UDC 004.8:004.032.26

CREDIBILISTIC ROBUST ONLINE FUZZY CLUSTERING IN DATA STREAM MINING TASKS

Shafronenko A. Yu. – PhD, Associate Professor at the Department of Informatics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Kasatkina N. V. – Dr. Sc., Professor, Division of doctoral and post-graduate training, National University of Food Technologies, Kyiv, Ukraine.

Bodyanskiy Ye. V. – Dr. Sc., Professor, Professor at the Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Shafronenko Ye. O. – Assistant at the Department of Media Engineering and Information Radio Electronic Systems, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

ABSTRACT

Context. The task of clustering-classification without a teacher of data arrays occupies an important place in the general problem of Data Mining, and for its solution there exists currently many approaches, methods and algorithms. There are quite a lot of situations where the real data to be clustered are corrupted with anomalous outliers or disturbances with non-Gaussian distributions. It is clear that "classical" methods of artificial intelligence (both batch and online) are ineffective in this situation. The goal of the paper is to develop a credibilistic robust online fuzzy clustering method that combines the advantages of credibilistic and robust approaches in fuzzy clustering tasks.

Objective. The goal of the work is online credibilistic fuzzy clustering of distorted data, using of credibility theory in data stream mining.

Method. The procedure of fuzzy clustering of data using credibilistic approach based on the use of both robust goal functions of a special type, insensitive to outliers and designed to work both in batch and its recurrent online version designed to solve Data Stream Mining problems when data are fed to processing sequentially in real time.

Results. Analyzing the obtained results overall accuracy of clustering methods and algorithm, proposed method similar with result of credibilistic fuzzy clustering method, but has time superiority regardless of the number observations that fed on clustering process.

Conclusions. The problem of fuzzy clustering of data streams contaminated by anomalous non-Gaussian distributions is considered. A recurrent credibilistic online algorithm based on the objective function of a special form is introduced, which suppresses these outliers by using the hyperbolic tangent function, which, in addition to neural networks, is used in robust estimation tasks. The proposed algorithm is quite simple in numerical implementation and is a generalization of some well-known online fuzzy clustering procedures intended for solving Data Stream Mining problems.

KEYWORDS: fuzzy clustering, distorted data, credibilistic fuzzy clustering, Data Stream Mining, robust function.

ABBREVIATIONS

FCM is a fuzzy c-means method;

SOM is self-organizing map;

CROFC is credibilistic robust online fuzzy clustering method.

NOMENCLATURE

X is a data set matrix;

N is number of observations;

R is space of input vectors;

n is number of attributes;

m is number of overlapping classes;

k is a number of the vectors-observation;

i is a number components of the vectors-observation;

$x(k)$ is a vector of observations;

$x_i(k)$ is a preprocessed original data;

l, j is a number of clusters;

μ is a membership level;

$\mu_j(k)$ is a membership level of k-th vector-observation to j-th cluster;

c is a centroid of cluster;

c_j is a centroid of j-th cluster;

d is a Euclidean distance;

d_p is a Minkowski distance;

J is a goal function;

L is Lagrange function;

$\eta(k)$ is learning-rate parameter;

$Cr_j(k)$ is fuzzy credibilistic membership level;

$\lambda(k)$ is indefinite Lagrange multiplier;

β is a fuzzifier;

β_i is parameter specifying the modification of function.

INTRODUCTION

The task of clustering-classification without a teacher of data arrays occupies an important place in the general problem of Data Mining, and for its solution there are currently many approaches, methods and algorithms [1–3].

A special place here is occupied by methods of fuzzy clustering, when it is a priori assumed that each observation can simultaneously belong to several or all classes at the same time with different levels of fuzzy membership, i. e. classes overlapping in the feature space [4, 5].

DOI 10.15588/1607-3274-2023-3-10

Fuzzy clustering methods can be conditionally divided into two large classes: probabilistic, among which the Fuzzy C-means algorithm (FCM) by J. Bezdek [4] was the most popular, and probabilistic. Each of these classes has its advantages and disadvantages, and to overcome these disadvantages, a so-called credibilistic approach was proposed [6, 7], which has already proved its effectiveness in solving a number of problems.

There are quite a lot of situations where the real data to be clustered are corrupted with anomalous outliers or disturbances with non-Gaussian distributions. This leads to the fact that traditional methods using quadratic metrics (Euclidian, Mahalanobis, etc.) do not provide the desired results. This led to the creation of robust clustering methods [8–10] resistant to these outliers and based on non-quadratic distances, while most of the known robust fuzzy clustering algorithms are based on a probabilistic approach.

It is appropriate to develop a credibilistic robust online fuzzy clustering (CROFC) method that combines the advantages of credibilistic and robust approaches in fuzzy clustering tasks and is designed to process data streams that arrive sequentially in real time.

The object of study is fuzzy clustering of data distorted by outliers.

The subject of study is procedure for fuzzy clustering of data distorted by outliers based on robust approaches in fuzzy clustering tasks.

The purpose of the work is to introduce robust online credibilistic method for fuzzy clustering of distorted data.

1 PROBLEM STATEMENT

The initial information for solving the clustering problem is an unlabeled sample of vector observations $X = \{x(1), x(2), ..., x(k), ..., x(N)\} \subset R^n$ where k-th observation number, in the sample when working in batch mode or index of the current discrete time, when solving Data Stream Mining tasks. The result of solving the problem of fuzzy clustering is the division of this sample into m overlapping classes-clusters with estimation of fuzzy membership levels $\mu_j(k)$ to each of the possible clusters c_j, $j = 1, 2, ..., m$.

2 REVIEWS OF THE LITERATURE

A special place in the general problem of Data Mining is occupied by tasks related to Data Stream Mining when the data to be processed, the tasks are not in the form of a batch, but are sequentially received for processing one by one, while the amount of this data is unknown a priori. The most characteristic example here is T. Kohonen's self-organizing maps (SOM) [11], which implements the traditional crisp algorithm of K-means clustering in an online version using the self-learning rule "Winner Takes All" (WTA). For fuzzy situations D.C. Park and J. Dagger [12] have proposed a recurrent version of FCM, and in [13] both probabilistic and probabilistic recurrent algorithms of online fuzzy clustering were considered.

Regarding credibilistic and robust approaches, in [14, 15] recurrent modifications of the credibilistic fuzzy clustering algorithm were proposed, and in [16, 17] recurrent robust procedures of fuzzy clustering designed for processing data streams in online mode where introduced.

3 MATERIALS AND METHODS

The most popular approach to solving this problem is related to the minimization of the objective function

$$J\left(\mu_j(k), c_j\right) = \sum_{k=1}^{N}\sum_{j=1}^{m}\mu_j^{\beta}(k) d_p^2\left(x(k), c_j\right)$$

with restrictions

$$\sum_{j=1}^{m}\mu_j(k) = 1,$$

$$0 < \sum_{j=1}^{m}\mu_j(k) < 1$$

where c_j – prototype-centroid of j-th cluster; β – parameter-fuzzifier (usually $\beta = 2$); $d_p^2\left(x(k), c_j\right)$ – the distance between $x(k)$ and c_j. Most often, this is the Minkowski distance

$$d_p\left(x(k), c_j\right) = \left\|x(k) - c_j\right\|_p = \left(\sum_{i=1}^{n}\left|x_i(k) - c_{ji}\right|^p\right)^{\frac{1}{p}},$$

the special case of which is the traditional Euclidean norm

$$d_2\left(x(k), c_j\right) = \left\|x(k) - c_j\right\|_2 = \left(\sum_{i=1}^{n}\left|x_i(k) - c_{ji}\right|^2\right)^{\frac{1}{2}}.$$

Using the standard procedure of non-linear programming, the Lagrange function is introduced for consideration

$$L\left(\mu_j(k), c_j, \lambda(k)\right) = \sum_{k=1}^{N}\sum_{j=1}^{m}\mu_j^{\beta}(k) d_p^2\left(x(k), c_j\right) +$$

$$+\sum_{k=1}^{N}\lambda(k)\left(\sum_{j=1}^{m}\mu_j(k) - 1\right) = \quad (1)$$

$$= \sum_{k=1}^{N}\left(\sum_{j=1}^{m}\mu_j^{\beta}(k) d_p^2\left(x(k), c_j\right) + \lambda(k)\left(\sum_{j=1}^{m}\mu_j(k) - 1\right)\right)$$

(here $\lambda(k)$ – the unknown Lagrange multiplier) and the system of Kuhn-Tucker equations

DOI 10.15588/1607-3274-2023-3-10

$$\begin{cases} \dfrac{\partial L\left(\mu_j(k), c_j, \lambda(k)\right)}{\partial \mu_j(k)} = 0, \\ \dfrac{\partial L\left(\mu_j(k), c_j, \lambda(k)\right)}{\partial \lambda(k)} = 0, \\ \nabla_{c_j} L\left(\mu_j(k), c_j, \lambda(k)\right) = \vec{0}, \end{cases} \quad (2)$$

solving which we get the result in the form:

$$\begin{cases} \mu_j(k) = \dfrac{\left(d_p^2\left(x(k), c_j\right)\right)^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}\left(d_p^2\left(x(k), c_l\right)\right)^{\frac{1}{1-\beta}}}, \\ \lambda(k) = -\left(\sum\limits_{l=1}^{m}\left(\beta d_p^2\left(x(k), c_l\right)\right)^{\frac{1}{1-\beta}}\right)^{1-\beta}, \\ c_j = \dfrac{\sum\limits_{k=1}^{N}\mu_j^{\beta}(k)x(k)}{\sum\limits_{k=1}^{N}\mu_j^{\beta}(k)}, \end{cases}$$

if $\beta = 2$ turns into the classical FCM algorithm of J. Bezdek [4]:

$$\begin{cases} \mu_j(k) = \dfrac{\left\|x(k) - c_j\right\|_2^2}{\sum\limits_{l=1}^{m}\left\|x(k) - c_j\right\|_2^2}, \\ \lambda(k) = -\sum\limits_{l=1}^{m}\left(\dfrac{\left\|x(k) - c_j\right\|_2^2}{2}\right)^{-1}, \\ c_j = \dfrac{\sum\limits_{k=1}^{N}\mu_j^2(k)x(k)}{\sum\limits_{k=1}^{N}\mu_j^2(k)}. \end{cases}$$

