirements to functional agents that are selected in the functional structure of control system under study.

Determining of characteristics for the existing structure of the net that can be used for solving the specific problem of related to remote control.

Determining of time delays emerging during the use of different protocols for information exchange within the net hired.

On the basis of the analysis performed distribution of the control blocks structure at the remote object is performed and in the control system.

The functional optimization is performed of control for the remote system at the expense of extra knowledge about characteristics of the system and environment.

Adding to the net structures of new information devices for receiving and processing information.

Creating of a developing net infrastructure with adding new intellectual features and possibilities.

Extra development of net information exchanging protocols and organization of new channels for information input.

For effective functioning of control systems with distributed information and executive resources in multitask net structures it is necessary to perform complex integrated considering of all elements of the control process and taking into account of the following recommendations:

– distribution of control information according to its value regarding the control goals and acceptable transmission delays;

– determining of specifications for data transmission protocols and for different types of information;

– determining expediency of usage for different channels and routs of information transmission;

– setting priorities of data meaningfulness and the sequences of its transmission in the information structures;

– creating the structures that provide for duplication or excessiveness of the data being transmitted;

– determining of structure for distribution of functional modules of the system and control object depending on possible intervals of control signal delay.

Besides taking into account of systemic requirements to information transmission additional possibilities can be hired for organizing of control structures that take into account perspective methods of control for high speed dynamic processes and structures for distribution of control means on the object and control system. All this helps to improve substantially quality of control for dynamic systems [7]. Fig. 1 shows examples of control organization for remote dynamic objects.

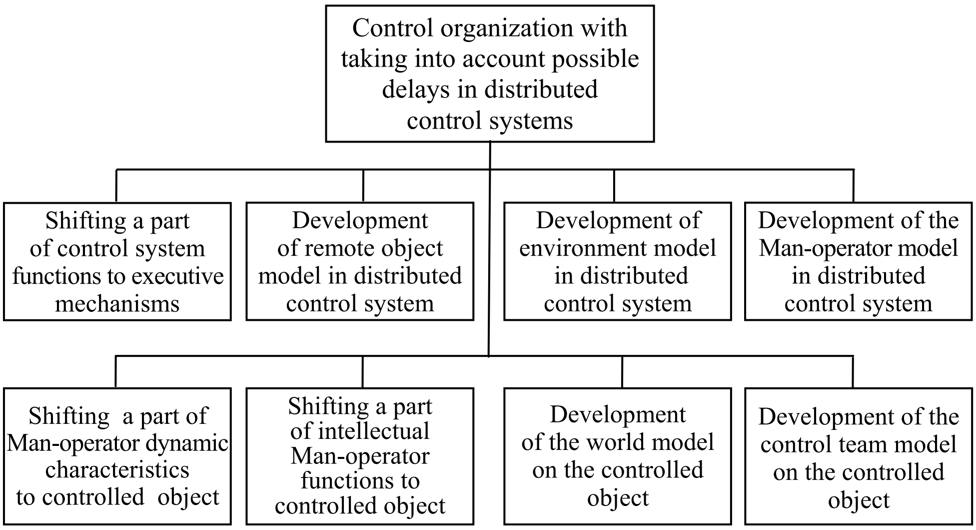


Fig. 1. Examples of control organization for remote dynamic objects

The complex usage of the methods developed for control of high-speed dynamic processes and taking into account of proposed recommendations regarding organization of information transmission in the distributed net structures allows for reaching general improvement of qualitative and quantitative effectiveness measures in the process of remote system control.

THE SYNTHESIS OF GOAL REACHING CONTROL OF A FLYING OBJECT IN CONFLICT AND ACCIDENT CONDITIONS

Consider example of constructing an airplane model and description of the airplane control in algorithms in the non-scheduled situations and equipment faults. Of high importance for successful solving of a flight safety problem are the questions of reliable and stable with regard to interferences of flight control when various goal maneuvers are performed. Also important are the questions of development control algorithms for providing the flights on an airplane with low stability reserve or in the case of unstable airplane.

The control algorithms being developed have to enhance the plane "vitality", reliability of its control system when maneuvers are performed. These algorithms also should provide for continuation of performing set tasks by an airplane when one or several control mechanisms exhibit fault regime. This problem can be solved at the expense of fault compensation by the control system thanks to reconfiguring of the control process by varying aerodynamic surfaces of an airplane.

Solving of the problems mentioned is proposed in the frames of a concept of constructing distributed control system. It is supposed that an airplane has various controlled aerodynamic surfaces, and control by executive mechanisms and surfaces can be performed separately. Flexible structure of control system provides the possibility for changing its configuration during flight, and use various functional regimes of aerodynamic surfaces to create aerodynamic forces of necessary configuration.

To reach stable movement of an airplane, automatic controllers for tangent angle, turn angle, and rove angle are used. The controllers should provide for reaching by airplane of the set orientation angle, performing maneuvers and the flight with given tangent, turn and roam angles. The automatic control of orientation angles provides for compensation of control mechanisms faults.

The algorithms development suggests on-board computer application in the control system. The on-board computer coordinates the functioning of control mechanisms and provides for performing in real time necessary computations and generation of control commands.

To solve the problems of performing goal tasks computer simulation of the control algorithms developed was performed for reaching the airplane control. To solve the simulation problem the flight of aerodynamically unstable airplane had been modeled in perturbed atmosphere. Simultaneously, it was modeled functioning of control system for aerodynamic surfaces of the airplane. Using the model created the testing of functioning of the algorithms developed was performed. The model developed allows for performing simulation of various modes and stages of flight as well as receive estimates of effectiveness of the algorithms functioning. The simulation allows for receiving information about possibilities and areas of application for each algorithm and control system as a whole.

Generally, the perturbed or controlled movement of an airplane is taking place in three-dimensional space because here simultaneously are taking place changings of parameters and variables that define direct and side movement.

When the amplitudes of parameter changings related to the movement of an airplane are small then equations of direct and side movement of the plane can be approximately studied separately. However, when the movement parameter changings are quite large such separation of the equations into direct and side movements starts to produce unacceptably large errors, and it is necessary to analyze complete system of equations describing spatial movement.

Development of the airplane dynamics model was performed on the purpose of simulation of various flight regimes and functioning of control system. To reach this goal complete system of equations describing spatial movement of an airplane was used.

The airplane movement takes place under influence of aerodynamic forces and moments produced by engines and gravitation forces. Generally, to define an airplane movement it is necessary to solve the following problems [8]:

– to find angular and linear velocities of an airplane movement, induced by the influence on it of all mentioned forces and moments;

– to determine the angles of the plane orientation relatively the flow that meet it, and coordinate axis linked to the Earth;

– to determine the shift of an airplane with respect to Earth.

Each of these tasks is supported by special set of differential equations. The model uses quite complete and precise equations of an airplane movement for performing necessary computations and allows for achievement of truthful results for modeling control system functions. In this case an airplane is considered as absolutely rigid body

The mass and inertia moments of airplane are considered as unchangeable during the time of modeling the plane and correspond to initial state of equilibrium flight.

It is also suggested that the plane configuration has symmetry plane and its masses are distributed symmetrically with respect to this plane.

To represent the movement model of unstable flying apparatus in three-dimensional space the work regarding development of stabilizing algorithms of the apparatus has been carried out with taking into consideration cross-channel links in the maneuvers are performed it is supposed that various external disturbances exist such as wind, atmosphere turbulence etc., as well as internal ones, such as poor precision of computations and sensors. To construct the trajectories of given movement into the goal area it is necessary to take into consideration the most substantial information characteristics of control object and environment, and have the possibility of continuous correction of the trajectory due to influence of random factors. Taking into account the general considerations given above, now present an example of control organization for flying apparatus at the stage of landing in accidental conditions and with influence of random wind changings.

Airplane control during the landing stage is the most complicated task of its control. That is why the problems of automation of airplane landing are given substantial attention. The system providing for the airplane landing for the case of engine fault has been developed as an compound part of integrated control system for complex dynamic objects that contains a program simulating movement of a flying apparatus in three-dimensional space.

The airplane landing in accident mode of engine fault puts ahead extra requirements to organization of functioning of interconnected system including pilot – automated control system – flying apparatus. Such systems, oriented to active interaction with man-operator in the extreme situation, should take into account specific features of the man-operator activities in such situations as well as accepting of incoming information by the man. This problem can be formulated as mutual coordination of goals and structure between automated part of the control system and activities of the man-operator. Such considering of a control system as integrated man-machine system can serve as a ground for determining general structure of control and for optimization of respective structural and functional solutions.

In the process of the system development it was accepted that man-machine system for controlling the object in critical regime of functioning requires taking into consideration for effective and reliable reaching the goals of control the following factors [9]:

– activity characteristics of a man-operator, special features of the process of accepting and processing the information by the man-operator as well as generating control actions;

– flexible structure of the algorithms developed and coordination of the algorithms structures with stable stereotypes of the man-operator;

– the possibility of operative switching the control functions between man and automaton;

– taking into consideration and compensation of various types and origin disturbances, dynamic analysis of situations, correction and synthesis of trajectories.

Computer simulation of the control algorithms developed using realistic airplane and environment model provided for the possibility of estimating effectiveness of reaching final goal in complicated navigation environment. For the task of emergency landing of a cargo plane it is possible to generate substantial enhancement of dimensions for the initial states area from which positive result can be reached in comparison with known trajectory control algorithms.

It was also developed and optimized regarding the types and content of control information the interface providing for interaction of a pilot with control system. The interface was coordinated with characteristics of the man-operator and his information requirements in critical situations.

ESTIMATION OF CONTROL EFFECTIVENESS

Estimation of decision effectiveness in general case should be received for the whole region of allowable initial conditions. As a rule the region is rather large and simple analysis (trying) of all points or variants requires substantial time and appropriate computational means. Taking into account the systemic methodology for specific realizations of a control process, generally, we managed to reduce the number of variants for determining effectiveness of analyzed decisions at the expense of considering continuity of the control process and conformity to the laws of systemic interactions. Continuity, as well as stability of the nature of control processes suggests that after changings in external conditions that influence the changings of object states, it is occurring relatively small change of final control. Usually, these values (changings) do not exceed some definite threshold in the limits of one given macro-situation. On the limit transitions between the situations (transition micro-situations) the control may change in the uneven form, and the changings in final results can substantially distinguish from inter-situational values. As an example of calculating such estimates it was considered control of the goal movement of an airplane with constraints regarding energy possibilities of the apparatus in perturbed environment. On the basis of control methodology considered above it was created a series of effective control systems for aviation and marine transportation means in complex navigation conditions, in conditions of equipment fault and strong influence of external disturbances.

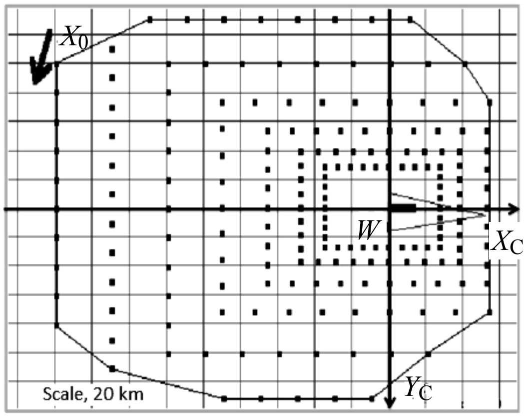


Fig. 2. The region of initial states for flying apparatus, from which successful landing is performed. The height echelon — 5000 m. Wind is favorable: X_0 — initial state of an apparatus; W — state of wind on the landing strip; X_C, Y_C — coordinate directions

Figs. 2–4 show examples of reachability regions for effectiveness estimating of control algorithms for cargo airplane, where the arrow indicates to initial course of flying apparatus, and sharp angle of the triangle points to the wind direction. The computer simulation of the control algorithms developed with realistic model of the airplane allowed for estimate effectiveness of reaching final goal in a complex navigation environment. The area of the initial positions of the airplane, from where a positive landing result is achieved, can be increased up to 2.5 times in comparison with the known methods of landing along a predetermined trajectory.

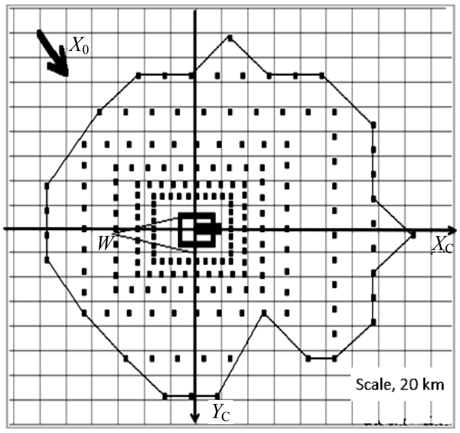


Fig. 3. The region of initial states for flying apparatus, from which successful landing is performed. Height echelon — 5000 м. Wind unfavorable: (X_0 — initial state of an apparatus, W — state of wind on the landing strip, X_C, Y_C — coordinate directions)

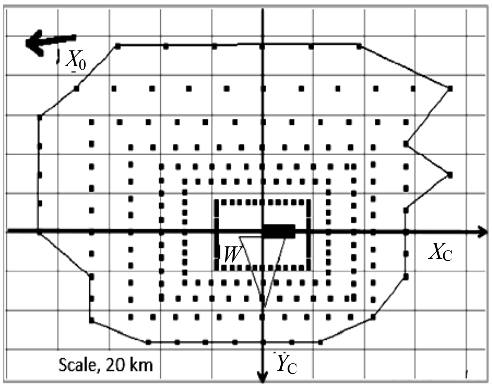


Fig. 4. The region of initial states for flying apparatus, from which successful landing is performed. Height echelon — 3000 m. Wind unfavorable: X_0 — initial state of an apparatus; W — state of wind on the landing strip; X_C, Y_C — coordinate directions

DIRECTIONS FOR PERSPECTIVE RESEARCH

The modern technical level of information processing tools and the experience of creating control systems capable of operating in automatic mode and effectively interacting with a person in critical situations makes it possible to identify promising areas of research and development.

It is proposed development of systems that include the sense notions, databases and knowledge bases, objects (systems) and their interactions (selected or general model of the world).

It is also proposed development of basic elements and processors that possess the features of adaptation, learning and repeated learning for organizing parallel data processing in stochastic environment.

We propose development of interaction architecture revealing actual interaction for modern computational means, basic elements and stochastic environment.

Development of structure and separate components for integrated and distributed control systems.

Reaching necessary level of automation various purpose of distributed databases and knowledge bases.

It is necessary to reach high quality of automatic and automated remote control for dynamic objects in real time.

Enhancement of intellectual possibilities of control systems and decision support systems with making use of distributed human and machine resources.

It is proposed development and application of hybrid man-machine intellectual complexes and systems for processing databases and knowledge data.

These are quite general directions that acquire in sophisticated up-to-date conditions special meaning. Development of multiprocessor computational means and cloud computing, storing knowledge and accessibility to large databases related to control systems and environments with making use of remote access provide for the new possibilities of creating modern control system with features of artificial intelligence. It is particular importance development and practical use of neuron-like structures and technologies of convolutional type and deep learning [3, 10, 11]. The active use of accessible databases related to various object types for neuron-like systems learning provided the possibility for getting results starting from 2014 of automatic object recognition that overcome the results of recognition received by man. On the basis of similar technologies new autonomous control systems are developed for transportation means, for machine translation using different languages, the systems for searching objects in the images and generation of their descriptions, the systems for autonomous carrying out dialogs in separate areas of activities, the systems for synthesis of stylized images, and many of other new developments.

CONCLUSIONS

In the process of organization of effectively functioning man-machine control system it can be distinguished the following basic stages of development control algorithms and structures.

First is the stage of creating adequate model of object and control environment. Here, it is necessary, as possible, to develop precise enough models. The limiting factors usually are the characteristics of environment noise, low observation precision of object variables, and impossibility to reach high precision of representation for complicated functional characteristics, available, for example, in the form of approximate table values.

On this stage the limit possibilities of a control object as a whole are studied, and the problems are solved touching upon stabilization of separate characteristics and development of automatic subsystems for them.

The second stage is usually devoted to development of basic control structure that is characterized by high stability and reliability of functioning. The structure has its specific logic of functioning for the basic set of realizable situations, logic of carrying out testing of basic nodes and logic for the usage of reserved and autonomous adaptive equipment.

The third stage includes development of means for representing control information and the means for monitoring functioning of the control system as a whole. For the second and third stages it is particularly important taking into consideration the goal requirements of a man-operator, his knowledge of functional characteristics of the system and the use of his former experience of work with typical control systems of definite class.

The fourth stage is devoted to implementation of extended matching of operator and control system in conditions of higher risk and non-trivial situations.

Besides, the possibilities are created for extended situational analysis and mutual adaptation between man-operator and control system being created.

The fifth stage is directed to providing for the control system with the means of interaction with man-operator using the language of natural representations and notions, providing the system with possibilities of realization creative elements of the control process.

The indicated general ideas regarding organization of intellectual control systems receive extended extra possibilities of implementation in modern engineering and informational conditions. Correct discretization of the information receiving processes, of the information transforming and control processes should be matched to the goal tasks of the system and interactions with man-operator. A detailed analysis of control processes on the relatively short time intervals (micro-situations) is coordinated well with the development of modern neural like systems in the form of multilayer high precision networks.

The appearance of new computational possibilities in the form of tensor processor and cloud services (computing), that provide for an access to such equipment, create new possibilities for introduction of intellectualization to information transformations and, consequently, constructing of intellectual control systems.

The practical development results are concentrated on control of moving dynamic systems, more precisely, on flying objects in critical regimes of flight.

It was also considered the problem of directing flying apparatus into given limited region with predetermined state characteristics in conditions of availability substantial external disturbances and limited control resources. The three-dimensional computer simulation model, environment and control algorithms were proposed that are distinguished by the high similarity measure with actual functioning of practically available systems.

The control algorithms developed are related well to the experience of pilots regarding control of flying apparatus in conditions of strong disturbances of various nature and with availability of substantial uncertainties in information data. These algorithms allow for making control decisions at each moment, construct optimal goal trajectory with taking into consideration current state characteristics of controlled object and environment.

The use of the micro-situational analysis and synthesis methods allows for automation in different spheres of practical human activities, especially in the directions where exist substantial uncertainties of data, and deficit of time for decision making.

The results achieved are directed to development of information technologies for the goal reaching systems of various classes, and have high practical meaning for solving wide class of problems regarding control of complex dynamic systems in the conflict and accident conditions. The results can also be applied in the analysis and anticipating of catastrophic events in economic and social systems and processes, in organizing of multilevel industrial productions in conditions of uncertainty and deficit of resources.

REFERENCES

1. M. Zgurovsky and Yu. Zaychenko, *Fundamentals of computational intelligence: System approach*. Springer, 2016, 275 p.
2. Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence*. Penguin, 2018, 380 p.

3. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*. Cambridge, MA, USA: The MIT Press, 2016, 800 p.
4. R.W. Beard and T.W. McLain, *Small Unmanned Aircraft: Theory and Practice*. Princeton, NJ: Princeton University Press, 2012, 320 p.
5. S.V. Melnykov, "Conformity to the laws of organizing control in ergatic systems," in *Cybernetics and computers*, issue 155, pp. 3–15, 2007.
6. S.V. Melnykov, "Conformity to the laws of organizing complex ergatic systems and creation of structure for remote control of dynamic objects," in *Cybernetics and computers*, issue 168, pp. 70–79, 2012.
7. S.V. Pavlova, Yu.P. Bogachuk, S.V. Melnykov and A.Yu. Gospodarchuk, "Organizing of effective data exchange in the network control systems for dynamic objects," in *Cybernetics and computers*, issue 177, pp. 43–53, 2014.
8. G.S. Bushgens and R.V. Studnev, *Dynamics of airplane. Dynamics of direct and side movement*. M.: Mashinostrojenie, 1979, 372 p.
9. V.V. Pavlov, *Synthesis of strategies in man-machine systems*. Kyiv: Vyshcha shkola, 1989, 162 p.
10. Neil Wilkins, *Artificial Intelligence: An Essential Beginner's Guide to AI, Machine Learning, Robotics, The Internet of Things, Neural Networks, Deep Learning, Reinforcement Learning and Our Future*. Bravex Publications, 2019, 112 p.
11. M. Zgurovsky and Yu. Zaychenko, *Big Data: Conceptual Analysis and Applications*. Springer Nature Switzerland AG, 2019, 306 p.

Received 13.06.2022

INFORMATIONON THE ARTICLE

Serhii V. Melnykov, ORCID: 0000-0003-2873-5730, International Scientific and Educational Center of Information Technologies and Systems, Ukraine, e-mail: psmail@i.ua

Petro M. Malezhik, ORCID: 0000-0001-6816-988X, Dragomanov National Pedagogical University, Ukraine, e-mail: p.m.malezhyk@npu.edu.ua

Aydin C. Gasanov, ORCID: 0000-0002-5821-0751, Dragomanov National Pedagogical University, Ukraine, e-mail: 0677937631@ukr.net

Petro I. Bidyuk, ORCID 0000-0002-7421-3565, Educational and Research Institute for Applied System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: pbidyuk_00@ukr.net

МЕТОДОЛОГІЧНІ АСПЕКТИ ІНТЕЛЕКТУАЛІЗАЦІЇ СИСТЕМ ОПЕРАТИВНОГО КЕРУВАННЯ ДИНАМІЧНИМИ ОБ'ЄКТАМИ / С.В. Мельников, П.М. Малежик, А.С. Гасанов, П.І. Бідюк

Анотація. Розглянуто питання організації інтелектуальних систем керування: визначення рівнів інтелектуалізації, послідовність проведення аналізу складових компонентів процесу керування, унесення в систему характеристичних властивостей інтелекту, особливості структурної організації, оцінювання міри інтелектуалізації, приклади практичної реалізації. Метою розгляду є висвітлення напрямів організації інтелектуальних систем керування, необхідності і корисності системного розгляду, що враховує вимоги поставленого цільового завдання, характеристики зовнішнього середовища, використовувані засоби отримання, перетворення інформації і відпрацювання керування, функціональні характеристики і досвід роботи людини-оператора. У результаті проведеного аналізу виділено рівні розвитку інтелекту системи керування, визначено етапи проведення інтелектуалізації системи, показано ефективність викладених представлень для вирішення практичних завдань.

Ключові слова: система керування, інтелектуалізація, організація систем керування, інтелектуальне керування, людино-машинні системи.

UDC 303.732.4 + 004.67 + 004.8
DOI: 10.20535/SRIT.2308-8893.2022.4.05

THE PROBLEM OF AUTOMATIC CLASSIFICATION OF PICTURES USING AN INTELLIGENT DECISION-MAKING SYSTEM BASED ON THE KNOWLEDGE GRAPH AND FINE-GRAINED IMAGE ANALYSIS

A.A. MARTYNENKO, A.D. TEVYASHEV, N.Ye. KULISHOVA, B.I. MOROZ

Abstract. In order to prevent the illegal export of paintings abroad, a museum examination using various methods for studying a work of art is carried out. At the same time, an analysis is also made of historical, art history, financial and other information and documents confirming the painting's authenticity — provenance. Automation of such examination is hampered by the need to take into account numerical values of visual features, quality indicators, and verbal descriptions from provenance. In this paper, we consider the problem of automatic multi-task classification of paintings for museum expertise. A system architecture is proposed that checks provenance, implements a fine-grained image analysis (FGIA) of visual image features, and automatically classifies a painting by authorship, genre, and time of creation. Provenance is contained in a knowledge graph; for its vectorization, it is proposed to use a graph2vec type encoder with an attention mechanism. Fine-grained image analysis is proposed to be performed using searching discriminative regions (SDR) and learning discriminative regions (LDR) allocated by convolutional neural networks. To train the classifier, a generalized loss function is proposed. A data set is also proposed, including provenance and images of paintings by European and Ukrainian artists.

Keywords: automatic multi-task classification, knowledge graph, attention mechanism, fine-grained image analysis, museum expertise, paintings, convolutional neural networks.

INTRODUCTION

The problem of art objects illegal export continues to be relevant, since they are a means of accumulating value. In Ukraine, normative documents that regulate the procedure for customs control and examination of cultural property have been adopted [1–6]. These documents establish and approve the procedure according to which it is possible to export values abroad, for which the Authority for Control over the Movement of Cultural Property and the Protection of the Cultural Heritage of Ukraine issued a certificate for the right to export. At the same time, basis for such a certificate is customs and museum expertise. The customs examination, first, aims to establish painting age, since, according to mentioned regulatory documents, antiques prohibited for export, include items over 100 years old. During the museum examination, the authenticity and authorship of painting is established, which, of course, also serves the purpose of dating the work of art.

A wide range of approaches used to establish paintings authenticity, include forensic, technological, attributive and other methods. They involve various forms of research, for example, study of artist's fingerprints, signatures, seals, lists of invoices confirming the painting sale, reproductions in books and catalogs, a description of the history of painting creation and ownership to the present moment (provenance). Experimental studies also include researching of paintings using microscopy, fluoroscopy, macrophotography, spectroscopy, etc.

Conducting such a comprehensive study takes a lot of time, requires the participation of dozens of highly qualified art historians, chemists, digital technology specialists.

Under conditions of a customs check, it is impossible to implement such an examination. Therefore, for prompt decision-making on a work of art exporting possibility, an intelligent decision-making system was proposed [7–9]. It provides for automatic identification of a painting and the establishment of its authenticity and value based on a photo. However, such an operational check is one part of the proposed two-stage procedure — it can only prevent the export of suspicious art values, but cannot replace a full museum expertize, which is the basis for a permit certificate.

It is possible to speed up an export permit by automating a full museum examination, which, from the point of view of machine learning, can be represented as a classification task. Research in this direction has been going on for many years, and in recent years, there has been a certain breakthrough associated with the use of deep networks, in particular, Convolutional Neural Networks (CNN). They are distinguished by ability to automatically generate vectors of non-obvious features, especially in image processing tasks, provide higher classification accuracy compared to other machine learning methods, and have high speed. Many works demonstrate that the application of CNN to automatic paintings classification gives positive results [10–14].

However, as noted, an important component of museum expertise is the study of painting's provenance, which is usually presented as textual descriptions of changing size. For CNN that analyze a painting image, provenance turns out to be useless. On the other hand, deep networks that have proved to be highly effective in word processing tasks, such as LSTM (Long Short-Term Memory), do not allow tracking signs of a correlation nature in two-dimensional signals – paintings photos.

Thus, the purpose of this work is to develop the architecture of an intelligent decision-making system based on deep networks for automatic classification of paintings, taking into account their provenance.

ANALYSIS OF AREAS OF RESEARCH AND STATEMENT OF THE PROBLEM

Deep networks are currently a common tool for solving a variety of data analysis problems: searching for objects in images and videos, automatic translation, handwriting recognition, processing streaming information. There are also examples of deep architectures use for solving various tasks of preserving cultural heritage in general [15–16], and paintings in particular. Thus, convolutional neural networks have been used to automatic paintings classification by author and artistic genre [10, 12, 14, 17, 18]. The initial data was digital paintings images, based on analysis of which CNN generates a response about painting authorship with high accuracy. At the same time, the classification attributes are formed automatically by the input layers of the CNN, and form internal descriptions – embeddings, "understandable" network parameters in numerical form. It is important to note that specially created datasets are used to train such networks, including tens of thousands of paintings images [19, 20]. However, the list of artists whose paintings are included in such sets is quite narrow and is limited to three to four dozen world-famous masters who worked during the 15th–20th centuries. In the course of training on such datasets, CNN studies the features of artists' writing and is able to quite accurately distinguish between paintings of different styles, but of the same master, or vice versa, of different authors, but of the same genre.

The works of artists included in the typical datasets used to train CNN are world masterpieces, their location is known, they are well protected. Therefore, the probability of their presentation for export from Ukraine is extremely small. However, there are a large number of paintings by lesser-known masters in the country that should be banned from export due to their significant value. These pictures were not used to train deep networks, so there is no guarantee that such a network will confirm the authorship with sufficient accuracy. An unequivocal help in this situation can be such an architecture that will allow using information about paintings provenance. In particular, in [19] it is proposed to use the Knowledge Graph for a provenance branched formalized description in the form of a graph structure available for further implementation using deep networks.

On the other hand, Ukrainian artists, whose works are of value and may be banned for export, worked for a much narrower period of time – during the 17th–20th centuries (those created no later than 1920 can be recognized as antiques). Due to historical circumstances, these paintings do not differ in genre and stylistic diversity, so it is possible that a network trained to distinguish between Renaissance and abstract art will not be able to accurately distinguish between landscapes painted in the style of 19th century classicism and early 20th century realism. The emerging field of machine intelligence Fine-Grained Image Analysis (FGIA) develops methods for analyzing sub-categories of images in a single meta-category [21, 22]. These methods are focused on finding more subtle and little noticeable, but significant (from authorship point of view) differences between images, which allow us to single out stable subclasses within the same class of objects.

In this paper, we propose to apply the Knowledge Graph to formalize provenance and use it as an attribute when categorizing paintings using Fine-Grained Image classification implemented in deep learning architecture.

METHODOLOGY

Representation of provenance

Information about painting creation history, its sale to past and current owners is an undoubted and weighty confirmation of authenticity, along with such characteristics as features of strokes, coloring, chemical composition of paints, primers, canvas and stretcher wood. This information can be quite scattered, since documents confirming it can be stored in various institutions, by different persons, or even be lost. Therefore, provenance data does not have a standardized format, and is represented most often by field's text entries such as [23]:

– author's name;
– artist's life years;
– picture name;
– picture creation date;
– technique (oil on canvas, oil on wood, watercolor, etc.);
– current location;
– URL link with a digital photo of the painting;
– form (painting, sculpture);
– type (portrait, still life, etc.);

– school (French, Dutch, etc.);

– era (years of the artist's work).

Examples of such records [23]:

TOULOUSE-LAUTREC, Henri de, "(b. 1864, Albi, d. 1901, Château Malromé, Langon)", Countess Adèle de Toulouse-Lautrec in the Salon of Malromé Château, 1887, "Oil on canvas, 54×45 cm", "Musée Toulouse-Lautrec, Albi", https://www.wga.hu/html/t/toulouse/2/1misc02.html, painting, portrait, French, 1851–1900.

UNKNOWN MASTER, German, (active 1490s in Nuremberg), Portrait of a Man, 1491, "Oil on linden panel, 37×20 cm", "Metropolitan Museum of Art, New York", https://www.wga.hu/html/m/master/zunk_ge/zunk_ge4a/portrman.html, painting, portrait, German, 1451–1500.

MONET, Claude, "(b. 1840, Paris, d. 1926, Giverny)", Monet's Garden at Argenteuil, 1873, "Oil on canvas, 61×82 cm", Private collection, https://www.wga.hu/html/m/monet/03/argent08.html, painting, landscape, French, 1851–1900.

In Ukraine, work is underway to catalog museum collections and draw up Scientifically Unified Passports. Although it is far from complete, it is being carried out in accordance with international experience (ICOM requirements, UNESCO Model export sertificate for cultural objects, etc.) [24].

A record about a painting can be represented as a graph, which in [19] is called the Knowledge Graph and displays elements of description of painting and the relationship between these elements. Taking into account Ukraine conditions, the structure is proposed to be modified (Fig. 1).

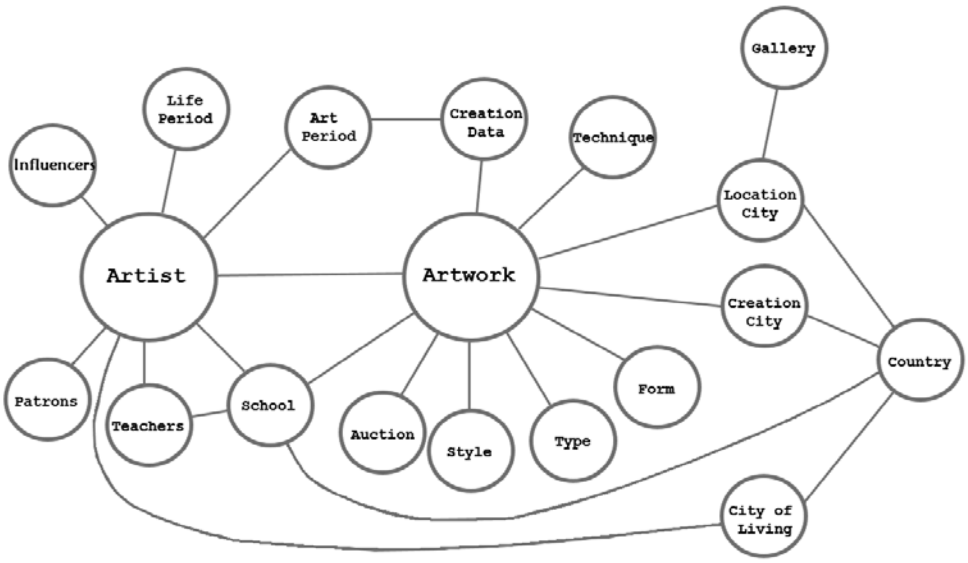


Fig. 1. Nodes and edges of a knowledge graph modeling metadata about a picture

Graph embeddings are extracted from model using an encoder – a pre-trained CNN that implements the node2vec transformation [25] and solves the task of classification a picture by provenance attributes represented as a graph model.

At present, networks such as ResNet50, ResNet101, ResNet152 demonstrate the highest classification accuracy in such problems [26, 27]. It is proposed to train network using a loss function [20]:

$$L_{\mathrm{Pr}\,ovenance}(p_j, u_j) = \left\| p_j - u_j \right\|_2^2 ,$$

where p_j — predicted embedding; u_j — ground truth context embedding; j — expertize object (painting).

Using Fine-Grained Image Analysis to paintings classification

Even experienced art historians sometimes make mistakes when determining the authorship or dating of works painted in the same style in a short time period. In addition to results of chemical, spectroscopic, and X-ray studies, the FGIA approach can help in solving this problem, which makes it possible to use information about difference in fine details of objects belonging to the same class. The main difficulty in approach implementing is preservation of information about regional features when network learns from hundreds of thousands of sample images. The attention mechanism allows finding the most significant regional features in images, and save information about them, despite large size of training datasets.

In [28], it is proposed to form special regions that store information about individual objects features belonging to subclasses – Searching and Learning Discriminative Regions (SDR, LDR).

Just as global features are extracted from images using a CNN and images are assigned to classes based on the mapping of these feature vectors, in discriminative regions, the deep network generates feature vectors within individual parts of images that are in some sense similar to each other. An example of such tasks is distinction between aircraft by type, birds by subspecies within the same family, and so on. The resulting vector includes both global features inherent in class images, and private (partial) inherent in subclass images.

Searching Discriminative Regions (SDR) are designed to search and locate particular features in an image. The scheme of search areas formation is shown in Fig. 2 [28].

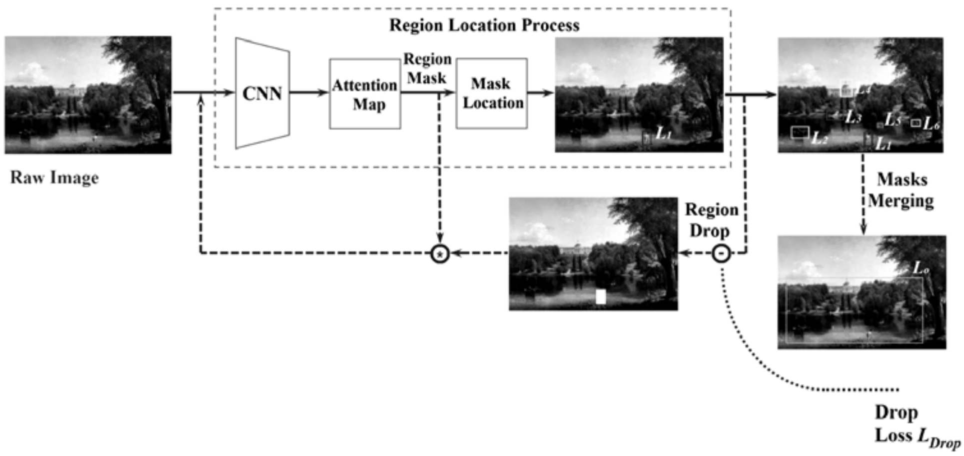


Fig. 2. Scheme of searching discriminative regions formation

At the heart of this system block operation is an attention-based search mechanism. The convolutional neural network here provides for search and selection of all possible features in image, and only thanks to the attention function does it become possible to search for heterogeneous characteristic features $L_1, L_2,..., L_n$ in one image, and use them as references for searching pictures in other images.

The network is trained to minimizing the objective function that describes the angular measure $\cos\theta_y$ of differences between actual categories logits (features) and their values predicted by network:

$$L_{arc} = -\log \frac{\exp(s(\cos\theta_y))}{\sum_{j=1, j\neq y}^{C} \exp(s(\cos\theta_j)) + \exp(s(\cos\theta_y))},$$

where C is the number of categories to classify. In this problem, it is equal to number of compared paintings in dataset.

When locating searching regions on image, it is necessary to minimize losses associated with excluded zones characteristics:

$$L_{Drop} = \sum_{i=1}^{n} L_{drop-arc}(C_d(em_{d_i})), \quad (1)$$

where em_{d_i} — image features generated by network that do not go beyond the network (embeddings) d_i, $i = 1,...,n$ associated with input picture; C_d — is a classifier that maps embeddings into classification objects (logits) categories features.

After searching regions finding, we need to compose their descriptions, taking into account their possible appearance in other dataset images. For this purpose, learning discriminative regions (LDR) are formed on basis of SDR, in accordance with scheme of Fig. 3.

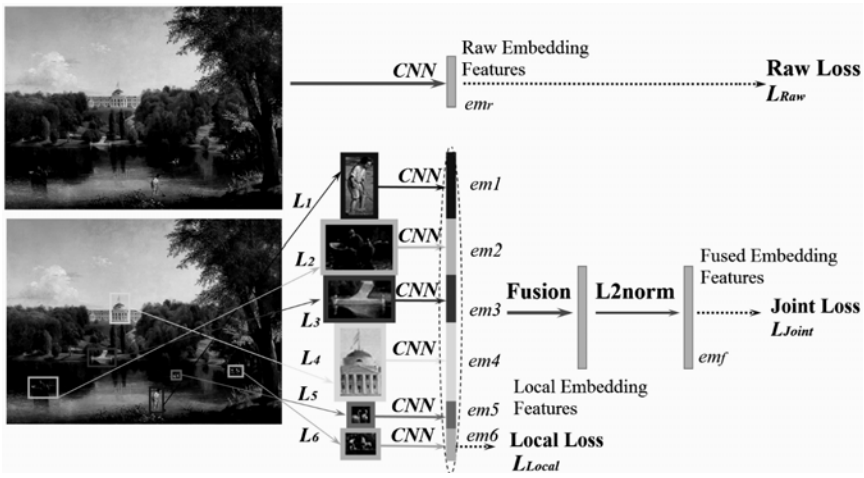


Fig. 3. Approach to LDR formation

By analogy with (1), loss functions are determined for training of convolutional networks that form embeddings $em_r, em_1, ..., em_6, ..., em_{d_m}, em_f$:

$$\begin{cases} L_{Raw} = L_{Raw-arc}(C_r(em_r)); \\ L_{Local} = \sum_{i=1}^{n} L_{Local-arc}(C_l(em_l)); \\ L_{Joint} = L_{Joint-arc}(C_f(em_f)), \end{cases}$$

where $em_r, em_1, ..., em_6, ..., em_{d_m}, em_f$ – embeddings associated with individual SDRs that are highlighted on input image; C_r, C_l, C_f – classifiers that map embedding features into category features (logits).

Local regional features (embeddings) are combined in Fusion module, forming a generalized vector em_f :

$$em_f = Fusion(em_r, em_1, ..., em_6, ..., em_{d_m}) .$$

Merge can be implemented by concatenation or convolution. In the second case, a 1D convolutional layer with H channels will need to be added to the network architecture, then dimension of generalized embedding vector will be $(n+2)H$.