For the online clustering procedure to find the saddle point of the Lagrange function (1), instead of directly solving the system of Kuhn-Tucker equations (2), the Arrow-Hurwitz-Uzawa algorithm [18] can be used, with the help of which we obtain a recurrent procedure [19]

$$\begin{cases} \mu_j(k) = \dfrac{\left(d_p^2\left(x(k), c_j(k)\right)\right)^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}\left(d_p^2\left(x(k), c_l(k)\right)\right)^{\frac{1}{1-\beta}}}, \\ c_j(k+1) = c_j(k) - \eta(k)\nabla_{c_j} L\left(\mu_j(k), c_j(k), \lambda(k)\right) = \\ = c_j(k) - \eta(k)\mu_j^{\beta}(k) d_p\left(x(k+1), c_j(k)\right) * \\ *\nabla_{c_j} d_p^2\left(x(k+1), c_j(k)\right) \end{cases}$$

(here $\eta(k)$ – self-learning parameter) which $\beta = 2$, $p = 2$ turns into GBFC Park-Dagger algorithm [12]

$$\begin{cases} \mu_j(k) = \dfrac{\left\|x(k) - c_j(k)\right\|_2^{-2}}{\sum\limits_{l=1}^{m}\left\|x(k) - c_l(k)\right\|_2^{-2}}, \\ c_j(k+1) = c_j(k) + \eta(k)\mu_j^2(k)\left(x(k+1) - c_j(k)\right). \end{cases} \quad (3)$$

Here it is interesting to notice that the second relation (3) should completely match the self-learning rule "Winner Takes More" (WTM) by T. Kohonen [11] in terms of structure, while the role of the neighborhood function here is performed by a set $\mu_j^2(k)$.

All the algorithms discussed above, based on the Minkowski metric, do not ensure the process of clustering robust properties, since they only "amplify" the influence of anomalous emissions present in the data sample.

Therefore, it is advisable to use distances that have robust properties, "suppressing" these emissions.

One of these distances can be based on a function [20, 21] of the form:

$$d^R\left(x(k), c_j\right) = \sum_{i=1}^{n}\beta_i \ln\left(\cosh\left(\frac{x_i(k) - c_{ji}}{\beta_i}\right)\right), \quad (4)$$

where β_i – the parameter specifying the modification of this function is usually accepted $\beta_i = 2$, $i = 1, 2, ...n.$

Introducing the robust objective function further

$$J^R\left(\mu_j(k), c_j\right) = \sum_{k=1}^{N}\sum_{j=1}^{m}\mu_j^{\beta}(k) d^R\left(x(k), c_j\right) =$$
$$= \sum_{k=1}^{N}\sum_{j=1}^{m}\mu_j^{\beta}(k)\sum_{i=1}^{n}\beta_i \ln\left(\cosh\left(\frac{x_i(k) - c_{ji}}{\beta_i}\right)\right)$$

and the corresponding Lagrange function

$$L^R\left(\mu_j(k), c_j, \lambda(k)\right) = \sum_{k=1}^{N}\sum_{j=1}^{m}\mu_j^{\beta}(k)\sum_{i=1}^{n}\beta_i \times$$
$$\times \ln\left(\cosh\left(\frac{x_i(k) - c_{ji}}{\beta_i}\right)\right) + \sum_{k=1}^{N}\lambda(k)\left(\sum_{j=1}^{m}\mu_j(k) - 1\right),$$

it is possible to write the system of Kuhn-Tucker equations, which, however, due to the complexity of the distance (4), does not have an analytical solution.

Therefore, the only solution here is to use the same Arrow-Hurwitz-Uzawa algorithm, which leads to the result and

DOI 10.15588/1607-3274-2023-3-10

$$\begin{cases} \mu_j(k) = \dfrac{\left(d^R\left(x(k), c_j(k)\right)\right)^{\frac{1}{1-\beta}}}{\sum\limits_{l=1}^{m}\left(d^R\left(x(k), c_l(k)\right)\right)^{\frac{1}{1-\beta}}}, \\ c_j(k+1) = c_j(k) - \eta(k)\dfrac{\partial}{\partial c_{ji}} L^R\left(\mu_j(k), c_j(k), \lambda(k)\right) = \\ = c_j(k) + \eta(k)\mu_j^{\beta}(k)\tanh\left(\dfrac{x_i(k) - c_{ji}(k)}{\beta_i}\right), \end{cases} \quad (5)$$

where function $\tanh(\bullet)$, which is usually used as an activation in many neural networks is used to suppress specifically anomalous outliers in the data.

Next, taking the value of the fuzzifier $\beta = 2$ (5) can be rewritten in a somewhat simplified form:

$$\begin{cases} \mu_j(k) = \dfrac{\left(d^R\left(x(k), c_j(k)\right)\right)^{-1}}{\sum\limits_{l=1}^{m}\left(d^R\left(x(k), c_l(k)\right)\right)^{-1}}, \\ c_{ji}(k+1) = c_{ji}(k) + \eta(k)\mu_j^2(k)\tanh\left(\dfrac{x_i(k) - c_{ji}(k)}{\beta_i}\right), \end{cases} \quad (6)$$

which is essentially a robust Park-Dagger algorithm (3) that suppresses outliers in the data using the function $\tanh(\bullet)$.

Based on this algorithm, it is easy to consider its credibilistic modification by supplementing (6) with a simple relation [6, 7]:

$$\begin{cases} \mu_j^r(k) = \dfrac{\mu_j(k)}{\sup \mu_l(k)}, \\ Cr_j(k) = \dfrac{1}{2}\left(\mu_j^r(k) + 1 - \sup \mu_l(k)\right). \end{cases} \quad (7)$$

Ratios (6), (7) define the credibilistic robust online fuzzy clustering algorithm, intended for use in systems for processing data streams distorted by various types of disturbances and arriving online.

4 EXPERIMENTS

For test the method of credibilistic robust online fuzzy clustering in data stream mining tasks was conducted test data sets of Nursery from the UCI repository.

Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools. It was used during several years in 1980's when there was excessive enrollment to these schools in Ljubljana, Slovenia, and the rejected applications frequently needed an objective explanation.

The final decision depended on three subproblems: occupation of parents and child's nursery, family structure and financial standing, and social and health picture of the family. The Nursery Database contains 12958 examples with the structural information removed, i.e., directly relates NURSERY to the eight input attributes: parents, has_nurs, form, children, housing, finance, social, health.

5 RESULTS

We compared the obtained results with classical FCM algorithm, probabilistic, possibilistic and credibilistic fuzzy clustering methods.

Table 1 – The overall accuracy of clustering methods and algorithm

Clustering algorithm	Overall accuracy		
	Highest	Mean	Variance
FCM	68.54	68.54	0.01
Credibilistic fuzzy clustering	67.98	67.98	0
Possibilistic fuzzy clustering	68.55	68.54	0.01
Probabilistic fuzzy clustering	68.48	68.48	0.01
CROFC	67.68	67.65	0

A comparative analysis of the quality of the clustering data was carried out according to the main characteristics of the quality ratings, such as the speed of data clustering and the average error.

Table 2 show the results of the algorithms proposed for comparison with different numbers of observations.

Table 2 – Comparative characteristics of the average error with different number of observations in percentage

Algorithm	50	Time	100	Time	150	Time
FCM	1.62	1.19	1.35	2.55	0.98	3.03
Probabilistic fuzzy clustering	1.66	1.62	1.32	2.72	0.99	3.12
Possibilistic fuzzy clustering	1.22	1.15	1.02	2.02	0.75	2.10
Credibilistic fuzzy clustering	0.69	1.02	0.49	1.33	0.14	1.41
CROFC	0.68	1.00	0.45	1.25	0.12	1.33

Analyzing the obtained results, it can be concluded that regardless of the size of the initial information submitted for processing by the proposed method for comparing performance and efficiency, it is not inferior in speed and quality of clustering in comparison with known algorithms and methods.

The comparative analysis is demonstrated on the diagrams of the dependence of error and time on the number of observations on Fig. 1.

DOI 10.15588/1607-3274-2023-3-10

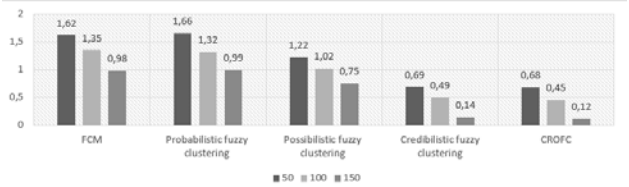


Figure 1 – Diagram of the dependence of the error on the number of observations (50, 100, 150)

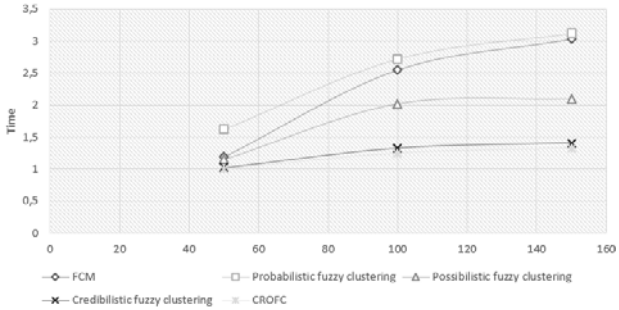


Figure 2 – Diagram of dependence of clustering time on the number of observations (50, 100, 150)

6 DISCUSSIONS

The result of clustering data set Nursery shown in Table 1 and Table 2. As the table shows, the prepositional credibilistic robust online fuzzy clustering in data stream mining tasks have shown good results.

As it can be seen in Fig. 1 the proposed method shows best result on diagram of the dependence of the error on the number of observations and Fig. 2, that demonstrate dependence of clustering time on the number of observations.

Analyzing the obtained results overall accuracy of clustering methods and algorithm, proposed method similar with result of credibilistic fuzzy clustering method, but has time superiority regardless of the number observations that fed on clustering process.

Due to its adaptability and robustness proposed method does not require a lot of time to process the data received in real time, and does not burden itself with intermediate calculations due to adaptability functions.

This is quite clearly demonstrated by the diagrams of the dependence of the clustering time on the number of observations and the dependence of the error on the number of observations.

CONCLUSIONS

The problem of fuzzy clustering of data streams contaminated by anomalous non-Gaussian distributions is considered. A recurrent credibilistic online algorithm based on the objective function of a special form is introduced, which suppresses these outliers by using the hyperbolic tangent function, which, in addition to neural networks, is used in robust estimation tasks. The proposed algorithm is quite simple in numerical implementation and is a generalization of some well-known online fuzzy clustering procedures intended for solving Data Stream Mining problems.

The scientific novelty of obtained results is that the method of credibilistic robust online fuzzy clustering in data stream mining tasks, that shows good results in comparative analyses with another methods, that “worked” with distorted data sets.

The practical significance of obtained results is that analyze properties of the propose methods of credibilistic fuzzy clustering of distorted data. The experimental results allow to recommend the proposed methods for use in practice for solving the problems of automatic clusterization of distorted data.

Prospects for further research methods of online robust credibilistic fuzzy clustering of distorted data in tasks of stream data mining.

ACKNOWLEDGEMENTS

The work is supported by the state budget scientific research project of Kharkiv National University of Radio Electronics “Development of methods and algorithms of combined learning of deep neuro-neo-fuzzy systems under the conditions of a short training sample”.

REFERENCES

1. Gan G., Ma Ch., Wu J. Data Clustering: Theory, Algorithms and Applications. Philadelphia, Pennsylvania: SIAM: 2007. – 455 p. doi: https://doi.org/10.1137/ 1.9780898718348
2. Abony J., Feil D. Cluster Analysis for Data Mining and System Identification. Basel, Birkhouser, 2007, 303 p.
3. Xu R., Wunsch D. C. Clustering. Hoboken N.J., John Wiley & Sons, Inc., 2009, 398 p.
4. Bezdek J. C. Pattern recognition with fuzzy objective function algorithms. New York, Springer, 1981, 253 p. DOI https://doi.org/10.1007/978-1-4757-0450-1.
5. Höppner F., Klawonn F., Kruse R., Runkler T. Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition. Chichester, John Wiley &Sons, 1999, 300 p.
6. Zhou J., Wang Q., Hung C.-C., Yi X. Credibilistic clustering: the model and algorithms, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2015, Vol. 23, № 4, pp. 545–564. DOI: https://doi.org/ 10.1142/S0218488515500245
7. Zhou J., Wang Q., Hung C. C. Credibilistic clustering algorithms via alternating cluster estimation, *Journal of Intelligent Manufacturing*, 2017, Vol. 28, pp. 727–738. DOI: https://doi.org/10.1007/s10845-014-1004-6.
8. Tsuda K., Senda S., Minoh M., Ikeda K. Sequential fuzzy cluster extraction and its robust against noise, *System and Computers in Japan*, 1997, 28, pp. 10–17.
9. Höppner F., Klawonn F. Fuzzy clustering of sampled functions, *19th Int. Conf. North American Fuzzy Information Processing Society (NAFIPS).* Atlanta, USA, 2000, pp. 257–255.
10. Georgieva O., Klawonn F. A clustering algorithm for identification of single clusters in large data sets, *Proc. 11th East – West Fuzzy Coll*. Zittau/Görlitz, FH, 2004, pp. 118–125.
11. Kohonen T. Self-Organizing Maps. Berlin, Springer, 1995, 362 p. DOI: 10.1007/978-3-642-56927-2.
12. Park D. C., Dagger I. Gradient based fuzzy c-means (GBFCM) algorithm, *IEEE International Conference on Neural Networks, 28 June – 2July,1984, proceedings.* Orlando, IEEE, 1984, pp. 1626–1631. DOI: 10.1109 / ICNN. 1994.374399.

DOI 10.15588/1607-3274-2023-3-10

13. Bodyanskiy Ye. Computational intelligence techniques for data analysis, *Lecture Notes in Informatics*. Bonn, Gesellschaft für Informatik, 2005, pp. 15–36.
14. Shafronenko A., Bodyanskiy Ye., Klymova I., Holovin O. Online credibilistic fuzzy clustering of data using membership functions of special type [Electronic resource], *Proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), April 27–1 May 2020*. Zaporizhzhia, 2020. Access mode: http://ceur-ws.org/Vol-2608/paper56.pdf.
15. Shafronenko A., Bodyanskiy Ye., Pliss I., Klymova I. Online Credibilistic Fuzzy Clustering Method Based on Cauchy Density Distribution Function, *2021 11th International Conference on Advanced Computer Information Technologies (ACIT): proceedings*. Deggendorf, Germany, IEEE, 2021, pp. 704–707. DOI: 10.1109/ACIT52158.2021.9548572
16. Bodyanskiy Ye., Gorshkov Ye., Kokshenev I., Kolodyazhniy V. Robust recursive fuzzy clustering algorithms, *Proc. 12th East West Fuzzy Coll 2005*. Zittau-Grölitz, FH, 2005, pp. 301–308.
17. Bodyanskiy Ye. Gorshkov Ye., Kokshenev I., Kolodyazhniy V. Outlier resistant recursive fuzzy clustering algorithms, *Ed. By B. Reusch «Computational Intelligence Theory and Applications» – Advances in Soft Computing*, Vol. 38. Berlin Heidelberg, Springer Verlag, 2006, pp. 647–652.
18. Arrow K. J., Hurwitz L., Uzawa H. Studies in Linear and Nonlinear Programming. Stanford University Press, 1958, 242 p.
19. Bodyanskiy Ye., Kolodyazhniy V., Stephan A. Recurcive fuzzy clustering algorithm, *Proc 10th East West Fuzzy Coll*, 2002. Zittau-Görlitz, HS, 2002, pp. 276–283.

Received 21.06.2023.
Accepted 17.08.2023.

УДК: 004.8:004.032.26

ДОСТОВІРНА РОБАСТНА ОНЛАЙН НЕЧІТКА КЛАСТЕРИЗАЦІЯ В ЗАДАЧАХ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ПОТОКІВ ДАНИХ

Шафроненко А. Ю. – канд. техн. наук, доцент кафедри інформатики Харківського національного університету радіоелектроніки, Харків, Україна.

Касаткіна Н. В. – д-р техн. наук, професор, відділ докторантури та аспірантури Національного університету харчових технологій, Київ, Україна.

Бодянський Є. В. – д-р техн. наук, професор, професор кафедри штучного інтелекту Харківського національного університету радіоелектроніки, Харків, Україна.

Шафроненко Є. О. - асистент кафедри медіаінженерії та інформаційних радіоелектронних систем, Харківський національний університет радіоелектроніки, Харків, Україна.

АНОТАЦІЯ

Актуальність. Задача кластеризації-класифікації без вчителя масивів даних займає важливе місце у загальній проблемі Data Mining, а для її вирішення існує на цей час безліч підходів, методів та алгоритмів. Існує достатньо багато ситуацій, коли реальні дані, що підлягають кластеризації, забруднені аномальними викидами або збуреннями з не Гаусівськими розподілами. Це веде до того, що традиційні методи, що використовують квадратичні метрики не забезпечують бажані результати. Метою статті є розробка достовірного робастного методу нечіткої кластеризації онлайн, який поєднує в собі переваги теорії довіри та робастних підходів у задачах нечіткої кластеризації.

Метод. Процедура нечіткої кластеризації даних з використанням достовірного підходу, заснованого на використанні як робастних цільових функцій спеціального типу, нечутливих до викидів, так і призначених для роботи як у пакетному режимі, так і в його повторюваній онлайн-версії, призначеній для вирішення проблем Data Stream Mining, коли дані надходять на обробку послідовно в режимі реального часу.

Результати. Аналізуючи загальну точність отриманих результатів методів і алгоритму кластеризації, запропонований метод подібний до результату достовірного методу нечіткої кластеризації, але має перевагу в часі незалежно від кількості спостережень, які були використані в процесі кластеризації.

Висновки. Розглянута задача нечіткої кластеризації потоків даних, забруднених аномальними викидами. Введено у розгляд рекурентний достовірний онлайн алгоритм, заснований на цільовій функції спеціального вигляду, що придушує ці викиди за допомогою використання функції гіперболічного тангенса, що крім нейронних мереж використовується у задачах робастного оцінювання. Запропонований алгоритм є достатньо простим у чисельній реалізації і є узагальненням деяких відомих онлайн процедур нечіткої кластеризації призначених для вирішення задач Data Stream Mining.

КЛЮЧОВІ СЛОВА: нечітка кластеризація, викривлені дані, достовірна нечітка кластеризація, Data Stream Mining, робастна функція.