Architecture of an automatic system for classifying paintings using Knowledge Graph and Fine-Grained Image Analysis

To solve the paintings classification problem, taking into account provenance in vector representation, and with possibility of distinguishing features of artists of the same genre, one time period, a system is proposed that is built using knowledge representation in the form of a graph structure, where feature extraction on images of paintings is performed using SDR and LDR .

The system general architecture is shown in Fig. 4. The main idea is to learn convolutional network to project metadata about picture and its fine-grained features into classification objects space. The solution is carried out in a multi-task mode due to concatenation of visual feature embeddings from original image, provenance embeddings from encoder, and fine-grained feature embeddings from SDR and LDR.

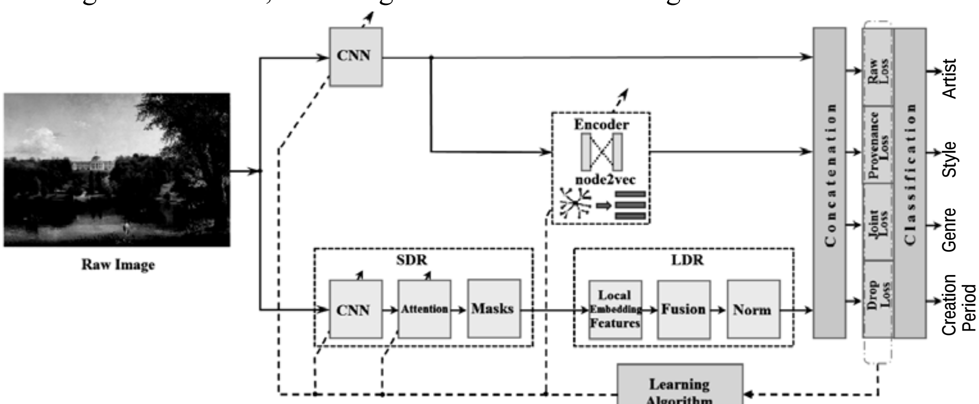


Fig. 4. Architecture of an automatic system for classifying paintings using Knowledge Graph and Fine-Grained Image Analysis

The system loss function in multitasking classification mode is defined as

$$L = (1-\gamma)\left(\lambda_{Raw} \sum_{j=1}^{N} L_{Raw-arcj}(C_r(em_r))\big(C_r(em_r)\big) + \right.$$

$$+ \lambda_{Joint} \sum_{j=1}^{N} L_{Joint-arcj}(C_f(em_f)) + \lambda_{Local} \sum_{j=1}^{N}\sum_{i=1}^{n} L_{Local-arcj}(C_l(em_i)) +$$

$$\left. + \lambda_{Drop} \sum_{j=1}^{N} L_{Drop-arcj}(C_f(em_{if})) \right) + \gamma \frac{1}{N} \sum_{j=1}^{N} L_{Provenance}(p_j, u_j) ,$$

where γ are error weights of system modules, λ_i are hyperparameters that takeinto account individual tasks contribution to classification result.

Datasets for solving the automatic paintings classification problem

Since provenance in this task is an integral part of initial information array, it is necessary to select data for system learning in an appropriate way. Many world-famous museums include metadata in verbal descriptions form when digitizing paintings. There are no detailed lists of all documents that verify the entire picture sale history in such descriptions, but even brief information about the time, place of creation, style, genre, school, etc. will increase accuracy of multitasking classification.

In this paper, it is proposed to use datasets [23, 29] that are freely available. They contain images of paintings by world masters who worked in 15th–20th centuries, in various techniques, styles and genres. In addition, these datasets contain brief information related to provenance.

To apply developed system in Ukraine, it is obviously necessary to supplement these sets with images of paintings by Ukrainian artists, for example, from the National Art Museum of Ukraine funds [30]. Metadata about these paintings and artists can be collected both on museum portal and on Wikipedia.

CONCLUSIONS

The paper considers the problem of paintings automatic classification using an intelligent decision-making system based on a knowledge graph and Fine-Grained Image Analysis. A solution is proposed in the form of a classifier based on convolutional neural networks with attention model, operating in a multitasking mode.

The architecture of system that performs visual features automatic detection and analysis of Fine-Grained features from picture image, provenance vector formation and picture identification by author, style, genre and time of creation has been developed.

To organize classifier learning, it is proposed to use the loss function based on the angular mismatch between intranet representations of classification objects.

It is proposed to select data for system training and validation from open access datasets, which contain both images of paintings and metadata with descriptions of provenance. In addition, it is proposed to use resources of Ukrainian museums to update the system in Ukraine.

REFERENCES

1. “Constitution of Ukraine dated June 28, 1996 No. 254k/96-BP,” *Bulletin of the Verkhovna Rada of Ukraine*, 1996.
2. “Law of Ukraine “Fundamentals of the Legislation of Ukraine on Culture” dated February 14, 1992 N 2117-XII,” *Bulletin of the Verkhovna Rada of Ukraine*, 1992.
3. “Law of Ukraine “On export, import and return of cultural values” of September 21, 1999 N 1068-XIV,” *Bulletin of the Verkhovna Rada of Ukraine*, 1999.
4. Resolution of the CMU of June 20, 2000 No. 983 “On the approval of the Regulation on the State Service of Control over the Movement of Cultural Values across the State Border of Ukraine”.
5. Letter of the State Customs Service dated 30.12.99 N 09/1619 “Regarding customs control over the movement of cultural values”.
6. Order of the Ministry of Culture and Tourism of Ukraine dated 11.15.2002 N 647 “On Approval of the List of State Institutions, Cultural Institutions, and Other Organizations Granted the Right to Conduct State Expertise of Cultural Values”.

7. A. Martynenko, V. Moroz, and I. Nulina, “An intelligent decision support system for cultural property identification,” *Computer-IntegratedTechnologies: Education, Science, Production*, 39, pp. 78–82, 2020. Available: http://cit-journal.com.ua/index.php/cit/article/view/126
8. A.A. Martynenko, A.D. Tevyashev, N.Ye. Kulishova, and B.I. Moroz, «System analysis of the problem of establishing the authenticity and authority of painting works,” *System Research and Information Technologies*, no. 1, pp. 1–16, 2022.
9. A.A. Martynenko, A.D. Tevyashev, N.Ye. Kulishova, B.I. Moroz, and A.S. Sergienko, “Automatic classification of paintings by year of creation,” *Radio Electronics, Computer Science, Control*, no. 2(61), pp. 80–89, 2022.
10. D. Pancaroglu, *Artist, style and year classification using face recognition and clustering with convolutional neural networks*; Eds.: D.C. Wyld et al., COMIT, SIPO, AISCA, MLIOB, BDHI, 2020, CS & IT, CSCP 2020, pp. 41–54. doi: 10.5121/csit.2020.101604.
11. N. Banar, M. Sabatelli, P.Geurts, W. Daelemans, and M. Kestemont, “Transfer Learning with Style Transfer between the Photorealistic and Artistic Domain,” *Electronic Imaging, Computer Vision and Image Analysis of Art*, pp. 41-1–41-9, 2021. Available: doi.org/10.2352/ISSN.2470-1173.2021.14.CVAA-041
12. E. Cetinic, T. Lipic, and S. Grgic, “Fine-tuning Convolutional Neural Networks for fine art classification,” *Expert Systems with Applications*, vol. 114, pp. 107–118, 2018. Available: doi.org/10.1016/j.eswa.2018.07.026
13. T. Chen and J. Yang, “A Novel Multi-Feature Fusion Method in Merging Information of Heterogenous-View Data for Oil Painting Image Feature Extraction and Recognition,” *Front. Neurorobot.*, 12, 2021. Available: doi.org/10.3389/fnbot.2021.709043
14. W. Zhao, D. Zhou, X. Qiu, W. Jiang, “Compare the performance of the models in art classification,” *PLoS ONE*, 16(3), 2021. Available: doi.org/10.1371/journal.pone.0248414
15. *Data Analytics for Cultural Heritage: Current Trends and Concepts*; Eds.: A. Belhi, A. Bouras, A. Kh. Al-Ali, A.H. Sadka. Springer, 2021, 279 p.
16. M. Fiorucci, M. Khoroshiltseva, M. Pontil, A. Traviglia, A. Del Bue, and S. James, “Machine learning for cultural heritage: A survey”, *Pattern Recognition Letters*, vol. 133, pp.102–108, 2020.
17. M. Kelek, N. Calik, and T. Yildirim, “Painter Classification Over the Novel Art Painting Data Set via The Latest Deep Neural Networks,” *Procedia Computer Science*, 154, pp. 369–376, 2019. doi: 10.1016/j.procs.2019.06.053.
18. M. Sabatelli, M. Kestemont, W. Daelemans, P. Geurts, “Deep transfer learning for art classification problems,” *Proceedings of the European Conference on Computer Vision (ECCV), 2018*. Available: https://doi.org/10.1007/978-3-030-11012-3_48
19. G. Castellano, G. Sansaro, and G. Vessio, “Integrating Contextual Knowledge to Visual Features for Fine Art Classification,” *arXiv:2105.15028v2 [cs.CV], 28 Sep 2021*.
20. N. Garcia, B. Renoust, and Y. Nakashima, “ContextNet: Representation and exploration for painting classification and retrieval in context,” *International Journal of Multimedia Information Retrieval*, 9(1), pp.17–30, 2020.
21. X.-S. Wei, J.-H. Luo, J. Wu, and Z.-H. Zhou, “Selective convolutional descriptor aggregation for fine-grained image retrieval,” *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2868–2881, 2017.
22. X. Zeng, Y. Zhang, X. Wang, K. Chen, D. Li, and W. Yang, “Fine-grained image retrieval via piecewise cross entropy loss,” *Image and Vision Computing*, vol. 93, p.103820, 2020.
23. *Open Access Images - National Gallery of Art*. Available: https://www.nga.gov/open-access-images.html
24. O.L. Kalashnikova, “Legal consequences of imperfect description and accounting of museum collections. International law and the law of the European Union,” *Actual problems of domestic jurisprudence*, no. 2, vol. 2, 2018. Available: http://apnl.dnu.in.ua/2_2018/tom_2/32.pdf

25. A. Grover and J.Leskovec, "node2vec: Scalable Feature Learning for Networks," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016*.
26. *Deep Residual Networks (ResNet, ResNet50) – 2022 Guide*. Available: https://viso.ai/deep-learning/resnet-residual-neural-network/
27. *RESNET50.* Available: https://pytorch.org/vision/main/models/generated/torchvision.models.resnet50.html#resnet50
28. K. Sun and J. Zhu, "Searching and Learning Discriminative Regions for Fine-Grained Image Retrieval and Classification," *IEICE TRANS. INF. & SYST.*, vol. e105–d, no.1, pp. 141–147, 2022.
29. *Best artworks of all time*. Available: https://www.kaggle.com/ikarus777/best-artworks-of-all-time/tasks
30. *National Art Museum of Ukraine*. Available: https://namumuseum.business.site/

Received 06.11.2022

INFORMATIONON THE ARTICLE

Andrii A. Martynenko, ORCID: 0000-0002-5033-4696, Dnipro University of Technology, Ukraine, e-mail: martynenko.andrey.a@gmail.com

Andriy D. Tevyashev, ORCID: 0000-0002-2846-7089, Kharkiv National University of Radio Electronics, e-mail: andrew.teviashev@nure.ua

Nonna Ye. Kulishova, ORCID: 0000-0001-7921-3110, Kharkiv National University of Radio Electronics, e-mail: nonna.kulishova@nure.ua

Boris I. Moroz, ORCID: 0000-0003-1290-5694, Dnipro University of Technology, Ukraine, e-mail: moroz.boris.1948@gmail.com

ПРОБЛЕМА АВТОМАТИЧНОЇ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ ЗА ВИКОРИСТАННЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ ГРАФА ЗНАНЬ І ТОЧНОГО АНАЛІЗУ ЗОБРАЖЕНЬ / А.А. Мартиненко, А.Д. Тевяшев, Н.Є. Кулішова, Б.І. Мороз

Анотація. Для запобігання незаконному вивезенню картин за кордон проводиться музейна експертиза з використанням різних методів дослідження твору мистецтва, зокрема аналіз історичних, мистецтвознавчих, фінансових та інших відомостей і документів, що підтверджують справжність картин – провенансу. Автоматизація такої експертизи ускладнюється необхідністю враховувати числові значення візуальних ознак, показників якості та словесні описи з провенансу. Розглянуто завдання автоматичної багатозадачної класифікації картин під час музейної експертизи. Запропоновано архітектуру системи, яка перевіряє провенанс, реалізує детальний аналіз (FGIA) візуальних ознак зображення та виконує автоматичну класифікацію картини за авторством, жанром та часом створення. Провенанс міститься у графі знань, для векторизації якого запропоновано використовувати енкодер типу graph2vec з механізмом уваги, а детальний аналіз пропонується виконувати за допомогою пошукових відмітних регіонів (SDR) та навчальних відмітних регіонів (LDR), що виділяються згортковими нейронними мережами. Для навчання класифікатора запропоновано узагальнену функцію втрат, а також набір даних, що включає провенанс та зображення картин європейських та українських художників.

Ключові слова: автоматична багатозадачна класифікація, граф знань, механізм уваги, дрібнодетальний аналіз зображень, музейна експертиза, твори живопису, згорткові нейронні мережі.

UDC 330.46:004.89
DOI: 10.20535/SRIT.2308-8893.2022.4.06

DATA MINING TOOLS FOR COMPLEX SOCIO-ECONOMIC PROCESSES AND SYSTEMS

T.V. OBELETS

Abstract. The paper considers discovering new and potentially useful information from large amounts of data that actualizes the role of developing data mining tools for complex socio-economic processes and systems based on the principles of the digital economy and their processing using network applications. The stages of data mining for complex socio-economic processes and systems were outlined. The algorithm of data mining was considered. It is determined that the previously used stages of data mining, which were limited to the model-building process, can be extended through the use of more powerful computer technology and the emergence of free access to large amounts of multidimensional data. The available stages of data mining for complex socio-economic processes and systems include the processes of facilitating data preparation, evaluation, and visualization of models, as well as in-depth learning. The data mining tools for complex socio-economic processes and systems in the context of technological progress and following the big data paradigm were identified. The data processing cycle has been investigated; this process consists of a series of steps starting with the input of raw data and ending with the output of useful information. The knowledge obtained at the data processing stage is the basis for creating models of complex socio-economic processes and systems. Two types of models (descriptive and predictive) that could be created in the data mining process were outlined. Algorithms for estimating and analyzing data for modeling complex socio-economic processes and systems in accordance with the pre-set task were determined. The efficiency of introducing neural networks and deep learning methods used in data mining was analyzed. It was determined that they would allow effective analysis and use of the existing large data sets for operational human resources management and strategic planning of complex socio-economic processes and systems.

Keywords: data mining, complex socio-economic systems, predictive modeling, neural networks, deep learning.

INTRODUCTION

The continuous scaling of data on the Internet is changing the way we interact in economic and social systems. Many users search, publish, and create new data daily, leaving a digital footprint that can help describe their behavior, decisions, and intentions. This highlights the role of developing data mining tools for complex socio-economic processes and systems based on the principles of the digital economy and their processing using network applications.

Analysis of recent research and publications. The most significant results in statistical analysis and applications for data mining were achieved in the works of R. Nisbet, H. Miner, O. Maimon, L. Rokach. Such scientists as H. Jiawei, M. Kamber, P. Jiang, H. Choi, H. Varian, and others devoted their research to the creation of concepts and development of data mining techniques. H. Xiong, H. Pandei, A. Kryzhevsky, I. Sutskever and Jeffrey E. Hinton explored machine-learning capabilities with deep convolution neural networks. Successful results in the field of artificial intelligence with deep learning were obtained in the works of J. Lekun, Y. Bengio, G. Hinton, J. Schmidhuber, and Ukrainian authors M. Lavrenyuk, N. Kusul, O. Novikov. The analysis of complex socio-economic systems was carried out by foreign authors in their works P. dos Santos, N. Wiener, J. Stefanovsky, and the issues of forecasting socio-economic processes were in the interests of Ukrainian scientists G. Prisenko and E. Ravikovych and others. At the same time, the progress of computing power and the availability of large amounts of multidimensional data make it necessary to develop data mining tools for complex socio-economic processes and systems.

The purpose of the article is to study the process of identifying new and potentially useful information from large amounts of data, outlining the stages of data mining for complex socio-economic processes and systems, and identifying appropriate tools in the context of the progress of computing power and the emergence of a large number of multi-dimensional data in the free-of-use.

Presentation of the main material of the study. As data mining has evolved as a professional activity, it is necessary to distinguish it from previous statistical modeling activities and broader knowledge discovery activities. Data mining is defined as the use of machine learning algorithms to find weak patterns of relationship between data elements in large and disordered datasets, which can lead to actions to increase benefits in one form or another (diagnostics, profit, prediction, management, etc.) [1].

Data mining is also called knowledge discovery in databases (Knowledge Discovery in Databases — KDD), i.e. the process of discovering new and potentially useful information from large amounts of data. The definition of data mining was initially limited to the modeling process, but over time the data analysis tools have included processes to facilitate data preparation, as well as evaluation and visualization of models [2] (Fig. 1).

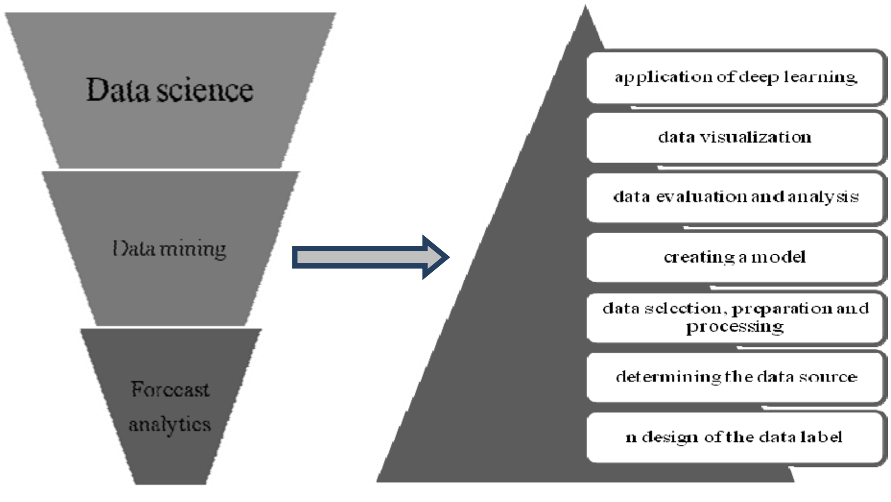


Fig. 1. Algorithm of data analytics and tools of data mining

The process of identifying knowledge in databases combines the mathematics used to identify patterns in the data with the whole process of data selection and the use of models to apply to other datasets and use the information for predetermined purposes. This process combines the development of business systems, statistical methods, and digital technologies to identify the structure of socio-economic processes and systems (relationships, patterns, associations, and basic functions), not just their statistical parameters (averages, weights, etc.) [1].

The data mining algorithm begins with the definition of objectives in the data matrix design process and ends with the introduction of the identified knowledge. At the stage of designing the data matrix, preparatory work is carried out, goals are defined and strategic ideas are formed, for the achievement of which the process of knowledge discovery in databases begins. Understanding the strategic goal, a clear understanding of the end-user of the data, and understanding the environment in which the data will be disseminated, is a prerequisites for an adequate process of datamining.

Data mining uses techniques from different fields of knowledge, such as statistics, machine learning, pattern recognition, database and storage systems, information search, visualization, algorithms, high-performance computing, and many application areas. Statistics examines the collection, analysis, interpretation or explanation, and presentation of data. Statistical models are widely used to model data and data classes. Multivariate graphical methods are used to research. analyze databases and present the results of data analysis [3].

Identification of data sources and searching for documents or information in documents is an important step in data mining. Documents can be text-based or multimedia can be on paper in archives and can be available electronically on the Internet. The main source of information for data mining for complex socio-economic processes and systems today is the Internet. Thus, Google Trends (GT) is an online tool that reports on the volume of search queries for a particular keyword or text. The use of GT data for the current forecast of social and economic variables was introduced in 2009 [4]. Social networking sites and blogs are specifically designed to encourage users to express their feelings and opinions, which can potentially be used to predict social variables. Websites and programs (transactional platforms, opinion platforms, and dissemination of information) created on the Internet by enterprises, public organizations, charitable foundations, or multinational corporations inform about their products, services, organizational structure, and intentions. In addition to providing information, websites are used for transactions, e-commerce, and online services. According to the big data paradigm, this wide variety of sources requires specific tools for processing them.

The selection, preparation and processing of data based on which the intellectual analysis will be carried out is a stage of creating opportunities. The data to be used to identify knowledge must meet the following requirements: the available data must be, firstly, reliable, secondly, up-to-date, thirdly, sufficient to present the information as fully as possible, and fourthly, optimally necessary so as not to overload research database systems and integrate all selected data into one set. The minimum set of available data, if necessary, can be extended with additional necessary data to identify nuances that will be taken into account when creating a model. Sometimes the presence of such minor accents can be fundamental to the success of the knowledge discovery process in databases. A large number of nuances provides more opportunities to create a multidimensional model that will allow the most complete consideration of the studied phenomena and perform intellectual analysis. However, storing, organizing, and managing large and complex databases requires large resources, which are planned in advance and often limited.

Research on database systems and data stores focuses on creating, maintaining, and using databases for organizations and end-users. Database systems are known for their high scalability in processing very large, relatively structured datasets. Choosing the optimal dataset should balance the requirements of sufficiency and necessity, so this stage of data mining creates the foundation for opportunities. In addition, the choice of data should be guided by their validity and relevance.

Preparing data for further processing increases the validity of the data set. Preparation includes sorting and filtering data that will eventually be used as input data. This involves cleaning up the data by removing missing values and informational noise. Noise removal increases the chances of performing data mining most efficiently.

Removing objects with noise is an important goal of data cleansing, as noise interferes with most types of data analysis. Most existing data cleaning methods focus on noise removal, which is the product of low-level data errors that are the result of an imperfect data collection process, but irrelevant data objects or of little relevance, can also significantly impede data analysis. Thus, if the goal is to improve data analysis as much as possible, these objects should also be considered noise, at least concerning to the main analysis. Therefore, there is a need for data cleansing techniques that eliminate both types of noise. Because datasets can contain a lot of noise, these methods should also be able to discard potentially much of the data [5].

Data processing is the process of converting raw data into useful information through electronic data processing, machining, or automated means. Data processing can take time depending on the complexity of the data and the amount of input data. The preparation step described above helps to make this process faster. Data processing is usually performed step by step: raw data is collected, filtered, sorted, analyzed, stored, and then provided in an accessible format, such as graphs, charts, and documents (Fig. 2).

The data processing cycle consists of a series of steps in which raw data (input data) enters the process to obtain useful information (output). Each step is performed in a certain order, but the whole process can be repeated cyclically. The output of the first data processing cycle can be stored and presented as input for the next cycle. If the data obtained in the processing process is not used as input data for the next processing cycle, this complete process cannot be considered a cycle and will remain a one-time activity for data processing and information obtaining.

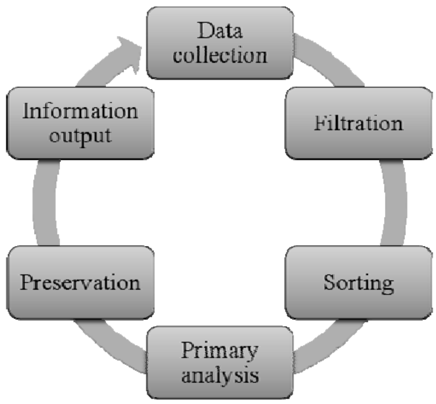


Fig. 2. Data processing cycle

Information is processed and analyzed databases. The information obtained at this stage can be useful and become the basis for the formation of knowledge. This is how the data processing phase discovers knowledge in databases. In the next stages, knowledge of social processes is identified, economic phenomena should be formed and presented in such a way as to create a model of complex socio-economic processes and systems. At the stage of data discovery and recognition of useful information and further knowledge of databases, mathematical, statistical methods, methods of artificial intelligence and machine learning are used.

The evolution of the Internet and social media has led to a huge explosion in the volume and complexity of data, so-called big data. Thus, data mining has also gone beyond traditional data modeling, such as regression and statistical models. Information theory offers tools to make formal conclusions about complex models of economic and social interaction. There are two main theoretical concepts of information that can help guide the observation of the relationship between the economic characteristics of a large number of people: entropy and mutual information. The concepts of entropy and mutual information make it possible to develop non-parametric characteristics of information associations present in the observed data generated by economic and broader social interactions [6].

Creating a model is a stage of data mining that requires special responsibility and diligence. Many algorithms can be used to model complex socio-economic processes and systems in accordance with a predetermined task. Different socio-economic processes and systems have different and complex causal relationships, so it is very important to determine the tactics of finding such connections and choose the optimal algorithm. This step involves choosing a specific method that will be used to find templates and data that most accurately describe the process or system under study. Currently, many algorithms are known to solve data mining problems: the method of reference vectors, the method of k-nearest neighbors, neural networks and decision trees.

When choosing a specific method, it is necessary to take into account the strategic goal of data mining and, accordingly, to determine the priority characteristics of the model that will be created at this stage. On the one hand, we can consider such a characteristic as the accuracy of the model, and on the other – clarity and simplicity of perception. To create a simpler model that should be intuitive, the best choice may be to use the decision tree method, which is one of the most popular methods of solving classification and forecasting problems This is a way of demonstrating rules in a hierarchical, sequential structure, where each object corresponds to a single node that provides the solution. The decision tree method should be used in cases where symbolic representation and good classification are required; the problem does not depend on many attributes; a modest subset of attributes contains relevant information; linear combinations of features are not critical; important learning speed [7].

To create an accurate model, it is appropriate to use neural networks - an extremely powerful method of modeling, which allows you to reproduce extremely complex relationships. For many years, linear modeling has been the main method of modeling in most areas, as it has well-developed optimization procedures. In problems where the linear approximation is unsatisfactory (and there are many of them), linear models work poorly. In addition, neural networks cope with the “curse of dimensionality”, which does not allow you to simulate linear relationships in the case of a large number of variables. The advantages of the neural network method are the following [8]:

– nonlinearity, neural networks are nonlinear;

– through controlled learning the network learns according to the examples: after receiving the primary information from the operator, the learning algorithm is started, which automatically perceives the data structure

– adaptability, i.e. the network can adapt its synaptic scales even in real time,

– response capability – in the context of template classification, the network not only provides template selection but also reliability of decision-making,

– fault tolerance due to massive interconnections,

– integrated large scale, i.e. its parallelism makes it potentially faster for certain tasks and thus captures complex patterns of behavior;

– homogeneity in analysis and design, i.e. the same notation is used in all in all areas related to neural networks,

– the analogy of neurobiology [9], in general, neural networks are self- adaptive and nonlinear methods that collect data and do not require specific assumptions about the basic model.

For each strategy of data mining and modeling in this process, there are several possible methods by which you can achieve your goals. The choice of a particular method is explained by the efficiency of the algorithm in a particular problem. Thus, at this stage of data mining, the most acceptable method of modeling is selected in accordance with the conditions. All these algorithms study the data and create models that are closest to the characteristics of the studied data of complex socio-economic systems and processes. Models created during data mining can be of two types: predictive or descriptive (Fig. 3).

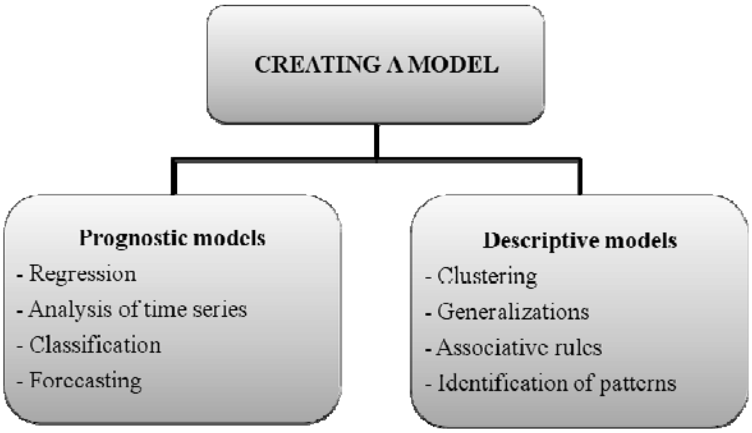


Fig. 3. Types of models, created in the process of data mining

The predictive model is a projection based on the data and information obtained in the earlier stages of data mining. As a rule, the forecast model is created on the basis of the directed analysis of data; that is, a top-down approach, where mappings from a vector input to a scalar output are obtained by applying a specific one. For example, predictive modeling can be performed using a variety of historical and statistical data. When creating a prognostic model in the process of data mining, the following tasks are performed: regression, time series analysis, classification, and forecasting [10].

The prognostic model is also known as a statistical regression. It is a monitoring method that involves explaining the relationship of several attribute values among themselves in similar elements and predicting the development of the model, that is, directional modeling based on these observations. As noted earlier, the two common methods of predictive modeling available in many data mining tools are neural networks and decision trees.

The descriptive model presents in a concise form the main characteristics of the data set. In essence, it is a collection of data points that allows you to study important aspects of a data set. As a rule, the descriptive model is created by indirect data analysis; that is, a bottom-up approach where the data "speaks for itself". Undirected data analysis finds patterns in the data set, but the patterns are interpreted by analysts. Data mining specialists determine the usability of the found templates. The most characteristic tasks of descriptive modeling are the following:

- clustering, i.e. decomposing or splitting a data set into groups;
- generalization as the process of providing summary information from data in an easier to understand form;
- association of rules – identification of causal relationships between different features in large data sets;
- sequence detection, which involves the identification of patterns of interest to researchers in the data.

Descriptive models and predictive models can (and often should) be used together in data mining. For example, it seems logical and appropriate to first look for patterns in the data using non-directional methods. These descriptive models can offer segments of data sets and ideas that improve the results of directional modeling when creating predictive models.

The modular design of neural network architecture facilitates the creation of models that simultaneously process data presented in different formats, such as creating text annotations from images, synthesizing language from the text, or through translation. This allows you to solve problems that go beyond traditional classification and regression, and is especially convenient when the data comes from different sources, which is often the case when working with big data. In addition, data obtained from different repositories or databases, presented in the form of object maps, can be reused in other contexts and, if necessary, further configured/ taught.

Evaluation and analysis of data. The purpose of any predictive modeling is to apply the model to new data. Forecasting models are useful only insofar as the quality of their prediction is adequate, therefore, the principle is not the process of creating a model as such, but the creation of a high quality model. Both predictive models and descriptive models have their evaluation criteria. For forecast models, the evaluation criterion is the accuracy of the forecast, measured by the size of the forecast error, i.e. the difference between the forecast and the actual value of the studied indicator. For descriptive models, it is more difficult to define obvious evaluation criteria, but they usually capture a discrepancy between the observed data and the proposed model. Thus, at this stage of data mining, different strategies for assessing the quality of models can be used.

Parametric methods for analyzing the accuracy of forecasts. According to the results of the ex-post-forecast, such indicators of forecast accuracy for m steps as the root mean square error are calculated, the root of standard error, mean absolute error, root of root mean square error in percentage, mean absolute relative error in percentage (MARE). The smaller the value of these values, the higher the quality of the forecast. In practice, these characteristics are used quite often. This approach gives good results, if in the period of the retro forecast there are no fundamentally new patterns. To create a prognostic model of complex socio-economic systems and processes, each time the forecast is built in a new situation, therefore, the comparison of the numerical accuracy of forecasts made at different points in time is not entirely correct. These considerations led to the use of non-parametric methods of analysis of the accuracy of forecasts [11].

Non-parametric methods of forecast accuracy analysis have two types of non-parametric criteria: label criterion and rank criterion. The criterion of labels for comparing the accuracy of two sequences of predictions is based on the percentage of cases when the method of determining the prediction A is better than the method B. Such a comparison is made for individual predictions of the same events (variables). If the ranks of their criteria are applied, the numerical characteristic of accuracy (absolute error when estimating one forecast, or root mean square error when considering a sequence of predictions) is replaced by ranks, which are then checked for significance. For example, if the sequences of predictions of indicators A and B are obtained using k methods, then first calculate the root mean square error, then the values are ranked from smallest to largest. Although non-parametric methods have their advantages, it is important to realize that they ignore some of the available information. Thus, the criteria of labels and ranks do not take into account the numerical values of errors [11].

Data visualization. Created models used in the process of data mining of complex socio-economic systems and processes including large and complex parameters. To solve the problem of size and complexity, the best methods are used to represent complex systems and data visualization (for example, advanced user interfaces). These technologies increase the level of abstraction, which helps users focus on the most important components and properties of complex models.

In the world of big data, data visualization tools and technologies are needed to analyze large and complex amounts of information and make decisions based on the intellectual analysis of this data. Data visualization is a graphical representation of information and data. Using visual elements such as charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, deviations, and patterns in data. Effective data visualization is a delicate balance between form and function. On the one hand, the simplest schedule can be both a very primitive transfer of information and a vision of the main core of information that is analyzed and must be presented and understood. On the other hand, the most complex visualization can overload information, and say about many details, but not convey the main essence of the message. Data and visual elements must work together to create a better understanding and awareness of information.

There is a choice of visualization methods for efficient and interesting data presentation. Common types of data visualization: charts, tables, graphs, maps, infographics, dashboards, etc.

Construction and illustration of relationships between different objects of the created models of complex socio-economic systems and processes can be done with the help of modern tools. Therefore, Draw.io is a free, intuitive browser-based flowchart builder where users can drag object shapes (including ellipses and parallelograms common to data models) onto a canvas, and then combine them into by means of through connecting lines. Lucidchart Chart Designer is similar to Draw.io, but it reproduces streams that are more complex and has more reliable data protection. SQuirreL is a free and open-source graphical tool supported by most major relational databases.

The most important trend in the field of data in recent years is the proliferation of data catalogs, largely due to privacy rules such as the GDPR and the CCPA (General Data Protection Regulation of 2016; California Consumer Privacy Act of 2018) [12]. This trend has not escaped the field of data mining and modeling. The line between data discovery tools and applications and data modeling tools is increasingly blurred, as exemplified by Amundsen, a metadata-based data discovery platform developed by Lyft [13].

Open source Metabase is a GUI tool with some useful analytics visualizations but does not support modeling tools [14]. Other notable data visualization tools include erwin, ER / Studio, SAP PowerDesigner, IBM InfoSphere Data Architect, and Microsoft SQL Server Management Studio, etc.

Application of deep learning. The knowledge and models of processes created in the socio-economic systems created in the process of data mining will be popular only if they can be included in other complex systems in order to predict their development. Forecasting analytics is interesting and useful in the context of the possibility of making changes to the simulated system and presenting the long-term consequences of these changes. The real structure of complex socio-economic systems is dynamic, data characteristics may change over time, new parameters may appear that were not foreseen in the model, and others may disappear. Therefore, this stage of application of deep learning determines the success and effectiveness of the whole process of data mining.

Progress in the field of deep learning has made it possible to use the powerful capabilities of artificial neural networks in this process. They are a universal tool for data mining and effective for learning based on data presented in various formats. For example, neural networks have demonstrated their effectiveness in performing certain tasks of object or image recognition [15–16]. Recent approaches have shown that deep learning can effectively learn based on data representations in variable length sequences (time series, sound, language, and text), graphs and networks, including social networks, natural language and even source code in computer programs [17–18].

Another aspect of deep neural networks that is closely related to big data is their ability to perform complex functional design. The problem of big data is often related to the difficulty of making reliable predictions when training data needs to be represented, for example, to identify successfully relevant classes of solutions. An important requirement for the use of deep learning methods is the availability of large samples of training, as insufficient training data causes the problem of "overfitting" when the model does not summarize the information obtained during training, but simply remembers it. In this case, the model shows good results on educational data but does not show such accuracy on unfamiliar data [19].

Previously, the search for useful representations had to be conducted by experts using manual design of characteristics or explicit methods of selection and construction of functions [20]. At the present stage of technology development, the most suitable architecture for data processing, which characterizes complex socio-economic processes and systems, and as a consequence to solve the problem of data mining are convolutional neural networks, because they are designed to process data in the form of multidimensional arrays [21].

Some neural models on the ethane of deep learning allow you to synthesize features in the form of hidden variables with certain desired properties. Neural networks help automate the tasks set at the beginning of the data mining process complex socio-economic systems: the construction of the characteristics of these systems becomes an integral part of the process of deep learning, closely related to the search in space for new hypotheses for the development of socio-economic processes.

CONCLUSIONS

Thus, the study of the data mining process showed that the expansion of the data analysis tools in connection with the powerful development of technologies, the formation of big data sets creates the ability to track, evaluate, simulate, and ultimately include key economic and social changes and trends in complex processes and systems. An important step that has increased the efficiency of data mining has been the inclusion of steps to facilitate data production as well as model evaluation and visualization.

The descriptive and predictive models generated by the mining process can and should be used together. The logical sequence of the model application, which will improve the results of the directional modeling, is seen primarily in the search for patterns in the data with the help of descriptive models, and already based on the obtained ideas of directional modeling when creating predictive models of complex socio-economic processes and systems.

At the present stage of technology development, machine learning is widely used in data mining to invent complex models and algorithms that serve to create descriptive and predictive models of complex socio-economic systems and processes. Machine learning gives computers the ability to "learn", recognize complex patterns and make intelligent decisions without explicit programming based on large data samples. These opportunities are the basic application of deep learning methods, designed to process data presented in the form of multidimensional arrays, and allow you to create models of complex socio-economic processes and take into account possible changes to design and manage the development of complex systems. That is, the use of the above tools allows you to perform successfully and efficiently the tasks of data mining of complex socio-economic processes and systems.

Therefore, digital tools are becoming relevant to maintain effective competitiveness, help model complex socio-economic processes and systems, effectively analyze and use existing large data sets for operational human resource management and strategic planning of complex socio-economic processes and systems.

REFERENCES

1. R.R. Nisbet, G. Miner, and K. Yale, "Chapter 2 – Theoretical Considerations for Data Mining," Editor(s): Robert Nisbet, Gary Miner, Ken Yale, *Handbook of Statistical Analysis and Data Mining Applications (Second Edition)*, Academic Press, 2018, pp. 21–37. Available: https://doi.org/10.1016/B978-0-12-416632-5.00002-5
2. O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*, 2nd ed. Springer, January 2010, 1285 p. Available: https://doi.org/10.1007/978-0-387-09823-4
3. H. Jiawei, M.Kamber, and J. Pei, *Data mining: concepts and techniques*, 3rd ed. Morgan Kaufmann Publishers, 2012, pp. 23–27.
4. H. Choi and H. Varian, *Predicting the Present with Google Trends*, 2009. Available: http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en//googleblogs/pdfs/google_predicting_the_present.pdf
5. H. Xiong, G. Pandey, M. Steinbach, and V. Kumar, "Enhancing data analysis with noise removal," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, iss. 3, pp. 304–319, 2006. Available: http://datamining.rutgers.edu/publication/ tkdehcleaner.pdf
6. P.L. dos Santos, and N. Wiener, "Indices of Informational Association and Analysis of Complex Socio-Economic Systems," *Entropy*, 21(4), 367, 2019. Available: https://doi.org/10.3390/e21040367
7. J. Stefanowski, *Discovering Decision Trees*. Institute of Computing Science. Poznań University of Technology, 2010, 45 p. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.176.1423&rep=rep1&type=pdf
8. E. Gómez-Ramos and F. Venegas-Martínez, "A Review of Artificial Neural Networks: How Well Do They Perform in Forecasting Time Series?" *Analítika, Revista de análisis estadístico*, vol. 6, no. 2, pp. 7–15, 2013.
9. R. Tadeusiewicz, *Neural networks: A comprehensive foundation: by Simon HAYKIN.* USA, New York: Macmillan College Publishing, 1995, 696 p.
10. U. Johansson, *Obtaining Accurate and Comprehensible Data Mining Models – An Evolutionary Approach.* Linköping, Sweden: Department of Computer and Information Science, Linköpings universitet, 2007, 272 p. Available: http://www.diva-portal.org/smash/get/diva2:23601/FULLTEXT01.pdf
11. G.V. Prisenko and E.I. Ravikovich, *Forecasting of socio-economic processes*: *Textbook*. K: KNEU, 2005, 378 p.

12. "Comparing privacy laws: GDPR v. CCPA," *Data Guidance and Future of Privacy Forum*, 42 p. Available: https://fpf.org/wp-content/uploads/2018/11/GDPR_CCPA_ Comparison-Guide.pdf
13. *Amundsen. Open source data discovery and metadata engine.* Available: https://www.amundsen.io/
14. *Metabase. Built for data*. Available: https://www.metabase.com/
15. A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, 25, pp. 1097–1105, 2012.
16. Jiquan Ngiam et al., "Multimodal deep learning," *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011.
17. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, 521(7553), pp. 436–444, 2015.
18. J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, 61(C), pp. 85–117, 2015.
19. M. Lavreniuk, N. Kussul, and A. Novikov, "Deep Learning Crop Classification Approach Based on Coding Input Satellite Data Into the Unified Hyperspace," *IEEE 38th International Conference on Electronics and Nanotechnology*, pp. 239–244, 2018.
20. K. Krawiec, "Evolutionary feature selection and construction," in S. Claude and G. Webb (Eds.) *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer, 2016.
21. M. Reichstein et al., "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, iss. 7743, pp. 195–204, 2019.

Received 16.05.2022

INFORMATIONON THE ARTICLE

Tetyana V. Obelets, ORCID: 0000-0002-1553-5150, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: obelectv@ukr.net

ІНСТРУМЕНТАРІЙ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ СКЛАДНИХ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ ТА СИСТЕМ / Т.В. Обелець

Анотація. Розглянуто процес виявлення нової та потенційно корисної інформації з великих обсягів даних, що актуалізує роль розроблення інструментарію інтелектуального аналізу даних для складних соціально-економічних процесів та систем на основі принципів цифрової економіки та їх оброблення за допомогою мережевих застосунків. Окреслено етапи інтелектуального аналізу даних для складних соціально-економічних процесів та систем. Розглянуто алгоритм інтелектуального аналізу даних. Визначено, що використовувані раніше етапи інтелектуального аналізу даних, які обмежувалися лише процесом побудови моделі, можуть бути розширені завдяки використанню більш потужної обчислювальної техніки та появи у вільному доступі великої кількості багатовимірних даних. До наявних етапів інтелектуального аналізу даних для складних соціально-економічних процесів та систем включено процеси полегшення підготовки даних, оцінювання та візуалізацію моделей, а також глибинне навчання. Визначено інструментарій інтелектуального аналізу даних для складних соціально-економічних процесів та систем у контексті технологічного прогресу та відповідно до парадигми великих даних. Досліджено циклічність оброблення даних; цей процес складається із серії кроків, починаючи із входу необроблених даних, закінчуючи виведенням корисної інформації. Отримані на етапі оброблення даних знання закладаються в основу створення моделей складних соціально-економічних процесів та систем. Окреслено два типи моделей (описову та прогностичну), що можуть бути створені у процесі інтелектуального аналізу даних. Визначено алгоритми оцінювання та аналізу даних моделювання складних соціально-економічних процесів та систем відповідно до заздалегідь поставленого завдання. Проаналізовано ефективність запровадження нейронних мереж та методів глибинного навчання, що застосовуються у процесі інтелектуального аналізу даних. Визначено, що вони дозволять ефективно аналізувати та використовувати наявні великі масиви даних як для оперативного управління людськими ресурсами, так і стратегічного планування розвитку складних соціально-економічних процесів та систем.

Ключові слова: інтелектуальний аналіз даних, складні соціально-економічні системи, прогностичне моделювання, нейронні мережі, глибинне навчання.

МЕТОДИ ОПТИМІЗАЦІЇ, ОПТИМАЛЬНЕ УПРАВЛІННЯ І ТЕОРІЯ ІГОР

UDC 004.8:004.032.26
DOI: 10.20535/SRIT.2308-8893.2022.4.07

CLUSTERIZATION OF VECTOR AND MATRIX DATA ARRAYS USING THE COMBINED EVOLUTIONARY METHOD OF FISH SCHOOLS

Ye. BODYANSKIY, A. SHAFRONENKO, I. PLISS

Abstract. The problem of clustering data arrays described in both vector and matrix forms and based on the optimization of data distribution density functions in these arrays is considered. For the optimization of these functions, the algorithm that is a hybrid of Fish School Search, random search, and evolutionary optimization is proposed. This algorithm does not require calculating the optimized function's derivatives and, in the general case, is designed to find optimums of multiextremal functions of the matrix argument (images). The proposed approach reduces the number of runs of the optimization procedure, finds extrema of complex functions with many extrema, and is simple in numerical implementation.

Keywords: combined optimization, fuzzy clustering, evolutionary algorithms, density functions, Fish School.

INTRODUCTION

The problem of clustering arrays of arbitrary nature observations is integral part of Data Mining, and more generally Data Science. To solve this problem it was proposed a lot of approaches that differ as a priori assumptions about the physical nature of data and problems solved by their basis, and the mathematical apparatus that was used [1–4]. From a computational point of view, the simplest are the so-called hierarchical methods and algorithms based on partitions [3], among of that we should mention the k-means procedure, that has become widespread to solve a variety of problems. It should be noted here that the most adequate mathematical apparatus for solving clustering problems are methods of computational intelligence [5–7] and, above all, artificial neural networks, fuzzy systems, evolutionary optimization and so-called hybrid systems of computational intelligence that connect these three areas. It is interesting to note that one of the most popular neural networks — self-organizing Kohonen maps [8] actually implements the k-means procedure, presented in recurrent form.

It should be noted that in the general case the solution of the clustering problem is significantly complicated if the original vectors (in the general case matrices) observations have a large variety are, distorted by perturbations and noises, contain outliers and omissions, the original arrays themselves or too large (Big Data) or too short, clusters can have a rather complex shape, and their number is a priori unknown.

In this case, the most effective (but also the most complex) are algorithms based on the analysis of data distribution densities, among which as one of the most "popular" are DENCLUE [9] and its modifications [10–12], which were proposed to solve clustering problems of large arrays of high-dimensional vector data, and the classes formed in the clustering process can have any complex shape. At the heart of these algorithms is the search for extremes — maxima in the data density functions in the analyzed array (multi-extremal optimization), and this function is formed as a superposition of kernel (bell-shaped) functions associated with each observation. In fact, this function is based on Parzen windows [13] and Nadaraya–Watson estimates [14, 15].

From a computational point of view, the clustering problem becomes of finding local extrema of the multiextrema function of the density vector argument using gradient procedures that are repeatedly run from different points in the original data set. It is clear that this takes a long time, because a priori it is not even known how many extremes the formed density function.

The process of finding these extremes can be accelerated by using the ideas of evolutionary optimization, that includes algorithms inspired by nature, swarms algorithms, population algorithms, etc. [16–18]. In this case, the search is conducted simultaneously by a group of agents acting either independently or in interaction, which can significantly speed up the process of finding extremes, each of which "corresponds" to one or another cluster that is being formed.

FORMATION OF THE DATA DISTRIBUTION DENSITY FUNCTION IN THE CLUSTERING ARRAY

The initial information for solving the clustering problem is traditionally an array of observation vectors $X = \{x(1), x(2),..., x(k),..., x(N)\}$, $x(k) = \{x_i(k)\} \in R^n$, the data are pre-centered on a hypercube so that $x(k) = \{x_{i,i_2}(k)\} \in R^{n_1 \times n_2}$. This situation can occur in the case of image array processing. The basic concepts on which DENCLUE is based are the influence function, the density function and the density attractors, which are essentially the local extremes of the density function. In the general case, the influence function for any vector observation $x(\bullet)$ from the original array X is a kernel bell shaped function $f^{x(\bullet)}(x)$, in this case the most popular is the traditional Gaussian one

$$f_G^{x(\bullet)}(x) = \exp\left(-\frac{d^2(x, x(\bullet))}{2\sigma^2}\right) = \exp\left(-\frac{\|x - x(\bullet)\|^2}{2\sigma^2}\right), \tag{1}$$

where $d^2(x, x(\bullet))$ — euclidean distance; σ^2 — parameter of the influence function width, due to the simplicity of calculating its derivatives.

In the matrix case, instead of the Euclidean one, we can use the Flobenius metric, and the influence function takes the form

$$f_G^{x(\bullet)}(x) = \exp\left(-\frac{d^2(x, x(\bullet))}{2\sigma^2}\right) = \exp\left(-\frac{Tr(x - x(\bullet))(x - x(\bullet))^T}{2\sigma^2}\right), \tag{2}$$

where $Tr(\bullet)$ — matrix trace symbol.

It is easy to see that (2) is a generalization of (1).

Based on the influence functions, formed the data density distribution function in the array X in the form

$$f^X(x) = \sum_{k=1}^{N} f\left(x, x(k)\right), \qquad (3)$$

which is essentially an estimate of Nadaraya–Watson. It's easy to see what the function $f^x(x)$ can take values in an interval $1 \le f^x(x) \le N,$ in this case the extrema values from this interval are accepted when the sample contains only one observation or all N observation observations coincide, i.e. there is only one cluster — a degenerate situation.

To find $m > 1$ clusters it's necessary to introduce some threshold $\xi > 1$, that allows to build really significant clusters by excluding anomalous observations and classes that contain too small data.

Actually, the process of cluster formation is associated with finding all extremes of the density function (3) using a gradient procedure

$$x^l = x^{l-1} + \eta^l \frac{\nabla f^X(x^l, x^{l-1})}{\left\| \nabla f^X(x^l, x^{l-1}) \right\|}, \; x_0 = x(k), \; l = 0,1,2,...; \; \forall k = 1,2,...,N, \qquad (4)$$

i.e. the number of runs of algorithm (4) is determined by the size of the training sample N. It is clear that with large N the process of clustering — finding local extrema can take a lot of time. Therefore, the proposed modifications of DENCLUE are associated with speeding up the process of finding local extrema (3) by modifying the gradient procedure (4) [10–12].

In the case, when observations $x(k)$ in dataset $X \in \left(n_1 \times n_2\right)$ are matrices, it is easy to consider the matrix version of the procedure (4):

$$x^l = x^{l-1} + \eta^l \Gamma^X(X, x^{l-1})(Tr\Gamma^X(X, x^{l-1})\Gamma^{xT}(X, x^{l-1}))^{-(1/2)},$$

where $\Gamma^X(X, x^{l-1}) = \left\{ \dfrac{\partial f^X(X, x^{l-1})}{\partial x_{i_1 i_2}} \right\} \in R^{n_1 \times n_2}.$

The gradient optimization process ends with a search m local extrema of function (3), with less value ξ, than more clusters can be formed.

It is possible to speed up the process of finding local extrema by using evolutionary optimization methods instead of gradient search, among which the so-called Fish schools search can be noted as quite efficient, numerically simple and fast [19–21], which should be modified to solve the clustering problem.

MODIFIED OPTIMIZATION METHOD BASED ON FISH SCHOOL

When using the methods of evolutionary optimization, which are essentially zero-order optimization methods, i.e. do not use derivatives, it is assumed that when finding the extrema of some function $f^x(x)$ the population of agents are used, each of them acts either independently or in interaction with others, with the movement of each q^{th} agent $(q = 1,2,...,Q)$ on l^{th} search iteration can be written as:

$$x_q^l = x_q^{l-1} + \eta_q^l Dir_q^l, \; q = 1,2,...,Q,$$

where $x_q^l = (x_{q1}^l, x_{q2}^l, ..., x_{qn}^l)^{\mathrm{T}}$; Dir_q^l — vector that specifies the direction of movement q^{th} agent on l^{th} search iteration.

In a large family of such methods should be noted the method based on of fish schools, where each agent of the population simulates the movement of an individual fish in the school [19–21].

The main advantage of this method is the sufficient efficiency of finding the global extrema of rather complex functions, which include the density function of data distribution in clustering problems.

The authors of the method introduce iterations related to the movement of the school: feeding and swimming.

The feeding operator is responsible for the weight of each fish as an element of school — the agent. The heavier the fish, the closer it is to the extreme — the maximum. The weight of each fish w_q is tuned according to the expression

$$w_q^l = w_q^{l-1} + \frac{f^x(x_q^l) - f^x(x_q^{l-1})}{\max\limits_p \{f^x(x_q^l) - f^x(x_q^{l-1})\}} \quad \forall q = 1,2,...,Q, \tag{5}$$

where

$$0 < w_q^l < w_{\max},\ w_l^0 = 0{,}5 w_{\max}.$$

The swimming operator describes both the individual movement of each fish and the collective movement of the school as a whole. Three types of movement are considered here: individual, instinctively — collective and collective volitional.

Individual movement is described by the relation

$$x_{qi}^l = \begin{cases} x_{qi}^l + \eta_q^l Rand\{0,1\}, \text{ if } f^x(x_q^l) > f^x(x_q^{l-1}); \\ x_q^{l-1} \text{ else,} \end{cases} \tag{6}$$

where $Rand\{0,1\}$ — evenly distributed in the interval $(0,1)$ random number. It should be noted that (6) is essentially a local random search with return, introduced by L. Rastrigin [22]. In fact, this is the procedure of "probing" the function $f^x(x)$ around the point x_q^{l-1}, in this case, in addition to (5), any other random search algorithm can be used here.

On the basis of probing the density function with the help of individual movement (5) the instinctive-collective movement in the direction of growth of this function is realized as:

$$x_q^l = x_q^{l-1} + \frac{\left(\sum\limits_{p=1}^{Q}(x_p^l - x_p^{l-1})\right)(f^x(x_q^l) - f^x(x_q^{l-1}))}{\sum\limits_{p=1}^{Q}(f^x(x_p^l) - f^x(x_p^{l-1}))}. \tag{7}$$

At this stage, there is a balanced averaging of individual movements, taking into account the "success" of each of the fish-agents.

And finally, collectively-volitional movement, when all the fishes of the school "pull" to the weighted center of gravity, if the cant goes to the extreme, and "run away" if the population moves in the wrong direction.

Considering the weighted center of gravity of the fish school

$$Bar^l = \frac{\sum_{p=1}^{Q} x_p^l w_p^l}{\sum_{p=1}^{Q} w_p^l}, \tag{8}$$

we can record this movement as

$$x_q^l = \begin{cases} x_q^l - \eta_q^l Rand\{0,1\} \dfrac{x_q^{l-1} - Bar^{l-1}}{\left\| x_q^{l-1} - Bar^{l-1} \right\|}, \text{ if } \sum_{p=1}^{Q} w_p^l > \sum_{p=1}^{Q} w_p^{l-1}; \\ x_q^l + \eta_q^l Rand\{0,1\} \dfrac{x_q^{l-1} - Bar^{l-1}}{\left\| x_q^{l-1} - Bar^{l-1} \right\|}, \text{ if } \sum_{p=1}^{Q} w_p^l < \sum_{p=1}^{Q} w_p^{l-1}. \end{cases} \tag{9}$$

To increase the efficiency of FSS, an additional breeding operator may be introduced, which allows the creation of new fish-agents that have improved characteristics compared to existing members of the school. To do this, we can use the ideas of evolutionary operations [23], among which from a computational point of view and efficiency — the credibility of finding the extrema can be noted sequential simplex method [24] and its modifications [25].

Let's form the school that containing $Q = n + 1$ fish-agents, but this number remains unchanged in the search process, i.e. the population $x_1^0, x_2^0, ..., x_Q^0$ generated randomly. In this population we find the "worst" fish x_{qworst}^0 , which has the lowest weight $w_{q\min}^0$ and the "best" fish x_{qbest}^0 with the greatest weight $w_{q\max}^0$. The main operation of the simplex movement is mapping x_{qworst}^0 through the center of gravity n fishes (without the worst), which can be written in the form

$$\bar{x}^0 = \frac{1}{n} \sum_{q=1}^{Q} (x_q^0 - x_{qworst}^0).$$

As a result of this operation, a new fish is created

$$x_q^{1*} = \bar{x}^0 + \alpha(\bar{x}^0 - x_{qworst}^0),$$

which replaces the worst individual in the school x_{qworst}^0. Thus creates a new population $x_1^1, x_2^1, ..., x_Q^1$. Here $0{,}5 \le \alpha \le 2$ — parameter that controls the shape of the school-simplex in the optimization process. In the case, when $\alpha = 1$ the mapping of the simplex through the center of gravity is realized $\bar{x}^0$ in the case if $f^x(x_q^{1*}) > f^x(x_{best}^0)$ accepted $\alpha = 2$, that is, the school is "stretched" in a favorable direction, if $f^x(x_q^{1*}) < f^x(x_{worst}^0)$ $\alpha = 0{,}5$ is accepted, that is, the simplex faces a relatively unfortunate direction. Thus the motion of the school-simplex can be described by relations

$$\begin{cases} \bar{x}^{l-1} = \dfrac{1}{n}\sum_{q=1}^{Q}(x_q^{l-1} - x_{qwost}^{l-1}); \\ x_q^l = \bar{x}^{l-1} + \alpha(\bar{x}_q^{l-1} - x_{qwost}^{l-1}), \end{cases} \quad (10)$$

then in the general case it is essentially an Nelder–Mead optimization algorithm [25]. Thus, in the process of finding the extreme, the worst fishes with the lowest weight are removed from the cant and new agents with higher weight are created.

Thus, the process of combined optimization of the density function (5)–(10) is essentially a combination of FSS, random search and evolutionary operating based on the Nelder–Mead method.

Since the problem under consideration is essentially a problem of multiextrema optimization, it is necessary to find a set of extremums, each of which is a centroid of a cluster. Therefore, the optimization problem must be solved repeatedly at different values σ^2 and ξ. When finding any of the extremes from the original sample X observations located directly in its vicinity are excluded. After this removal, the proposed procedure of combined evolutionary optimization is repeated until all extrema centroids are found.

EXPERIMENTAL RESEARCH

The experimental research was conducted on two databases, such us Page blocks and Spam base and two test multiextrema functions. The description of datasets shown in Table 1 and test multiextrema functions in Table 2.

Table 1. Data set description

Dataset	Instances	Attributes	Clusters
Page blocks	5472	10	5
Spam base	4601	57	2

Table 2. Test multiextreme functions

Name of function	Formulas	Domain	Step
Rastrigin	$f(x) = 20 + x^2 + y^2 - 10\cos(2\pi x) + \cos(2\pi y)$	[–5.12;5.12]	0.01
Griewangk's	$f(x) = \dfrac{1}{4000}x + \dfrac{1}{4000}y - \cos\left(\dfrac{x}{\sqrt{1}}\right)\cos\left(\dfrac{x}{\sqrt{2}}\right) + 1$	[–30; 30]	0.1

Due the fact, that Rastrigin's and Griewangk's functions has a lot if local extreme points in its search area, as shown on Fig. 1, *a* and Fig. 2, *a*, we add 514 agents.

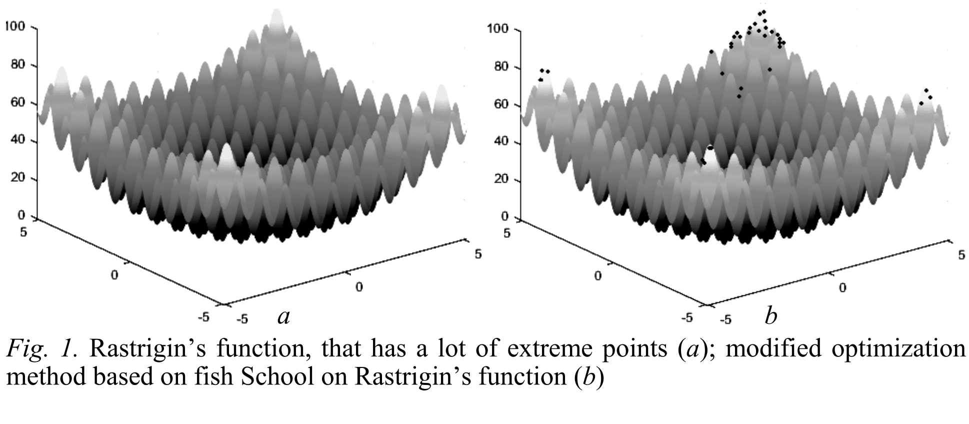


Fig. 1. Rastrigin's function, that has a lot of extreme points (*a*); modified optimization method based on fish School on Rastrigin's function (*b*)

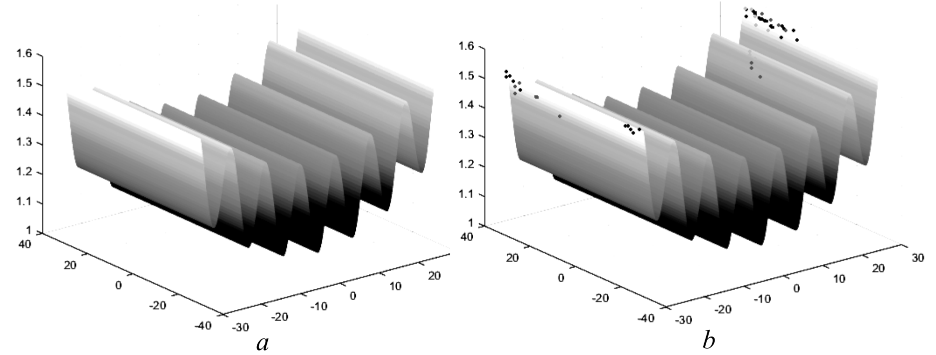


Fig. 2. Griewangk's function, that has a lot of extreme points (*a*); modified optimization method based on fish School on Griewangk's function (*b*)

In Page blocks dataset was presents classified blocks of the page layout in a document that has been detected by a segmentation process. Spam base dataset also extracted from the UCI Machine Learning Repository and describes e-mail classified as spam or not spam.

The accuracy comparison of the well known optimization algorithms such as Fish School (FSS) and Cat Swarm (CSO) and proposed Modified Optimization Method Based on Fish School (OMFS).

Table 3. Accuracy comparison

Data	Accuracy	OMFS	FSS	CSO
Rastrigin	Mean	**190.46**	189.65	**190.46**
	Best	**195.83**	195.59	**195.83**
Griewangk's	Mean	**3.65**	3.41	3.65
	Best	**4.82**	4.12	4.81
Page blocks	Mean	**951.47**	951.01	951.15
	Best	**959.64**	959.43	959.55
Spam base	Mean	**291.77**	291.17	**291.77**
	Best	**299.84**	299.48	299.64

From obtained result, that shown in Table 3 we see, that Modified Optimization Method Based on Fish School generally perform better than the original algorithm. The convergence process of hybrid algorithm demonstrate in Fig. 3.

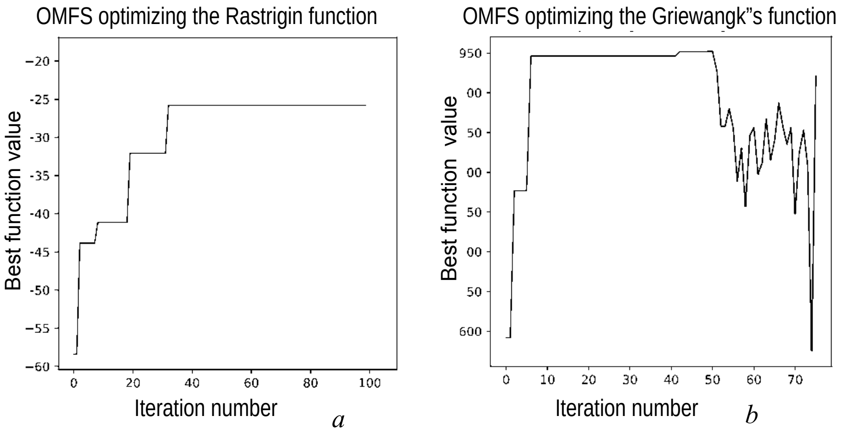


Fig. 3. Modified Optimization Method Based on Fish School on test function: Rastrigin's function (*a*) and Griewangk's function (*b*)

To evaluate the performance of clustering method used the several validity metrics: Dunn Index (DI) — high value indicates a better clustering; Davies-

Bouldin Index (DBI) — the smallest value indicates the better clustering; Cluster Accuracy (CA) — the high value indicates the best clustering quality.

For comparison proposed method classification of vector and matrix data sets based on combined optimization of distribution functions (CODF) against classical DENCLUE algorithm and DENCLUE-IM for big data clustering (Table 4).

Table 4. The comparison algorithms according to their validity metric

Data	Measures	Spam base	Page blocks
CODF DENCLUE DENCLUE-IM	DI	**0.835** 0.789 0.831	**0.721** **0.721** 0.693
CODF DENCLUE DENCLUE-IM	DBI	0.768 0.867 **1.041**	**0.764** 0.864 1.041
CODF DENCLUE DENCLUE-IM	CA	**0.718** **0.805** 0.701	**0.920** **0.920** 0.911

All these results conclude that proposed method classification of vector and matrix data sets based on combined optimization of distribution functions has an acceptable clustering performance.

CONCLUSION

The problem of clustering data arrays that are described in both vector and matrix forms based on the optimization of data distribution density functions in these arrays is considered. For optimization of these functions — local extrema search we have proposed the hybrid of Fish School Search algorithm, random search and evolutionary optimization. This algorithm does not require the calculation of derivatives of the function, which is optimized and in the general case is designed to find the maxima of multiextrema functions of the matrix argument (images).

The proposed approach allows to reduce the number of the optimization procedure runs, allows to find the extremes of complex shape functions and is easy in numerical implementation.

REFERENCES

1. G. Gan, Ch. Ma, and J. Wu, *Data Clustering: Theory, Algorithms and Applications.* Philadelphia, Pennsylvania: SIAM, 2007, 455 p.
2. J. Abonyi and D. Feil, *Cluster Analisis for Data Mining and System Identification*. Basel, Birdhouses, 2007, 303 p.
3. R. Xu and D.C. Wusch, *II - Clustering*. Hoboken, N.J.: John Willey Sons, Inc., 2009, 341 p.
4. C.C. Aggarwal, *Data Mining: Text Book*. Springer, 2015.
5. A.P. Engelbrecht, *Computational Intelligence An Introducion*. John Willey& Sons, 2007, 597 p.
6. L. Rutkowski, *Computational Intelligence Methods and Techniques*. Berlin Heidelberg: Springer-Verlag, 2008, 514 p.
7. A. Kroll, *Computational Intelligence. Eine Einfürung in Problelme, Methoden and Tchnische Anwendungen*. München: Oldenbourg Verlag, 2013, 428 p.
8. T. Kohonen, *Self-Organizing Maps*. Berlin: Springer, 1995, 362 p. doi: 10.1007/978-3-642-56927-2.
9. A. Hinneburg and D.A. Keim, ”An efficient approach to clustering in large multimedia databases with noise,” *Proc. 4th Int. Conf. in Knowkedge Discovery and Data Mining (KDD 98)*, N.Y.: AAAI Press, 1998, pp. 58–65.
10. A. Hinneburg and H.-H. Gabriel, ”DENCLUE 2.0: Fast clustering based on kernel density estimation,” *Proceedings of the 7th International Symposium on Intelligent Data*

Analysis, New York: Springer Science + Business Media, 2007, pp. 70–80. doi:10.1007/978-3-540-74825-0_7.

11. A. Hinneburg and D.A. Keim, "A general approach to clustering in large databases with noise," *Knowledge and Information Systems*, 5 (5), pp. 387–415, 2003.
12. H. Rehioni, A. Idrissi, M. Abourezq, and F. Zegrary, "DENCLUE-IM: A new approach for big data clustering," *Procedia Computer Science*, 83, pp. 560–567, 2016.
13. E. Parzen, "On estimation of a probably density function and mode," *The Annals of Math Statistics.*, 33, no. 3, pp. 1065–1076, 1962.
14. E.A. Nadaraya, "On nonparametric estimation of density function and regressiion curves," *Theory of Probab. Appl.*, 10, pp. 186–190, 1965.
15. G.S. Watson, "Smooth regression analysis," *Sankhya: The Indian Journal of Statistics, Ser. A.*, 26, no. 4, pp. 359–372, 1964.
16. J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proc. IEEE Int. Conf. on Neural Networks*, Perth, Australia, 1995, pp. 1942–1948.
17. A. Eiben and J. Smith, *Introduction to Evolutionary Computing*. Heidelberg: Springer, 2003.
18. A.P. Karpenko, "Population algorithms for global continious optimization," *Review of new and little-known algorithms. Supplement to the journal "Information Technologies"* №7/2012, 32 p.
19. C.J.A. Bastos-Filho, F.B. Lima Neto, A.J.C.C. Lins, A.I.S. Nascimento, and M.P. Lima, "Fish School Search," *Nature-Insperiod Algorithms for Optimization*. Berlin Hedelberg: Springer Verlag, 2009, SCI 193, pp. 261–277.
20. Jr. G.M.Cavalcanti, C.J.A. Bastos-Filho, F.B. Lima Neto, and R.M.C.S. Castro, "A hybrid algorithm based on fish school search and particle swarm optimization for dynamic problems," *Proc. Int. Conf. in Swarm Intelligence (ICSI), 2011*, vol. 2, pp.543–552.
21. A. Janecek and Y. Tan, "Feeding the fish-weight update strategies for the fish school seach algorithm," *Lecture Notes in Computer Scince*. Berlin Heidelberg: Springer-Verlag, 2011, vol. 6729, Part II, pp. 553–562.
22. L.A. Rastrigin, *Random search in adaptation processes*. Riga: Zinatne, 1973, 132 p.
23. Y.E.P. Box, "Evolutionary operation: A method for increasing industrial productivity," *Applied Statistics*, 6, pp. 81–101, 1957.
24. W. Spendley, G.R. Hext, and F.R. Himswath, "Sequential application of simplex design in optimization and evolutionary operation," *Tehnometrics,* 4, pp. 441–461, 1962.
25. J.A. Nelder and R. Mead, "A simplex method for function minimization," *Computer J.*, 7, pp. 308–313, 1965.

Received 30.06.2022

INFORMATION ON THE ARTICLE

Yevgeniy V. Bodyanskiy, ORCID: 0000-0001-5418-2143, Kharkiv National University of Radio Electronics, Ukraine, e-mail: yevgeniy.bodyanskiy@nure.ua

Alina Yu. Shafronenko, ORCID: 0000-0002-8040-0279, Kharkiv National University of Radio Electronics, Ukraine, e-mail: alina.shafronenko@nure.ua

Iryna P. Pliss, ORCID: 0000-0001-7918-7362, Kharkiv National University of Radio Electronics, Ukraine, e-mail: iryna.pliss@nure.ua

КЛАСТЕРИЗАЦІЯ ВЕКТОРНИХ ТА МАТРИЧНИХ МАСИВІВ ДАНИХ ІЗ ВИКОРИСТАННЯМ КОМБІНОВАНОГО ЕВОЛЮЦІЙНОГО МЕТОДУ РИБНИХ ЗГРАЙ / Є.В. Бодянський, А.Ю. Шафроненко, І.П. Плісс

Анотація. Розглянуто задачу кластеризації масивів даних, що описано як у векторній, так і матричній формі на основі оптимізації функцій щільності розподілу даних у цих масивах. Для оптимізації цих функцій – пошуку локальних екстремумів запропоновано алгоритм, що є гібридом Fish School Search, випадкового пошуку та еволюційної оптимізації. Цей алгоритм не потребує обчислення похідних функції, що оптимізується, і у загальному випадку призначений для відшукання максимумів багатоекстремальних функцій матричного аргумента (зображень). Запропонований підхід дозволяє скоротити кількість запусків процедури оптимізації, знаходити екстремуми функцій складної форми та є простим у числовій реалізації.

Ключові слова: комбінована оптимізація, нечітка кластеризація, еволюційні алгоритми, функція щільності, Fish School.

UDC 519.87:004.02
DOI: 10.20535/SRIT.2308-8893.2022.4.08

ON SOME METHODS FOR SOLVING THE PROBLEM OF POWER DISTRIBUTION OF DATA TRANSMISSION CHANNELS TAKING INTO ACCOUNT FUZZY CONSTRAINTS ON CONSUMPTION VOLUMES

E.V. IVOKHIN, L.T. ADZHUBEY, P.R. VAVRYK, M.F. MAKHNO

Abstract. The article deals with the mathematical formulation of the problem of optimal distribution of the power of data transmission channels in information and computer networks with a three-level architecture and fuzzy restrictions on consumption volumes. An efficient algorithm has been developed for solving the problem, the peculiarity of which is the inability to meet the end user's needs at the expense of the resources of different suppliers. A standard solution method based on a fuzzy optimization problem of mathematical programming is considered. A constructive variant of finding a solution based on the backtracking method is proposed. Computational experiments have been carried out. The developed approach was used to determine the optimal configuration of a three-level information and computer network with a given number of communication servers.

Keywords: data transfer, power distribution, fuzzy constraints, optimal solution, backtracking algorithm.

INTRODUCTION

The tasks of finding optimal solutions arise in the process of development and practical implementation of methods for effective management of various organizational, technological and information systems.

An important characteristic of optimization problems is the desire to find the optimal solution (optimality principle). In practice, there are a number of constraints that do not allow finding such a solution. In these cases, the question is raised of finding not optimal, but rational (compromise, effective) solutions that satisfy the problem statement. It is often necessary to find a compromise between the effectiveness of solutions and the cost of finding them. Serious difficulties arise when solving optimization problems under conditions of incomplete information, as well as in the case when random or subjective factors (parameters) play a significant role.

One of the applied problems in which there can be uncertainty in setting the parameters is the problem of distributing the limited capacities of data transmission channels between different nodes of the Internet providers network. Suppose that there is a local computer network of an enterprise (organization, educational institution) that provides users with access to the Internet. User access to the global network and obtaining the necessary information is carried out using several communication servers located on the territory of the information and computer center of the enterprise and connected by high-speed external communication channels with Internet providers. The bandwidth levels of the servers are within the bandwidth (bandwidth) of the local network (for example, 1Gb per second).

It is assumed that the network implements the conditions for efficient channel switching (relative to their bandwidth), which are provided by programmable network devices (communication servers, routers). The structure of the network and the information distributed in it in the general case can be very diverse. In this case, we consider the problem of distribution of limited capacities with the following constraints:

- information is distributed from the provider to subscribers (nodes) through switching servers via communication channels with a bandwidth that takes into account the specified bandwidth;
- each network subscriber is serviced by one switching server;
- the throughput of receiving information for switching nodes and subscribers is limited both from above (fundamental limitations of the provider's capabilities) and from below (the minimum need for subscribers to receive information).

The problem of determining the bandwidth of an external connection is considered, which makes it possible to maximize the total bandwidth of user communication channels by changing the total power of communication servers, taking into account both the needs and wishes of subscribers (users) and the capabilities of the information and computing center.

The solution of the formulated problem was considered in [1–7] on the basis of solving problems of optimal resource allocation. The problems of efficient use of a homogeneous resource were considered using the example of time distribution in the form of a classical problem of distributing resources of a given volume over a set of categories (works) [8]. The setting of such tasks consists in finding a cost plan for the available resource (such a resource is most often considered time) for the execution of a group of tasks, in which the total (final) use of the resource is optimal.

In a number of papers [9–11] to find a solution, an approach is proposed that uses multi-index problems of the transport type [11]. In the noted works, meaningful formulations of such applied problems are given and their mathematical models are constructed.

When solving applied multi-index optimization problems, special interest is given to formulations related to the class of problems of integer linear programming [11]. One of the approaches to the development of algorithms for solving such optimization problems is the use of streaming methods. Known efficient flow algorithms [12] make it possible to construct solution methods that have acceptable estimates of computational complexity compared to estimates of general methods for solving linear programming problems.

Solutions obtained on the basis of models of three-index transport problems [10] allow solving the problem of distribution of a homogeneous resource for cases where the cost and resource consumption factors are known a priori.

In [13,14], a model of a two-level production and transport problem was considered with a criterion that takes into account the optimal cost indicators for the production and transportation of resources, the volumes of production and consumption of which are given.

PROBLEM FORMALIZATION OF THE DATA TRANSMISSION CHANNELS OPTIMAL DISTRIBUTION POWER

An information and computer network is considered, including N_1 data transmission channels (global network providers), N_2 communication servers and N_3

end users (subscribers). We denote by A_i^+, $i=\overline{1,N_1}$, the values of the maximum bandwidth of the data transmission channel that provider *i*, $i=\overline{1,N_1}$, is able to provide; B_j^+, $j=\overline{1,N_2}$ — the value of the maximum bandwidth of the data transmission channel that the communication node *j*, $j=\overline{1,N_2}$, can provide; C_k^-, C_k^+, $k=\overline{1,N_3}$ — values of the minimum and maximum bandwidth of the data transmission channel, which must be provided to the subscriber *k*, $k=\overline{1,N_3}$; t_k — throughput of the *k*-th subscriber station, $k=\overline{1,N_3}$. Then, assuming that the power distribution of communication channels satisfies the conditions of additivity and proportionality, we can consider the problem of distributing a limited homogeneous resource (bandwidth of communication channels) with transport-type constraints in order to find the optimal data transmission plan. This ensures the effective functioning of the system for providing users with Internet access, which consists in finding the optimal values of data transmission bandwidths T_i of the *i*-th information provider (provider), $i=\overline{1,N_1}$, and the optimal values of the bandwidths t_k of using local communication channels of the *k*-th user, $k=\overline{1,N_3}$.

Formally, the statement of problem can be written as

$$\max t_1;\ \max t_2;\ \ldots \max t_{N_3},$$

with the following constraints

$$\sum_{k=1}^{N_3} t_k \le \sum_{i=1}^{N_1} A_i^+;$$

$$t_k \le B_j^+,\ j=\overline{1,N_2},\ k=\overline{1,N_3};\ \ C_k^- \le t_k \le C_k^+,\ k=\overline{1,N_3};$$

$$\sum_{j=1}^{N_2} B_j^+ \le \sum_{i=1}^{N_1} A_i^+ \le \sum_{k=1}^{N_3} C_k^+ .$$

SOLVING METHODS OF THE PROBLEM OF DATA TRANSMISSION CHANNELS OPTIMAL DISTRIBUTION POWER TAKING INTO ACCOUNT FUZZY CONSTRAINTS ON CONSUMPTION VOLUMES

Let's assume that the needs of network subscribers to increase the speed of obtaining one or another amount of information are known. The wishes (preferences) of subscribers are set regarding a possible increase in consumption volumes (bandwidths) for transmitting information from the provider to the user node. To implement the changes, it is necessary to update the capacities of the switching servers of the network by deploying new, more powerful computers or by increasing the number of existing servers. In other words, it is necessary to conduct a study on updating the resources of the server park of the information and computing center, which makes it possible to increase the total bandwidth of a group of switching servers. At the same time, the value of the total capacity of servers, both in the case of an increase in the capacity of the existing fleet of computers, and in the case of an increase in the number of servers, is assumed to be the same.

If the values of consumption parameters are random variables with known distribution functions, then it can be solved by stochastic programming methods. However, in practice these parameters are often unknown and only the range of possible values can be determined for them. A problem of this type can be called a problem with multiple values of the coefficients. Within the framework of this problem, it makes no sense to talk about maximizing the objective function, since the values of this function are not numbers, but sets of numbers. In this case, it is necessary to find out what preference relation this function generates on the set of alternatives, and then determine which products should be considered rational in the sense of this preference relation.

The next step on the way of detailing and refining the model considered here is the description of the problem parameters in the form of fuzzy sets (numbers) [15]. Additional information is introduced into the model in the form of a membership function of these fuzzy sets. These functions can be considered as a way for an expert to approximate his unformalized idea of the real value of a given parameter. Membership function values are the weights that experts assign to the various possible values of this parameter.

Fuzzy sets are a mathematical model of object classes with fuzzy or blurry boundaries. In other words, an element can have some degree of membership in the set, and it is intermediate between full membership and complete non-membership.