ЛІТЕРАТУРА

1. Gan G. Data Clustering: Theory, Algorithms and Applications/ G. Gan, Ch. Ma, J. Wu. – Philadelphia, Pennsylvania: SIAM, 2007. – 455 p. DOI: https://doi.org/10.1137/1.9780898718348
2. Abony J. Cluster Analysis for Data Mining and System Identification / J. Abony, D. Feil. – Basel : Birkhouser, 2007. – 303 p.
3. Xu R. Clustering/ R. Xu, D. C. Wunsch. – Hoboken N. J. : John Wiley & Sons, Inc., 2009. – 398 p.
4. Bezdek J. C. Pattern recognition with fuzzy objective function algorithms / J. C. Bezdek – New York : Springer, 1981. – 253 p. DOI https://doi.org/10.1007/978-1-4757-0450-1.
5. Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition/ [F. Höppner, F. Klawonn, R. Kruse, T. Runkler]. – Chichester : John Wiley &Sons, 1999. – 300 p.
6. Credibilistic clustering: the model and algorithms / [J. Zhou, Q. Wang, C.-C. Hung, X. Yi] // International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. – 2015.

DOI 10.15588/1607-3274-2023-3-10

– Vol. 23, № 4. – P. 545–564. DOI: https://doi.org/10.1142/S0218488515500245
7. Zhou J. Credibilistic clustering algorithms via alternating cluster estimation / J. Zhou, Q. Wang, C. C. Hung // Journal of Intelligent Manufacturing. – 2017. – Vol. 28. – P.727–738. DOI: https://doi.org/10.1007/s10845-014-1004-6.
8. Sequential fuzzy cluster extraction and its robust against noise / [K. Tsuda, S. Senda, M. Minoh, K. Ikeda] // System and Computers in Japan. – 1997. – 28. – P. 10–17.
9. Höppner, F. Fuzzy clustering of sampled functions / F. Höppner, F. Klawonn // 19th Int. Conf. North American Fuzzy Information Processing Society (NAFIPS) – Atlanta, USA, 2000. – P. 257–255.
10. Georgieva O. A clustering algorithm for identification of single clusters in large data sets / O. Georgieva, F. Klawonn // Proc. 11th East – West Fuzzy Coll. – Zittau/Görlitz : FH, 2004. – P. 118–125.
11. Kohonen T. Self-Organizing Maps/ T. Kohonen. – Berlin: Springer, 1995. – 362 p. doi: 10.1007/978-3-642-56927-2.
12. Park D. C. Gradient based fuzzy c-means (GBFCM) algorithm / D. C. Park, I. Dagger // IEEE International Conference on Neural Networks, 28 June – 2July,1984: proceedings. – Orlando : IEEE, 1984. – P. 1626–1631. DOI: 10.1109 / ICNN. 1994.374399.
13. Bodyanskiy Ye. Computational intelligence techniques for data analysis / Ye. Bodyanskiy // Lecture Notes in Informatics. – Bonn : Gesellschaft für Informatik, 2005. – P. 15–36.
14. Online credibilistic fuzzy clustering of data using membership functions of special type [Electronic resource]/ [A. Shafronenko, Ye. Bodyanskiy, I. Klymova, O. Holovin] // Proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), April 27–1 May 2020. – Zaporizhzhia, 2020. – Access mode: http://ceur-ws.org/Vol-2608/paper56.pdf.
15. Online Credibilistic Fuzzy Clustering Method Based on Cauchy Density Distribution Function / [A. Shafronenko, Ye. Bodyanskiy, I. Pliss, I. Klymova] // 2021 11th International Conference on Advanced Computer Information Technologies (ACIT): proceedings. – Deggendorf, Germany: IEEE, 2021. – P.704–707. DOI: 10.1109/ACIT52158.2021.9548572
16. Robust recursive fuzzy clustering algorithms / Ye. Bodyanskiy, Ye. Gorshkov, I. Kokshenev, V. Kolodyazhniy // Proc. 12th East West Fuzzy Coll 2005 – Zittau-Grölitz, FH, 2005. – P. 301–308.
17. Outlier resistant recursive fuzzy clustering algorithms / [Ye. Bodyanskiy, Ye Gorshkov, I. Kokshenev, V. Kolodyazhniy] // Ed. By B. Reusch «Computational Intelligence Theory and Applications» – Advances in Soft Computing – Vol. 38. – Berlin Heidelberg, Springer Verlag, 2006. – P. 647–652.
18. Arrow K. J. Studies in Linear and Nonlinear Programming / K. J. Arrow, L. Hurwitz, H. Uzawa. – Stanford University Press, 1958. – 242 p.
19. Bodyanskiy Ye. Recurcive fuzzy clustering algorithm/ Ye. Bodyanskiy, V. Kolodyazhniy, A. Stephan // Proc 10th East West Fuzzy Coll. 2002, – Zittau-Görlitz, HS, 2002. – P. 276–283.

DOI 10.15588/1607-3274-2023-3-10

УДК 004.9

ТЕХНОЛОГІЯ СЕНТИМЕНТ-АНАЛІЗУ ВІДГУКІВ КОРИСТУАЧІВ СИСТЕМ Е-КОМЕРЦІЇ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Тчинецький С. А. – магістр кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

Поліщук Б. О. – магістр кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

Висоцька В. А. – канд. техн. наук, доцент, доцент кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

АНОТАЦІЯ

Актуальність. Взаємодія між компанією та цільовою аудиторією досліджується вже століттями. З самого початку комерційних відносин, стосунки надавача послуг та отримувача цінувалися чи не понад усе. Торгівля побудована на довірі та повазі. Імідж підприємця часто є важливішим ніж товар, який він продає. За багато сотень років, взаємини торговця і покупця, підприємця та клієнта не втратили важливості і в час масової диджиталізації якість відносин компанії та цільової аудиторії різного розміру та професійна підтримка зворотного зв'язку з клієнтами часто визначають успіх е-бізнесу. Для цього необхідні додаткові інструменти та інформаційні технології для допомоги бізнесменам слідкувати за можливостями розвитку е-бізнесу в певній локації, а також встановлювати зворотній зв'язок з користувачами за допомогою соціальних мереж та ЗМІ. Такі інструменти допоможуть суттєво розширити бачення ринкових можливостей для е-бізнесу, з'ясує – в які з них є сенс інвестувати, а на які не варто витрачати час. Також побачити, яка ідея має майбутнє і яку бізнес-модель потрібну реалізувати/підтримувати/розвивати для стрімкого розвитку територіального/ міжрегіонального е-бізнесу. Також допоможе розібратися, які важелі мають найбільший ефект для зміни політики бізнесу: що не чіпати, а що змінити, щоб забезпечити високу швидкість в реалізації задуму на основі аналізу відповідних результатів досліджень, наприклад, отримувати: прямий фідбек від клієнтів, динаміку зміни загальної задоволеності або зацікавленості цільової аудиторії та переваги/недоліки від користувачів за допомогою NLP-аналізу; підтримку розвитку е-бізнесу відносно локацій знаходження їхнього підприємства та найкращі напрями розвитку; – графіки розвитку бізнесу (покращення/погіршення) залежно від змісту коментарів.

Метою дослідження є розробка інформаційної технології підтримки розвитку е-бізнесу за допомогою аналізу локацій знаходження бізнесу, опрацювання фідбеку від користувачів, аналізу та класифікації відгуків клієнтів в режимі реального часу з соціальних мереж: Twitter, Reddit, Facebook та інші за допомогою методів глибокого навчання та Natural Language Processing українсько- та англомовних текстів.

Метод. Для аналізу відгуків користувачів та клієнтів використано NLP-методи. Серед методів реалізації основних функцій класифікації англомовних новин використані такі методи машинного навчання, як: наївний Баєсів класифікатор, логістична регресія та метод опорних векторів. Для класифікації україномовних відгуків від користувачів використано алгоритм Наївного Байєса, оскільки він добре показує себе на малих обсягах даних, простий у тренуванні та експлуатації та добре працює з текстовими даними. Наївний класифікатор Байєс є дуже хорошим варіантом для нашої системи і з розрахунку того, що кількість відгуків у датасеті є меншою порівняно з середніми показниками.

Результати. Розроблено модель машинного навчання для аналізу та класифікації україномовних та англомовних відгуків від користувачів систем е-комерції.

Висновки. Створена модель показує відмінні результати класифікації на тестових даних. Загальна точність сентиментальної моделі для аналізу україномовного контенту є доволі задовільною, 92.3%. Найкраще з завданням аналізу впливу англомовних новини на фінансовий ринок впорався метод логістичної регресії, який показав точність 75,67%. Безперечно, це не є бажаним результатом, проте це найбільший показник із усіх розглянутих. Дещо гірше зі завданням впорався метод опорних векторів (SVM), який показав точність 72,78%, що є дещо гіршим результатом за той, який було отримано завдяки методу логістичної регресії. І найгірше зі завданням впорався метод наївного баєсового класифікатора, який отримав точність 71,13%, що є меншою за отриману у двох попередніх методах.

КЛЮЧОВІ СЛОВА: NLP, text pre-processing, сентимент-аналіз, відгук, коментар, е-комерція, е-бізнес, машинне навчання, контент аналіз.

АБРЕВІАТУРА

БД – база даних;
ЗМІ – засоби масової інформації
ІС – інтелектуальна система;
ІТ – інформаційна технологія;
ШІ – штучний інтелект;
ПО – предметна область;
ML – machine learning;
NLP – natural language processing.

НОМЕНКЛАТУРА

S – система аналізу та класифікації відгуків;
I – множина вхідних даних;
O – множина вихідних даних;
R – основні правила опрацювання вхідних даних;
U – параметри опрацювання вхідних даних;
N – машинне навчання;
α – оператор скачування вхідних даних;
β – оператор опрацювання вхідних даних;
γ – оператор збереження вхідних даних;
μ –оператор видалення шуму в даних;
χ – оператор пошуку ключових слів;
ω – оператор машинного навчання ІС на достовірних текстових даних;

DOI 10.15588/1607-3274-2023-3-11

λ – оператор класифікації відгуків;
i_1 – множина даних ідентифікації;
i_2 – множина вхідного текстового контенту;
i_3 – множина шаблонів/правил NLP;
i_4 – множина відфільтрованих відгуків;
o_1 – маркований/тегований текст відгуків;
o_2 – колекція пропозицій аналізу відгуків;
o_3 – множина класифікованих відгуків;
r_1 – правила алгоритму взаємодії;
r_2 – NLP-правила;
r_3 – правила алгоритму машинного навчання;
r_4 – правила алгоритму класифікації відгуків;
r_{5j} – правила алгоритму j-тої NLP-задачі;
r_6 – правила алгоритму подання результатів;
u_1 – множина рівнів доступу;
u_2 – множина вимог доступу;
u_3 – множина NLP-вимог;
u_4 – множина метрик машинного навчання;
u_5 – множина вимог класифікації відгуків;
u_{6j} – множина вимог розв'язку j-тої NLP-задачі;
α_1 – оператор збору текстового тематичного контенту з Google за певний період часу;
α_2 – оператор збору текстового тематичного контенту з Twitter за певний період часу;
α_3 – оператор збору текстового тематичного контенту з Facebook за певний період часу;
α_4 – оператор збору текстового тематичного контенту з Reddit за певний період часу;
α_5 – оператор завантаження власних даних;
ϕ_1 – оператор пошуку за ключовими словами;
ϕ_2 – оператор сентимент аналізу відгуків;
ϕ_3 – оператор розрахунку рівня популярності запитів;
ϕ_4 – оператор узагальнення тексту;
ϕ_5 – оператор пошуку оптимальних локацій.

ВСТУП

Бізнес відіграє ключову роль в економіці кожної країни. Так в Україні малий та середній бізнес забезпечує близько 64% доданої вартості, 81,5% зайнятих працівників у суб'єктів господарювання та 37% податкових надходжень в 2021 році [1]. Із-за війни в України велика частина малого та середнього бізнесу біла або ліквідована (особливо на окупаційних територіях), або переїхала, або перейшла на чатової/повністю в сферу електронної торгівлі. Великою проблемою е-бізнесу є те, що вони не мають достатньої інформації про можливості розвитку у певних локаціях та не мають зворотного зв'язку з їхніми споживачами. Або ця інформація надходить із запізненням або неповна, або з надлишковим шумом. В умовах війни варто також говорити не тільки про розвиток е-бізнесу, а й про його відновлення, адже багато підприємств зупиняються або взагалі руйнуються у зв'язку з війною. В таких умовах необхідні додаткові інструменти та інформаційні технології для допомоги бізнесменам слідкувати за можливостями розвитку е-бізнесу в певній локації, а також встановлювати зворотній зв'язок з користувачами за допомогою соціальних мереж та ЗМІ. Такі інструменти допоможуть суттєво розширити бачення ринкових можливостей для е-бізнесу, з'ясує – в які з них є сенс інвестувати, а на які не варто витрачати час. І врешті решт, побачити, яка ідея має майбутнє і яку бізнес-модель потрібну реалізувати/підтримувати/розвивати для стрімкого розвитку територіального/міжрегіонального е-бізнесу. Також допоможе розібратися, які важелі мають найбільший ефект для зміни політики бізнесу: що не чіпати, а що змінити, щоб забезпечити високу швидкість в реалізації задуму на основі аналізу відповідних результатів досліджень, наприклад, отримувати:

– прямий фідбек від клієнтів, динаміку зміни загальної задоволеності або зацікавленості цільової аудиторії та переваги/недоліки від користувачів за допомогою NLP-аналізу.

– підтримку розвитку е-бізнесу відносно локацій знаходження їхнього підприємства та найкращі напрями розвитку.

– графіки розвитку бізнесу (покращення/погіршення) залежно від змісту коментарів.

Метою дослідження є розробка інформаційної технології аналізу україномовних та англомовних відгуків користувачів-клієнтів на сайтах е-комерції, дописів та новин в соцмережах та ЗМІ на основі методів опрацювання природної мови та технології машинного навчання для просування, адаптації та подальшого розвитку відповідного е-бізнесу.

Для досягнення поставленої мети необхідно вирішити такі завдання:

– дослідження та порівняння аналогів;

– порівняння та дослідження сучасних методів NLP як лематизація і стемінг, вилучення ключових слів, аналіз настроїв, узагальнення тексту, мішок слів та токенізація;

– розробити модель системи класифікації відгуків клієнтів та новин з достовірних джерел для ідентифікації емоційне забарвлення тексту українсько-англійською мовами на основі класифікатора Naive Bayes;

– здійснити експериментальну апробації розробленої системи сентимент аналізу інформаційного простору як зворотна реакція цільової аудиторії для підтримки е-бізнесу в Україні.

Об'єкт дослідження – процеси аналізу емоційного забарвлення текстового контенту відгуків цільової аудиторії на товари/послуги е-комерції.

Предмет дослідження – методи та засоби сентимент-аналізу англомовного та україномовного текстового контенту відгуків користувачів.

1 ПОСТАНОВКА ПРОБЛЕМИ

Необхідно розробити таку систему, яка покликана спростити спілкування клієнтів та компаній, особливо для тих компаній, які не можуть собі дозволити повноцінний центр підтримки. Особливість цієї системи полягатиме у використанні NLP-алгоритмів

DOI 10.15588/1607-3274-2023-3-11

для скорочення витрат на обслуговування клієнтів за рахунок скорочення кількості активних працівників в компанії. На заміну людській силі прийде алгоритм штучного інтелекту, який сам класифікуватиме відгуки та скарги клієнтів і визначатиме потрібні дії для них. Систему аналізу тональності інформаційного простору як зворотна реакція цільової аудиторії для підтримки та розвитку е-бізнесу подано кортежем:

$$S = < I, O, R, U, N, \alpha, \beta, \gamma >,$$

де $I = \{i_1, i_2, i_3, i_4\}$, $O = \{o_1, o_2, o_3\}$, $R = \{r_1, r_2, r_3, r_4\}$, $U = \{u_1, u_2, u_3, u_4, u_5\}$.

Основними процесами ІС є «Збір відгуків», «NLP відгуків», «Машине навчання» та «Класифікація відгуків». Процес збору відгуків із соціальних мереж опишемо суперпозицією:

$$C_{AU} = \mu \circ \beta \circ \alpha,\ C_{AU} = \mu(\beta(\alpha(i_1, i_2, i_4), r_1, u_1), u_2).$$

Процес «NLP відгуків» ІС граматичної корекції опишемо суперпозицією: $C_{CU} = \chi \circ \beta \circ \alpha$, тобто

$$C_{CU} = \chi(\beta(\alpha(C_{AU}, i_2, i_3, i_4), r_1, u_3), r_2).$$

Процес машинного навчання на достовірних даних ІС граматичної корекції опишемо суперпозицією:

$$C_{UL} = \omega \circ \gamma \circ \beta \circ \alpha,\ C_{UL} = \omega(\gamma(\beta(\alpha(C_{CU}, i_2), i_3), u_4), r_3).$$

Процес «Класифікація відгуків» ІС на основі машинного навчання опишемо суперпозицією:

$$C_{US} = \lambda \circ \gamma \circ \beta \circ \alpha,\ C_{US} = \lambda(\gamma(\beta(\alpha(C_{US}, i_2), i_4), u_5), r_4).$$

2 АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

Взаємодія між компанією та цільовою аудиторією досліджується вже століттями. З самого початку комерційних відносин, стосунки надавача послуг та отримувача цінувалися чи не понад усе. Торгівля побудована на довірі та повазі. Імідж підприємця часто є важливішим ніж товар, який він продає. За багато сотень років, взаємини торговця і покупця, підприємця та клієнта не втратили важливості і в час масової диджиталізації якість відносин компанії та цільової аудиторії різного розміру та професійна підтримка зворотного зв'язку з клієнтами часто визначають успіх е-бізнесу [1].

Взаємодія між компаніями та клієнтами – це складні стосунки, які дуже необхідно підтримувати в хороших тонах для компаній. Саме тому, що більше пів століття тому почали відкриватися центри підтримки клієнтів з арміями агентів, які допомагали покупцям. Проте час не стояв на місці, і вже зараз ці величезні центр не є ні корисними, ні вражаючими. Кожна компанія повинна тепер мати свій центр підтримки клієнтів. Проте, такі центри коштують дорого і, в часи стартапів і компаній, які з'являються і зникають однаково швидко, створювати домашній центр підтримки клієнтів з найманим персоналом невигідно. Зараз, в час глобальної диджиталізації та ще більшого прискорення руху життя, мати центри підтримки клієнтів, який оперує на базі агентів – невигідно. Адже швидкості бізнесу зростають, а з нею зростає й кількість нових клієнтів. Проте більше клієнтів – це не лише збільшення прибутку.