Traditional (ordinary) set theory can be viewed as a special case of fuzzy set theory. An ordinary subset A of a set X can be represented as a fuzzy set, which is given by the characteristic function $\chi_A : X \to \{0,1\}$:

$$\chi_A(x) = \begin{cases} 0: & x \notin A; \\ 1: & x \in A. \end{cases}$$

In accordance with the idea of Zadeh [15], a fuzzy subset of a given universal set X is formulated as follows.

Definition 1. A fuzzy subset $\widetilde{A}$ of the universal set X, is a collection of pairs $\widetilde{A} = \{(\mu_{\widetilde{A}}(x), x)\}$, where $\mu_{\widetilde{A}}(x) : X \to [0,1]$ is the mapping of the set X into the unit segment [0,1], which is called the membership function of the fuzzy set.

The value of the membership function $\mu_{\widetilde{A}}(x)$ for an element $x \in X$ is called the degree of membership. The interpretation of the degree of membership $\mu_{\widetilde{A}}(x)$ is a subjective measure of how much an element $x \in X$ corresponds to a concept, the meaning of which is formalized by a fuzzy set $\widetilde{A}$.

Let $X = R^1$ is a universal set.

Definition 2. [16] A fuzzy triangular number (triplet) $\widetilde{A}$ is an ordered triplet of numbers *(a, b, c)*, $a \le b \le c$, defining a membership function $\mu_{\widetilde{A}}(x)$ of the form:

$$\mu_{\widetilde{A}}(x) = \frac{x-a}{b-a}, \ x \in [a,b]; \ \mu_{\widetilde{A}}(x) = \frac{c-x}{c-b}, \ x \in [b,c]; \ x \notin [a,c]. \qquad (1)$$

A fuzzy triangular number of the form (a,b,b), called a left fuzzy triangular number, is determined by the membership function of the form:

$$\mu_{\tilde{A}}(x)=0,\ x<a;\ \ \mu_{\tilde{A}}(x)=\frac{x-a}{b-a},\ \ x\in[a,b];\ \ \mu_{\tilde{A}}(x)=1,\ \ x>b,$$

and the fuzzy triangular number of the form *(b, b, c)*, called the right fuzzy triangular number, is the membership function $x \in R^n$:

$$\mu_{\tilde{A}}(x)=1,\ \ x<b;\ \ \mu_{\tilde{A}}(x)=\frac{c-x}{c-b},\ \ x\in[b,c];\ \ \mu_{\tilde{A}}(x)=0,\ x>c. \qquad (2)$$

After such clarification, we can proceed to the next statement of the problem of fuzzy mathematical programming [17]. A linear view model is specified

$$\sum_{j=1}^{n}\tilde{c}_j x_j \to \max, \qquad (3)$$

in which the values of the coefficients $\tilde{c}_j$, $j=\overline{1,n}$, are given fuzzy in the form of fuzzy sets of given universal sets. In addition, there are constraints:

$$\sum_{j=1}^{n}\tilde{a}_{ij}x_j \le \tilde{b}_i,\ i=\overline{1,m},\ x_j \ge 0,\ j=\overline{1,n}, \qquad (4)$$

and the values of the coefficients $\tilde{a}_{ij}$, $\tilde{b}_i$, $i=\overline{1,m}$, $j=\overline{1,n}$, are also described in the form of the corresponding fuzzy sets. It is required to make a rational choice of a solution $x \in R^n$, that, in a certain sense, maximizes the given fuzzy linear form (3).

We call such a statement of the fuzzy optimization problem a linear programming problem with fuzzy parameters. One of its variants is a problem with fuzzy resource constraints on the right side.

Consider now a linear programming problem with a given goal function

$$\max_x \sum_{j=1}^{n} c_j x_j \qquad (5)$$

and fuzzy constraints on resources of the form

$$\sum_{j=1}^{n} a_{ij}x_j \le \tilde{b}_i,\ \ i=\overline{1,m},\ \ x \ge 0;\ x \in R^n, \qquad (6)$$

where the right parts of constraints (6) are given as fuzzy right triangular numbers with corresponding membership functions of the form (1). Here, the allowable deviations determine the values of the boundary changes of the model resources.

This formulation does not restrict the general form of optimization problems with fuzzy constraints [18, 19]. Indeed, one can consider a linear programming problem with fuzzy resources in the form of an optimization problem for the goal function (5) in the presence of a system of mixed constraints:

$$\sum_{j=1}^{n} a_{ij}x_j \ge \tilde{b}_i,\ \ i=\overline{1,m_1};\ \ \sum_{j=1}^{n} a_{ij}x_j \le \tilde{b}_i,\ \ i=\overline{m_1+1,m_2};$$

$$\sum_{j=1}^{n} a_{ij}x_j = \tilde{b}_i,\ \ i=\overline{m_2+1,m},$$

where the right parts of the first m_1 constraints are given by left fuzzy triangular numbers $\tilde{b}_i=(b_i-b_i^0, b_i, b_i)$, $b_i^0 \ge 0$, $i=\overline{1,m_1}$, the right parts of the next group of constraints are given by right fuzzy triangular numbers $\widetilde{b}_i = (b_i, b_i, b_i + b_i^0)$, $b_i^0 \ge 0$, $i = \overline{m_1 + 1, m_2}$, and the right parts of the last $m - m_2$ constraints are given by fuzzy triangular numbers $\widetilde{b}_i = (b_i - b_i^l, b_i, b_i + b_i^r)$, with allowable deviations $0 \le b_i^l \le b_i$, $b_i^r \ge 0$, $i = \overline{m_2 + 1, m}$.

This LP model can be rewritten in the form (4)–(5) by replacing the first m_1 conditions with the next constraints $\sum_{j=1}^{n} (-a_{ij}) x_j \le -\widetilde{b}_i$, $\widetilde{b}_i = (-b_i, -b_i, -b_i + b_i^0)$, $i = \overline{1, m_1}$, and the last $m - m_2$ conditions — with a system of constraints $\sum_{j=1}^{n} (-a_{ij}) x_j \le -\widetilde{b}_i$, $\widetilde{b}_i = (-b_i, -b_i, -b_i + b_i^l)$; $\sum_{j=1}^{n} a_{ij} x_j \le \widetilde{b}_i$, $\widetilde{b}_i = (b_i, b_i, b_i + b_i^r)$; $i = \overline{m_2 + 1, m}$. Thus, we can assume that the general form of a linear programming problem with fuzzy resource constraints on the right side is given by model (5)–(6).

The optimization problem under consideration can be solved as a parametric linear programming problem [20]. This method is universal, not always taking into account the specifics of the task.

We use an approach based on the defuzzification of problem (5)–(6). To do this, we calculate the optimal values of the levels of the objective function Z_l and Z_u by solving two linear programming problems:

$$Z_l = \max_x \sum_{j=1}^{n} c_j x_j \tag{7}$$

with constraints

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, \quad i = \overline{1, m}, \quad x \ge 0; \; x \in R^n; \tag{8}$$

$$Z_u = \max_x \sum_{j=1}^{n} c_j x_j \tag{9}$$

under condition

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i + b_i^0, \quad i = \overline{1, m}, \; x \ge 0; x \in R^n. \tag{10}$$

Let be $L = \min(Z_l, Z_u)$, $U = \max(Z_l, Z_u)$. The fuzzy set of optimal values of problem (5)–(6) specified in R^n (we denote it by $\widetilde{G}$) is described by the membership function of the form:

$$\mu_{\widetilde{G}}(x) = \begin{cases} 0, & \sum_{j=1}^{n} c_j x_j < L; \\ (\sum_{j=1}^{n} c_j x_j - L)/(U - L), & L \le \sum_{j=1}^{n} c_j x_j \le U; \\ 1, & \sum_{j=1}^{n} c_j x_j > U, \end{cases} \tag{11}$$

and the fuzzy sets of each constraint (we denote them by $\widetilde{F}_i$, $i=\overline{1,m}$) from (6) are determined by the membership functions:

$$\mu_{\widetilde{F}_i}(x)=\begin{cases}1, & \sum_{j=1}^{n} a_{ij}x_j < b_i;\\ (b_i+b_i^0-\sum_{j=1}^{n} a_{ij}x_j)/b_i^0, & b_i \le \sum_{j=1}^{n} a_{ij}x_j \le b_i+b_i^0,\ i=\overline{1,m};\\ 0, & \sum_{j=1}^{n} a_{ij}x_j > b_i+b_i^0.\end{cases} \qquad (12)$$

Based on the definition of the Bellman-Zadeh fuzzy solution [21], the fuzzy linear programming problem (5)–(6) can be written in the form of an optimization problem of the following form: find the value of the parameter $\lambda \in [0,1]$, that is the solution of the linear programming problem

$$\max_x \lambda$$

with constraints

$$\mu_{\widetilde{G}}(x) \ge \lambda;\quad \mu_{\widetilde{F}_i}(x) \ge \lambda;\quad x \ge 0.\ \ i=\overline{1,m}. \qquad (13)$$

Substituting (11) and (12) into (13), we write the final form of the optimization problem

$$\max_x \lambda$$

with constraints

$$\lambda(U-L)-\sum_{j=1}^{n} c_j x_j + L \le 0;$$

$$\sum_{j=1}^{n} a_{ij}x_j \le b_i + b_i^0 - \lambda b_i^0,\quad i=\overline{1,m};\quad x\ge 0,\quad 0\le\lambda\le 1. \qquad (14)$$

This problem is a classical linear programming problem, for finding solutions to which any variant of the simplex method can be applied.

Let us assume that in the formulation of the problem of distributing the power of data transmission channels, the current values of the throughput of communication channels of each subscriber *k,* C_k, $k=\overline{1,N_3}$, are known, and the values of C_k^+, $k=\overline{1,N_3}$, determine the values of the bandwidths that are planned by users as a result of updating communication equipment. Obviously, it is possible to fully satisfy the expansion of the bandwidth of subscriber channels only under the condition $\sum_{j=1}^{N_2} B_j^+ \ge \sum_{k=1}^{N_3} C_k^+$.

Formally, the statement of problem can be written as

$$\max t_1;\ \max t_2;\ \ldots \max t_{N_3}, \qquad (15)$$

with the following constraints

$$t_k \in \operatorname{supp} \tilde{t}_k = [C_k, C_k^+], \ k = \overline{1, N_3}; \ \sum_{k=1}^{N_3} t_k \le \sum_{i=1}^{N_1} A_i^+;$$

$$t_k \le B_j^+, \ j = \overline{1, N_2}, \ k = \overline{1, N_3}; \ \sum_{j=1}^{N_2} B_j^+ \le \sum_{i=1}^{N_1} A_i^+ \le \sum_{k=1}^{N_3} C_k^+ . \qquad (16)$$

We will assume that the capacities of communication channels available to users satisfy the conditions $\sum_{k=1}^{N_3} C_k \le \sum_{k=1}^{N_3} t_k \le \sum_{i=1}^{N_1} A_i^+$, and the values of the possible expansion of the channel capacity are determined by right-hand fuzzy triangular numbers in the form (C_k, C_k, C_k^+), $k = \overline{1, N_3}$, with linear membership functions (2).

This problem is a multiobjective optimization problem. To solve it, methods are used that allow finding a compromise (effective) solution by reducing the problem to a single-criterion one in the form of a convolution of criteria or to a sequence of single-criteria optimization problems [22]. In the case of fuzzy constraints, each such problem can be reduced to an optimization problem of the form (13) or (14) with subsequent solution by the method proposed above.

Taking into account the specifics of the obtained problem, the most rational method is the sequential introduction of constraints [22]. A characteristic feature of this method, which makes it possible to use it to find an effective solution, is the sequential (at each step) introduction of constraints on the width of the communication channel, at which unsatisfactory values of the criteria are achieved.

Following the search methodology, at each algorithm's step $p = 1, 2, \ldots$, an "ideal assessment" $t^{*(p)} = (t_1^{*(p)}, t_2^{*(p)}, \ldots, t_{N_3}^{*(p)})$, $p = 1, 2, \ldots$, is formed, where $t_k^{*(p)}$, $k = \overline{1, N_3}$, are the optimal values of each of the criteria (19) $\max t_k$, $k = \overline{1, N_3}$, on a given range of acceptable values G_p, $G_1 = \{t_k = C_k^+; k = \overline{1, N_3}\}$, $G_{p+1} = \{t_k \in G_p; k = \overline{1, N_3} \mid t_s \ge \xi_s\}$, $s \in \{1, 2, \ldots, N_3\}$ is the number of the criterion, the value of which is the least consistent with the compromise solution. It is clarified to what level ξ_s the value of this criterion should be changed, and a search for a new solution is performed, taking into account the additional constraint.

This method allows solving the problem of efficient distribution of channel capacities, taking into account fuzzy constraints on consumption volumes, however, to use it at each step, it is necessary to evaluate the compliance of the current solution with a certain "ideal" solution, which, as a rule, is formed with the participation of an expert. In addition, the solution procedure turns out to be cumbersome, leading to the multiple solution of optimization problems of the form (7)–(10) and the construction of a Bellman-Zade fuzzy solution (14).

Additionally it is easy to formalize this process by applying the back tracking solution search procedure [23].

From the condition of the problem of optimizing the distribution of channel powers, taking into account fuzzy constraints on consumption volumes (15)–(16), it follows that

$$\sum_{k=1}^{N_3} C_k^+ \ge \sum_{j=1}^{N_2} B_j^+ \ge \sum_{k=1}^{N_3} C_k .$$

Obviously, in this case, it is impossible to allocate the maximum expected power of communication channels to all subscribers. We will look for a solution on rational distribution based on the scheme of the back tracking algorithm.

Algorithm.

Step 0. Without loss of generality, we will assume that the order of users is ordered in non-increasing order of the planned capacities of communication channels. We put the required values in the initial solution $t_k = C_k^+$, $k = \overline{1, N_3}$.

Step $s = 1, 2, \ldots$ *s=1,2,...* We check the fulfillment of condition

$$\sum_{k=1}^{N_3} t_k \le \sum_{j=1}^{N_2} B_j^+ . \tag{17}$$

If inequality (17) is satisfied, the algorithm terminates, otherwise:

a) determine the q, $q \in [1, N_3]$, largest (first of N_3) values t_k, $k = \overline{1, N_3}$;

b) decrease the values t_k, $k = \overline{1, q}$, by $\Delta t > 0$: $t_k = t_k - \Delta t$, $k = \overline{1, q}$.

Obviously, the total demand in this case decreases.

Change $s = s + 1$ and move on to the next step.

RESULTS OF COMPUTATIONAL EXPERIMENTS

The algorithm proposed above for finding a solution in the problem of rational distribution of the power of communication channels, taking into account fuzzy constraints on consumption volumes (15)–(16), was used to calculate the values of throughput resources in a network with 1 Internet provider, 2 (3, 4) routers (communication servers) and 17 end users (collective subscribers).

The bandwidth of user connections to communication servers was initially 350, 250, 250, 245, 180, 180, 165, 165, 160, 145, 140, 140, 140, 120, 110, 80, 80 Mb/s (total capacity 2900 Mb/s). In order to expand consumer traffic, it is proposed to upgrade equipment in the form of a possible increase in the number of servers or/and increase their capacity. The bandwidth of the communication channel with the provider remains constant and equals 10 Gb/s. The total throughput capacity of communication servers after the upgrade is planned to be 3 Gb/s.

To determine the rational distribution of the size of communication channels, consumers were asked to determine the required size of connections to communication servers. Based on the given amount of traffic, it was planned to use 2, 3 or 4 servers with a total capacity of 3 Gb/s.

Computational experiments on the efficient distribution of the power of communication channels were carried out using the above algorithm for the classical solution of optimization problems with fuzzy constraints on consumption levels (fuzzy approach) and the algorithm using the backtracking approach. In the latter case, both a consistent uniform decrease in consumer requests by the value $\Delta t > 0$ (app1) and a proportional decrease in the values of requests were applied, taking into account the required volumes of traffic increase (app2).

The results of the numerical experiments performed are shown in Table.

Table 1. The results of numerical experiments on the efficient distribution of the power of communication channels

Consumers																	
p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{17}	Sum
Init Power, Mb/s (*Max Sum Power*=2900 Mb/s)																	
350	250	250	245	180	180	165	165	160	145	140	140	140	120	110	80	110	2900
Plan Power, Mb/s (*Max Sum Power*=3000 Mb/s)																	
370	275	275	260	195	185	180	175	165	155	150	150	145	125	115	90	115	3100
Results for K communication servers, Mb/s																	
Approuch: ***app1***																	
*CommunicationPower*K*=1500×2																	
363	268	268	253	188	183	173	168	158	148	143	143	138	118	108	85	108	2995
*CommunicationPower*K*=1000×3																	
359	264	264	249	184	179	169	164	154	144	139	139	134	114	105	83	105	2934
*CommunicationPower*K*=750×4																	
357	262	262	247	182	177	167	162	152	142	137	137	132	112	107	89	107	2914
Approuch: ***app2***																	
*CommunicationPower*K*=1500×2																	
354	254	254	254	184	184	174	169	164	149	149	149	144	124	114	89	114	2999
*CommunicationPower*K*=1000×3																	
352	252	252	252	182	182	172	167	162	147	147	147	142	122	112	87	112	2967
*CommunicationPower*K*=750×4																	
350	250	250	250	180	180	170	165	160	145	145	145	140	120	110	85	110	2935
Approuch: ***fuzzy***																	
*CommunicationPower*K*=1500×2																	
361	266	266	251	186	181	171	166	160	150	146	144	141	121	110	83	110	2985
*CommunicationPower*K*=1000×3																	
362	267	267	252	187	182	173	167	160	150	147	142	140	120	110	82	110	2989
*CommunicationPower*K*=750×4																	
355	260	260	245	180	175	165	160	150	150	145	140	135	118	108	79	108	2900

As follows from the results obtained, the application of the proposed algorithm made it possible to obtain the most efficient (close to optimal) solutions in the considered distribution problem for a configuration with two communication servers with a maximum bandwidth of 1500 Mb/s. The best solution to the problem using the method of efficient channel power distribution, taking into account fuzzy constraints, was obtained for the connection option with 3 routers. At the same time, it slightly differs from the solution with two servers, which suggests that the best option in the considered distribution problem is the variant with two communication servers. It should also be noted that the solution based on the algorithm using the return scheme does not require significant computational resources, which allows us to speak about the constructiveness of the method. The resulting solution was used as the basis for the technical modernization of equipment to ensure the operation of network subscribers.

DISCUSSION AND CONCLUSIONS

Several remarks should be noted. First, the amount Δt of change in the power of communication channels in the backtracking algorithm, which is set at the beginning of the work, depends on the values of the optimized data transfer volumes and affects the rate of convergence of the algorithm. The choice of small values Δt leads to a more accurate rational distribution of powers, but slows down convergence. Otherwise, for large values Δt, the solution is reached faster, but its quality in terms of the obtained volumes, as a rule, turns out to be lower.

In addition, in the proposed version of the algorithm, a rational solution is sought at the expense of the most demanding subscribers in terms of volume. Obviously, the search procedure can be restructured to use other similar principles or to evenly distribute the redundancy of the total traffic request among all network users.

The problem of optimal power distribution of communication channels in information-computer networks with a three-level architecture is considered. Approaches for its solution are studied, the problem statement with fuzzy constraints on the consumption volumes of end users is considered. A fuzzy optimization problem is formulated, which allows taking into account the interval specified volumes for the connection values. A variant of solving fuzzy optimization problems in the case of using fuzzy numbers is proposed. A multi-objective problem of efficient power distribution of communication channels with fuzzy constraints is formulated. A variant of the algorithm with a return is proposed, which allows solving the obtained problem. The approach is illustrated by a number of numerical examples of a problem with a given number of end users and different allowable bandwidths of communication servers.

The results obtained were analyzed, which made it possible to make a decision on the method of upgrading the communication equipment. The proposed approach based on the method using the return scheme turned out to be a constructive way to solve the problem considered in the article.

REFERENCES

1. L.G. Afraimovich and M.Kh. Prilutskii, “Multiindex Resource Distributions for Hierarchical Systems,” *Automation and remote control*, vol. 67, no. 6, pp. 1007–1016, 2006.
2. M.Kh. Prilutskii, “Many-Criteria Distribution of a Homogeneous Resource in Hierarchical Systems,” *Automation and Remote Control*, no. 2, pp. 24–29, 1996.
3. M.Kh. Prilutskii, “Distribution of a Homogeneous Resource in Tree-Structured Hierarchical Systems,” *Proc. III Int. Conf. Identification of Systems and Control Problems (SICPRO’2000)*, pp. 2038–2049.
4. M.Kh. Prilutskii and A.G. Kartomin, “Flow Algorithms for Allocation of Resources in Hierarchical Systems,” *Investigated in Russia*, 39, pp. 444–452, 2003. Available: http://zhurnal.ape.relarn.ru/articles/2003 /039.pdf
5. A. Schrijver, *Theory of Linear and Integer Programming*. Wiley, 1998, 484 p.
6. E.G. Zhdanova, *Schedule theory*. M., 2000, 136 p.
7. D.W. Pentico, “Assignment problems: A golden anniversary survey,” *European Journal of Operational Research*, vol. 176, pp. 774–793, 2007. doi: 10.1016/j.ejor.2005.09.014.
8. E. Ivokhin and M. Makhno, “On an approach to construction of structured fuzzy sets and their application for description of fuzzy time response,” *Journal of Automation and Information Science*, 49(10), pp. 55–63, 2017. doi: 10.1615/JAutomatInfScien. v49.i10.60.
9. F.C.R. Spieksma, “Multi index assignment problems: complexity, approximation, applications,” *Nonlinear Assignment Problems. Algorithms and Applications*; P.M. Pardalos, L.S. Pitsoulis (Eds.). Dordrecht: Kluwer Academic Publishers, 2000, pp. 1–11. doi: 10.1007/978-1-4757-3155-2_1.
10. Y.Crama and F.C.R. Spieksma, “Approximation algorithms for three-dimensional assignment problems with triangle inequalities,” *European Journal of Operational Research*, vol. 60, pp. 273–279, 1992. doi: 10.1016/0377-2217(92)90078-N.
11. L.G. Raskin and I.O. Kirichenko, *Multi-index problems of linear programming*. M.: Radio and communication, 1982, 240 p.
12. J.B. Orlin, “A Faster strongly polynomial minimum cost flow algorithm,” *Operations research*, vol. 41, iss. 2, pp. 338–350, 1993. doi: 10.1145/62212.62249.

13. Z. Lukac, D. Hunjet, and L. Neralic, "Solving the production-transportation problem in the Petroleum Industry," *Revista Investigacion Operacional*, vol. 29, iss. 1, pp. 63–70, 2008.
14. E. Ivokhin, D. Apanasenko, and V. Navrodskiy, "About production-transport problem reduction to the two-level problem of discrete optimization and its application," *Bulletin of Taras Shevchenko Nathional University of Kiev: Ekonomics*, no. 3(198), pp. 48–53, 2018. doi: 10.17721/1728-2667.2018/198-3/6.
15. L.A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965. doi: 10.1016/S0019-9958(65)90241-X.
16. R.E. Bellman and L.A. Zadeh, "Local and fuzzy logics," *Modern Uses of Multiple-Valued Logics*; edited by J.M. Dunn, G. Epstein, D. Reidel. Dordrecht-Holland, 1977, pp. 103–165. doi: 10.1007/978-94-010-1161-7_6
17. D. Dubois, "Linear programming with fuzzy data," *Analysis of Fuzzy Information*; J.C. Bezdek (ed.). Boca Raton: CRC Press, 1987, vol. 3: Applications in Engineering and Science, pp. 241–263.
18. Bablu Jana and Tapan Kumar Roy, "Multi-Objective Fuzzy Linear Programming and Its Application in Transportation Model," *Tamsui Oxford Journal of Mathematical Sciences*, 21(2), pp. 243–268, 2005.
19. H.J. Zimmermann, "Fuzzy programming and linear programming with several objective functions," *Fuzzy Sets and System*, no. 1, pp. 45–55, 1978. doi: 10.1016/0165-0114(78)90031-3
20. D.B. Yudin and E.G. Golshtein, *Linear programming*. M.: Nauka, 1969, 424 p.
21. R.E. Bellman and L.A. Zadeh, "Decision making in a fuzzy environment," *Management Science*, no. 17, pp. 141–164, 1970. doi: 10.1287/mnsc.17.4.B141.
22. O.F. Voloshin and S.O. Mashchenko, *Models and methods for decision making*. K.: "Kiev University", 2010, 336 p.
23. D. Watson, *A Practical Approach to Compiler Construction*. Springer, 2017, 254 p.

Received 02.05.2022

INFORMATION ON THE ARTICLE

Eugene V. Ivokhin, ORCID: 0000-0002-5826-7408, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: ivohin@univ.kiev.ua

Larisa T. Adzhubey, ORCID: 0000-0002-8103-9657, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: adzhubey@ukr.net

Petro R. Vavryk, ORCID: 0000-0003-4989-7544, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: petro.vavryk@knu.ua

Mykhailo F. Makhno, ORCID: 0000-0001-9694-2200, Taras Shevchenko National University of Kyiv, Ukraine, e-mail: makhnom@gmail.com

ПРО ДЕЯКІ МЕТОДИ РОЗВ'ЯЗАННЯ ЗАДАЧІ РОЗПОДІЛУ ПОТУЖНОСТІ КАНАЛІВ ПЕРЕДАВАННЯ ДАНИХ З УРАХУВАННЯМ НЕЧІТКИХ ОБМЕЖЕНЬ НА ОБСЯГИ СПОЖИВАННЯ / Є.В. Івохін, Л.Т. Аджубей, П.Р. Ваврик, М.Ф. Махно

Анотація. Розглянуто математичну постановку задачі оптимального розподілу потужностей каналів передавання даних в інформаційно-комп'ютерних мережах з трирівневою архітектурою та нечіткими обмеженнями на обсяги споживання. Розроблено ефективний алгоритм розв'язання задачі, особливістю якої є неможливість забезпечувати запити кінцевого споживача за рахунок ресурсів різних постачальників. Розглянуто стандартний метод розв'язання на основі нечіткої оптимізаційної задачі математичного програмування. Запропоновано конструктивний варіант пошуку розв'язку на основі методу з поверненням. Проведено обчислювані експерименти. Розроблено підхід, використаний для визначення оптимальної конфігурації трирівневої інформаційно-комп'ютерної мережі із заданою кількістю комунікаційних серверів.

Ключові слова: передавання даних, розподіл потужності, нечіткі обмеження, оптимальний розв'язок, алгоритм з поверненням.

UDC 620.179.1.001.5
DOI: 10.20535/SRIT.2308-8893.2022.4.09

DYNAMIC CERTIFICATION AND ASSESSMENT OF THE BUILDINGS LIFE CYCLE UNDER REGULAR EXPLOSIVE IMPACTS

O.M. TROFYMCHUK, I.I. KALIUKH, V.A. DUNIN, S.Y. KYRASH

Abstract. Today in Ukraine, there is no single legalized, generally accepted methodology (at the level of a Ukrainian building standard) for dynamic certification of buildings and structures. A unified approach is proposed as such a technique. It includes four components: visual inspection of buildings; experimental studies of the dynamic response of buildings or structures to explosive effects; mathematical modeling of the stress-strain state of the object under study; synthesis of the results of visual inspection; experimental studies and numerical simulation in order to generalize them systematically. As an approbation, the deterioration of the resource of reinforced concrete structures of residential buildings under the conditions of constant mass industrial explosions with a capacity of 500 to 700 tons in the quarry of Southern GZK (Mining and Processing Plant) in the city of Kryvyi Rih, Ukraine, has been studied. Based on the processing of numerous experimental data and the results of mathematical modeling, a probabilistic model for predicting the deterioration of the technical condition of reinforced concrete structures of the Center for Children and Youth Creativity "Mriya" has been obtained. Calculations of the risks of destruction of the building's load-bearing elements for its vulnerable areas made it possible to clarify its service life. It decreased by ~ 30 years compared to the standard in 2012.

Keywords: shock waves, experiment, risk, dynamic certification, life cycle of buildings and structures.

INTRODUCTION

Due to the insufficiently developed network of roads and land transport in the USSR and for the economic feasibility reasons, the residential neighborhoods were usually built near industrial zones. That ensured minimal time expenditures for workers transportation to the enterprises in those zones. At the same time, the nature of the works performed in industrial zones and their impact on the housing stock were minimally taken into account during housing development design and construction. The housing stock buildings in the city of Kryvyi Rih (Dnipropetrovsk region, Ukraine) can be mentioned as a typical example of man-made influences during blasting works in open iron ore quarries. The buildings and structures dynamic certification is the primary stage of works to ensure the necessary and economically feasible level of construction projects dynamic stability in conditions of industrial explosions, moral wear and physical tearing. Two main goals of the buildings and structures dynamic certification include:

1) comparative assessment of the actual seismic resistance of buildings with the specified territory seismicity (the determination of the buildings seismic resistance deficiency);

2) identification of the most dangerous projects that require priority strengthening, repurposing or demolition of the buildings.

The issues of the buildings dynamic certification, physical and dynamic wear impact and earthquake-resistant construction economic feasibility were given a serious attention at the state level in the relevant resolutions and other documents of the Cabinet of Ministers of Ukraine [1, 2], as well as considered in the numerous works of Ukrainian and foreign researchers including S.V. Medvedev, V.I. Keilis-Borok, I.E. Itskov, Y.M. Eisenberg, M.A. Cornell, K. Oliveira, I. Idriss and others [3–10]. The known scientists A. Alonso-Rodriguez, M. Barla, H. Burton, N. Casagli, F. Catani, C. Del Ventisette, W. Frodella, G. Gigli, G. Kampas, G. Lollino, G. Luzi, M. Martinelli, R. Piccioni, J. Stewart, Y. Wang and others should be mentioned among modern foreign researchers in the field of structures dynamic analysis [11–21].

The existing certification methods can be conditionally divided into the following three groups: expert assessment methods, analytical calculation methods, and technical diagnostics methods.

Their comparative analysis shows that, in addition to advantages, the existing certification methods have several disadvantages [22, 23]. When using the expert assessment method, the statistical consistency degree of expert assessments and their confidence intervals remain uncertain. Although this method, in an opinion of several researchers [2, 22, 23], is the cheapest and the most common when the actual seismic resistance of buildings is continuously assessed. At the same time, according to other researchers [24], it gives a rather high error.

The analytical calculation methods related with DBN [25, 26] reflect the inherent contradictions in the normative calculation procedures (conditional design loads or linear elastic models use). Their advantages include taking into account the actual physical and mechanical characteristics of the structures material and the possibility of considering the building structures physical tearing. However, they are labor-intensive and require a lot of time.

The technical diagnostics methods allow to detect and localize an anomaly of the building dynamic structure, but at the same time they cannot establish its cause and reveal by means of the structures opening the respective defects of structural elements and their connection nodes. Usually, the analysis of the building dynamic structure is carried out at the microdynamic level of impact, at which the structures and their connections clearly operate in the elastic stage. This does not allow to take into account the influence of structural and physical nonlinearity and stiffness parameters degradation on the assessment of the building seismic resistance.

Due to the described above advantages and disadvantages, specific for each of the three approaches, the methodological problems of the dynamic certification can be reduced to two main problems [24].

The first problem is the correct definition of the criterion for assessing the inspected building dynamic stability reserve including:

1) the expert assessment reflecting the degree of project compliance with structural requirements [25];

2) the computational and analytical assessment of seismic resistance corresponding to conditional seismic loads [25];

3) the results of the building technical diagnostics [27].

The second problem is the determination of the necessary level of influence, at which the building dynamic structure including the microdynamic level at the elastic stage of the structures operation or the load level corresponding to the building structures operation beyond the elastic limit should be investigated.

Thus, it can be concluded that in Ukraine today there is no unified dynamic certification method legalized at the level of DBN [26, 28, 29] or standard. In view of this, the methods improvement for the buildings inspection to assess their actual seismic resistance and determine the residual resource with an allowance for the structures physical tearing, is an urgent task.

The paper proposes to synthesize the methods into a unified VEMS method consisting of four parts: **v**isual inspection of buildings – the first part; experimental studies of the dynamic response of buildings or structures to explosive actions – the second part; mathematical modeling of the stress-strain state of the project under study – the third part; **s**ynthesis of the results of visual inspection, experimental studies and numerical modeling with an aim of their systematic generalization – the fourth part.

Visual examination. Today, the man-made impact of industrial explosions on the existing housing stock and socio-cultural assets of Kryvyi Rih has significantly increased. This is primarily due to the expansion of the iron ore quarries industrial zone. For example, the open pit of the PivdGZK (Southern Mining and Processing Plant) for the iron ore extraction in the city of Kryvyi Rih was founded in 1952 and reached a depth of about 250 m. At ground level, the development of minerals in the quarry is carried out on the area outlined by an oval with the major and minor axes of 4 km and 3 km, respectively (sanitary and protective zone during explosions is 700 m). The general appearance of the quarry of the Kryvyi Rih Southern Mining and Processing Plant is presented in Fig. 1.

Fig. 1. General view of the PivdGZK quarry with the adjacent infrastructure of the Inhulets district of the Kryvyi Rih city

The boundary of the housing area of the Inhulets district of the city located to the east and southeast of the quarry passes at an 800–900 m distance from its eastern edge. Thus, during blasting operations at the quarry upper horizons, the sanitary protection zone directly approaches to the residential development territory. The seismic impacts on buildings located in the sanitary protection zone are periodic in nature. Until recently, explosions at the PivdGZK were carried out once every 2 weeks. At the same time, the maximum mass of explosives for carrying out explosions in the PivdGZK quarry ranged from 490 to 652 tons during the observations period in 2008–2012.

For the observations during the blasting works in the quarry the following objects are selected: secondary school No. 40, the Center for Children and Youth Creativity (the Center), the Church of the Nativity of the Theotokos, and three one-storey buildings measuring $\approx 7 \times 8\,\mathrm{m}$. The listed structures do not cover the comprehensive spectrum of buildings according to their structural design, but they fully correspond to their socio-cultural purpose. The School, the Center and the Church of the Nativity of the Theotokos are, in addition, the buildings of mass concentration of people and must satisfy the increased safety requirements during their operation. The assessment of their current state under the seismic effects from explosions, predictive assessment of their durability (resource) and, therefore, of the people safety include the stated task in the category of actual problems.

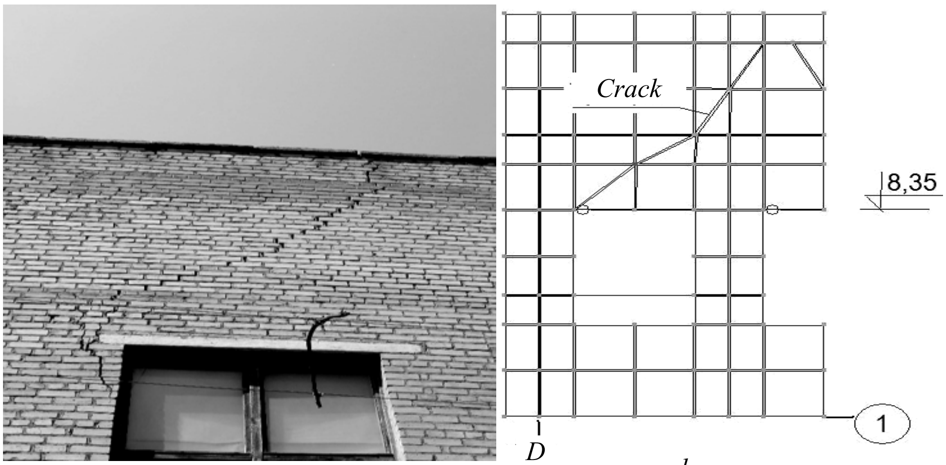
In the photographs (Fig. 2) the fragments of the Church of the Nativity of the Theotokos at 14, Obrucheva St. in Kryvyi Rih, located between two iron ore quarries are shown.

Fig. 2. Instances of buildings damages due to industrial explosions in Kryvyi Rih (photographs by Iurii Kaliukh)

Due to its proximity to the quarries, the building is exposed to a double "dose" of man-made seismic actions from industrial explosions in two quarries at once, as can be seen from the recently installed and already cracked glass at the entrance (Fig. 2, *a*) and vertical cracks in the wall (Fig. 2, *b*). In Fig. 2, *c* the church priest presents vertical cracks on his house located on the territory of this church. In Fig. 2, *d* the damage to the load-bearing walls of the secondary school No. 40 building, which is located next to the iron ore quarry, is shown. By the cracks nature, it can be assumed that the school is in an emergency state resulting from the regular and constant man-made tremors posing a serious threat, which to the students at the school are exposed every day during classes.

Many of the observed buildings have a long-term service life. The design and executive documentation of some buildings has now been lost. Therefore, all three methods (visual, experimental and theoretical) for the assessment of the buildings and their load-bearing elements technical condition should be used. The Center of Creativity building in the Kryvyi Rih city could be considered as an example. The choice of this building for the detailed experimental and theoretical research, as well as for the residual resource assessment was also reasoned by the partial availability of design materials. The building design was developed in 1960. The Center of Creativity was built in the 70s during the period of intensive extraction of iron ore in the quarry, therefore the building has been subjected to explosive influences, starting from the construction stage.

In Fig. 3 the area of the load-bearing wall along the axis 1 with an oblique crack starting from the window opening corner at the second floor, and its modeling via the building computer model using the finite element method are shown. Cracks with openings smaller than the above one were not modeled.



a *b*

Fig. 3. Photo recorded view of the Center bIuilding wall along the axis 1 and an arrow indicating the crack in the damaged area (*a*) and its representation in the computer model N 2 (*b*)

Experimental studies. It is possible to consider in more detail the features of dynamic diagnostics of building structures (floor slabs) with small-amplitude vibrations (from fractions of a mm to several mm) using high-sensitivity sensors that can record and single out the buildings vibrations arising due to the background dynamic loads from industrial explosions. In the 2008–2012 period the Center building and soil dynamic examinations were carried out according to the developed methodology, which envisaged the selection and composition of vibration measuring equipment and the development of the vibration sensors arrangement schemes and methods for recording, storing and processing the received vibration signals. In Fig. 4 the scheme for recording the shock waves propagation is shown.

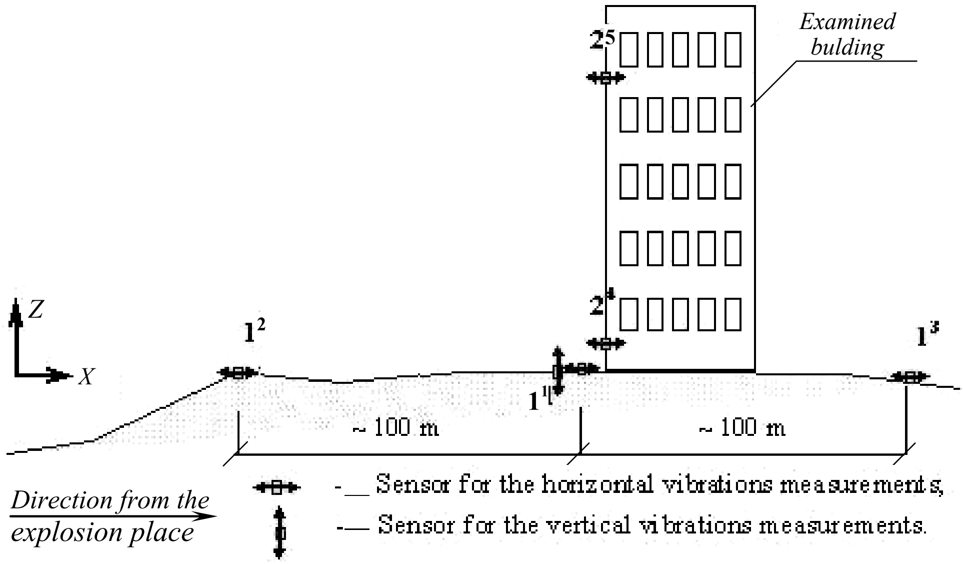


Fig. 4. Sensors arrangement according to the diagsrams 1 and 2

A typical spectrum of the time signal of vibration acceleration and the amplitude spectrum of the examined project reinforced concrete floors vibrations are presented in Fig. 5.