З іншого боку на сьогодення соціальні мережі займають велике, можливо навіть надто велике, місце в житті пересічної сучасної людини, потенційного клієнта конкретного е-бізнесу. Та швидкість, з якою новина може розлетітися по соціальних мережах – захоплює та лякає водночас. І саме в такому середовищі компаніям доводиться спілкуватися з клієнтами. Ціна поганого обслуговування клієнтів, в тому числі і підтримка, може бути надто велика. Саме тому важливо мати якісний, ефективний центр підтримки клієнтів. Саме центри підтримки клієнтів часто визначають ставлення загальної маси до компанії. Відношення компанії до цільової аудиторії збільшує не лише утримання е-бізнесу, а ще й слугує як безкоштовна реклама: якщо клієнту сподобалось товар/послуга та обслуговування – він скоріше порекомендує його бізнес іншим або залишить коментар/відгук в соціальній мережі.

Підтримка клієнтів – один з найважливіших аспектів багатьох підприємств та компаній. Проте, це не так вже і легко. Для ефективного центру підтримки клієнтів необхідно багато витрат – зарплата агентів, їхні робочі місця, інструктаж агентів. Це все – витрати. І для багатьох компаній ці витрати стають надто великими. Все більше компаній віддають перевагу фірмам-посередникам, які спеціалізуються на спілкуванні з цільовою авдиторією конкретного е-бізнесу. Це також вимагає певних витрат та часу на співпрацю та навчання персоналу під конкретний-е-бізнес. В сучасний час диджиталізації саме заміна таких call-центрів та фірм-посередників інструментом у вигляді інформаційної системи взаємодії з клієнтами та аналізу коментарів та новин на основі методів машинного навчання та NLP може стати успішним рішенням ведення бізнесу. NLP дозволяє застосовувати алгоритми машинного навчання для тексту та мови. Наприклад, ми можемо використовувати NLP, щоб створювати системи на кшталт розпізнавання мовлення, узагальнення документів, машинного перекладу, виявлення спаму, розпізнавання іменованих сутностей, відповіді питання, автодоповнення, предиктивного введення тексту тощо [4]. Завдяки новітнім і/або класичним алгоритмам, наприклад, тест Тюрінга [5], система може конкурувати з провідними компаніями на ринку аутсорсингу та, потенційно, змінити правила взаємодії з клієнтами. Тоді й невеликі компанії зможуть запросто утримувати лише кількох агентів, проте мати таку ж якість підтримки, як і гіганти їх індустрії з багатократними бюджетами, наприклад, на основі технології моделювання, синтезування та розпізнання мовлення [6]. Зараз також є дуже актуальною проблема вирішення задач NLP для

DOI 10.15588/1607-3274-2023-3-11

слов'янських мов, особливо української мови на фоні війни в Україні (наприклад, для ідентифікації фейків та пропаганди навіть актуально для е-бізнесу – приклад буде чи не буде війна в Тайвані змінює цінову політику на всі цифрові девайси), яка б дозволила слов'янським країнам якісно користуватись такими NLP рішеннями як: генерація тексту; аналіз настроїв; узагальнення тексту; та інші.

Аутсорсинг є стратегічним рішенням компанії для зниження витрат та підвищення ефективності бізнесу шляхом найму фізичної/юридичної особи для виконання відповідних завдань [7]. Аутсорсинг підтримки клієнтів досить поширена практика (наприклад, Sykes [8], Sensee [9], Serco [10], Teleperformance [11]), тому ринок аутсорсингових компаній, що спеціалізуються на спілкуванні з клієнтами досить обширний. В ньому знайдеться рішення практично для будь-якого е-бізнесу. Проте, якщо створити стартап як аналог виконання хоча б частини завдань відповідних аутсорсингових компаній, що буде більш заощадливим або ефективнішим, то це сильно підірве уже встановлений ринок. Проаналізувавши різні компанії та послуги, розроблено набір характеристик та критеріїв оцінювання для системи взаємодії з клієнтами:

– Доступ підтримки цілодобово – оцінюється наявність/відсутність підтримки цілодобового зв'язку;

– Швидкість зворотного зв'язку – скільки в середньому між усіма каналами годин потрібно для надання першої відповіді клієнту;

– Конфіденційність, ціна та кількість мов;

– Кількість агентів – значення кількості агентів не повинне бути надто високим та не надто низьким;

– Розташування та розміри офісу – розташування офісу повинне дозволяти охоплення якнайбільшу кількість клієнтів, розміри офісу повинні забезпечувати робоче місце для усіх агентів компанії;

– Кількість доступних каналів зв'язку;

– Можливості вхідного/вихідного зв'язку, телемаркетингу, активного збору відгуків;

– Глибина підтримки – на скільки агенти можуть допомогти клієнту тут і зараз.

Ще одним напрямом збору інформації та настроїв, яка впливає на розвиток е-бізнесу певного сектору, є платформи відстежування світових/регіональних медіа та друкованих ЗМІ, соціальних, онлайнових, цифрових і телерадіокомпаніях як Carma Media Monitoring, Repustate. Patient Voice [12–13], Siri [14], Grammarly [15], Klevu Smart Search [16] тощо. Зазвичай продукти, які використовують NLP в бізнесі є дуже зручними, але обмеженість в функціоналі не дає користувачам повністю покрити свої потреби. Тому в розроблюваному продукті потрібно залучити всі переваги аналогових продуктів, розширити функціонал продукту, який би покривав всі потреби клієнтів та основне виправив би недоліки продуктів-аналогів. Найкращим аналогом є Repustate, саме він має бути основним конкурентом, якого потрібно обійти. Даний продукт залучає велику кількість NLP методів, як і передбачається в розроблюваному продукті. Всі інші продукти розглянуті вище, зроблені за допомогою методів NLP та є лідерами у своїх сферах, тож маючи їх досвід можна залучити їхні підходи у якості розширення функціоналу для розроблюваного продукту, що зробить його лідером на ринку продуктів, які залучають NLP.

3 МАТЕРІАЛИ ТА МЕТОДИ

Для створення серйозного та процвітаючого е-бізнесу у будь-якій галузі із взаємодією з клієнтом необхідно приділяти час і увагу обслуговуванню цих клієнтів. Зрештою, команди служб обслуговування клієнтів щодня напряму взаємодіють із потенційно вашими клієнтами [2]. Це може принести як найбільше вигоди, так і найбільших збитків. Коли обслуговування клієнтів є пріоритетним, компанії отримують масу переваг: більше лояльних клієнтів, більше позитивних відгуків і більше доходу. Ось чому так важливо зосередитися на обслуговуванні клієнтів. Надання підтримки клієнтам може зайняти багато часу та енергії, тому традиційне обслуговування клієнтів часто розглядається як центр витрат. Керівники компаній знають, що їм потрібно надавати послуги, але вони бачать це як «витрати на ведення бізнесу». Проте спілкування з клієнтами може бути настільки ж прибутковим, як і розробка самого продукту. Обслуговування клієнтів – це не лише вартість ведення бізнесу. Це важлива частина загального досвіду клієнтів. Проте хороша підтримка клієнтів може привести до великих витрат що ніколи не є добре, особливо для менших компаній, чи таких, що лише починають комерційний шлях. Саме тому, все більше компаній [3] починають передавати проблеми з організацією та підтримки хорошого та ефективного сервісного центра іншим, аутсорсинговим компаніям чи стартапам. Отже є актуальним аналіз напрямів побудови інформаційної технології підтримки розвитку е-бізнесу України за допомогою аналізу локацій знаходження бізнесу, опрацювання фідбеку від користувачів, аналізу та класифікації відгуків клієнтів в режимі реального часу з соціальних мереж: Twitter, Reddit, Facebook та інші за допомогою методів глибокого навчання та NLP українсько- та англомовних текстів (рис. 1).

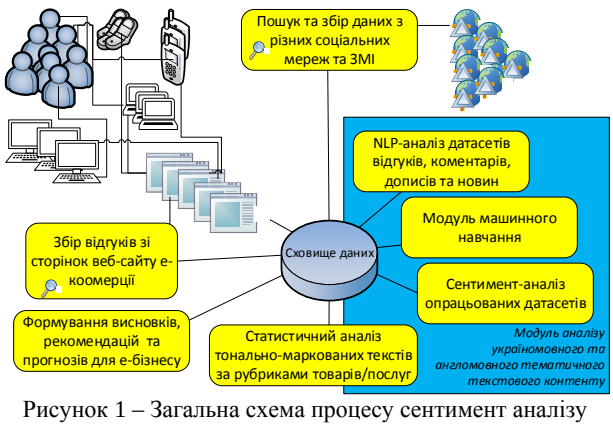


Рисунок 1 – Загальна схема процесу сентимент аналізу інформаційного простору

DOI 10.15588/1607-3274-2023-3-11

Тобто аналіз реакцій на товари-послуги через аналіз коментарів, відгуків на них на сайтах, в профілях соцмереж та паралельно новин в пабліках на подібні товари-послуги тощо. Вхідні дані – україномовний та англомовний контент з власних сайтів конкретного е-бізнесу, з профілів соцмереж постійних клієнтів та профілю самої фірми і паралельно з достовірних джерел ЗМІ, де можливі новини щодо цих подібних товарів-послуг, наприклад будівництва тощо. Необхідно розробити підхід для аналізу зворотної реакції цільової аудиторії для українського е-бізнесу, бо в сучасних умовах більше виживає на території України саме електронна торгівля. Тобто воюємо з часом – швидко та оперативно автоматично зібрати та проаналізувати реакцію цільової аудиторії для можливості скерувати бізнес. У час війни, постійного відключення в тому числі не за графіком світла, бізнес мусе адаптуватися швидко без використання стандартних для мирного часу інструментів та технік, в тому числі збору даних наприклад для прогнозування що буде актуальніше та краще реалізовуватися. Для якої саме аудиторії (вік, стать, регіон тощо). Це має бути технологія опрацювання вже зібраних даних з достовірних джерел для витягування певних реакцій (сентимент-аналіз, тональність відгуку позитивна, нейтральна чи негативна наприклад на товар). Коментарі зазвичай на сайтах українці пишуть або українською або російською. Другу не розглядали принципово. Українська не лише складна та багатогранна. Просто користувачі часто або не грамотні, або випадково пишуть з помилками, або застосовують суржик у залежності від регіону користувача в тому числі англійські слова вставляють англійською або транслітерацією. Та ще з помилками. А ось ті ж самі користувачі особливо молодь в соцмережах часто пишуть відгуки англійською. Тому і комбінація двома мовами та з різних джерел. Це подібно як бот збирає дані з достовірних джерел та фільтрує, а потім формує датасет (це детально в статі не описано, бо багато є подібних публікацій в тому числі авторів). В статті акцент надано лише на процес опрацювання датасетів двома мовами на основі NLP та машинного навчання. І акцент більше на сентимент-аналіз, що витягнути примітивну емоцію в тексті на конкретний або товар, або вид товару або послуги від наприклад конкретних користувачів. Щоб далі можна було аналізувати та прогнозувати на основі рекомендацій та зібраної статистики наприклад реакцію загальну на категорію товарів від певного класу цільової аудиторії.

NLP поєднує обчислювальну лінгвістику зі статистичними моделями, моделями машинного навчання та глибокого навчання. Разом ці технології дозволяють комп’ютерам опрацьовувати людську мову у вигляді тексту або голосових даних і «розуміти» її повне значення, враховуючи наміри та настрої мовця чи письменника [17]. NLP стало важливим бізнес-інструментом для виявлення прихованих даних із каналів соціальних мереж. Аналіз настроїв може аналізувати мову, що використовується в публікаціях у соціальних мережах, відповідях, оглядах тощо, щоб витягти ставлення та емоції у відповідь на продукти, рекламні акції та події – інформацію, яку компанії можуть використовувати в дизайні продуктів, рекламних кампаніях тощо. Також доречно NLP використати для класифікації відгуку клієнта. Єдина зовнішня дія яка потрібна для запуску роботи системи – це написання клієнтом відгуку. Цей відгук може бути написаний на будь-якій платформі: від соціальних мереж до Google Maps. Специфіка кількості та яких саме платформ узгоджує компанія, яка використовує систему. Після того, як клієнт написав свій відгук, система стягує цей відгук з визначеної платформи до власного сховища. Таким чином будується банк відгуків, які можна використати в подальших ітераціях моделі системи. Коли відгук стягнуто та записано до сховища, система проводить операцію класифікації відгуку. Це означає, що система визначає чи новий відгук позитивний чи негативний, перевіряє чи потрібна якась дія стосовно цього відгуку та яке слово з відгуку найточніше описує відгук загалом. Після успішної класифікації, залежно від результатів, система зберігає відгук в ще одне сховище для ведення архіву та передає інформацію далі до агентів, якщо це потрібно.

Оскільки ресурс планується бути онлайновим, для взаємодії з користувачем використовуватиметься доступні йому девайси. Коли користувач заходить на якусь з визначених платформ, він повинен натиснути відповідну кнопку, щоб залишити відгук. Після того, як користувач надіслав свій відгук, відгук автоматично стягується контролером системи (рис. 2).

Контролер передає цей відгук до Сховища, яке виконує зберігання сирого відгуку. Після того, як Сховище провело збереження, він надсилає статус відгука назад до Контролера для логування. Тоді, коли Контролер отримав зворотне повідомлення від Сховища, він надсилає відгук до Класифікатора. Класифікатор проводить класифікацію відгуку. Тоді, вже класифікований відгук надсилається назад Класифікатору, який, в свою чергу, надсилає інформацію про класифікований відгук Сховищу, щоб воно знову зберегло відгук, проте уже в опрацьованому вигляді. Після збереження відгуку, Сховище знову надсилає статус збереження Контролеру системи, де той продовжує потік, а саме, надсилає відгук Агенції. Агенція, залежно від того, що передбачив Класифікатор, або надсилає відгук далі до агентів, або закінчує шлях цього відгуку.

Система постійно моніторить доступні платформи на наявність нових відгуків. Цикл перевірки нових відгуків триває доти, поки не буде знайдений хоча б один новий відгук на будь-якій платформі. Якщо ж буде знайдено новий відгук, система вибивається з циклу та починає активну роботу.

Спочатку новий відгук зберігається у сховище. У сховище поступає будь-який відгук, який пройшов попередній етап, тому, можливо, що в сховищі можуть бути наявні однакові, або близькі за значенням та структурою відгуки. В будь-якому разі,

DOI 10.15588/1607-3274-2023-3-11

коли новий відгук поступає у систему і він записується в сховище – система передає новий відгук вниз по воронці та повертається до моніторингу нових відгуків. Завдяки цьому нові відгуки не будуть накопичуватись, що є важливим для швидкості опрацювання усіх відгуків. Після збереження та передачі відгука далі, йде найбільш затратний дія з усієї системи – класифікація. Тут відбуваються усі основні обрахунки системи, через що це є критична точка для ефективності системи. Важливо оптимізувати цю діяльність. Після класифікація, залежно від результатів, відгук або передається агентам для подальших дій, або відправляється в сховище для можливого подальшого використання, такого як аналіз, архівування, покращення та ітерація моделей класифікації. Якщо система вирішила, що відгук потребує дії, вона відправляє його агентам. Агенти повинні вирішити проблему, яку піднімає відгук так, як тільки це можливо.

Людська мова вражаюче складна і різноманітна. Люди висловлюються нескінченними способами, як усно, так і письмово. Існують не тільки сотні мов і діалектів, але й у кожній мові є унікальний набір граматичних і синтаксичних правил, термінів і сленгу. Коли люди пишуть, то часто роблять помилки, скорочують слова або пропускають розділові знаки. Також є регіональні акценти, бурмотіння, заїкання та запозичені терміни з інших мов, в тому числі в українській [18]. Усі бізнес дані містять багато корисної інформації, ідей, а NLP може швидко допомогти компаніям отримати їх. Інструменти NLP опрацьовують дані в режимі реального часу, 24/7, і застосовують одні й ті самі критерії до всіх даних, тому отримані результати є точними – і не містять невідповідностей. Після того, як інструменти NLP зможуть зрозуміти, про що йдеться в тексті, і навіть виміряти такі речі, як настрої, компанії можуть почати розставляти пріоритети та організовувати свої дані таким чином, щоб відповідати їхнім потребам [19]. Перед застосуванням методів машинного навчання будь-який текст англійською чи українською, або їх суміш має пройти попереднє опрацювання методами NLP, зокрема або частково в залежності від мети та типу задачі з врахуванням особливостей методу:

1) Тематичний аналіз – витягування сенсу із тексту шляхом визначення повторюваних тем [20];

– Тематичне моделювання може виводити шаблони та групувати подібні вирази без необхідності визначати теги теми або навчати дані заздалегідь;

– Класифікація тексту або виділення теми з тексту.

2) Аналіз настроїв – визначення того, чи є текст позитивним, негативним чи нейтральним на основі інших методів NLP і машинного навчання, щоб призначити зважені оцінки настроїв об’єктам, темам, темам і категоріям у реченні або фразі [21];

3) Виявлення намірів використовує машинне навчання та NLP, щоб пов’язувати слова/вирази з певним наміром. Наприклад, модель машинного навчання може дізнатися, що слова купити або придбати асоціюються з наміром придбати [22];

4) Вилучення ключових слів – техніка аналізу тексту, яка автоматично витягує з тексту найбільш вживані та найважливіші слова/вирази [23–24];

5) Лематизація – групування різних флективних форм слова для подальшого аналізу як єдиного елементу та на відмінну від стемінгу привносить контекст до слів, тобто зв’язує слова зі подібними значеннями в одне слово; використовують позиційні аргументи як вхідні дані, наприклад, чи є слово прикметником, іменником чи дієсловом [25–26];

6) Стемінг використовується для видалення суфіксів зі слів і в кінцевому підсумку отримати так звану основу слова, що дозволяє стандартизувати слова до їхньої основи незалежно від їх перегинів наприклад для кластеризації або класифікації тексту та пошуку [25–26];

7) Токенізація – це спосіб поділу фрагмента тексту на менші одиниці (токени) та використовують у традиційних методах NLP (Count Vectorizer), так і в архітектурах на основі розширеного глибинного навчання (Transformers); маркери можуть бути словами, символами або підсловами (n-грам) [27];

8) Машинний переклад – завдання автоматичного перетворення однієї природної мови в іншу, збереження значення введеного тексту та створення вільного тексту мовою виведення [28–29];

9) Узагальнення тексту – семантичне скорочення тексту, видаливши неважливий текст і перетворивши той самий текст у меншу семантичну текстову форму без видалення семантичної структури тексту [30]; визначення важливих фраз у документі та використання їх для виявлення відповідної інформації для додавання в резюме є критичною роботою для узагальнення на основі вилучення [31].