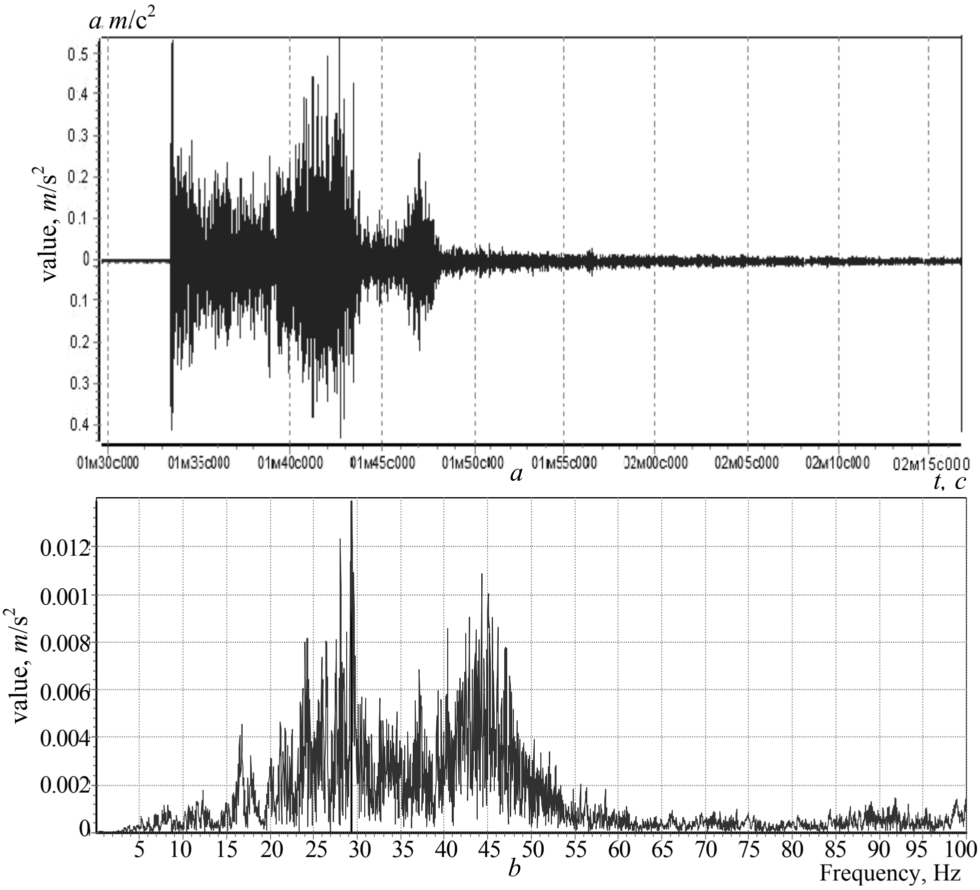


Fig. 5. The time-dependent signal of the radial component of the soil horizontal acceleration near the Center for Children's Creativity building (*a*) and its amplitude spectrum (*b*) during an explosion with a power of 650 tons, 2012

Based on the obtained experimental data concerning the actual levels of acceleration and vibration velocity of the soil and Center building structures the following conclusions can be drawn:

1. The values of the soil vibration accelerations at the buildings bases recorded during explosions are in the range [0.0238 m/s^2 to 0.643 m/s^2].

2. The values of the soil vibration velocity at the buildings bases recorded during explosions are in the range of 0.0004–0.015 m/s [31], which corresponds to 1–4 points on the S.V. Medvedev's seismic intensity scale [30].

3. The spectra analysis of the soil accelerations during explosions allows to find out that the predominant frequencies are in the range of 1–54 Hz. This confirms the possibility of building structures (floor and walls) vibrations in a mode close to resonance. In addition, according to recommendations [31], for the exclusion of the building foundations settlements during explosions, the soil acceleration should be limited to 0.15 m/s^2.

Mathematical modeling. For the further research and examination of the Center building condition, its computer model was developed based on the finite element method (FEM) with the use of the LIRA-9.6 software package [32]. The LIRA software package is a multifunctional software package for calculation, research and design of structures for various purposes, which has more than 40 years of history of creation, development and use in the scientific research and practice of structures design.

The FEM is the theoretical basis of the LIRA software package. The FEM implemented version uses the principle of possible motions:

$$a(u,v) = (f,v)\,, \tag{1}$$

where u is the desired exact solution; v is any possible displacement; $a(u,v)$ and (f,v) are the possible actions of external and internal forces (1).

The area occupied by the structure is divided into finite elements Ω_r; the nodes and their degree of freedom L_i (nodes displacements and rotation angles) are assigned. The degrees of freedom have corresponding them basic (coordinate or approximating) functions μ_i that differ from zero only at the elements corresponding nodes and correspond to the following equality:

$$L_j\mu_i = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases} \tag{2}$$

The approximate solution U_h must satisfy the main kinematic conditions and is calculated in the form of a linear combination of basic functions

$$U_h = \sum_{i=1}^{N} u_i\mu_i\,, \tag{3}$$

where u_i are the numbers; N is the number of degrees of freedom.

Substituting U_h instead of u and μ_i ($i = 1..N$) instead of v in (2) allows to obtain a system of FEM equations shown in the elements of the block-diagram in Fig. 6.

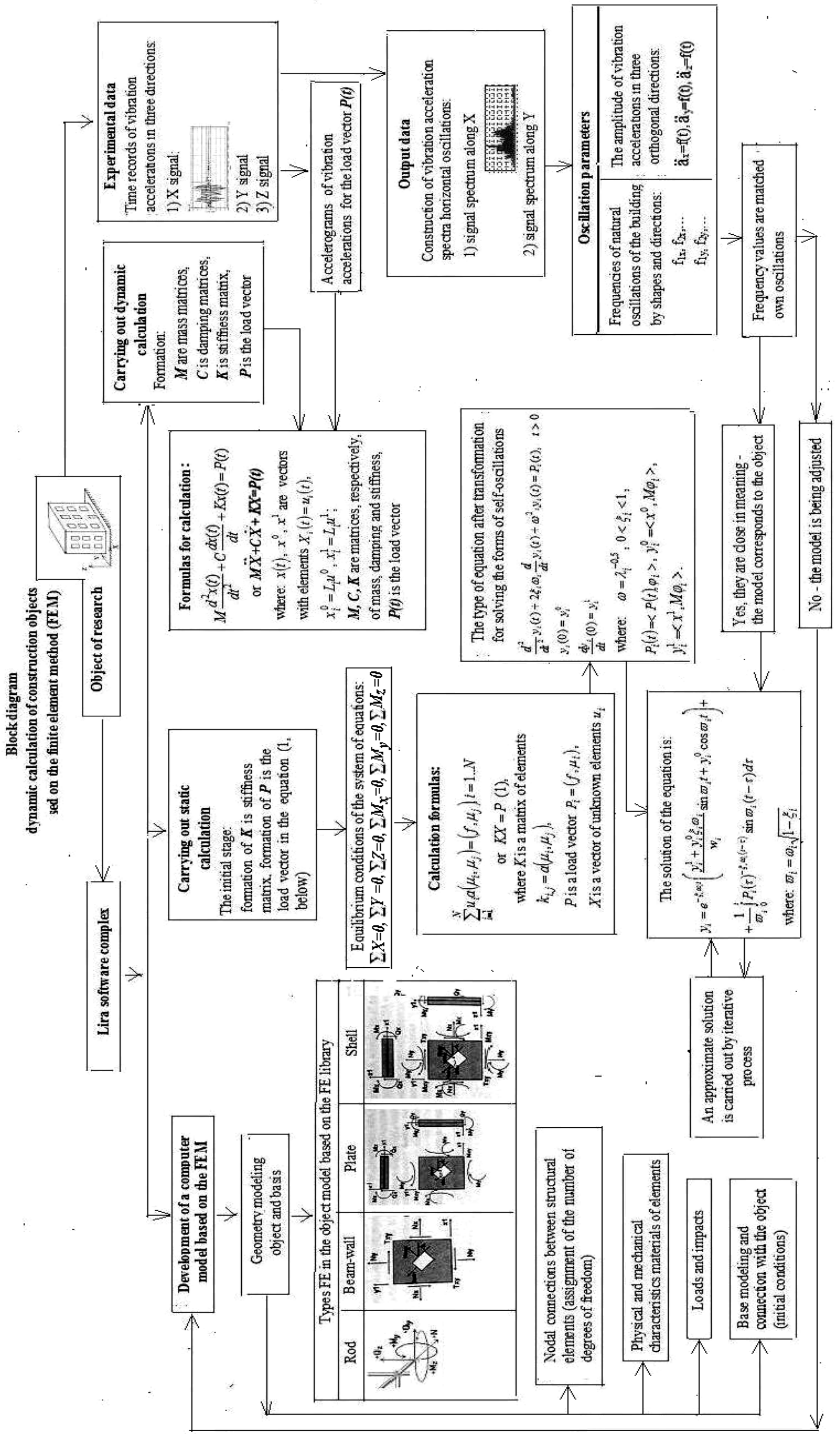


Fig. 6. Block-diagram of construction projects dynamic analysis based on the finite element method (FEM)*c*

Thus, the FEM use reduces the problem to a system of linear algebraic equations. An important advantage of the described method is that the ***K*** matrix and the ***P*** vector are obtained by summing the corresponding elements of the stiffness matrices and load vectors constructed for the individual finite elements. The Finite Element Library (FEL) contains elements to model the operation of various types of structures including the rod elements; the quadrangular and triangular elements of a planar problem such as slabs or shells; the spatial problem elements, for instance, tetrahedron, parallelepiped or trihedral prism. In addition, there are various special elements in the FEL that model the connection of finite stiffness, elastic compliance between nodes, or elements specified by the numerical stiffness matrix. All finite elements in the FEL are theoretically justified and the energy and displacement error assessments are obtained for them.

In the dynamic analysis, the principle of possible FEM displacements is changed in comparison with the static calculation according to formula (1) and takes the following form:

$$b(u,v)+c(u,v)+a(u,v)=(f(t),v)\,,\ t>0\,, \tag{4}$$

where u is the desired exact solution; v is any possible displacement; $b(u,v)$ and $c(u,v)$ are the possible actions of inertial and damping forces, $a(u,v)$ and $(f(t),v)$ are the possible actions of external and internal forces. The problem of dynamic analysis is formulated in the form of a variational equality with partial derivatives:

$$b\left(\frac{\partial^2 u}{\partial t^2},v\right)+C\left(\frac{\partial u}{\partial t},v\right)+a(u,v)=(f(t),v),\quad t>0\,;$$

$$u(0)=u^0\,;\ \frac{\partial u}{\partial t}(0)=u^1\,, \tag{5}$$

where $u_i=u(t)$ is the exact solution; u^0 and u^1 are the initial values of displacement and velocity; other values are the same as in the static problem.

The method of the dynamic problem solution implemented in the LIRA software envisages combining the FEM with the decomposition according to the natural vibrations shapes. The solution (5) is searched in the following form:

$$U_h=\sum_{i=1}^{N}u_i(t)\mu_i\,, \tag{6}$$

where $u_i(t)$ are the scalar functions; μ_i are the basic functions of the corresponding static problem.

Substituting, by analogy, U_h instead of u and μ_j ($j=1..N$) instead of v in (5), a system of ordinary differential equations given for the block-diagram elements for dynamic analysis in Fig. 6 is obtained. The equation written in a matrix form is also given there. In the above equation from the block-diagram, the mass matrix M is determined by the elements $m_{i,j}=b(\mu_i,\mu_j)$, and the damping matrix C — by $c_{i,j}=c(\mu_i,\mu_j)$, respectively. The stiffness matrix K is determined by the elements $k_{i,j}=a(\mu_i,\mu_j)$ and the loading vector $P(t)$ is determined by the elements $P_i=(f(t),\mu_i)$ by analogy with a static problem.

$$M\frac{d^2x(t)}{dt^2}+C\frac{dx(t)}{dt}+Kx(t)=P(t), \quad (7)$$

where $x(t)$, x^0, x^1 are the vectors with the elements $X_i(t)=u_i(t)$, $x_i^0=L_iu^0$, $x_i^1=L_iu^1$.

This method is known as semi-discrete approximation. Its error (the difference between U and U_h) in terms of the potential and kinetic energy is assessed in both compatible and incompatible cases by a value proportional to h^τ.

System (7) is calculated by the method of decomposition by the natural vibrations shapes. Let λ_i, φ_i be the eigenvalue problem solution; $<M\varphi_i,\varphi_i>=1$, where the symbol $<,>$ denotes the scalar product:

$$K\varphi=\lambda M\varphi. \quad (8)$$

The eigenvalue problem (8) is calculated by the subspace iteration method. Considering that in (7) $x(t)=\sum_{i=0}^{N}y_i(t)\varphi_i$, the orthogonality of the function φ_i will ensure (under certain assumptions about the matrix *C*) the system (7) disintegration into independent equations with respect to $y_i(t)$, which are presented in the block-diagram for the natural vibrations shapes solution (6).

The equation solution with respect to $y_i(t)$ is as follows:

$$y_i=e^{-\xi_i\omega_i t}\left(\frac{y_i^1+y_i^0\xi_i\omega_i}{w_i}\sin\varpi_i t+y_i^0\cos\varpi_i t\right)+\frac{1}{\varpi_i}\int_0^t P_i(\tau)^{-\xi_i\omega_i(t-\tau)}\sin\varpi_i(t-\tau)d\tau,$$

where $\varpi_i=\omega_i\sqrt{1-\xi_i}$. The vectors of the inertial forces $S_i(t)$ are calculated by the formula $S_i(t)=\omega_i^2y_i^{(t)}M\varphi$. The values $S_{i,0}=_t^{\max}\{|\omega_i^2y_i(t)|\}$ are used for calculations. Their values are selected depending on the loads type including wind, seismic, impulse, impact, or harmonic loads.

The time-dependent dynamics generally involves the four loads specifying as follows:

The first load is a static load to the structure. For example, dead load of the structure or self-weight with technological load etc. (it is not mandatory for specifying and may be absent). The static influences are specified in the form of concentrated forces and moments both at the scheme nodes (nodal loads) in the directions of the global and local coordinate systems axes, and at the elements (local loads) in the directions of the local or global coordinate system.

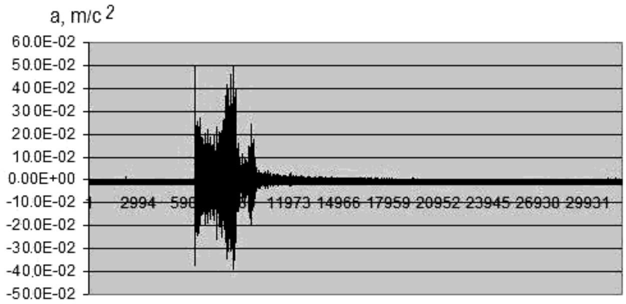
The second load characterizes the inertial characteristics of the structure (mass distribution). It can be collected from the first load, from the elements density, by concentrated masses specifying in the second load, or by the listed options combination.

The third load establishes the active dynamic load for the structure. In the LIRA-9.6 software package [32], four types of dynamic loads are implemented: a broken line with arbitrary segments (the quantity of points pairs and the "time – value" points pairs themselves are specified); a sinusoidal load specified as $A\sin(\omega t+\varphi)$; an accelerogram; a broken line with uniform segments (the points quantity, start time, sampling step, scaling factor and broken line value are specified). The dynamic effects are specified as nodal loads acting along the axes of the global or local coordinate systems. The weight of the structure mass is specified as the dead load of structures, equipment etc., while the use of both local and nodal loads is allowed.

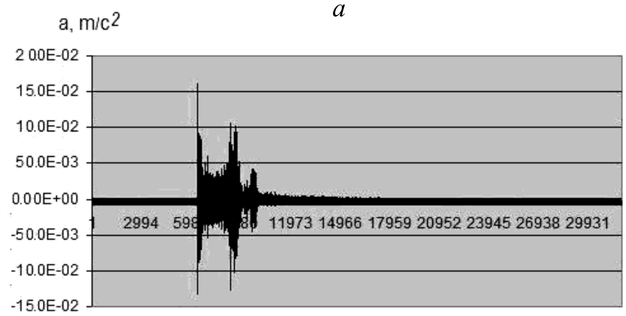
In the fourth load, the structure damping characteristics are specified (they are not mandatory for specifying and may be absent). The load direction is not important, because all values are taken according to their absolute values and in the process of integration the damping forces will be directly proportional to the velocities.

The boundary conditions in the design scheme can be specified directly for the node, as well as modeled using finite stiffness connections. The latter is especially effective if it is necessary to define responses in overlapping connections. After specifying two to four loads in the sequence indicated above, for successful execution of the calculation it is necessary to specify the parameters for the displacements equations integration and specify just which calculation results are required by the user: only displacements, displacements and forces, or displacements, forces and calculated combinations of forces.

Loads and actions are set according to the DBN [33] and seismic effects are taken into account based on the DBN [25]. It should be noted that the actual soil accelerograms of radial and tangential direction affecting the building at angles of 45° and 135°, respectively, are accepted as the external seismic effects on the building under consideration. The accelerograms were recorded near the building on 05/16/2012 during a massive explosion of 652 tons in the PivdGZK quarry. Graphs of digitized accelerograms of the radial *R* and tangential *T* directions are plotted in Fig. 7.



a



b

Fig. 7. Calculated accelerograms of radial *R* (*a*) and tangential *T* (*b*) components of the seismic effect on the Center building

The accelerograms feature the following main parameters: 4 ms discretization step, 245 s total impact duration, 17 s explosion duration, 0.5 m/s maximum accelerogram value for the seismic impact radial component R, and 0.15 m/s maximum accelerogram value for the seismic impact tangential component T. Two design schemes of the building are created for conducting the theoretical studies: scheme 1, in which there is no damage, and scheme 2 with the main damages in the load-bearing walls of the Center building (Fig. 3, *b*):

- the design scheme 1 reflects the original building condition at the time of its commissioning (≈1969–1971);
- the design scheme 2 reflects the building condition at the time of the inspection (May 2013).

The digital values of the stress-strain state dynamic parameters of the Center building computer model are the main criteria for checking the model correctness (comparison was made with the results of experimental and visual inspections of the Center building). The directions of the seismic effects on the building during explosions are shown in Fig. 8, where the Center building computer model is presented.

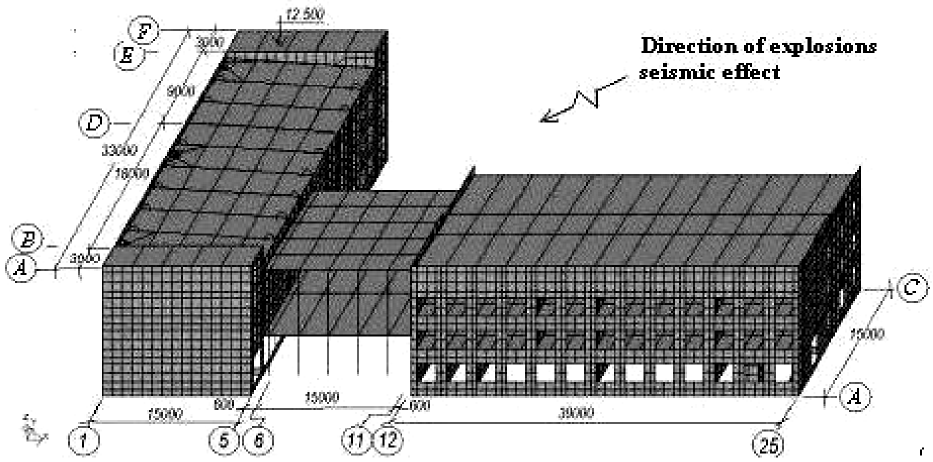


Fig. 8. The general view of the computer model of the Center for Children and Youth Creativity "Mriya" building in the Kryvyi Rih city

The design scheme 1 is created based on the following types of finite elements: rod elements of type 10 (six degrees of freedom at the node) and plate elements of types 41, 42 and 44 (six degrees of freedom at the node). The building stabilization is done at the base level. For nodes of plate elements adjacent to the base the angular displacements are allowed. The column nodes at the level of the base are constrained. The calculations at that stage are performed in a linear version.

In Fig. 3, *b* the section of the load-bearing wall along axis 1 with an oblique crack passing from the window opening corner at the second floor and its simulation in the building computer model are shown. Cracks with openings smaller than the shown one (Fig. 3, *b*) are not modeled. The crack is modeled in the LIRA-9.6 software package [32] with the use of the finite element method for dividing the nodes and elements of the building computer model connected along the crack line and combining these nodes displacements in the direction of existing connections in the building computer model.

The building dynamic parameters in the design scheme 2 are accepted for the further analysis if they are close to the experimental data, or if necessary, they are corrected by iterative calculations to refine the design scheme. Also, the dynamic parameters of the computer model are the main criteria for confirmation its similarity to the real building. When the dynamic analysis of each of the mentioned design schemes (1 and 2) was performed, the number of specified vibrations shapes corresponded to 10.

Table shows the design values of the natural vibrations frequencies for the design scheme 2 of the building with damage, as well as experimentally recorded values.

For that condition, the experimental values of the natural vibrations frequencies of the Center of Creativity building for the first main shapes of vibrations in the directions of the numerical and letter axes were obtained during experimental vibrodynamic surveys conducted on 16.05.2012.

Table. Dynamic characteristics of the Center building computer model 2

Numbering of vibration shapes	**Frequency of vibrations, Hz (design)**	**Frequency of vibrations, Hz (experimental)**	**Relative error, %**	**Vibration period, s**	**Sum of modal masses during seismicity, %**	
					for *X*	for *Y*
1	4.561	4,25	7,3	0.219	55.570	12.725
2	4.693	-“-		0.213	59.979	65.489
3	5.161	-“-		0.194	59.997	65.607
4	5.355	-“-		0.187	60.006	65.609
5	5.669	-“-		0.176	60.046	65.635
6	5.712	-“-		0.175	60.046	65.639
7	6.219	6,7	7,2	0.161	75.127	72.458
8	6.246	-“-		0.160	76.368	73.025
9	6.678	-“-		0.150	76.927	74.082
10	6.751	-“-		0.148	76.934	74.092

Note. The vibration shapes determining the parameters of the building stress-strain state are indicated by the filling.

During the vibrodynamic surveys of the Center building, based on the experimental data an amplitude spectrum of vibration accelerations was obtained (Fig. 5), according to which the experimental values of natural vibrations frequencies f_i (and periods T_i) of the building by the shapes were obtained:

— for X direction: $f_1 = 4.50\,\text{Hz}\ (T_1 = 0.235\,\text{s})$, $f_2 = 6.70\ \text{Hz}\ (T_1 = 0.149\,\text{s})$;

— for Y direction: $f_1 = 4.50\,\text{Hz}\ (T_1 = 0.222\,\text{s})$, $f_2 = 5.40\ \text{Hz}\ (T_2 = 0.185\,\text{s})$.

In Table for the building computer model 2 the values of natural vibrations frequencies and periods according to shapes 1, 2, 7 are close to their experimental values (a relative error does not exceed 7.3%). Thus, the Center building computer model 2 with the crack simulation in the upper part of the building load-bearing wall along the axis 1 was verified as a starting point for mathematical modeling of the stress-strain state of the Center building.

Along with the dynamic parameters assessment, the developed computer model 2 performed an assessment of the building stress-strain state under dynamic influences. The maximum displacements reaching 0.0503 m during the dynamic explosion influences on the building in the *X* direction were obtained for the wall along the axis 1 in the area in the B–I axes (Fig. 7). The deformed state of the building wall along the axis 1 in the *X* direction was graphically displayed in 3D in Fig. 9.

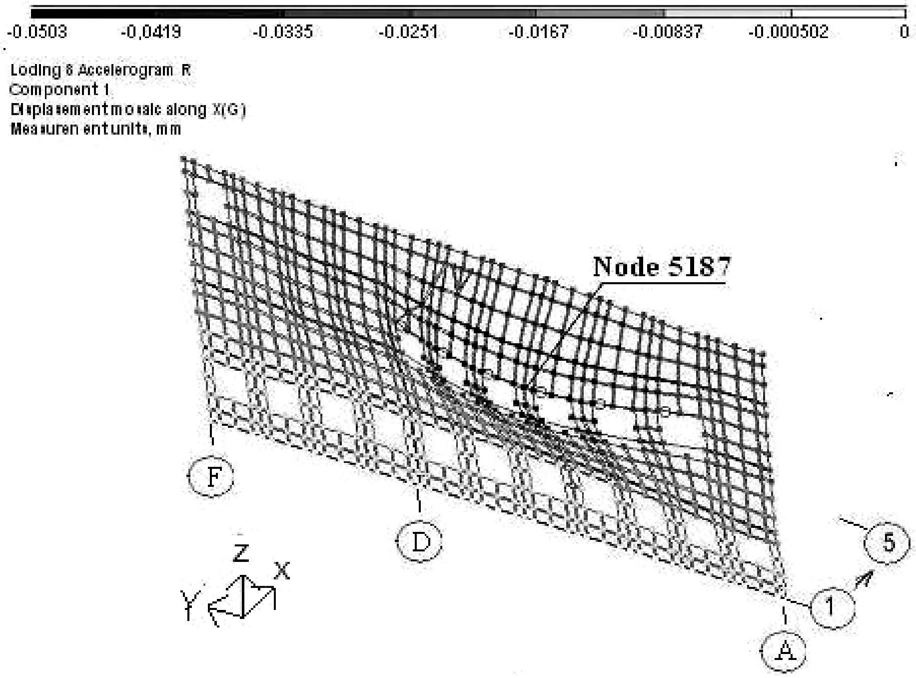


Fig. 9. The deformed state of the building wall along the axis 1 under seismic effects in the *X* direction

The auditorium space of the Center for Creativity on the second floor of the building is limited by the presented area of the load-bearing brick wall along the axis 1 with the largest deformations under seismic effects, which is the most dangerous for spectators. The reinforced concrete covering beams at 10.85 elevation and the upper belt of the transverse frame along the I axis at 11.80 elevation that bear the covering elements, are the most responsible structural elements of the building roof, therefore they determine its durability (resource). The destruction risks calculation was performed for those elements to determine the reduction of those elements resource under the seismic loads influence. The calculation results are given in the paper [34].

System analysis. The final diagram of the load-bearing capacity changes for a separate load-bearing element and the entire Center building is presented in Fig. 10 based on the synthesis of visual inspection, experimental studies and mathematical modeling materials.

Each section of the diagram of the building bearing capacity changes is constructed for a time interval of about 30 years and covers the entire period of the design building operation of ≈100 years. The diagram represents four categories of the building technical condition defined according to [34]. Graph 2 of the building bearing capacity changes during the operation time T is a smooth curve described by the theoretical dependence $P = P_0 - \Delta P(T, T^2)$. The points, where the vertical line drawn from the middle of the interval of each section of the building bearing capacity changes diagram and graph 2 intersect, determine the values and positions of the diagram sections for the corresponding categories of the building technical condition.

That dependence 2 does not describe existing defects and effects. But if they are taken into account, it becomes possible to assess the building technical condition and, according to the given diagram, to assess its bearing capacity.

At any moment the bearing capacity of a separate load-bearing element of the building under the external forces actions (for instance, seismic effects) should a priori be greater comparing to the building itself and its serviceability should be longer. Graph 1 in Fig. 10 corresponds to that conclusion. Accordingly, graphs 3 and 4 compared to graph 2 a priori demonstrate the decrease of the bearing capacity of the building and of its service life when there are seismic effects (graph 3) or they are combined with a damage (graph 4).

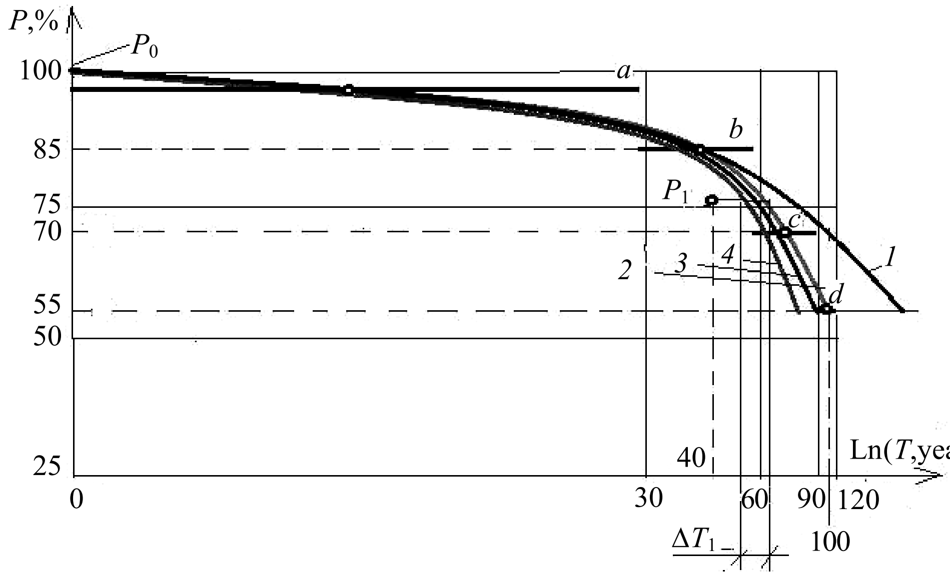


Fig. 10. Comparative graph of the load-bearing capacity changes depending on the period of operation for the load-bearing structural element of the Center building and for the entire building. Segments of straight lines *a*, *b*, *c*, *d* are the elements of the diagram of the entire building load-bearing capacity changes for four categories of technical condition during the entire period of the building operation: building in *a* — normal condition, *b* — building in a satisfactory condition, *c* — building unsuitable for normal operation , *d* — building in an emergency condition; *1* — graph of the load-bearing capacity changes in a separate load-bearing structural element of the building under seismic effects, *2* — graph of the load-bearing capacity changes in the building as a whole without seismic effects, *3* — graph of the load-bearing capacity changes in the building without damage during seismic effects, *4* — graph of the load-bearing capacity changes in the building with damages during seismic effects; point P_1 characterizes the technical condition of the Center building at the time of research (2012) after 40 years of its operation

At the initial stage of the building or the load-bearing element operation (within 30 years) the load-bearing capacity changes over time are insignificant, even under the seismic loads effects and damages occurrence. The reduction of the building service life in this case will be $\Delta T \approx 0$ (identity of graphs 1, 2 and 4 at this area).

As the period of building operation increases, the effect of seismic loads from industrial explosions and its combination with the building structures damages cause the development of detected defects and, accordingly, the accelerated decrease of the building bearing capacity. According to the DSTU, the building emergency condition is determined by the reduction of the building bearing capacity to a value of 55% of the initial value (100%).

The building technical condition determination is carried out based on the visual and instrumental inspections of load-bearing and envelope structures. Clarification of the building bearing capacity value is carried out using the computer models analysis with taking into account the results of the above-mentioned inspections.

The building inspections for its technical condition determining and appropriate calculations performing for the bearing capacity assessment allow to determine the service life reduction ΔT at the time of the inspections according to graphs 2 and 4. In Fig. 10 the value of the building bearing capacity P_1 for the operation period of 40 years is shown. During that period of the building operation the value ΔT_1 undergoes lowering.

For the bearing capacity emergency value of P=55%, the design building service life will be reduced to the maximum value ΔT_{max}. In this case, the calculated building service life compared to the design $T_{пр}$ =100 years will be reduced to the $T=T_{пр}-\Delta T_{max}$ value.

To prevent a significant reduction of the building service life in the presence of damages and seismic effects, it is necessary to increase the frequency of visual and instrumental inspections after 40 years of building operation. The building bearing capacity clarification based on the building computer model relevant calculations should preferably be carried out taking into account the new inspection data for the period of building operation of 60 and 70 years.

The reliability and in-depth analysis of the obtained data for the likely continuation of experimental studies of the Center building must be ensured by repeating, as a prerequisite, the schemes, quantitative composition, as well as the types of vibration transducers and their arrangement on the building structures, which were used during the primary study. If necessary, the layouts of vibration transducers arrangement can be supplemented based on the results of visual inspections.

As a result, it can be seen from Fig. 10 [35; 36] that the design age of the Center building differs from the actual one by ~ 30 years. Thus, for a comprehensive assessment of the life resource of the Center of Creativity building, it is necessary to take into account the technical condition of all its elements and the results of all types of surveys, including the instrumental and visual inspections, as well as the results of mathematical modeling.

The defects revealed by the visual inspection in combination with calculated and experimental data allow to characterize the general technical condition of the Center building as unsuitable for normal operation already in the nearest future. With the growth of internal defects in the load bearing reinforced concrete structures, the risk of their destruction will increase, which will accelerate the building aging and its resource significant reduction (increase of ΔT_1).

CONCLUSIONS

1. Based on research and experimental data obtained in 2005–2012 the intensity of the dynamic effects caused by explosions at the PivdGZK quarry on the

housing stock of the city of Kryvyi Rih is assessed to be within 2–3 (4) points range (the velocity of soil vibrations near buildings varies within 0.002–0.004 m/s). At the same time, the maximum mass of explosives for explosions in the PivdGZK quarry during the 2008–2012 observation period is from 490 to 652 tons.

2. The use of only technical means (use of decelerators during explosions) for the reduction of the level of seismic effects on buildings and soil during explosions weighing 500–650 tons in the PivdGZK quarry does not allow to significantly reduce their intensity.

3. Based on a systematic combination of numerous experimental data and mathematical modeling results, the technical condition deterioration forecast for the Center building reinforced concrete structures in the conditions of permanent explosive impacts from the PivdGZK quarry is obtained. The risks of the Center building load-bearing elements destruction are calculated for all vulnerable zones, which allows to predict and assess the service life resource that decreased compared to the normative one by ~ 30 years.

4. The rate of the Center building life resource decline is determined. The defects detected during the visual inspections and vibrodynamic surveys and the obtained calculation characteristics (2012–2013) allow to characterize the general technical condition of the Center building as unsuitable for normal operation in the nearest future.

5. A systematic assessment of the buildings and structures life resource should take into account the technical condition of all their elements and the results of all types of surveys, including the instrumental and visual inspections, mathematical modeling and their data final synthesis.

6. The experimental and theoretical method of the life resource assessment can be successfully used for the dynamic certification of buildings and structures in Ukraine and is already applied by the SE NDIBK experts in the comprehensive assessment of destroyed and damaged buildings in the cities of Borodianka, Bucha, Vorzel, Gostomel, Irpin, Chernihiv etc., where combat operations have already stopped.

REFERENCES

1. Cabinet of Ministers of Ukraine. *Resolution On the unified national system of prevention and response to man-made and natural emergency situations*. 1998.
2. President of Ukraine. *Decree On the concept of protection of the population and territories in the event of threats and emergency situations*. 1999.
3. *Rules for certification and assessment of the actual seismic resistance of military buildings and structures: VSP - 22-01-95*. M., 1996, 43 p.
4. B.Ye. Paton, “Resource problems of structures, buildings and equipment in Ukraine,” *All-Ukrainian scientific and practical conference on Reconstruction of buildings and structures. Experience and problems*, Kyiv, NDIBK, 2001.
5. Ya.M. Aizenberg, *Development of concepts and norms of anti-seismic design*. M.: VNIINTPI, 1977, 70 p.
6. Ya.M. Aizenberg, “Seismic zoning in the building codes of the CIS countries,” *Seismic resistance of structures. Building safety*, no. 3, 2001.
7. V.V. Meleshko, A.A.Bondarenko, A.N. Trofimchuk, and R.Z. Abasov, “Elastic waveguides: history and the state of the art. II,” *J. Math. Sci*., vol. 167, no. 2, pp. 197–216, 2010. Available: https://doi.org/10.1007/s10958-010-9915-z
8. M.U. Ashimbaev and I.E. Itskov, “Standards for the design of high-rise buildings in seismic areas,” *Interdepartmental Science and Technology Coll. “Building structures*”, vol. 64, pp. 555–558, 2006.
9. A.N. Trofimchuk, “Unsteady Oscillations of a Liquid-Saturated Poroelastic Soil Layer,” *International Journal of Fluid Mechanics Research*, vol. 29, issue1, 2002.

Available: https://www.dl.begellhouse.com/journals/71cb29ca5b40f8f8,6f0c9aea60dcc7ce,61c3aaba422d63b3.html

10. I. Kaliukh, V. Senatorov, O. Khavkin, T. Kaliukh, and K. Khavkin, "Experimental and analytic researches on technical state, design and operation of reinforced concrete anti-landslide structures for seismic dangerous regions of Ukraine," *Proceedings of the fib Symposium on Engineering a Concrete Future: Technology, Modeling and Construction, Tel-Aviv, 2013*.
11. A. Alonso-Rodriguez, N. Nikitas, J. Knappett, G. Kampas, I. Anastasopoulos, and R. Fuentes, "System identification of tunnel response to ground motion considering a simplified model," *Frontiers in Built Environment*, vol. 4, pp.1–11, 2018.
12. M. Balducci et al., "Design and built of a ground reinforced embankment for the protection of a provincial road (Assisi, Italy) against rockslide," *Proc. XXIV conv. Naz. Geotecnica, AGI, Napoli, 22–24 June 2011*.
13. M. Barla, F. Antolini, and S. Dao, "Il monitoraggio delle frane in tempo reale," *Strade e Autostrade*, vol.107, pp. 154–157, 2014.
14. I. Kaliukh, G. Fareniuk, O. Trofymchuk, I. Fareniuk, and Y. Berchun, "Identification of defects in reinforced concrete piles based on multi-wave reflection", in *W. Derkowski, P. Gwoździewicz, Ł. Hojdys, P. Krajewski (Eds). Proceedings fib Symposium, 2019: Concrete – Innovations in Materials, Design and Structures, Fédération Internationale du Béton (fib) – International Federation for Structural Concrete*.
15. N. Casagli, F. Catani, C. Del Ventisette, and G. Luzi, "Monitoring, prediction, and early warning using ground-based radar interferometry," *Landslides*, 7(3), pp. 291–301, 2010.
16. W. Frodella et al., "A method for assessing and managing landslide residual hazard in urban areas," *Landslides*, 15(2), pp. 183–197, 2018.
17. E. Intrieri et al., "Application of an ultra-wide band sensor-free wireless network for ground monitoring," *Engineering Geology*, vol. 238, pp. 1–14, 2018.
18. G. Lollino and A. Chiara, "UNESCO World Heritage sites in Italy affected by geological problems, specifically landslide and flood hazard," *Landslides*, 3(4), pp. 311–321, 2006.
19. M. Martinelli, A. Burghignoli, and L. Callisto, "Dynamic response of a pile embedded into a layered soil," *Soil Dynamics and Earthquake Engineering*, vol. 87, pp. 16–28, 2016.
20. M. Shokrabadi, H.V. Burton, and J.P. Stewart, "Impact of Sequential Ground Motion Pairing on Mainshock-Aftershock Structural Response and Collapse Performance Assessment", *Structural Engineering*, 144 (10), 04018177, 2018.
21. Y. Wang and E.M. Rathje, "Application of a Probabilistic Assessment of the Permanent Seismic Displacement of a Slope", *Journal of Geotechnical and Geoenvironmental Engineering*, 144 (6), 04018034, 2018.
22. S.N. Savin, "Diagnosis of building structures of buildings and structures by the method of free vibrations," *NDIBK collection of scientific papers "Construction in seismic regions of Ukraine*", no. 64, pp. 395–400, 2006.
23. V.A. Zarenkov, I.D. Zakharov, S.N. Savin, and A.F. Shnitkovski, *Modern methods of technical diagnostics of building structures of buildings and structures*. St.-P.: RDK-print, 2000.
24. L.P. Berzhynskaia, *Reliability of regional types of buildings under seismic impacts (on the example of the Baikal region): Technical Science candidate's thesis*. Ulan-Ude: East Siberian State Technological University, 2006.
25. Construction in seismic regions of Ukraine: DBN V.1.1-12:2006. 2006.
26. Scientific and technical support of construction projects: DBN V.1.2-5:2007. 2007.
27. I. Kaliukh, V. Senatorov, N. Marienkov, O. Trofymchuk, K. Silchenko, and T. Kalyukh, "Arrangement of deep foundation pit in restricted conditions of city build-up in landslide territory with considering of seismic loads of 8 points. Geotechnical Engineering for Infrastructure and Development," *Proceedings of the XVI European Conference on Soil Mechanics and Geotechnical Engineering, ECSMGE, 2015*, 3, pp. 955–959.
28. O. Trofymchuk, O. Lebid, O. Klymenkov, Y. Berchun, and V. Berchun, "Dynamic certification of landslide protection structures in a seismically hazardous region of Ukraine. Experimental and analytical research," in *Earthquake Geotechnical Engineering for Protection and Development of Environment and Constructions, Silvestri & Moraci (Eds). Associazione Geotecnica Italiana. Proceedings of the 7th International*

Conference on Earthquake Geotechnical Engineering, ICEGE 2019, June 17–20, 2019, Rome, Italy.

29. O. Trofymchuk, I. Kaliukh, K. Silchenko, V. Polevetskiy, V. Berchun, and T. Kalyukh, "Use of Accelerogram of Real Earthquakes in the Evaluation of the Stress-Strain State of Landslide Slopes in Seismically Active Regions of Ukraine," *Engineering Geology for Society and Territory* , vol. 2. Cham: Springer, 2015. Available: https://doi.org/10.1007/978-3-319-09057-3_236
30. S.V. Medvedev, *Seismic of rock explosions*. M.: Nedra, 1964.
31. *Rationing of vibration of structures in the USSR and abroad*; Issue 1. M., 1990, 60 p.
32. *LIRA 9.2. User's manual. Basics: Tutorial*. Kyiv, 2005, 60 p.
33. "Loads and actions: DBN V.1.2-2:2006," *Minbud of Ukraine*, Kyiv, 2006.
34. *Guidelines for inspection of buildings and facilities for identification and evaluation of their technical condition: DSTU-N B.1.2-18:2016*. Kyiv, 2017.
35. I. Kaliukh, I. Syvko, K. Chorna, V. Shokarev, I. Shokarev, and A. Shokarev, "Assessment of Residual Service Life for the Residential Buildings under Systematical Industrial Explosions," *IEEE 2nd International Conference on System Analysis & Intelligent Computing* (*SAIC, 2020*). doi: 10.1109/SAIC51296.20s20.9239172.
36. V.A. Dunin, *Influence of industrial explosions on durability of constructions of buildings in the conditions of Kryvyi Rih: Technical Science candidate's thesis*. Kyiv: The State Research Institute of Building Constructions, 2021.

Received 14.04.2022

INFORMATION ON THE ARTICLE

Oleksandr M. Trofymchuk, ORCID: 0000-0003-3358-6274, Institute of telecommunications and global information space of NAS of Ukraine, Ukraine, e-mail: itc_nasu@ukr.net

Iurii I. Kaliukh, ORCID: 0000-0001-7240-4934, Institute of telecommunications and global information space of NAS of Ukraine, The State Enterprise "The State Research Institute of Building Constructions", Ukraine, e-mail: kalyukh2002@gmail.com

Volodymyr A. Dunin, The State Enterprise "The State Research Institute of Building Constructions", Ukraine

Sergiy Y. Kyrash, The State Enterprise "The State Research Institute of Building Constructions", Ukraine

ДИНАМІЧНА СЕРТИФІКАЦІЯ ТА ОЦІНЮВАННЯ ЖИТТЄВОГО ЦИКЛУ БУДІВЕЛЬ ЗА РЕГУЛЯРНОГО ВИБУХОВОГО ВПЛИВУ / О.М. Трофимчук, Ю.І. Калюх, В.А. Дунін, С.Ю. Кураш

Анотація. Нині в Україні не існує єдиної узаконеної загальноприйнятої методики (на рівні українського будівельного стандарту) динамічної паспортизації будівель і споруд. Запропоновано єдиний підхід у значенні такої методики. Він містить чотири компоненти: візуальний огляд будівель; експериментальні дослідження динамічної реакції будівель або споруд на вибухові дії; математичне моделювання напружено-деформованого стану досліджуваного об'єкта; узагальнення результатів візуального огляду, експериментальних досліджень та числового моделювання з метою їх систематичного узагальнення. Як апробацію наведено погіршення ресурсу залізобетонних конструкцій житлових будинків в умовах постійних масових промислових вибухів потужністю від 500 до 700 тонн в кар'єрі Південного ГЗК (ГЗК) м. Кривий Ріг, Україна. На основі оброблення численних експериментальних даних та результатів математичного моделювання отримано ймовірнісну модель прогнозування погіршення технічного стану залізобетонних конструкцій Центру дитячої та юнацької творчості «Мрія». Розрахунки ризиків руйнування несучих елементів будівлі для її вразливих ділянок дали змогу уточнити термін її експлуатації. Він зменшився приблизно на 30 років порівняно зі стандартом 2012 року.

Ключові слова: ударні хвилі, експеримент, ризик, динамічна паспортизація, життєвий цикл будівель і споруд.

UDC 519.652
DOI: 10.20535/SRIT.2308-8893.2022.4.10

APPLICATION OF BEAM THEORY FOR THE CONSTRUCTION OF TWICE DIFFERENTIABLE CLOSED CONTOURS BASED ON DISCRETE NOISY POINTS

I. ORYNYAK, D. KOLTSOV, O. CHERTOV, R. MAZURYK

Abstract. The smoothing of measured noisy positions of discrete points has considerable significance in various industries and computer graphic applications. The idea of work consists of the employment of the technique of beam with spring supports. The local coordinates systems are established for beam straight line segments, where the initial angles between them are accounted for in the conjugation equations, which provide the angular continuity. The notions of imaginary points are introduced, the purpose of which is to approach the real length of the smoothed contour to the length of the straight chord. Several examples of closed denoised curve reconstruction from an unstructured and highly noisy 2D point cloud are presented.

Keywords: spline, elastic beam, spring support, closed contour, imaginary point, noisy data.

INTRODUCTION

Two cases are discerned at geometrical computer modeling of plane curves or closed contours – their design or restoration from the set of measured points. In each case the provision of the curves smoothness is a necessary requirement of their construction.

During geometrical design the important prerequisite is an aesthetical appeal and C^2 continuity of a curve. The most popular are the methods which are based on the Bezier curves, rational Bezier curves, B-splines and rational B-splines, or so-called NURBS curves [1]. For their construction it is needed to define the enumerated system of so-called control points. These curves pass exactly through the first and last points, while all other points are only attracting the curve to themselves. To increase or decrease the "weight" of attraction of the given control point the rational splines are applied.

The problem of restoration of closed curves from the noise points has a big significance in a computational geometry. It is applied in image analysis, in reverse engineering, for restoration of the trajectories and forces acting on the moving material points, etc. [2]. As an example, mention the extraction of silhouettes from sensed depth images.

Very often the restoration of the geometry is performed by fitting the points for the prescribed analytical figure (circle, ellipse, helix) with unknown characteristics, orientation and center of coordinates. In this case the functional of error (sum of deviations) is specified, which usually is iteratively solved by least square method [3]. When the number of points is too big it is necessary to perform the preliminary smoothing. Then the weighting Gaussian probability function is usually applied, which makes the closer positions of the considered points [4]. Yet its application might lead to unpredictable loss of the useful information (over smoothing). Furthermore, the drawback of similar techniques is that the real noise of input points (mean deviation) is not taken into account, even if it is known in advance [4].

Historically the first splines which exactly pass through the given measured points were solutions based on the Strength of Materials beam technique [5] and actually used the known equation of three moments [6]. To decrease the deviation of the curves from the given points it was suggested to use the solutions for initially pre-tensed beams [7]. Later on, the theory of beam splines was supplemented by possibility to use the elastic springs (supports) placed in the points of measurements. The increase in compliance of these springs lead to larger smoothing of the resulting curve [8].

The most drawback of all beam based smoothing techniques is that the curve is presented in unique global coordinate system, i.e., the solution of kind $y = f(x)$ is searching for. Sush presentation can be effective for drawing the statistical splines, where the independent variable x is determined exactly, while the functions y can be presented in any scale. For geometrical splines both coordinated has the similar meaning and both are determined with errors. Usual approach is a parametrical presentation of both coordinates and execution the independent smoothing with respect to each coordinate.

Significant contribution to application of the theory of beams to the geometrical smoothing are works of present authors [9, 10], where the initial straight segments between the neighboring points predetermine the local coordinate and vector basis which are adjacent to each other at a certain angle. Namely, the goal of the beam deformation was the smoothing of these angles [9, 10]. Note, that idea of application of initial straight segments and their vectorial basis is the main one in so-called corotational approaches in geometrical nonlinear analysis of the spatial beams, when the position of the deformed geometry is essentially deviate from the initial one [11].

The peculiarity and main idea of given work consists in that beam spring splines are calculated iteratively with accounting for the big deviation and change of local basis vector system (tangent and normal) [12]. Besides the notion of fictitious points is introduced here which are not related with the real (measured) points. These points are placed on calculated contour between the real points for decreasing the angle between the neighboring segments.

GOVERNING EQUATIONS

Consider the simplest model of the bending deformation of beam, so called Euler-Bernoulli beam [6], its equations and parameters. Beam is described by the generalized vector of state $\vec{Y}(s)$, which characterized by set of four parameters at each local point of length coordinate s, i.e., $\vec{Y}(s) = column\ \{W(s); \theta(s); M(s); Q(s)\}$. Here, instead of some function and its derivatives, as usually accepted in mathematics, the following conventional parameters of the theory of beam are used. So, we operate by notions of: $W(s)$ is the deviation or displacement, the positive value of which is directed toward the normal; $\theta(s)$ is the angle of rotation, the positive direction coincides with rotation from the tangent vector to normal one, i.e., is in clockwise direction; $M(s)$ is the bending moment; $Q(s)$ is transverse force. Their positive directions are chosen in such manner that in all differential equations the sign «+» is used. Thus, the following dependences between the beam parameters are used:

$$\frac{dW(s)}{ds} = \theta(s); \quad \frac{d\theta(s)}{ds} = \frac{M(s)}{EI}; \quad \frac{dM(s)}{ds} = Q(s); \quad \frac{dQ(s)}{ds} = q_m,$$

where q_m is given outer distributed force, EI is so called rigidity of the beam section. The solution of the system of governing equations is presented in form suitable for application of transfer matrix method, and is the following:

$$(\vec{Y}(s)) = \begin{bmatrix} 1 & s & \frac{s^2}{2EI} & \frac{s^3}{6EI} \\ 0 & 1 & \frac{s}{EI} & \frac{s^2}{2EI} \\ 0 & 0 & 1 & s \\ 0 & 0 & 0 & 1 \end{bmatrix} (\vec{Y}_0) + q_m \begin{pmatrix} \frac{s^4}{24EI} \\ \frac{s^3}{6EI} \\ \frac{s^2}{2} \\ s \end{pmatrix}, \tag{1}$$

where $\vec{Y}_0 = \vec{Y}(s=0)$ is the vector of state at initial point of the segment. Note that here we take that $EI \equiv 1$, and $q_m = 0$.

The set of measured positions of enumerated points $B_i(X_{bi}, Y_{bi})$ is a required input information, thus each point has number i, it is characterized by two coordinates position which are measured with some error characterized by some prescribed statistical deviation σ.

At each iteration number, j, the geometry of analyzed curve is presented as a set of points $A_i^j(X_i^j, Y_i^j)$, where each point A_i^j is related with initial point B_i with the same index i. Note, that before first iteration the positions of points $A_i^{j=0}$ coincide with those of measured points B_i. Usually, we will omit the upper index j in designation of points A_i. Consequent points A_i are connected by straight segments. These segments form certain angles between themselves, $\psi_{i-1,i}$, the positive directions of which is counted clockwise (Fig. 1). Thus, the beginning and end of each segment i, at each iteration, j, have some positions, which coincide with positions of points A_i (Fig. 1). The vectorial length of each segment is denoted as $\vec{L}_i$:

$$\vec{L}_i = \vec{A}_{i+1} - \vec{A}_i = (X_{i+1} - X_i)\vec{i} + (Y_{i+1} - Y_i)\vec{j}.$$

Each of these segments is related with local tangent vector $\vec{t}_i$ and normal vector $\vec{n}_i$, which is rotated clockwise with respect to former one. Thus, if the tangent vector $\vec{t}_i$ is given by:

$$\vec{t}_i(s) = \frac{\vec{L}_i}{\left|\vec{L}_i\right|} = a_i\vec{i} + b_i\vec{j} = \text{const}\,.$$

Then normal vector $\vec{n_i}$ is found from the expression:

$$\vec{n}_i(s) = c_i\vec{i} + d_i\vec{j} = b_i\vec{i} - a_i\vec{j} = \text{const}\,.$$

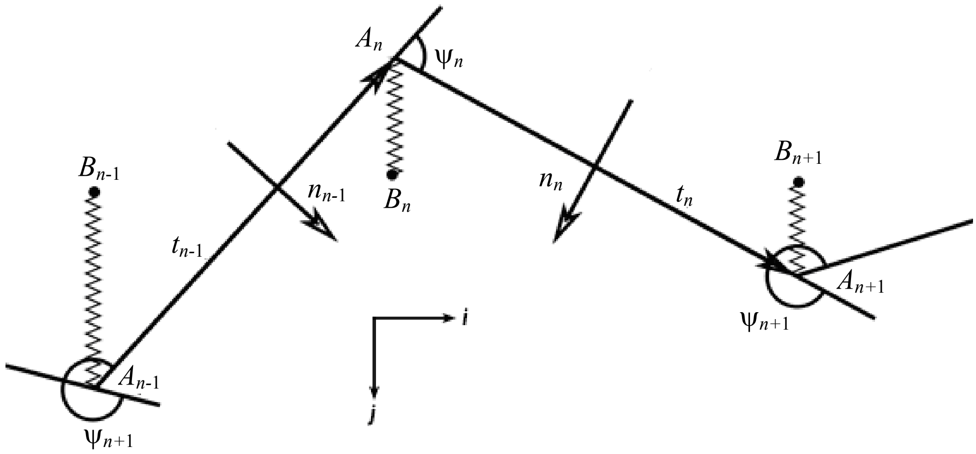


Fig. 1. The model of discrete beam segments on elastic springs (supports)

The important constituent of the modelling is accounting for the angle between two segments, name it as a misalignment angle. It must be smoothed at the next iteration. This angle is denoted as $\psi_{i-1,i}$, and its direction is clockwise, Fig 1, and calculated by the formulas of the vector and scalar products of the unit vectors:

$$\sin(\psi_{i,i-1}) = \vec{t}_i * \vec{t}_{i-1}\,; \qquad \cos(\psi_{i,i-1}) = \vec{t}_i\,\vec{t}_{i-1}\,.$$

We use both definitions of angle to get the correct values of it in each quadrant of the coordinate system. Note, that if both signs are positive, then angle lies in first quadrant, if both are negative – in third quadrant, and so on.

Formulate the main parameters, unknowns and equations which will be used in calculation.

Introduce the following designations for main parameters (unknowns at each iteration). At each segment, i, introduce the vector of state in initial point, thus $\vec{Y}_{i,0}(s=0)$, and vector of state at the last point of segment, thus $\vec{Y}_{i,e}(s=s_i)$. Evidently, these two vectors of state are related by Connection equations (1), at $s = s_i = \left|\vec{L}_i\right|$.

The peculiarity of given approach consists in consideration of initial noisy points as the fixed positions of elastic supports, where the additional force of interaction is proportional to the deviation of the position of the contour point $\vec{A}_i$ from the support point B_i. This deviation is determined before the next iteration and designated as Π_i. Other peculiarity consists in accounting for the angle of misalignment [10]. So, additionally to the Connection equations, write the Conjugation equations, which relate the vectors of state at the end of previous segment with that at the beginning of the next one:

$$W_{i+1,0} = W_{i,1}; \qquad (2)$$

$$\theta_{i+1,0} = \theta_{i,1} - \psi_{i,i+1}; \qquad (3)$$

$$M_{i+1,0} = M_{i,1}; \qquad (4)$$

$$Q_{i+1,0} = Q_{i,1} - C_i(W_{i,1} - \Pi_i), \qquad (5)$$

where C_i is the rigidity of support (characteristic of spring), and Π_i is some conditional distance, the value of which before the first iteration is taken as $\Pi_i = 0$.

Equation (5) can lead to some computational errors when C_i is a very big. As alternative presentation of (5) we use relation

$$(W_{i,1} - \Pi_i) = D_i(Q_{i+1,0} - Q_{i,1}), \qquad (6)$$

where the notion of the support's compliance D_i is introduced, which is inverse to the rigidity $D_i = C_i^{-1}$.

Comment these equations. First one (2) is intended for that displacements (positions) at the border between two segments should be equal — C^0 continuity. Second one (3) is intended for that angle of tangent to the next segment should be on value of $\psi_{i,i+1}$ lesser than the angle of tangent at the end of the previous one – actually this gives the angle continuity requirement of the deformed contour — C^1 continuity. Third condition (4) is a condition of continuity of second derivatives. The values of calculated $W_{i,1}$ and Π_i, the latter is being determined before each iteration by the special procedure, have the signs. The positive sign means that value is directed toward direction of vector $\vec{n}_i$. Note that at intermediate values of D_i and C_i the equations (5) and (6) are equivalent. If $C_i \gg 1$, then equation (6) should be used, which in limit case of $C_i \to \infty$, or $D_i \to 0$, actually leads to equation: $W_{i,1} = \Pi_i$. Similarly, there are no restrictions in applying of (6), when $C_i \to 0$, or $D_i \to \infty$, then equation (5) will be reduced to $Q_{i+1,0} = Q_{i,1}$.

Thus, if the whole number of measured points is equal to N, then we introduce N segments (the last one for closed contours connect the point $i = N$ with point $i = 1$). This means that at whole we have $2*4*N = 8N$ unknowns (two sets of four ones at the beginning and at the end of each segment). For determination of them there are N sets of Connection equations (1), thus $4N$ ones at the whole, and similarly $4N$ Conjugation equations (2) – (5).

MAIN IDEA OF ALGORITHM

Describe the procedure of consequent rebuilding of the smoothed contour at each iteration. The main result of calculation is deviation of points A_i from the previous positions, and they are designated as $\vec{W}_i$. According to Conjugation requirement (2):

$$\left|\vec{W}_i\right| = \left|\vec{W}_{i+1,0}\right| = \left|\vec{W}_{i,1}\right| .$$

Yet, evidently the displacements of beam, $W(s)$ are directed toward the normal to each straight segment, which are different for two adjacent segments. Thus, the vectorial displacements $\vec{W}_{i+1,0}$ and $\vec{W}_{i,1}$ are different:

$$\vec{W}_{i+1,0} = \left|\vec{W}_{i+1,0}\right| \vec{n}_{i+1} \neq \vec{W}_{i,1} = \left|\vec{W}_{i,1}\right| \vec{n}_i .$$

To provide the C^0 continuity of deformed contour we make the following original enhancement: namely, introduce the notion of continuous normal vectors to the calculated deformed geometry. They are derived by rotation of initial normal vectors on the calculated angle of deformation $\theta_i(s)$. We name them as deformed normal, denote them as $\overrightarrow{n\theta}_{i+1}$ and calculate by the following formula:

$$\overrightarrow{n\theta}_i(s) = \vec{n}_i \cos\theta(s) - \vec{t}_i \sin\theta(s) = c\theta_i(s)\vec{i} + d\theta_i(s)\vec{j} , \tag{7}$$

where

$$\begin{pmatrix} c\theta_i(s) \\ d\theta_i(s) \end{pmatrix} = \begin{pmatrix} \cos(-\pi/2 - \theta(s)) & -\sin(-\pi/2 - \theta(s)) \\ \sin(-\pi/2 - \theta(s)) & \cos(-\pi/2 - \theta(s)) \end{pmatrix} \begin{pmatrix} a_i \\ b_i \end{pmatrix} .$$

Evidently, that due to condition (3) the deformed normal are continuous along the whole contour including the borders between adjacent elements, i.e. $\overrightarrow{n\theta}_{i+1}(s=0) = \overrightarrow{n\theta}_i(s=s_i)$. Then treating the calculated vectorial displacements as $\vec{W}_i(s) = W(s) \cdot \overrightarrow{n\theta}_i(s)$, we get that positions of two adjacent border points $\vec{A}_i^j$ will be continuous.

Accounting for continuity of displacements $W(s)$ and angle of deformation $\theta(s)$ calculate the position of each inner point of segment $\vec{R}_i^j(s)$, which is presented as the sum of initial position and calculated deformed position:

$$\vec{R}_i^j(s) = \vec{A}_i^{j-1} + \vec{t}_i \cdot s + W(s) \cdot \overrightarrow{n\theta}_i(s) . \tag{8}$$

Thus, formula (8) is a main formula to determine the positions of all points after each iteration. It may happen, that derived figure satisfy the requirements to the smoothed contour and we can stop the calculations immediately. But if the derived contour does not satisfy to the requirements, we have two options. First is a trivial one. We can start again the procedure from the first iteration, where $A_i^{j=0} = B_i$, but different value of rigidity C_i can be chosen: the smaller it is, the smoother contour will be. This option is applicable when the deviation of measured points is small as compared with distance between them.

More complicated is the second option which requires the creation of additional refining procedures, and which is the main objective of our analysis. Consider that first iteration is already performed at initial (big) values of C_i, take them to be of the same value and designate as C_S. This results in derivation of new positions of points A_i. Connect these points, form new segments and derive new angles of misalignments. Take other (smaller) values of C_s and perform the next iteration.

Important to evaluate whether the given value of C_s is big or small. Evidently the value of C_s can be considered as a small one, when the calculated displacements W are small as compared with segment length. Theory of beams [6] states, that following property characterizes the dimensionless rigidity of spring, $\overline{C}_S = \dfrac{C_S \cdot L^3}{6EI}$. The spring is rigid if $\overline{C}_s > 1$, and is flexible if $\overline{C}_s < 1$. Thus, we initially chose some approximate mean value of length of all segments, L, and then we chose some initial big value of the rigidity, for example, $\overline{C}_s = 400$. Then real C_s is calculated as

$$C_s = \frac{6}{L^3}\overline{C}_S ,$$

because $EI = 1$.

To complete the preparation for the next iteration it is necessary to find the preliminary force of the spring tension, which is characterized by some conditional distance Π_i. If consider Π_i as the absolute distance between points $\vec{B}_i$ and $\vec{A}_i$, then some collisions could occur. It may happen that some point $\vec{B}_i$ lays on (or is very close to) the new contour, but is situated very far from related calculated point $\vec{A}_i$. In such treatment: a) the value of Π_i could be big, thus the large preliminary force will arise, although point $\vec{A}_i$ is very close to contour; b) it is not possible to define the sign of Π_i (“+” or “-”), which would lead to unpredictable results at the next iteration.

Then, value of Π_i is defined as the shortest distance from point $\vec{B}_i$ for both segments $\overrightarrow{A_{i-1}A_i}$ and $\overrightarrow{A_iA_{i+1}}$ with accounting for their signs. These two distances are designated as Π_{1i} and Π_{2i}, and are determined as: $\overrightarrow{A_iB_i} \cdot \vec{n}_{i-1} = \Pi_{1i}$; and $\overrightarrow{A_iB_i} \cdot \vec{n}_i = \Pi_{2i}$. The smallest by module of them is chosen as Π_i with accounting of its sign, so Π_i is taken with sign «+», if point B_i is situated at the right side from the respected segment and «-» otherwise.

Some cases of choosing the value of Π_i are shown on Fig. 2.

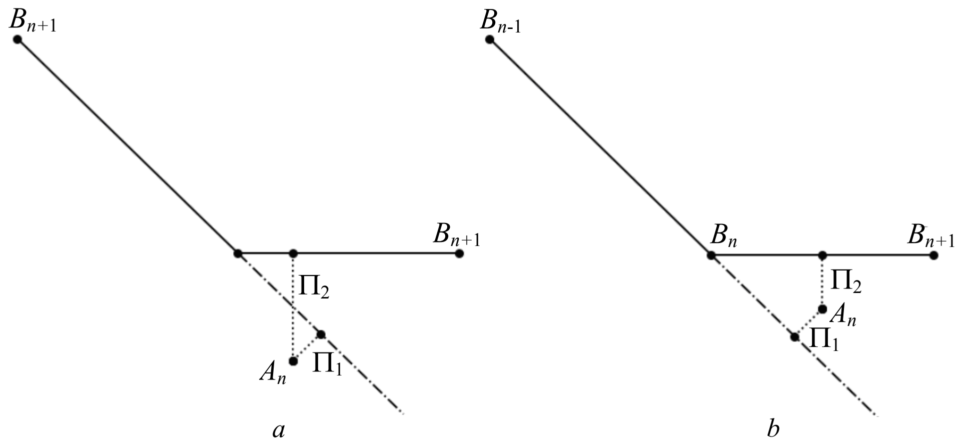


Fig. 2. Some cases of finding the distances from control point A to two straight segments: *a* — one distance is positive and other is negative; *b* — both distances are positive

REFINEMENT OF LENGTH AND CURVATURE AND IMAGINARY POINTS

Before j-th iteration we operate only by the positions of border points A_i^{j-1}. The main result of calculation at j-th iteration is position of each point s, where $0 \le s \le s_i$, of the segment (8). Analyze the change of length of infinitely small straight element Δs of the segment after deformation. Rewrite the expression (8) with accounting for (7):

$$\vec{R}_i^j(s) = \vec{A}_i^{j-1} + \vec{t}_i^{\,j-1} X(s) + \vec{n}_i^{\,j-1} Y(s)\,, \tag{9}$$

where

$$X(s) = (s - W(s)\sin\theta(s))\,,\ Y(s) = W(s)\cos\theta(s)\,.$$

These formulas give the position of each point s. Accounting for known formula of differential geometry get the deformed length and curvature at each point s. Write the auxiliary expressions which are derived by differentiation of (9):

$$\dot{X}(s) = 1 - \dot{W}(s)\sin\theta(s) - W(x)\cos\theta(s)\dot{\theta}(s)\,;$$

$$\dot{Y}(s) = \dot{W}(s)\cos\theta(s) - W(s)\sin\theta(s)\dot{\theta}(s)\,.$$

They allow to find elongation, $\varepsilon(s)$, (relative change of length) of each point s:

$$\varepsilon(s) = \frac{\Delta s_d}{\Delta s} = \sqrt{(\dot{X})^2 + (\dot{Y})^2}\,,$$

where Δs_d is the deformed length of initial element Δs. The curvature, $\kappa_d(s)$, of the curve (9) in each point is given by known expression:

$$\kappa_d(s) = \frac{\ddot{X}\dot{Y} - \ddot{Y}\dot{X}}{\left(\sqrt{(\dot{X})^2 + (\dot{Y})^2}\right)^3} = \frac{\ddot{X}\dot{Y} - \ddot{Y}\dot{X}}{(\varepsilon(s))^3} \tag{10}$$

Expand the components of numerator (10) with accounting for known differential dependences (1):

$$\ddot{X}(s) = M(s)\sin\theta(s) + 2\theta(s)\cos\theta(s)M(s) -$$

$$- W(s)\sin\theta(s)(M(s))^2 + W(s)\cos\theta\cdot Q(s)\,; \qquad (11)$$

$$\ddot{Y}(s) = M(s)\cos\theta(s) - 2\theta(s)\sin\theta(s)M(s) -$$

$$- W(s)\cos\theta(s)(M(s))^2 - W(s)\sin\theta\cdot Q(s)\,. \qquad (12)$$

So, expressions (10) together with (11) and (12) allow to find the real curvature of curve (9).

Now introduce the notion of imaginary points. Theory of beams is based on assumption that calculated angles are small, i.e., $\theta \approx \sin\theta \approx \mathrm{tg}\,\theta$. If the number of points A_i is small, then misalignment angles are large (Fig. 1). Thus, calculated angles are also large which lead to some inaccuracy. To decrease these angles of misalignment it is suggested to introduce the imaginary support points. These supports have zero rigidities $C_i = 0$. Nevertheless, the number of segments becomes larger and angles of misalignments become smaller. In details, this crucial idea of the method will be illustrated on testing examples.

RECOVERY OF FIGURES FROM EXACTLY MEASURED POINTS

Example 1. Start with example when several points of the circle are given exactly. Let we have four points $B_1(x_1 = -2, y_1 = 1)$; $B_2(2,1)$; $B_3(2,-1)$; $B_4(-2,-1)$. Evidently, they are situated on the circle of radius $R = \sqrt{5} \approx 2.236$, thus the curvature κ is equal to $\kappa = 0.4472$.

When all points are given exactly (imaginary points are not exact) there is no necessity to perform the iteration procedure, because the result is obtained at once. The fourth Conjugation equation is used in form: $W_{i,1} = 0$.

The calculation results are shown on Fig. 3. Due to small number of input points the calculated figure is not continuous. This relates to the tangent C^1 and curvature (moment) C^2 continuity. The calculated values of moment (curvature) are constant and equal to $\kappa = 0.5235$, which is 17% higher than accurate value (Fig. 3, *b*). Note, that axis of abscise is related with the lengths of initial straight segments, formed by 4 input points and the whole length is equal to 4+2+4+2=12. The above values of $\kappa = 0.5235$ is the formal solution by the beam theory. The real curvatures are calculated by formulas (10). The graph of real curvatures is shown on Fig. 3, *c*, which extremal values notably differ from exact and beam ones.

To improve the results, employ the idea of imaginary points. Introduce one imaginary point at the middle of each calculated graph section between two real points (Fig. 3, *a*). Denote them as B_{1-2}, B_{2-3}, B_{3-4}, B_{4-1}. Determine their coordinates. Connect consequently points B_1, B_{1-2}, $B_2 \dots B_4$, B_{4-1}, B_1 and get the new graph for eight initial points (Fig. 4).

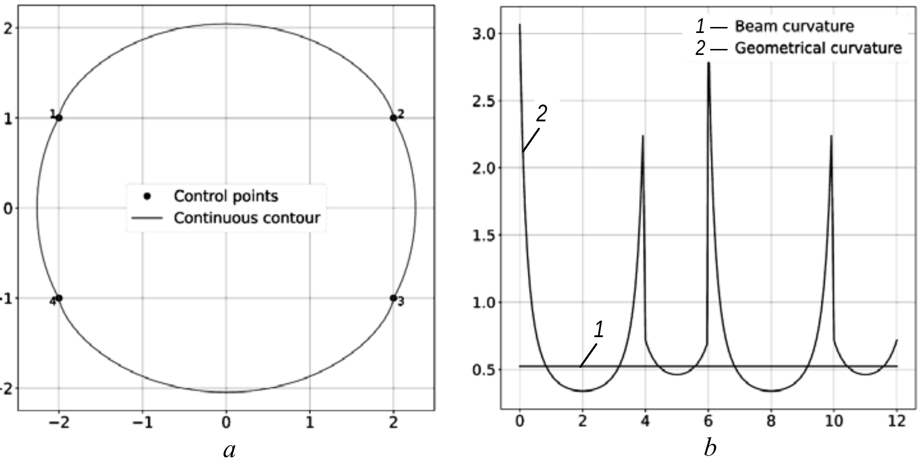


Fig. 3. Recovery of circle from 4 exactly measured points given as the vertices of the rectangle: *a* — calculated figure; *b* — beam and geometrical curvatures

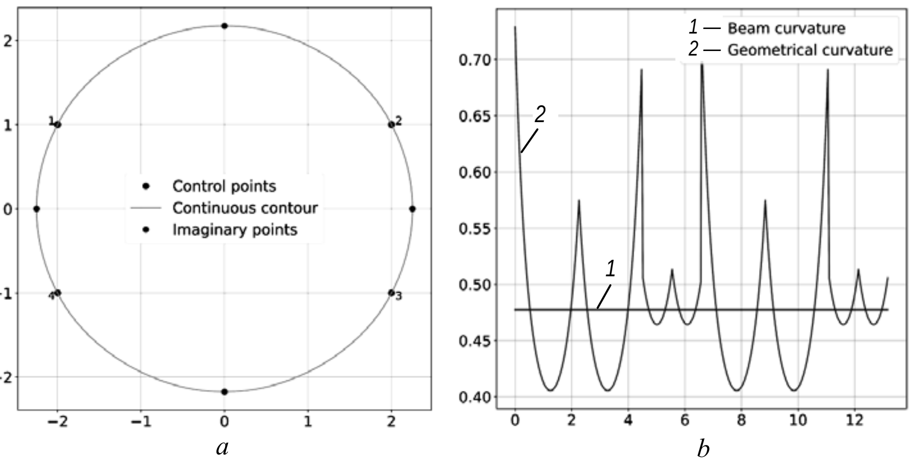


Fig. 4. Recovery of circle by four exact points and four imaginary ones: *a* – restored graph; *b* – recalculated curvatures by formula (10)

As we see the contour is more smooth, angular misalignments are almost invisible. The calculated beam curvatures are equal to 0.4773, which are 6.7% higher than exact ones. Yet the extremal geometrical curvatures given by formula (10) are still in 1.5 times higher than accurate value.

Therefore, insert additional 24 imaginary points – 3 points between the already used points – real and imaginary ones (Fig. 4). So, at the whole we get 4 real and 28 imaginary points. The results of calculation are shown on Fig. 5. The beam curvatures coincide with exact one (Fig. 5, *b*). The real curvatures after 3rd iteration are shown on Fig. 5, *a*. In this case they better correspond to the exact ones and the difference does not exceed 2%.

One may get the false impression that the accuracy of reproduction of calculated contour is related to the accuracy of preliminary placement of imaginary points. To show that crucial is only the fact of their insertion rather than place of placement, make the following test. Consider the contour obtained from

32 points (4+28), Fig. 5, *a*. For convenience consequently renumerate them starting from real point B_1. Artificially shift points N4, N14, and N26 outwardly, as shown on Fig. 6, *a*. From these points, we will form an initial polygon, which is the initial input geometry.

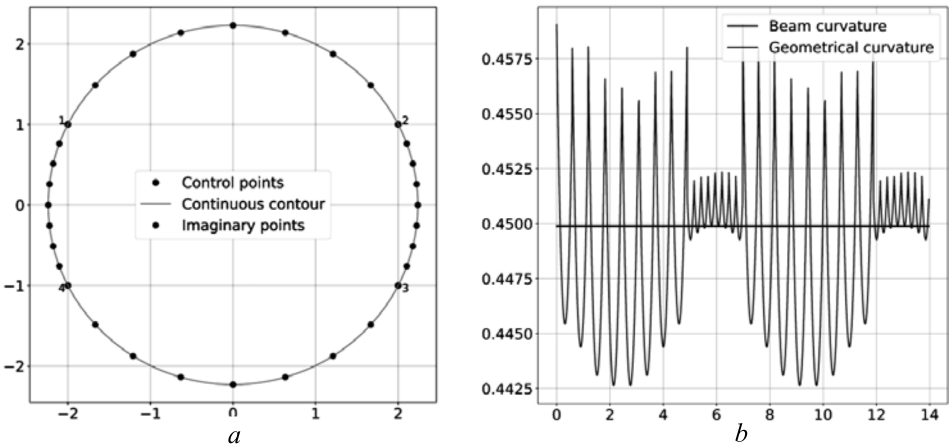


Fig. 5. Recovery of circle by 4 real and 28 imaginary points: *a* — calculated graph; *b* — recalculated curvatures by formula (10)

The calculated contours for some iterations are shown on Fig. 6. In particular, after first iteration the contour contains local loops, Fig. 6, *b*. After second iteration the contour resemble a slightly oval contour, Fig. 6, *c*. The contour become better with each subsequent iteration and after 6th iteration it is almost indiscernible from the ideal circle. After 9th iteration the correctly calculated curvatures are very close to the exact ones and differ with them only by 1–2%. Thus, availability of 32 points (real or imaginary) is almost enough to get the ideal circle. This means, that in overage $360/32 \approx 12^\circ$ (and lesser) angular misalignment can be effectively smoothed by our modified beam approach.

Example 2. Consider the circle formed by unevenly placed points. In practice the measurements are often performed with some restrictions, when the available to measurements is only some segment of the circle. Set three points (Fig. 7), namely $B_1(x_1 = -\sqrt{3}/2,\ y_1 = 1/2)$; $B_2(0,1)$; $B_3(\sqrt{3}/2, 1/2)$. Evidently these points belong to the circle of radius 1, and curvature is also 1. Note, that these points embrace only the 1/3 of the full circumference. Due to unevenness of their placement, the angular misalignments between the straight segments are: between the first and second ones – angle is 60°, between second and third -120°, and third and first ones – 120°. When the angle is larger than 90°, the sense of beam theory is lost – angle does not only essentially differ from its tangent, but they have the different sighs. Therefore, from the start insert only one imaginary point, for example, $B_{3-1}(0,-5)$. The results are shown on Fig. 7. Evidently, the contour contains the loop and do not resemble the circle at all.

Consider the calculated contour Fig. 7, *b* as initial one, and insert between the real points B_1 and B_2 and point B_2 and point B_3 15 additional imaginary points, also insert 26 points between the points B_3 and point B_{3-1}, the same is between B_{3-1} and B_1. The calculation is performed at several iterations, until the stabilization of the geometry (Fig. 8). The derived geometry, which contains 3 real and 83 imaginary points, is very close to the ideal circle, Fig. 8, *a*. As to beam curvatures they differ only by 0.1% from exact ones, while the geometrical curvatures are slightly worse – the difference reach up to 0.3%.

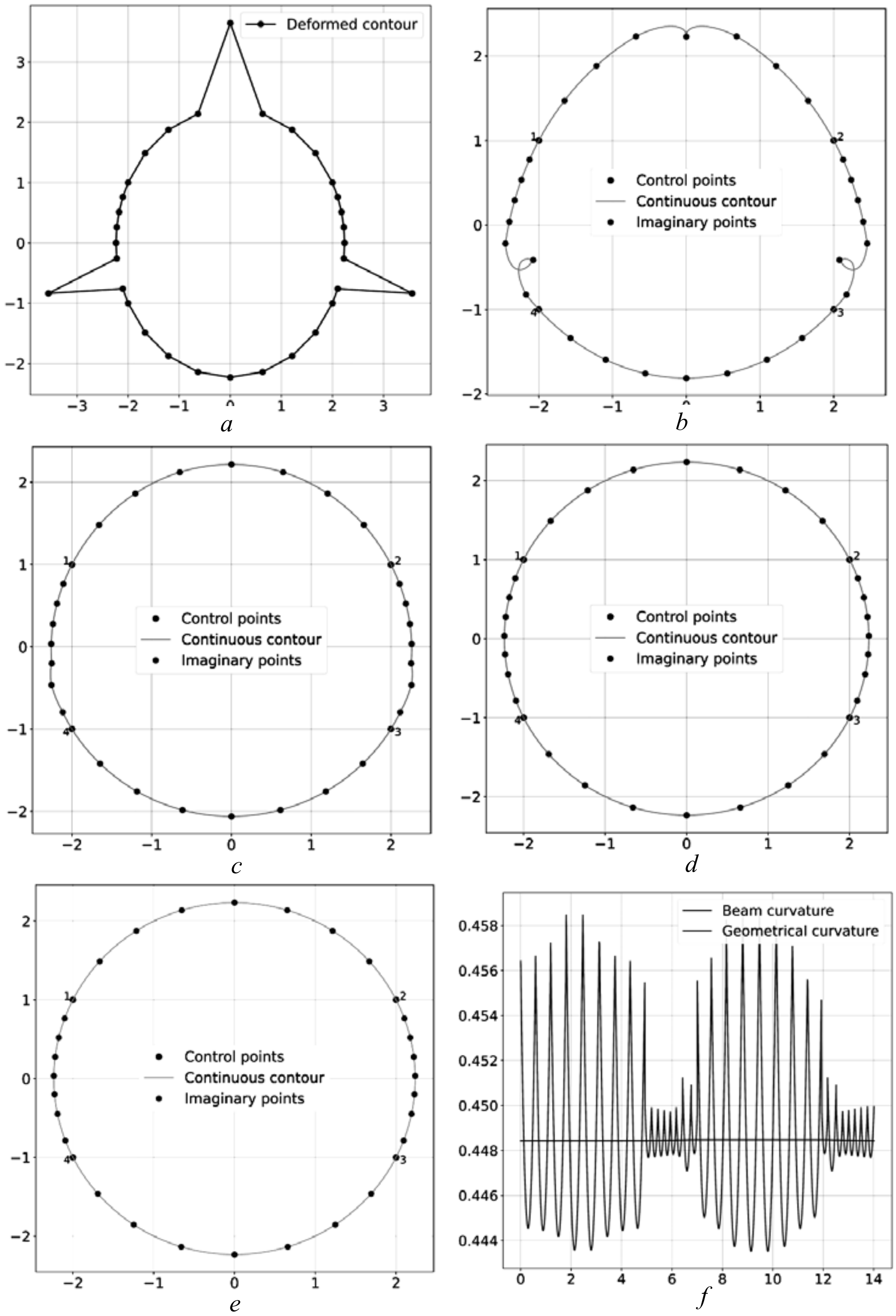
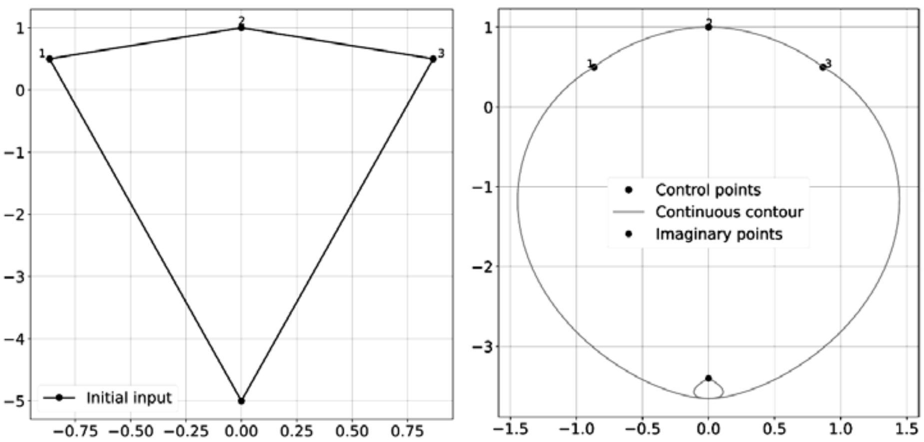
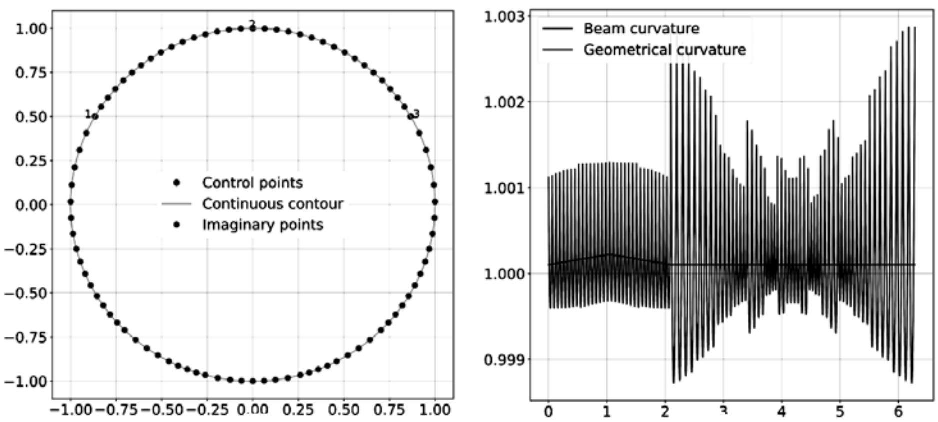


Fig. 6. Recovery of circle from 4 real and 28 imaginary points, three of which were essentially shifted: *a* — initial placement of points; *b* — contour after 1st iteration; *c* — after 2nd iteration; *d* — after 6th iteration; *e* — after 9th iteration; *f* — calculated by (10) curvatures after 9th iteration



a *b*

Fig. 7. Recovery of the circle from 3 real points placed at the upper part of circle and one imaginary point: *a* — input point; *b* — calculated contour with loop



a *b*

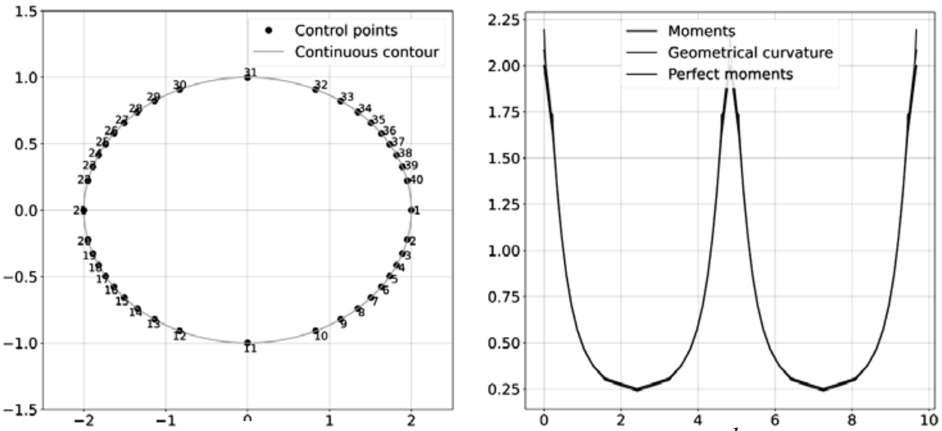
Fig. 8. Recovery of circle from 3 real points and 83 imaginary ones: *a* — calculated contour; *b* — beam and geometrical (10) curvatures

Example 3. Here we consider a more complicated figure – ellipse. In contrast to the circle, its curvature is not a constant. So, it is interesting to explore the ability of the method to restore the figures with variable curvature. Use the equation of ellipse with respect to parameter φ, where the coordinates of any point are given by equations: $x = 2 * \cos(\varphi)$ and $y = 1 * \sin(\varphi)$. The curvature of ellipse is given by the following expression:

$$\kappa(\varphi) = \frac{a * b}{\sqrt{a^2 - (a^2 - b^2)(\cos(\varphi))^2}^{\,3}}. \tag{13}$$

Note, that in point $\varphi = 0$ $\kappa(0) = \dfrac{a}{b^2} = 2$, and value of $\kappa(\pi/2) = \dfrac{b}{a^2} = \dfrac{1}{4}$. So, the curvature changes in 8 times within the range of 90 degrees.

Generate 40 points (10 points for each quadrant), between which the curvature changes proportionally, i.e., in $k = \sqrt[10]{8}$ times. These points (their parametric values of φ_i) are easily generated from equation (13). Fig. 9 shows the positions of 40 generated ellipse points and corresponding calculated contour and its beam and geometrical curvatures. In general, we see a very good correspondence, with the exception of points which lays on axis x, where the real curvature is the largest. Here the deviation of geometrical curvature from the exact one attains 10%. Of course, these results can be improved by taking more input points in the vicinity of point $\varphi = 0$. Yet, in general, the results are very optimistic, even for figure with quickly changing curvature.



a *b*

Fig. 9. Recovery of ellipse from exactly placed discrete points: *a* — input points and calculated contour; *b* — graph of exact curvature of ellipse and its calculated beam and geometrical curvatures

RECOVERY OF THE CIRCLE FROM NOISY POINTS

Example 4. Consider again the ideal circle of radius 1. Take 40 evenly distributed points on it. Then add the noise to each coordinate of point (X and Y) according to uniform continuous distribution with random parameters ranged from lower boundary $a = -0.3$ to upper boundary $b = 0.3$. The noise is characterized by theoretical statistical deviance:

$$\sigma_{theor} = \frac{(b-a)}{2\sqrt{3}}.$$

So, in our case $\sigma_{theor} = 0.173$.

The input random points are given in Table .Their sequence does not resemble the ordered one, where each next enumerated point is placed in clockwise direction from the previous one. So, their placement sometimes looks as chaotic one, which complicates the calculations. The results of calculations are shown on Fig. 10. A large number of closed loops are formed at first iteration. Then, with decreasing of the spring rigidity their number decreases, and the reason is the overlapping of the points which form the loops. The loops are completely unraveled at 15th iteration when $\overline{C} = 0.473$. With subsequent iterations the contour more and more resembles a perfect circle. The calculations are ceased at 40th iteration when $\overline{C} = 3.3310^{-5}$.

Introduce the notion of conditional empirical calculation deviance, σ_{emp}, i.e., root mean square deviation of input points B_i from the calculated contour, which we approximately define by formula

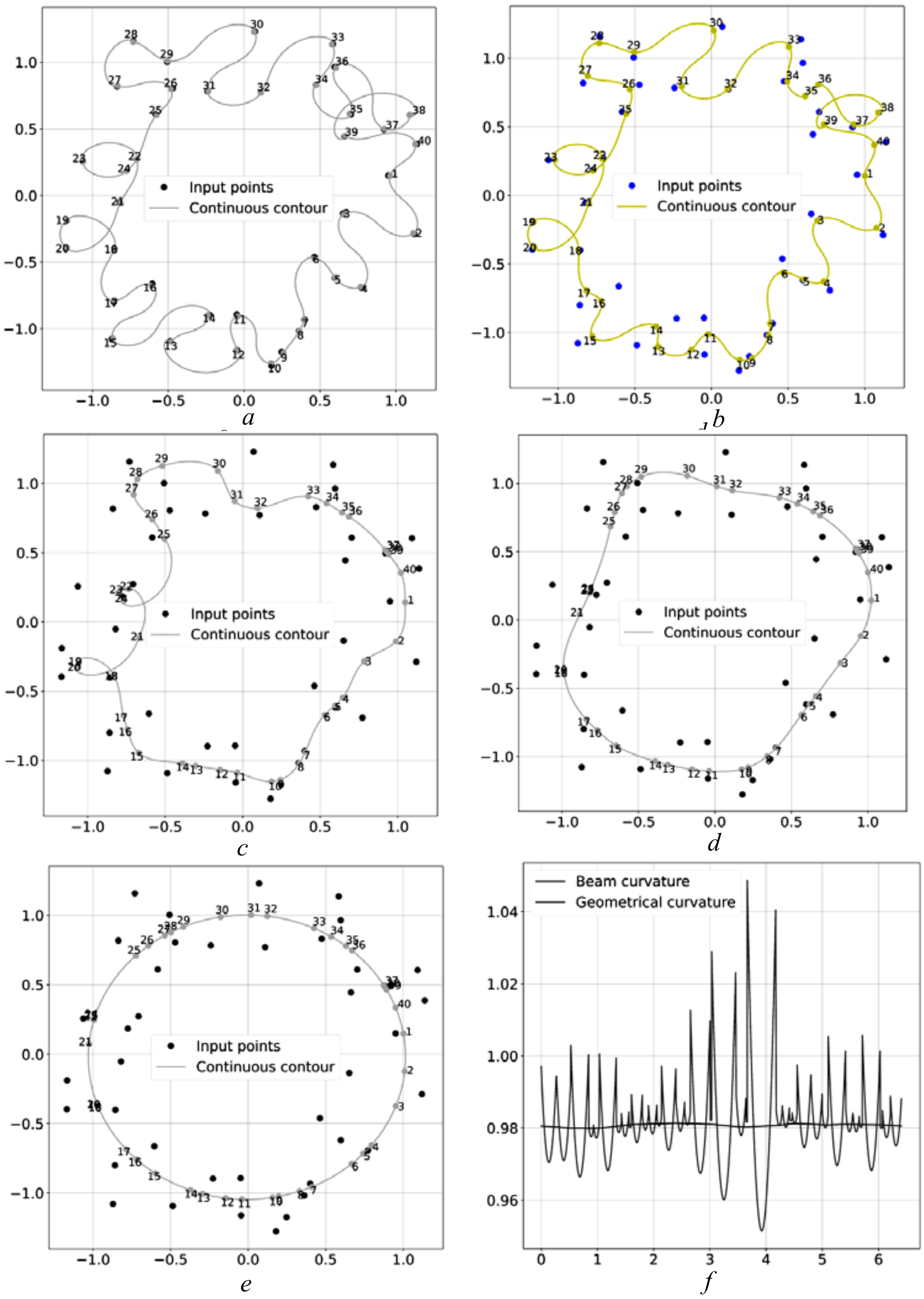


Рис. 10. Recovery of circle from 40 noised points generated by uniform distribution law with an extremal deviation equal to ± 0.30: *a* — input points and contour at 1st iteration; *b* — contour after 5th iteration, $\overline{C}=21.66$; *c* — after 12th iteration, $\overline{C}=1.49$; *d* — after 15th iteration, $\overline{C}=0.47$; *e* — after 40th iteration $\overline{C}=3.3310^{-5}$; *f* — the beam and geometrical (10) curvatures after 40th iteration

Generated 40 random points with respect initial 40 evenly chosen points on the circle

N	*X*	*Y*	N	*X*	*Y*	N	*X*	*Y*	N	*X*	*Y*
1	0.9616	0.0000	11	0.0727	-1.2507	21	-0.9969	0.0611	31	-0.0193	0.8941
2	0.7032	0.0773	12	-0.1389	-1.0679	22	-1.2485	0.3154	32	-0.0225	0.8435
3	0.9809	-0.4040	13	-0.5283	-0.7405	23	-0.9942	0.1106	33	0.3933	0.8832
4	0.8522	-0.4137	14	-0.4458	-0.9472	24	-1.1331	0.3298	34	0.4438	1.0902
5	0.7612	-0.6313	15	-0.7771	-1.0927	25	-1.0327	0.6022	35	0.5909	0.9511
6	0.6053	-0.7451	16	-0.5359	-0.8588	26	-0.6491	0.6211	36	0.6392	0.6346
7	0.4106	-0.6431	17	-0.5966	-0.8475	27	-0.7522	0.5364	37	0.9852	0.2956
8	0.5256	-0.8696	18	-0.8945	-0.1577	28	-0.6898	1.1809	38	0.9390	0.6324
9	0.1888	-0.6788	19	-0.7431	-0.0267	29	-0.4768	0.9159	39	0.7484	0.1707
10	0.0165	-0.9612	20	-1.2399	0.0237	30	-0.2465	0.9901	40	1.1081	0.2060

$$\sigma_{emp} = \sqrt{\frac{\sum_1^N \Pi_i^{\,2}}{N}} \,. \tag{14}$$

Calculate this value by formula (14) at the last (40th) iteration. Empirical deviation is equal to $\sigma_{emp} = 0.151$, which is very close to the theoretical deviance. Such good correspondence is due to the big number of input points, and that conditional distance Π_i reflects the essence of the distance to the contour. Underline, that above example illustrates that attained empirical deviance at the given iteration σ_{emp} can be used as a condition of termination of smoothing process, provided that theoretical deviance is known.

EXTRACTION OF SILHOUETTES FROM NOISY POINTS [4]

Example 5. The problem of recovery of silhouettes from noisy dense points was considered in work [4]. Among other, this work is remarkable that it gives a lot of sets of artificially generated random points from continuous silhouettes, yet the original silhouettes were not provided. Unknown is also the principle of the random points generation. In the title of respected files, which were supplemented to work [4], there are some numerical values of errors which can be treated as deviance of input points. Other problem with these data is the numeration of points. Our method works with numerated points and does not change their numeration. But as is shown in Example 4 when points are very dense noisy, the renumeration is desirable. Yet, in this work we will not apply the renumeration procedure, to show that our method is very effective even without it.

Consider first figure from work [4] — butterfly. Input noisy data are given in the file (butterfly_2percent_noise.txt"), where 164 points are presented. The mentioned in the file name noise is given in percent. Due to that all pictures are given in absolute coordinates which range approximately from -0.5 to $+0.5$ as for x as for y, thus we assume that deviance was calculated relatively to the value of 0.5. Thus, we take that theoretical deviance is equal to $\sigma_{\text{theor}} = 0.02 0.5 = 0.01$.

The results of calculation are given on Fig. 11. Initial input points are presented on Fig. 11, *a*. The results of calculation after 1st iteration are shown on Fig. 11, *b*, where three local loops are presented, and the empirical deviance is

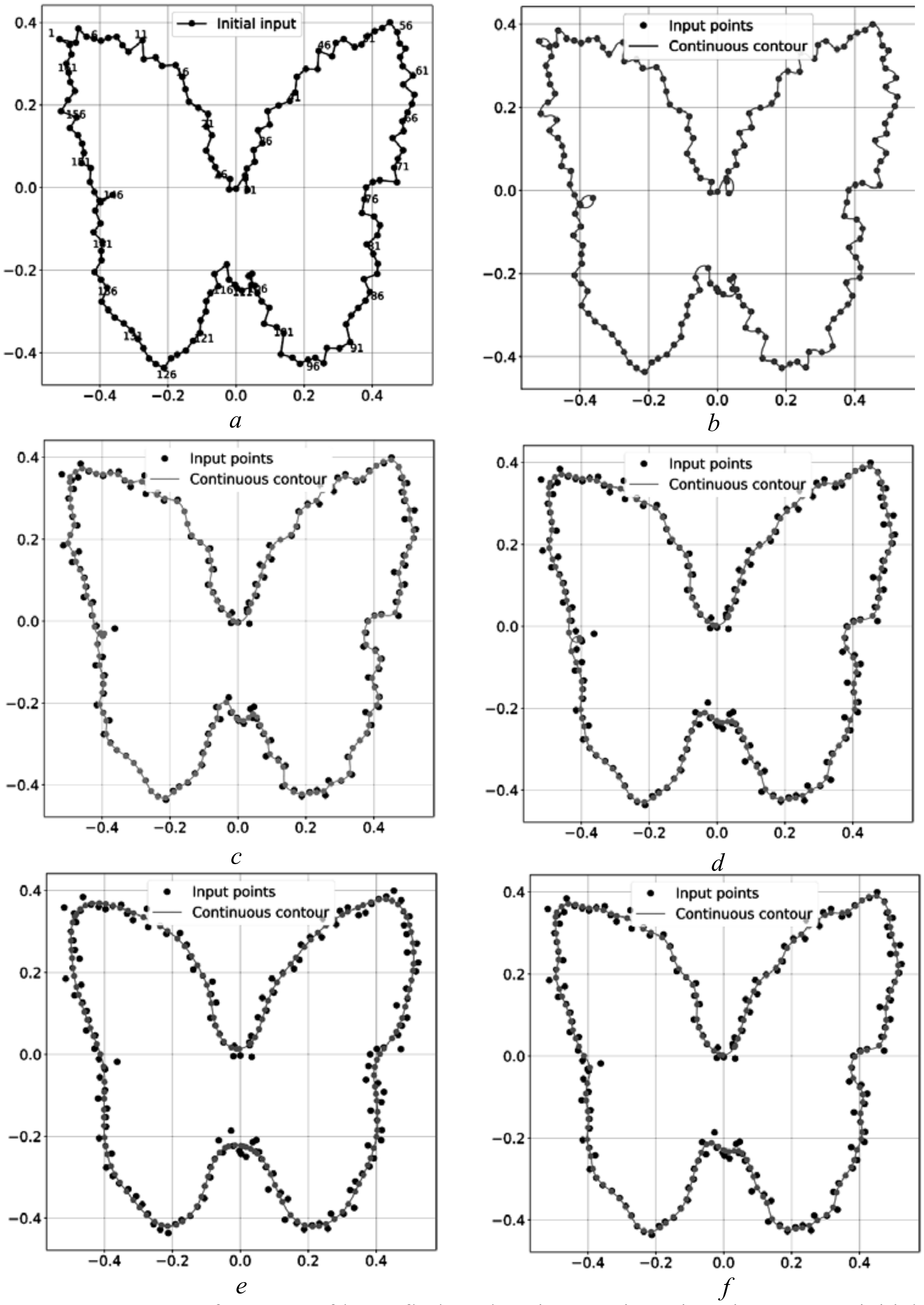


Fig. 11. Recovery of contour of butterfly based on input points given in [4]: *a* — initial points; *b* — contour after 1st iteration; *c* — after 15th iteration, $\overline{C} \approx 1.64$; *d* — after 20th iteration, $\sigma_{emp} \approx 0.010$, $\overline{C} \approx 0.377$; *e* — after 30th iteration, *f* — returning to the same rigidity as at 20th iteration, $\sigma_{emp} \approx 0.010$

very small and equal to $\sigma_{emp} = 0.00029$. After 15th iteration the contour still has one small loop (at the left side of picture) and calculated deviance becomes $\sigma_{emp} = 0.00672$, which is less than assumed theoretical value. Interesting to notice, that the contour doesn't look smooth enough. At 20th iteration ($\overline{C} = 0.377$), the calculated value $\sigma_{emp} = 0.01060$, which is very close to the criterial value. Yet the contour still contains one local loop; to unravel it perform the several additional calculations. At 30th iteration $\sigma_{emp} = 0.01378$, which is larger than theoretical one, which testifies about the oversmoothing of the contour, what is subjectively visually confirmed. It is interesting to note, that we can return from 30th iteration's rigidity to the 20th iteration's rigidity (which is equal to $\overline{C} = 0.377$). In this last case the calculated value of $\sigma_{emp} = 0.01059$, yet the local loop has disappeared.

Consider the data for crab [4], ("crab_2percent_noise.txt"), Fig. 12, where 284 points are given. The theoretical deviance is taken to be $\sigma_{theor} = 0.01$. The larger number of points might lead to the increase of the number of local loops at first iterations. The results at first iterations are presented on Fig. 12, *b*, where input points and calculated contour points almost coincide. The value of calculated deviance $\sigma_{emp} = 0.00032$ is very small due to large rigidity of supports, yet contour has four local loops. At 15th iteration the contour still has some visual misalignments, and the calculated deviance is equal to $\sigma_{emp} = 0.01051$, which is close to the theoretical one. Besides, contour still has one loop. At 18th iteration the calculated and theoretical deviances are almost the same, but contour still contain the loop. At 30th iteration the calculated $\sigma_{emp} = 0.01716$, which is larger than the theoretical one, and visually the contour looks like the oversmoothed one. Therefore, we start to change the supports rigidity in inverse order. Fig. 12, *f* shows the calculated contour when the rigidity $\overline{C} \approx 1.22$, which is the same as in direct order 16th iteration. Here the value of $\sigma_{emp} = 0.01049$, which is very close to the theoretical one. So, contour of Fig. 12, *f* can be considered as the good result of smoothing.

Consider the figure of dolphin [4] ("dolphin_2percent_noise.txt"), Fig. 13, 179 points, $\sigma_{theor} = 0.01$. There are 3 local loops at 1st iteration, Fig. 13, *b*. Calculated deviance after 15th iteration at $\overline{C} = 1.64$ is equal to $\sigma_{emp} = 0.0086$, which is slightly less than the theoretical one. At 18th iteration $\sigma_{emp} = 0.0102$ and is close to theoretical one. At 30th iteration the deviance $\sigma_{emp} = 0.0165$, which testifies that contour is oversmoothed, and this is visually confirmed. Increasing the iteration number (decreasing the rigidity) lead to additional oversmoothing. As example, consider the contour obtained after 60th iteration at $\overline{C} = 210^{-5}$, where $\sigma_{emp} = 0.08902$, which is one order higher than the theoretical one. Evidently, the peculiarity of the method is that at very small number of rigidity it eventually gives the perfect circle. So, the criteria of termination should be specified.

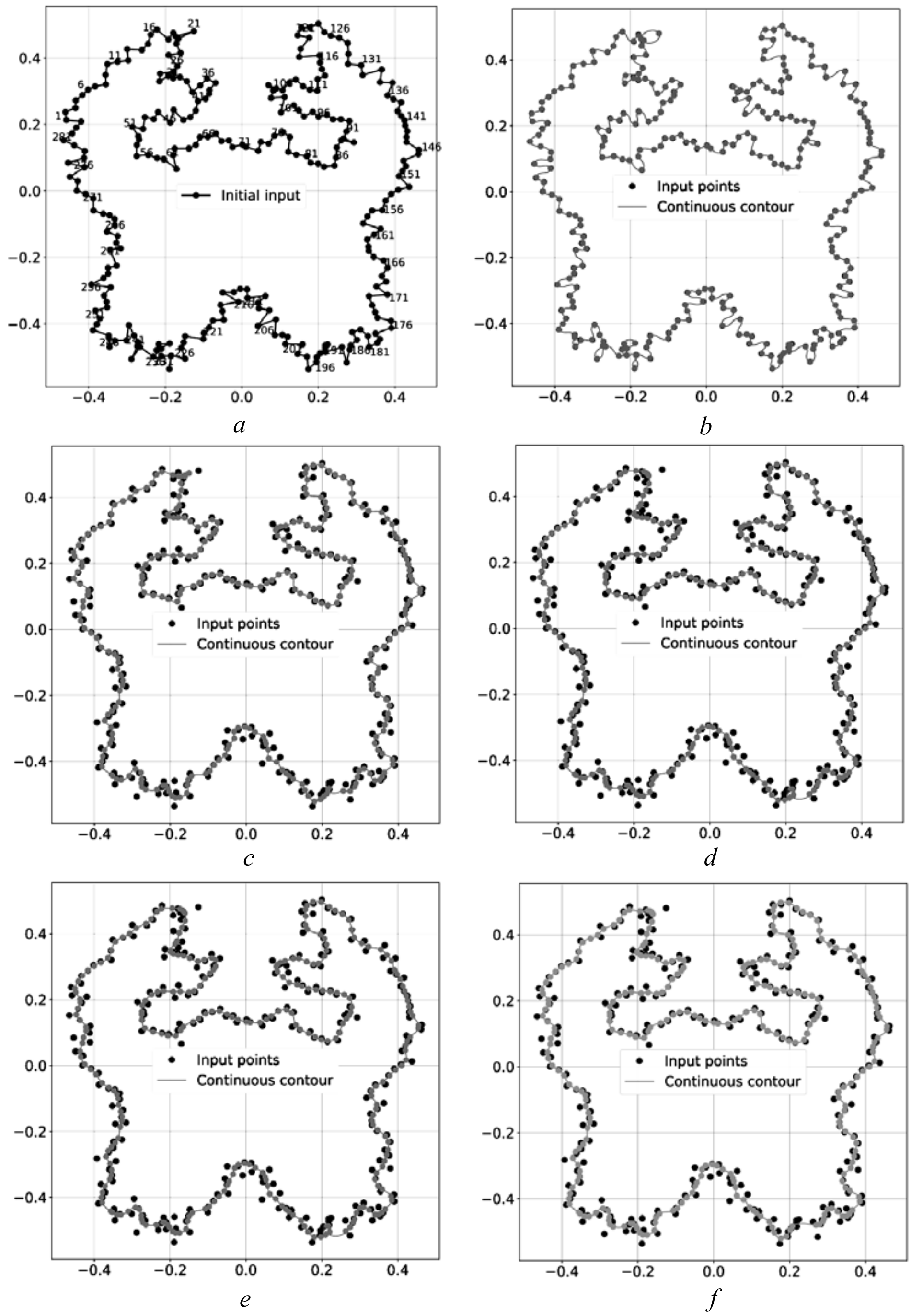


Fig. 12. Recovery of the contour of crab based on data of [4]: *a* — nitial points; *b* — contour after 1st iteration; *c* — after 15th iteration; *d* — after 18th iteration, $\sigma_{emp} \approx 0.010$; *e* — after 30th iteration; *f* — reverse increasing of rigidity till the value as in direct smoothing after 16th iteration, $\sigma_{emp} \approx 0.010$

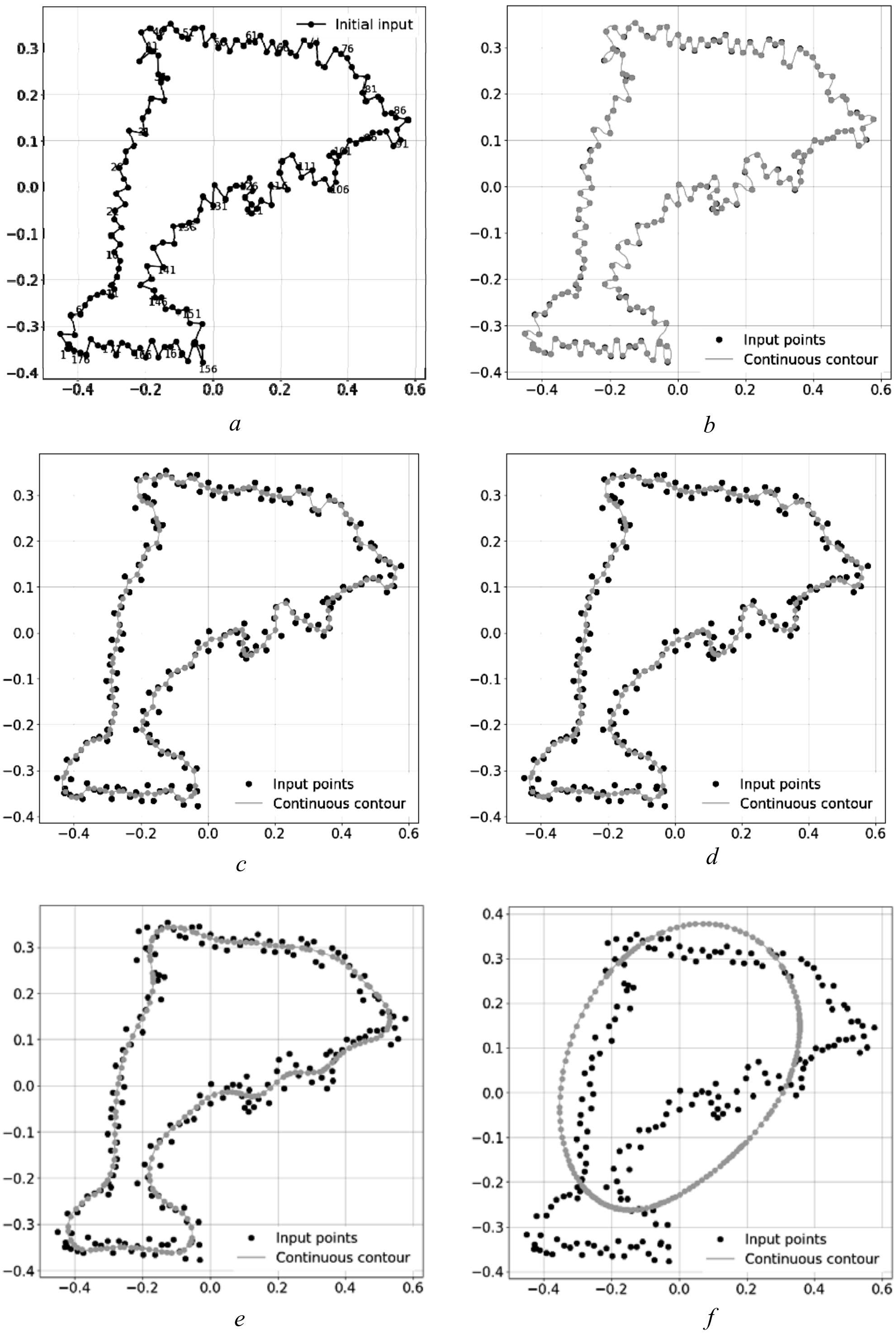


Fig. 13. Recovery of dolphin contour based on data of work [4]: *a* — input data; *b* — contour after first iteration; *c* — 15th iteration, $\sigma_{emp} = 0.0086$; *d* — 18th iteration, $\sigma_{emp} = 0.0102$; *e* — 30th iteration $\sigma_{emp} = 0.0165$; *f* — 60th iteration, $\sigma_{emp} = 0.08902$

CONCLUSIONS

Based on the authors' experience in the field of structural mechanics the principally new methods of smoothing the noisy data to get the continuous closed contours is proposed. The following results are obtained:

Based on the theory of straight beams, which lays on discrete elastic supports with finite rigidity, the general governing equations are formulated. They include 4 Connection equations for each straight beam segment and 4 Conjugation equations, which should "match" the end of previous segment with the beginning of subsequent one. The last one among other serves to smooth the angle of misalignment between two neighboring segments, which is attained by corresponding requirements to the angular deformations of segments at their border.

The notion of imaginary points is suggested, which broke the segment on two smaller one but do not envisage the insertion of the additional support between them. They are introduced when the angular misalignment between two adjacent segments is large. Their availability is very instrumental in providing the C^2 continuity of the smoothed contour.

The algorithm of refinement of calculated positions of the points of segment is proposed. It accounts for by rotating the initial normal vector to the segment on the calculated deformation angle in each point. This serves to provide the continuity of the contour points. The iteration process of refinement of the contour which consists in consequent process of the decreasing the support rigidities is proposed. Its efficiency is confirmed to analysis of very dense and noisy input points, where the local loops may occur at initial stages of algorithm.

The notion of experimental (calculated) statistical deviance from the initial input points is suggested. It is shown that for attainment the visually best smoothing this value should be close with the theoretical (generated) deviance of input points. So, this value can be used for formulation of the criteria of formal termination of smoothing procedure.

REFERENCES

1. D.F. Rogers, *An Introduction to NURBS With Historical Perspective*. San Diego, CA: Academic Press, 2001.
2. Mark S. Nixon and Alberto S. Aguado, *Feature Extraction & Image Processing for Computer Vision*. Academic Press, 2013, 609 p. doi. 10.1016/C2011-0-06935-1.
3. T. Strutz, *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond*. Wiesbaden, Germany: Vieweg, 2010.
4. Stefan Ohrhallinger and Michael Wimmer, "StretchDenoise: Parametric Curve Reconstruction with Guarantees by Separating Connectivity from Residual Uncertainty of Samples," *PG '18: Proceedings of the 26th Pacific Conference on Computer Graphics and Applications: Short Papers, October 2018*, pp. 1–4. Available: https://doi.org/10.2312/pg.20181266
5. Ahlberg J. Harold, Edwin Norman Nilson, and Joseph Leonard Walsh, "The theory of splines and their applications," *Mathematics in Science and Engineering*, 1967.
6. Stephen P. Timoshenko, Strength of materials. Part 1: Elementary theory and problems, 1955.
7. D.G. Schweikert, "An interpolating curve using a spline in tension," *J. Math. Physics*, 45, pp. 312–317, 1966.

8. P.H. Wagner, X. Luo, and K.A. Stelson, "Smoothing curvature and torsion with spring splines," *Computer-Aided Design*, vol.27, no. 8, pp. 615–626, 1995. doi: 10.1016/0010-4485(95)99798-D.
9. I. Orynyak, I. Lokhman, and A. Bohdan, "The Spring Splines Procedure with Prescribed Accuracy for Determination of the Global (Pipe Centerline) as well as the Local (Dent) Curvatures," *Proceedings of IPC 2012 9th International Pipeline Conference, September 24–28, Calgary, Alberta, Canada, IPC2012-90127*. doi: 10.1115/IPC2012-90127.
10. I.V. Orynyak, I.V. Lokhman, and A.V. Bogdan, "Determination of curve characteristics by its discrete points measured with an error and its application to stress analysis for buried pipeline," *Strength of Materials*, 44(3), pp. 268–284, 2012. doi: 10.1007/s11223-012-9380-7.
11. M.A. Crisfield, "A consistent co-rotational formulation for non-linear, three-dimensional, beam-elements," *Computer Methods in Applied Mechanics and Engineering*, vol. 81, issue 2, pp. 131–150, 1990. doi: 10.1016/0045-7825(90)90106-V.
12. I. Orynyak, R. Mazuryk, and A. Orynyak, "Basic (discontinuous) and smoothing up (conjugated) solutions in transfer matrix method for static geometrically nonlinear beam and cable in plane," *Journal of Engineering Mechanics*, 146(5), 2020. doi: 10.1061/(ASCE)EM.1943-7889.0001753.

Received 21.05.2022

INFORMATIONON THE ARTICLE

Igor V. Orynyak, ORCID: 0000-0003-4529-0235, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: igor_orinyak@yahoo.com

Dmytro R. Koltsov, ORCID: 0000-0002-0396-7255, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: koltsovdd@gmail.com

Oleg R. Chertov, ORCID: 0000-0003-0087-1028, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: chertov@i.ua

Roman V. Mazuryk, ORCID: 0000-0003-4309-824X, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: r.mazuryk.ua@gmail.com

ЗАСТОСУВАННЯ ТЕОРІЇ БАЛОК ДЛЯ ПОБУДОВИ ПЛОСКИХ ДВІЧІ ДИФЕРЕНЦІЙОВАНИХ ЗАМКНУТИХ КРИВИХ ЗА НЕТОЧНИМИ ДИСКРЕТНИМИ ДАНИМИ / І.В. Ориняк, Д.Р. Кольцов, О.Р. Чертов, Р.В. Мазурик

Анотація. Згладжування дискретних точок, заміряних з певною похибкою, має велике значення в різних технічних застосуваннях та комп'ютерній графіці. Ідея роботи полягає в застосуванні методів теорії балок на пружних опорах. Установлюються локальні системи координат для прямолінійних відрізків балок, де кути неспіввісності згладжуються відповідними умовами в рівняннях спряження. Для наближення довжини отриманого контуру до довжини прямолінійних балок уводиться поняття умовних точок, що розміщуються між точками замірів. Наводиться ряд прикладів реконструкції реальних контурів по заміряних дискретних точках із заданою і невідомою похибками.

Ключові слова: сплайн, пружна балка, опора, замкнутий контур, уявна точка, неточні дані.

УДК 517.9:534.1
DOI: 10.20535/SRIT.2308-8893.2022.4.11

ТИПОВІ ТА УЗАГАЛЬНЕНІ ПЕРЕХОДИ ДО ДЕТЕРМІНОВАНОГО ХАОСУ НЕТИПОВИХ АТРАКТОРІВ НЕІДЕАЛЬНИХ ДИНАМІЧНИХ СИСТЕМ

О.Ю. ШВЕЦЬ

Анотація. Розглянуто деякі прикладні нелінійні неідеальні динамічні системи п'ятого порядку, які застосовуються для опису коливань сферичних маятників та у гідродинаміці. Побудовано максимальні атрактори, як регулярні, так і хаотичні, таких систем. Обговорено різноманітні біфуркації максимальних атракторів. Установлено перехід до детермінованого хаосу для максимальних атракторів за типовими сценаріями Фейгенбаума та Манневілля–Помо. Досліджено імплементацію сценарію узагальненої переміжності для хаотичних максимальних атракторів таких систем. Виявлено ознаку реалізації сценарію узагальненої переміжності.

Ключові слова: неідеальна динамічна система, регулярний максимальний атрактор, хаотичний максимальний атрактор, переміжність типова та узагальнена.

ВСТУП

Під час дослідження динамічних систем великий інтерес викликають питання виникнення хаотичних граничних множин у таких системах. Тобто виникнення усталених режимів динамічної поведінки, за яких поведінка детермінованої динамічної системи стає у деякому сенсі непередбачуваною. Причому ця непередбачуваність залежить виключно від внутрішніх властивостей динамічної системи, а не від тих чи інших зовнішніх хаотичних збуджень цієї системи. Якщо динамічна система є дисипативною, то граничними множинами будуть атрактори різних типів. Причому такі множини можуть бути як регулярними (положення рівноваги, граничні цикли, інваріантні тори), так і хаотичними. Перший хаотичний атрактор був побудований у знаменитій праці Лоренца 1963 р. [1] і надалі отримав назву «метелик Лоренца».

Із плином часу було розроблено понятійну базу для класифікації різних типів хаотичних атракторів. Натепер розглядаються гіперболічні, квазіхаотичні, приховані, самозбуджені, рідкісні хаотичні атрактори [2–4].

Усі перелічені атрактори є ізольованими граничними множинами. Тобто, якщо $\mathcal{A}$, $\mathcal{B}$ — два різні атрактори, то

$$\inf_{\substack{x\in\mathcal{A}\\ y\in\mathcal{B}}} \rho(x,y) > 0, \qquad (1)$$

де ρ — деяка метрика, як правило, евклідова, у фазовому просторі динамічної системи.

Однак, крім існування таких «типових» атракторів, у деяких динамічних системах існують граничні множини, які не задовольняють умову ізольованості (1). Граничні множини такої динамічної системи можна об'єднати у сім'ї за спільними властивостями, серед яких неізольованість і спільний спектр ляпуновських характеристичних показників (ЛХП). Серед інших такі сім'ї можуть ще й мати властивість «притягання» до себе інших траєкторій (яка виокремлює атрактор серед інших граничних множин у випадку її ізольованості). У випадку сім'ї властивість «притягання» буде виявлятись у вигляді, що кожна траєкторія з деякої відкритої множини прямує не до всієї сім'ї граничних множин як такої, а до деякого її представника. Такі сім'ї утворюють так звані максимальні атрактори, означення яких наведено у працях [5, 6]. Зауважимо, що максимальні атрактори не є атракторами у класичному сенсі.

ПОСТАНОВКА ЗАВДАННЯ ТА МЕТА РОБОТИ

Чільне місце серед динамічних систем займають неідеальні системи або системи з обмеженим збудженням. Уперше такі системи були розглянуті Зоммерфельдом на початку XX ст. [7]. Але як усталений науковий напрям теорія систем з обмеженим збудженням утворилася після публікації монографії Кононенка [8], у якій він увів чітку аксіоматику та побудував математичні моделі для широкого кола завдань. Теорія систем з обмеженим збудженням досліджує взаємодію коливальних систем з джерелами збудження їх коливань. У межах цієї теорії передбачається, що джерела збудження коливань мають потужність, порівнянну до потужності, споживаної коливальним навантаженням. У цьому випадку робота джерела енергії залежить від режиму коливального навантаження і вплив джерела не може бути виражений у вигляді заданої явної функції часу. Тоді як у традиційному математичному моделюванні коливальної системи розглядаються ідеалізовані джерела збудження необмеженої потужності.

У багатьох випадках «ідеальний» підхід в корені хибний, що на практиці призводить до грубих помилок в описі динаміки як коливальної системи, так і джерела збудження [9–15].

Відкриття детермінованого хаосу стимулювало появу нового напряму в теорії систем з обмеженим збудженням, пов'язаного з пошуком хаотичних режимів взаємодії коливальних систем з джерелами збудження. Особливий інтерес становлять ті хаотичні режими, поява яких зумовлена нелінійною взаємодією коливальної системи з джерелом збудження, а не з їх автономними властивостями. У публікаціях [16, 17] описано виникнення хаотичних усталених режимів у низці детермінованих неідеальних динамічних систем. У цих системах хаос принципово неможливий без урахування взаємодії коливної системи з джерелом збудження її коливань.

Розглянемо систему звичайних диференціальних рівнянь:

$$\frac{dy_1}{d\tau} = Cy_1 - \left[y_3 + \frac{A}{2}(y_1^2 + y_2^2 + y_4^2 + y_5^2) \right] y_2 + B(y_1 y_5 - y_2 y_4) y_4 + 2y_2 ;$$

$$\frac{dy_2}{d\tau} = Cy_2 + \left[y_3 + \frac{A}{2}(y_1^2 + y_2^2 + y_4^2 + y_5^2) \right] y_1 + B(y_1 y_5 - y_2 y_4) y_5 + 2y_1 ;$$

$$\frac{dy_3}{d\tau} = D\left(y_1 y_2 + y_4 y_5 \right) + Ey_3 + F ; \qquad (2)$$

$$\frac{dy_4}{d\tau} = Cy_4 - \left[y_3 + \frac{A}{2}(y_1^2 + y_2^2 + y_4^2 + y_5^2) \right] y_5 - B(y_1 y_5 - y_2 y_4) y_1 + 2y_5 ;$$

$$\frac{dy_5}{d\tau} = Cy_5 + \left[y_3 + \frac{A}{2}(y_1^2 + y_2^2 + y_4^2 + y_5^2) \right] y_4 - B(y_1 y_5 - y_2 y_4) y_2 + 2y_4 .$$

Це нормальна нелінійна система диференціальних рівнянь п'ятого порядку. Тут y_1, y_2, y_3, y_4, y_5 — фазові змінні; τ — безрозмірний час; A, B, C, D, E, F — деякі параметри. Система рівнянь (2) має достатньо широке застосування у математичному моделюванні систем з обмеженим, за Зоммерфельдом–Кононенком, збудженням.

Так, якщо $A = \frac{1}{4}$, $B = -\frac{3}{4}$, ця система описує коливання сферичного маятника, точка опори якого збуджується по вертикалі електродвигуном обмеженої потужності. У цьому випадку фазові змінні y_1, y_2, y_4, y_5 визначають положення маятника у просторі, а фазова змінна y_3 визначає швидкість обертання вала електродвигуна. Відповідно параметри C, D, E, F залежать від довжини і власної частоти маятника, рушійного моменту електродвигуна, внутрішнього моменту сил опору обертанню ротора електродвигуна, коефіцієнта демпфування, кута нахилу статичної характеристики електродвигуна тощо. Детальне виведення системи рівнянь (2) для випадку сферичного маятника з обмеженим збудженням виконано у праці [18].

Систему рівнянь (2) можна використовувати також для опису коливань вільної поверхні рідини циліндричного бака частково заповненого рідиною, платформа якого збуджується по вертикалі електродвигуном обмеженої потужності. У загальному випадку система «бак з рідиною – електродвигун» достатньо складна нескінченно вимірна динамічна система. Але у частинному випадку параметричного резонансу система п'ятого порядку (2) дозволяє з високою точністю описати коливання вільної поверхні рідини. Звичайно фізичний зміст фазових координат і параметрів значно відрізняється від випадку сферичного маятника. Так у випадку системи «бак з рідиною – електродвигун» фазові змінні y_1, y_2 і y_4, y_5 — коефіцієнти розкладу амплітуд коливань вільної поверхні рідини, відповідно за першою і другою основними домінантними модами; фазова змінна y_3 пропорційна швидкості обертання вала електродвигуна; C — коефіцієнт сил в'язкого демпфування; D — коефіцієнт пропорційності вібраційного моменту; E — кут нахилу статичної характеристики електродвигуна. Параметри A і B є константами, які залежать від радіуса бака і висоти налитої в нього рідини. Детальні пояснення з приводу обґрунтування математичної моделі системи «бак з рідиною – електродвигун» та обчислення параметрів цієї системи наведено у працях [19, 20].

Метою роботи є побудова максимальних атракторів системи (2) та дослідження різноманітних біфуркацій такого типу атракторів.

МЕТОДИКА ВИКОНАННЯ ЧИСЛОВИХ РОЗРАХУНКІВ І ОСНОВНІ ЧИСЛОВІ РЕЗУЛЬТАТИ

Система диференціальних рівнянь (2) суттєво нелінійна, тому знаходження її розв'язків та дослідження динамічної поведінки цієї системи, у загальному випадку, можна проводити тільки числовими методами. Опис деяких з таких методів наведено у працях [21, 22], а методику їх застосування стосовно систем з обмеженим збудженням роз'яснено у працях [23, 24].

Передусім відзначимо, що положення рівноваги системи рівнянь (2) проаналізовано у праці [25]. Коротко нагадаємо результати цієї роботи. Система (2) має єдине ізольоване положення рівноваги:

$$y_1 = 0,\ \ y_2 = 0,\ \ y_3 = -\frac{F}{E} = 0,\ \ y_4 = 0,\ \ y_5 = 0. \qquad (3)$$

Достатні умови асимптотичної стійкості за Ляпуновим такого положення рівноваги можуть бути отримані за допомогою теореми Л'єнара–Шіпара [26]. Крім одного ізольованого положення рівноваги система (2) має нескінченну кількість неізольованих положень рівноваги. Ці положення рівноваги утворюють замкнені лінії у фазовому просторі. Кожна точка цих особливих ліній буде положенням рівноваги. Якщо положення рівноваги ізольоване, положення рівноваги (3) є асимптотично стійким, то воно буде єдиним традиційним атрактором системи. Усі інші, можливі, граничні множини не будуть атракторами у традиційному сенсі цього терміна.

Для з'ясування можливого вигляду граничних множин системи (2) проведемо числові дослідження її динамічної поведінки. Спочатку розглянемо випадок, коли ця система описує нелінійну взаємодію бака з рідиною й електродвигуна. Для виконання числових розрахунків покладемо:

$$A = 0.56,\ \ B = -1.531,\ \ C = -0.8,\ \ D = -4.6,\ \ F = 0.6. \qquad (4)$$

Параметр E (кут нахилу статичної характеристики електродвигуна) розглядатимемо як біфуркаційний. Будемо досліджувати поведінку системи, якщо $-2.2 < E < -1.75$. За таких вибраних значень параметрів у системі виникають надзвичайно цікаві граничні множини, які мають такі загальні властивості. Кожна з таких множин являє собою деяку сім'ю нескінченної кількості траєкторій у фазовому просторі системи (2). Ці траєкторії неізольовані одна від одної, при цьому вони не мають спільних точок і не є положеннями рівноваги. Траєкторії сім'ї мають властивості «притягання», що споріднює таку сім'ю з атрактором. У той же час неізольованість траєкторій сім'ї вказує на те, що такі сім'ї траєкторій не є атракторами у традиційному сенсі поняття «атрактор». Причому зауважимо, що такі граничні множини можуть бути як регулярними, так і хаотичними [25].

Розглянемо деякі конкретні приклади таких сімей. Припустімо, що $E = -1.792$. У цьому випадку виникає гранична множина, яка складається з нескінченної кількості замкнених траєкторій (циклів), усі з яких існують одночасно. Усі цикли розташовані як завгодно близько один до одного, тобто не ізольовані. Однак такі цикли не мають точок дотику або перетину. Кожна така замкнена траєкторія сама по собі є граничною множиною. Це зумовлено тим, що майже будь-яка траєкторія, яка починається в деякій достатньо великій області фазового простору, прямує до одного з циклів сім'ї. Але жоден із циклів не є атрактором у традиційному розумінні цього терміна. Отже, кожен із цих циклів не є граничним. Крім того, кожен окремий цикл має один і той же період, однаковий спектр ЛХП. Старший ляпуновський показник для всіх циклів сім'ї дорівнює нулю. Переріз Пуанкаре кожного з циклів складається з однакової скінченної кількості точок.

На рис. 1 побудовано проекції фазового портрета трьох представників такої сім'ї траєкторій, які нанесено відтінками чорного кольору. Такі граничні множини [5, 6] називаються максимальними атракторами. Отже, на рис. 1 зображено представники періодичного максимального атрактора.

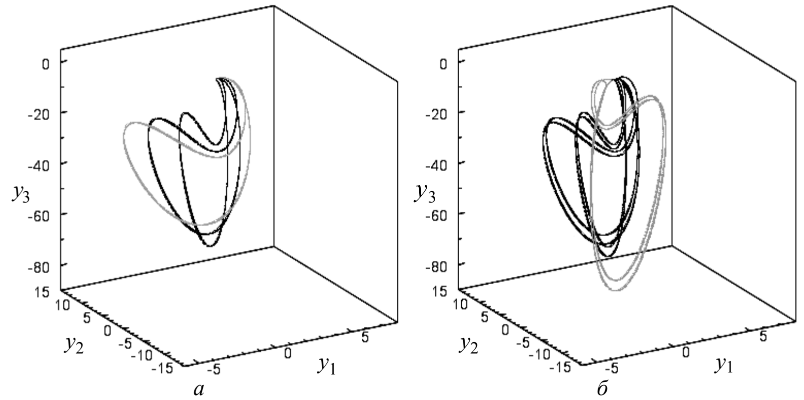


Рис. 1. Проекції фазових портретів, якщо $E = -1.792\,(a)$ і $E = -1.781\,(б)$.

Великий інтерес становлять біфуркації максимальних атракторів, зокрема переходи від регулярних максимальних атракторів до хаотичних максимальних атракторів. Для традиційних атракторів такі послідовності біфуркацій переходу від регулярного атрактора до хаотичного атрактора називаються сценаріями переходу до хаосу. Двома основними сценаріями таких переходів є сценарій Фейгенбаума [27, 28] та сценарій Манневілля–Помо [29, 30]. Останнім часом виявлено нові сценарії переходів до хаосу, які узагальнюють уже відомі сценарії [31, 32]. Виявляється подібні послідовності біфуркацій притаманні і максимальним атракторам.

На рис. 2, *а* побудовано фазопараметричну характеристику (біфуркаційне дерево) для одного з представників максимальних атракторів системи (2) за значень параметрів за формулами (4). Зауважимо, що такі фазопараметричні характеристики для будь-якого представника максимального атрактора якісно подібні і відрізняються тільки несуттєвими кількісними відмінностями. Маємо типове біфуркаційне дерево. Окремим гілкам цього біфуркаційного дерева відповідають періодичні граничні множини, а густо чорним — хаотичні граничні множини. Можна побачити точки розгалуження гілок біфуркаційного дерева, які вказують на можливі типи різних біфуркацій.

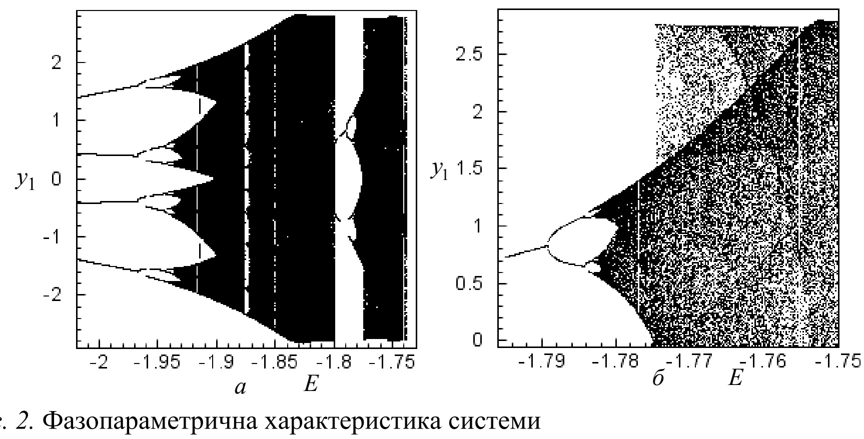


Рис. 2. Фазопараметрична характеристика системи

Для більш детального вивчення можливих біфуркацій на рис. 2, *б* побудовано збільшений фрагмент біфуркаційного дерева, який дозволяє ідентифікувати тип біфуркацій. Чітко можна бачити точки розгалуження окремих гілок біфуркаційного дерева, із проходженням яких спостерігаються біфуркації подвоєння періодів циклів, що утворюють максимальний атрактор. Так, якщо $E=-1.792$, максимальний атрактор складається з нескінченної кількості циклів, кожний з яких має порівняно просту однотактну структуру. Зі збільшенням значень біфуркаційного параметра E починається каскад біфуркацій подвоєння періодів циклів сімей.

На рис. 1, *б* та на рис. 3, *а* побудовано періодичні максимальні атрактори, які виникають після першої та другої біфуркацій подвоєння періоду. Як і раніше відтінками чорного кольору зображено по три представники відповідних періодичних максимальних атракторів. Із кожною біфуркацією вдвічі збільшується тактність усіх циклів максимального атрактора. Спочатку вони стають двотактними, а потім — чотиритактними. Усі цикли мають нульовий старший ляпуновський показник. Також після кожної біфуркації подвоюється кількість точок у перетинах Пуанкаре. Такий нескінченний каскад біфуркацій подвоєння періоду завершується виникненням хаотичного максимального атрактора, якщо $E=-1.776$.

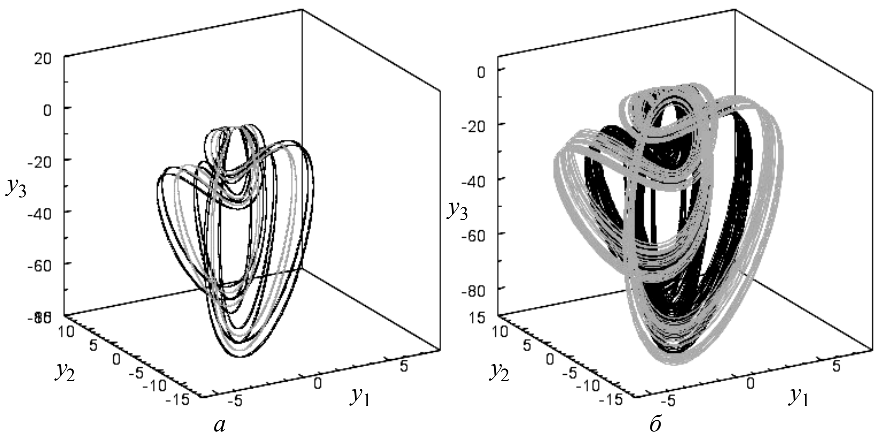


Рис. 3. Проекції фазових портретів, якщо $E=-1{,}7845$ (*а*) і $E=-1.776$ (*б*)

Побудовані проекції трьох представників нескінченної сім'ї неперіодичних траєкторій виниклого хаотичного максимального атрактора показано на рис. 3. Про виникнення хаотичної граничної множини свідчить поява додатного старшого ляпуновського показника у всіх траєкторій, які утворюють максимальний атрактор. Виникла сім'я містить нескінченну кількість хаотичних траєкторій. Відомо, що традиційний хаотичний атрактор складається із нескінченної кількості нестійких траєкторій. На перший погляд ця сім'я являє собою об'єднання нескінченної кількості хаотичних атракторів. Проте кожен член сім'ї не є атрактором у традиційному сенсі. Тут, як і раніше, для визначення такого об'єднання можна запропонувати поняття максимального атрактора [5, 6]. Усі траєкторії максимального хаотичного атрактора мають однаковий спектр ляпуновських характеристичних показників, старший з яких додатний. Отже, як і для традиційних атракторів, можна вводити поняття спектра ЛХП максимального атрактора в цілому, а не якоїсь його окремої траєкторії. Перетини Пуанкаре кожної з траєкторій сім'ї є структурно подібні хаотичні множини, що складаються із нескінченної кількості точок.

Таким чином, перехід до хаосу для максимальних атракторів відбувається через нескінченний каскад біфуркацій подвоєння періодів циклів. Такий сценарій для традиційних атракторів називається сценарієм Фейгенбаума [27, 28]. Отже, попри те, що максимальні атрактори не є атракторами у класичному розумінні цього поняття, перехід до хаосу для максимальних атракторів може здійснюватися за природною аналогією типового сценарію Фейгенбаума.

Із подальшим зростанням біфуркаційного параметра E спостерігається перехід від хаотичного максимального атрактора одного типу до хаотичного атрактора максимального іншого типу. Такий перехід типу «хаос–хаос» відбувається за сценарієм узагальненої переміжності. Цей, порівняно новий, сценарій є узагальненням типового сценарію переміжності Манневілля–Помо [29, 30]. Різні варіанти сценарію узагальненої переміжності описано у працях [31, 32]. Зазначимо, що сценарій узагальненої переміжності спочатку був виявлений для традиційних атракторів. Але цей нетиповий сценарій реалізується також для максимальних атракторів.

Ознакою можливої реалізації узагальненої переміжності послуговує вигляд біфуркаційного дерева в околі точки $E \approx -1.775$ (див. рис. 2 (*б*)). Після проходження цієї точки в бік зростання біфуркаційного параметра E значно збільшується площа густо чорної ділянки на біфуркаційному дереві. Це свідчить про виникнення у системі (2) хаотичного атрактора нового типу.

На рис. 4, *а* побудовано проекції фазових портретів трьох представників хаотичного максимального атрактора іншого типу, який існує у системі (2), якщо $E = -1.774$. Перехід до такого хаотичного максимального атрактора відбувається за сценарієм узагальненої переміжності [32]. Для ілюстрації імплементації такого сценарію побудовано розподіли інваріантної міри для двох типів хаотичних максимальних атракторів. Рис. 4, *в* побудований, якщо $E = -1.776$, а рис. 4, *г* — якщо $E = -1.774$. Після проходження точки біфуркації зі зростанням параметра E хаотичний максимальний атрактор, зображений на рис. 3, *б*, зникає і в системі виникає новий хаотичний максимальний атрактор, зображений на рис. 4, *а*. Рух траєкторій по новому хаотичному атрактор включає дві фази, груболамінарну та турбулентну. Груболамінарній фазі на рис. 4, *г* відповідає більш темна ділянка, яка за формою нагадує зниклий хаотичний атрактор. Турбулентній фазі переміжності відповідають більш світлі ділянки на рис. 4, *г*. Чергування фаз груболамінарна – турбулентна повторюється незліченну кількість разів. Час переходу із груболамінарної фази у турбулентну і знову назад у груболамінарну непередбачуваний. Такий перехід відбувається на всіх траєкторіях сім'ї, яка утворює хаотичний максимальний атрактор за одного і того ж значення біфуркаційного параметра.

Імплементацію сценарію узагальненої переміжності можна виявити, аналізуючи перетини Пуанкаре. На рис. 4, *б* побудовані перерізи Пуанкаре представників хаотичного максимального атрактора, якщо $E = -1.776$ (сірі точки) і $E = -1.774$ (чорні точки більшого розміру). В обох випадках перетини утворюють квазістрічкові хаотичні множини. Локалізація квазістрічкової множини представника хаотичного максимального атрактора за $E = -1.776$ збігається з локалізацією перетину Пуанкаре відповідного груболамінарній фазі представника хаотичного максимального атрактора за $E = -1.774$ (чотири ділянки на цьому рисунку, на яких сірі точки розташовуються посеред чорних). Чорні точки на рис. 4, *б* відповідають турбулентній фазі максимального аттрактора, за $E = -1.774$.

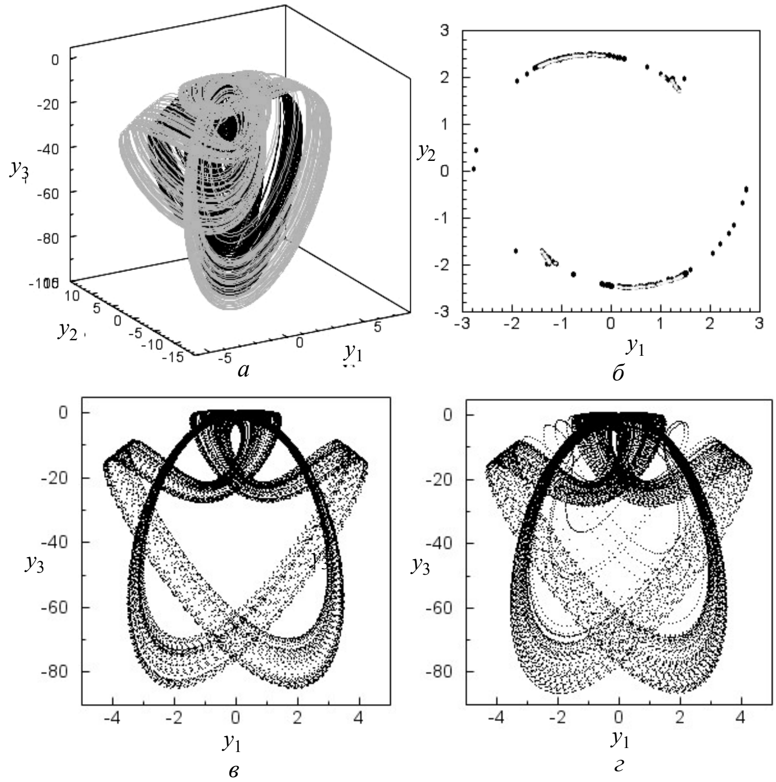


Рис. 4. Проекції фазових портретів, якщо $E = -1.774$ (*а*); перерізи Пуанкаре (*б*); розподіли природної інваріантної міри, якщо $E = -1.776$ (*в*) і $E = -1.774$ (*г*)

На завершення розглянемо реалізацію ще одного типового сценарію переходу до хаосу, а саме переміжності [29, 30], для нетипових максимальних атракторів. Цей розгляд, для різноманітності, проведемо в іншому вікні періодичності $-1.89 < E < -1.88$ біфуркаційного дерева. У малому правому півоколі точки $E = -1.89$ у системі (2) існує періодичний максимальний атрактор (сім'я циклів). Проекцію розподілу природної інваріантної міри побудовано для одного з представників максимального періодичного атрактора за $E = -1.886$ (рис. 5, *а*). Цикли цього атрактора мають більш складну структуру на відміну від розглянутих вище. Якщо $E = -1.89$, сім'я циклів зникає і в системі виникає хаотичний максимальний атрактор. Виникнення хаосу відбувається за одну жорстку біфуркацію. На рис. 5, *б* показано проекцію розподілу природної інваріантної міри, побудовану за значення

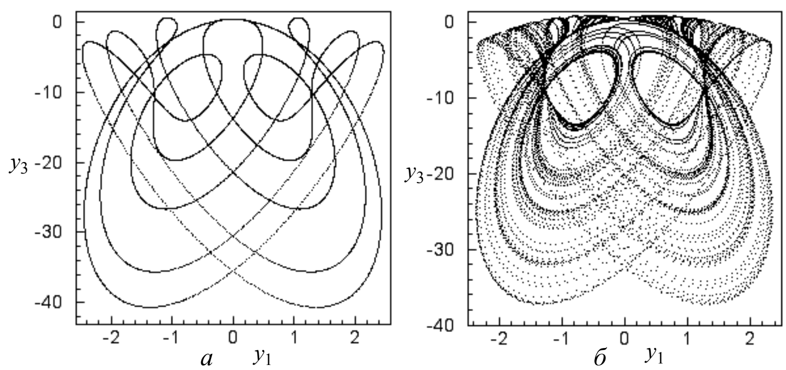


Рис. 5. Розподіли природної інваріантної міри за значень $E = -1.886$ (*а*) і $E = -1.89$ (*б*)

$E = -1.89$. Порівняльний аналіз розподілів (рис. 5) наочно свідчить про виникнення хаосу через переміжність за Манневіллем–Помо. На рис. 5, *б* можна помітити більш жирно прокреслені лінії, які за формою нагадують контури зниклого граничного циклу. Це ламінарна фаза переміжності. Більш світлі ділянки всередині цього рисунка є турбулентною фазою переміжності. Як зазначалося, перехід «цикл–хаос» відбувається за значення $E = -1.89$ для всієї сім'ї циклів періодичного максимального атрактора.

Усі числові розрахунки виконано за значень параметрів, що відповідають системі «бак з рідиною–електродвигун». Але подібні типи максимальних атракторів і сценарії переходу до хаосу характерні і для системи «сферичний маятник–електродвигун». Деякі результати для біфуркацій по іншому параметру частково отримано у праці [25].

ВИСНОВКИ

У роботі проаналізовано перехід до детермінованого хаосу в деяких важливих для застосувань неідеальних динамічних системах. Основними граничними множинами розглянутих систем будуть максимальні атрактори. Установлено, що незважаючи на те, що максимальні атрактори не є атракторами у традиційному розумінні цього терміна, перехід до хаосу для максимальних атракторів відбувається за типовими і нетиповими сценаріями, описаними для традиційних атракторів.

ЛІТЕРАТУРА

1. E.N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.*, **20**, pp. 130–141, 1963. Available: https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2
2. *Handbook of Applications of Chaos Theory*; Edited By C.H. Skiadas and Char. Skiadas. Chapman and Hall/CRC, 2016, 952 p.
3. G. Leonov and N. Kuznetsov, *Nonlinear mathematical models of phase-locked loops. Stability and Oscillations*. Cambridge Scientific Publisher, Cambridge, 2014.
4. N. Kuznetsov, "Hidden attractors in fundamental problems and engineering models," *Lect. Notes Electr. Eng.*, **4**, pp. 13–25, 2016.
5. J. Milnor, "On the concept of attractor," *Commun. Math. Phys.*, **99**, pp. 177–195, 1985. Available: https://doi.org/10.1007/BF01212280
6. V.S. Anishchenko and T.E. Vadivasova, *Lectures on Nonlinear Dynamics*. Research Center Regular and Chaotic Dynamics, 2011, 516 p.
7. A. Sommerfeld, "Beitrage zum dynamischen Ausbau der Festigkeitslehre," *Physikalische Zeitschrift*, **3**, pp. 266–271, 1902.
8. V.O. Kononenko, *Vibrating system with a limited power-supply*. London: Iliffe, 1969, 236 p.
9. T.S. Krasnopolskaya, "Acoustic chaos caused by the Sommerfeld effect," *J. Fluids Struct.,* **8**(6), 803–815, 1994. doi: 10.1016/s0889-9746(94)90300-x
10. T.S. Krasnopolskaya and G.J.F. van Heijst, "Wave pattern formation in a fluid annulas with a vibrating inner shell," *Journal of Fluid Mechanics*, vol. 328, pp. 229–252, 1996. Available: https://doi.org/10.1017/S0022112096008701
11. M.F. Dimentberg et al., "Dynamics of an Unbalanced Shaft Interacting with a Limited Power Supply," *Nonlinear Dynamics*, **13**, pp. 171–187, 1997.
12. M. Dimentberg and C. Bucher, "Combinational parametric resonance under imperfectly periodic excitation," *J. Sound Vibr.,* **331**(19), pp. 4373–4378, 2009. doi: 10.1016/j.jsv.2012.04.025
13. R.F. Ganiev and T.S. Krasnopolskaya, "The Scientific Heritage of V.O. Kononenko: The Sommerfeld–Kononenko Effect," *Journal of Machinery Manufacture and Reliability*, **47**(5), pp. 389–398, 2018. Available: https://doi.org/10.3103/S1052618818050047
14. A.A. Alifov, "About mixed forced, parametric and self-oscillations by limited excitation and delayed elasticity," *PNRPU Mechanics Bulletin*, 3, pp. 12–19, 2020.
15. S. Manikandan and P. Kokil, "Stability analysis of network-controlled generator excitation system with interval time-varying delays," *Automatika*, **62**(1), pp. 65–75, 2021. doi: 10.1080/00051144.2020.1860390

16. A. Shvets and S. Donetskyi, “Transition to deterministic chaos in some electroelastic systems,” *Springer Proceedings in Complexity*, pp. 257–264, 2019. DOI 10.1007/978-3-030-15297-0_23.
17. E.D. Pechuk et al., “Influence of the Heart Rate on Dynamics of Cardiorespiratory System,” *Springer Proceedings in Complexity*, pp. 211–216, 2020. doi: 10.1007/978-3-030-39515-5_17
18. T.S. Krasnopol’skaya and A.Yu. Shvets, “Chaotic oscillations of a spherical pendulum as an example of interaction with energy source,” *Int. Appl. Mech.*, **28**(10), pp. 669–674, 1992.
19. T.S. Krasnopol’skaya and A.Y. Shvets, “Parametric resonance in the system: Liquid in tanks +electric motor,” *Int. Appl. Mech.*, **29**(9), pp. 722–730, 1993.
20. T.S. Krasnopolskaya and A.Yu. Shvets, “Chaotic surface waves in limited power—supply cylindrical tank vibrations,” *J. Fluids Struct.*, **8**(1), pp. 1–18, 1994. doi: 10.1006/jfls.1994.1001
21. S.P. Kuznetsov, *Dynamic Chaos*. M.: Fizmatlit, 2006, 292 p.
22. T.S. Krasnopolskaya and A.Yu. Shvets, *Regular and chaotic dynamics of systems with limited excitation*. M.: Research Center Regular and Chaotic Dynamics, 2008, 280 p.
23. A.Yu. Shvets and T.S. Krasnopolskaya, “Hyperchaos in piezoceramic systems with limited power supply,” *Solid Mechanics and its Applications*, 6, pp. 313–322, 2008. doi: 10.1007/978-1-4020-6744-0_27
24. A.Yu. Shvets, “Deterministic chaos of a spherical pendulum under limited excitation,” *Ukr. Math. J.*, **59**, pp. 602–614, 2007. doi: 10.1007/s11253-007-0039-7
25. A. Shvets and S. Donetskyi, “New Types of Limit Sets in the Dynamic System “Spherical Pendulum—Electric Motor,” *Advanced Structured Materials*, 157, pp. 443–455, 2021. Available: https://doi.org/10.1007/978-3-030-75890-5_25
26. A. Li’enard and M.H. Chipart, “Sur le signe de la partie r’eelle des racines d’une quation alg’ebrique,” *J. Math. Pures Appl.*, **10** (4), pp. 291–346, 1914.
27. M.J. Feigenbaum, “Quantitative universality for a class of nonlinear transformations,” *J. Stat. Phys.*, **19**(1), pp. 25–52, 1978. Available: https://doi.org/10.1007/BF01020332
28. M.J. Feigenbaum, “The universal metric properties of nonlinear transformations,” *J. Stat. Phys.*, **21**(6), pp. 669–706, 1979.
29. P. Manneville and Y. Pomeau, “Different ways to turbulence in dissipative dynamical systems,” *Physica D. Nonlinear Phenom.*, **1**(2), pp. 219–226, 1980.
30. Y. Pomeau and P. Manneville, “Intermittent transition to turbulence in dissipative dynamical systems,” *Comm. Math. Phys.*, **74**(2), pp. 189–197, 1980. Available: https://doi.org/10.1007/BF01197757
31. A.Yu. Shvets and V.A. Sirenko, “Scenarios of transitions to hyperchaos in nonideal oscillating systems,” *J. Math. Sci.*, **243**(2), pp. 338–346, 2019. doi: 10.1007/s10958-019-04543-z
32. A. Shvets, “Overview of Scenarios of Transition to Chaos in Nonideal Dynamic Systems,” *Springer Proceedings in Complexity*, pp. 853–864, 2021. Available: https://doi.org/10.1007/978-3-030-70795-8_59

Надійшла 27.07.2022

INFORMATION ON THE ARTICLE

Aleksandr Yu. Shvets, ORCID: 0000-0003-0330-5136, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: aleksandrshvetskpi@gmail.com

TYPICAL AND GENERALIZED TRANSITIONS TO DETERMINISTIC CHAOS FOR ATYPICAL ATTRACTORS OF NON-IDEAL DYNAMIC SYSTEMS / A.Yu. Shvets

Abstract. Some applied nonlinear, non-ideal dynamic systems of the fifth order, which are used to describe the oscillations of spherical pendulums and in hydrodynamics, are considered. Maximal attractors, both regular and chaotic, of such systems are constructed. Various bifurcations of maximal attractors are discussed. The transition to deterministic chaos is established for maximal attractors in typical Feigenbaum and Manneville–Pomeau scenarios. The implementation of the generalized alternation scenario for chaotic maximum attractors of such systems is investigated. A sign of the implementation of the scenario of generalized alternation has been revealed.

Keywords: non-ideal dynamic system, regular maximal attractor, chaotic maximal attractor, typical and generalized intermittency.

ВІДОМОСТІ ПРО АВТОРІВ

Аджубей Лариса Трохимівна,

доцент, кандидат фізико-математичних наук, доцент кафедри обчислювальної математики Київського національного університету імені Тараса Шевченка, Україна, Київ

Бідюк Петро Іванович,

професор, доктор технічних наук, професор кафедри математичних методів системного аналізу ІПСА КПІ ім. Ігоря Сікорського, Україна, Київ

Бодянський Євгеній Володимирович,

професор, доктор технічних наук, професор кафедри штучного інтелекту, науковий керівник ПНДЛ АСУ Харківського національного університету радіоелектроніки, Україна, Харків

Ваврик Петро Русланович,

студент кафедри системного аналізу та теорії прийняття рішень Київського національного університету імені Тараса Шевченка, Україна, Київ

Гасанов Айдин Сардар огли,

доцент, доктор технічних наук, професор кафедри комп'ютерної і програмної інженерії Національного педагогічного університету імені М.П.Драгоманова, Україна, Київ

Дунін Володимир Андрійович,

кандидат технічних наук, науковий співробітник Державного науково-дослідного інституту будівельних конструкцій, Україна, Київ

Івохін Євген Вікторович,

професор, доктор фізико-математичних наук, професор кафедри системного аналізу та теорії прийняття рішень Київського національного університету імені Тараса Шевченка, Україна, Київ

Калюх Юрій Іванович,

професор, доктор технічних наук, провідний науковий співробітник, Інституту телекомунікацій і глобального інформаційного простору НАН України, Київ

Кольцов Дмитро Романович,

випускник магістратури кафедри прикладної математики КПІ ім. Ігоря Сікорського, Україна, Київ

Кулішова Нонна Євгенівна,

доцент, кандидат технічних наук, професор кафедри медіасистем та технологій Харківського національного університету радіоелектроніки, Україна, Харків

Кураш Сергій Юрійович,

науковий співробітник Державного науково-дослідного інституту будівельних конструкцій, Україна, Київ

Любчик Леонід Михайлович,

професор, доктор технічних наук, викладач кафедри комп'ютерної математики та аналізу даних Національного технічного університету «Харківський політехнічний інститут», Україна, Харків

Мазурик Роман Володимирович,

аспірант кафедри прикладної математики КПІ ім. Ігоря Сікорського, Україна, Київ

Малежик Петро Михайлович,

кандидат фізико-математичних наук, доктор педагогічних наук, доцент кафедри комп'ютерної і програмної інженерії Національного педагогічного університету імені М.П.Драгоманова, Україна, Київ

Мартиненко Андрій Анатолійович,

старший викладач кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка», Україна, Дніпро

Махно Михайло Федорович,

кандидат технічних наук, асистент кафедри системного аналізу та теорії прийняття рішень Київського національного університету імені Тараса Шевченка, Україна, Київ

Мельников Сергій Вікторович,

кандидат технічних наук, старший науковий співробітник відділу екологічних цифрових систем Міжнародного науково-навчального центру інформаційних технологій та систем НАН України, Київ

Мороз Борис Іванович,

професор, доктор технічних наук, член-кореспондент Академії прикладної електроніки, професор кафедри програмного забезпечення комп'ютерних систем Національного технічного університету «Дніпровська політехніка», Україна, Дніпро

Обелець Тетяна Володимирівна,

кандидат економічних наук, старший викладач кафедри економічної кібернетики КПІ ім. Ігоря Сікорського, Україна, Київ

Ориняк Ігор Володимирович,

професор, доктор технічних наук, професор кафедри прикладної математики КПІ ім. Ігоря Сікорського, Україна, Київ

Пишнограєв Іван Олександрович,

доцент, кандидат фізико-математичних наук, доцент кафедри штучного інтелекту КПІ ім. Ігоря Сікорського, Україна, Київ

Плісс Ірина Павлівна,

старший науковий співробітник, кандидат технічних наук, провідний науковий співробітник ПНДЛ АСУ Харківського національного університету радіоелектроніки, Україна, Харків

Тевяшев Андрій Дмитрович,

професор, доктор технічних наук, завідувач кафедри прикладної математики Харківського національного університету радіоелектроніки, Україна, Харків

Ткаченко Іванна Олексіївна,

лаборант Світового центру даних з геоінформатики і сталого розвитку КПІ ім. Ігоря Сікорського, Україна, Київ

Трофимчук Олександр Миколайович,

професор, доктор технічних наук, директор Інституту телекомунікацій та глобального інформаційного простору НАН України, Київ

Чертов Олег Романович,

професор, доктор технічних наук, завідувач кафедри прикладної математики КПІ ім. Ігоря Сікорського, Україна, Київ

Шафроненко Аліна Юріївна,

доцент, кандидат технічних наук, доцент кафедри інформатики Харківського національного університету радіоелектроніки, Україна, Харків

Швець Олександр Юрійович,

професор, доктор фізико-математичних наук, професор кафедри математичної фізики та диференціальних рівнянь КПІ ім. Ігоря Сікорського, Україна, Київ

Ямковий Клим Сергійович,

асистент кафедри комп'ютерної математики та аналізу даних Національного технічного університету «Харківський політехнічний інститут», Україна, Харків

Abdelbadeeh M. Salem,

Professor, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

EL-Sayed A. El-Dahshan,

Professor, Department of Physics, Faculty of Science, Ain Shams University, Cairo, Egypt.

Islam Hegazy,

Assistant Professor, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

Mahmoud M. Bassiouni,

Assistant Lecturer, Faculty of computer and information technology, Egyptian E-Learning University (EELU), El-Giza, Egypt

Nouhad Rizk,

Inst. Associate Professor of computer Science, Director of Undergraduate Studies, Computer Science Department, Houston University, USA