Для навчання моделей використовується дані, зібрані з відгуків у Google Maps на різного роду закладів: ресторани, готелі, кафе, магазини, тощо. В датасет входить відгук, записаний у вигляді стрічки, до якого класу відноситься цей відгук до класу позитивних відгуків, чи класу негативних відгуків, а також до якого класу відноситься цей відгук стосовно потреби у допомозі/діях. Загалом, у датасеті наявні три показники. Відгуки написані українською мовою, що суттєво ускладнює завдання. Також, загалом, в датасеті наявно приблизно 500 рядків даних. В рамках функцій цього процесу використаний словник української мови користувача GitHub DICT_uk, де зібрані більше мільйону українських слів, значення, приналежність до частин мови та більше [32]. Для класифікації опрацьованого тексту виберемо [33–34]:

– Naive Bayes Classifier – це група дуже простих алгоритмів класифікації, які базуються на теоремі Байєса; всі атрибути датасету є незалежні і що жодна з них не впливає на будь-яку іншу; є швидким та потребує мало даних для тренування, також, має хорошу тенденцію роботи з текстами при NLP;

– Підтримка векторних машин (SVM) – алгоритм, що використовується для класифікації та для

DOI 10.15588/1607-3274-2023-3-11

регресійних задач; розбиває дані в дві півплощини з найкращим можливим результатом, тобто, знаходить таку лінію на площині даних, що ділить ці дані на два класи; є швидкість тренування, висока точність та велика кількість можливих застосувань;

– Decision Tree – алгоритм розбиває датасет на малі підсети даних та вибудовує асоціативне дерево рішень для кожного з них; використовується для побудови моделі для передбачення цільових значень, де правила передбачення вибудовуються на основі попередніх даних; є простий та легкий в розумінні та реалізації з здатністю пояснення складних моделей за допомогою чітких візуалізацій, проте, легко піддається до оверфітингу та погано працює з нечисловими значеннями, також показує погані результати з малою кількістю даних.

Для розробки використаємо мову програмування Python [35–36] та його бібліотеки та фреймворки Flask [37], FastAPI [38] та NLTK [39], також для інтерфейсу буде застосовуватись javascript та його бібліотека React. Для ілюстраціїзміну даних на екрані, а не чекав повного опрацювання застосуємо message-broker Kafka [40]. Для створення кваліфікуючої частини системи класифікації відгуків, використаємо мову Python та середовище програмування Jupyter Notebook. Для реалізації алгоритму використаємо sklearn, а саме sklearn.naive_bayes.GausianNB. В рамках проекту також використовуються бібліотеки Python [35–36]: Numpy (робота моделей), Pandas (зберігання та трансформації даних), Re (маніпуляція з стрічками) та NLTK (з tokenize функція TreebankWordTokenizer для токеназації слів у реченнях) та Sklearn (машинне навчання).

4 ЕКСПЕРИМЕНТИ

Систему сентименту-аналізу відгуків користувачів подано кортежем:

$$S = < I, O, R, U, N, \alpha_i, \beta, \gamma, \mu, \omega, \chi, \phi_j, \lambda >,$$
$$i=[1,5], j=[1,5],$$

де $I = \{i_1, i_2, i_3, i_4\}$, $O = \{o_1, o_2, o_3\}$, $R = \{r_1, r_2, r_3, r_4, r_{5j}, r_6\}$, $U = \{u_1, u_2, u_3, u_4, u_5, u_{6j}\}$.

Основними процесами ІС є «Збір відгуків з різних джерел», «NLP відгуків», «Машине навчання» та «Класифікація відгуків за параметрами та ключовими словами». Додатковими процесами ІС є «Пошук за ключовими словами», «Сентимент аналіз», «Популярність запитів», «Узагальнення тексту», «Пошук оптимальних локацій», «Збір відгуків з Google», «Збір відгуків з Reddit», «Збір відгуків з Twitter», «Збір відгуків з Facebook», «Виведення отриманих результатів в вигляді звіту», «Виведення оттрманих результатів в вигляді графіку».

Процес формування колекції тематичних відгуків користувачів із різних джерел з попереднім опрацюванням (видалення дублів, інформаційного шуму, форматуванням за шаблоном) опишемо суперпозицією: $C_{AU} = \gamma^\circ\lambda^\circ\gamma^\circ\mu^\circ\beta^\circ\alpha_i$, $i=[1,5]$, тобто

$$C_{AU} = \gamma(\lambda(\gamma(\mu(\beta(\alpha_i(i_1, i_2, i_4), r_1, u_1), u_2)), r_4)).$$

Дані отримуються в режимі реального часу, тому вгадати з препроцисінгом майже нереально, але можна покращити отримані від користувачів дані, наприклад з соціальної мережі Twitter інтегруються так звані сирі дані із великим обсягом сміття, не потрібного для дослідження (багато юнікод символів). Для цього на основі пакету re треба очистити дані від юнікоду. Регулярний вираз є послідовністю символів для визначення шаблону пошуку в тексті, наприклад, для операцій типу "знайти" або "знайти і замінити" над рядками або для перевірки введених даних [50]. При очищенні даних, тобто видалення інформаційного шуму з тестовго контенту, для кожного отриманого посту замінимо юнікод символи за допомогою RegEx та патерну "[^\x00-\x7F]+".

Для покращеної роботи з даним текстом токенізуємо пости за допомогою wordtokenization або regexptokenization (regexptokenization працює краще, бо вилучає зайві розділові знаки). за допомогою лемматизації на англомовного текстового контенту та основі стемінгу для україномовного контенту.

Процес «NLP відгуків» ІС для інтегрованого тематичного контенту опишемо суперпозицією:

$$C_{CU} = \gamma^\circ\phi_j^\circ\chi^\circ\beta^\circ\alpha_i,\ i=[1,5],\ j=[1,5],\ \text{тобто}$$
$$C_{CU} = \gamma(\phi_j(\chi(\beta(\alpha_i(C_{AU}, i_2, i_3, i_4), r_1, u_3), r_2), r_{5j}, u_{6j})).$$

Методи NLP доцільно реалізувати за допомогою Python та відповідних бібліотек: Nltk (завантаження датасетів та імпорт класів Tokenizer), Re (regexp), SentimentIntensityAnalyzer (сентимент аналіз), WordNetLemmatizer (лемматизація речень на слова), PorterStemmer (стеммінг), Stopwords (словник стоп слів), Heapq.nlargest (визначає список n найбільших елементів в датасеті).

Процес машинного навчання та тренування на достовірних корпусах текстів для вдосконалення аналізу тематичних відгуків опишемо суперпозицією:

$$C_{UL} = \omega^\circ\gamma^\circ\phi_j^\circ\beta^\circ\alpha_i,\ i=[1,5],\ j=[1,5],\ \text{тобто}$$
$$C_{UL} = \omega(\gamma(\phi_j(\beta(\alpha_i(C_{CU}, i_2), i_3), r_{5j}, u_{6j}), u_4), r_3).$$

Процес «Виведення отриманих результатів» ІС на основі машинного навчання опишемо суперпозицією:

$$C_{US} = \upsilon^\circ\gamma^\circ\beta^\circ\alpha_i,\ C_{US} = \upsilon(\gamma(\beta(\alpha_i\,(C_{US}, i_2), i_4), u_5), r_6).$$

Опис розгорнутого сценарію прецеденту за стандартом RUP (рис. 2–3) [41–49]:

DOI 10.15588/1607-3274-2023-3-11

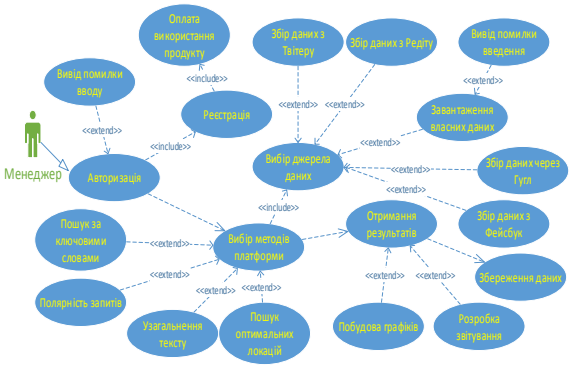


Рисунок 2 – Діаграма варіантів використання

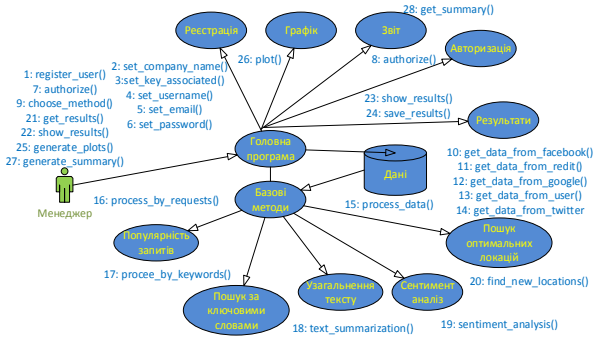


Рисунок 3 – Діаграма кооперацій

1) Менеджер отримує результат сентимент-аналізу множини відгуків відповідно до тематичного запиту на категорію товарів/послуг тощо;

2) Основний користувач системи сентимент-аналізу множини відгуків – менеджер е-комерції.

3) Передумови прецеденту (preconditions): підтримка збору тематичного контенту з Google та соціальних мереж за вимогою та їх опрацювання;

4) Основний успішний сценарій Менеджера: входить в систему → реєструється/авторизується → якщо вперше або закінчився термін, оплачує підписку → вибирає методи/ джерела → отримує результати;

5) Альтернативні потоки (рис. 3–4), коли Менеджер:

– не може ввійти в систему: повідомлення про помилку → повернення користувача на початок.

– задав некоректні дані: отримує повідомлення про помилку, що дані введені некоректно → повторне надсилання даних надсилають в систему.

– вибирає методи (рис. 3): Пошук за ключовими словами; Сентимент аналіз; Популярність запитів; Узагальнення тексту; Пошук оптимальних локацій.

– обирає джерела: Google; Reddit; Twitter; Facebook;

– вибирає розширені результати через Побудову графіків та Розробку звітування.

6) Postconditions: Менеджер отримав результати;

7) Спеціальні системні вимоги – це забезпечити надійність передачі даних, зручним інтерфейсом, цілодобову підтримку та швидке опарцювання запиту.

8) Список необхідних технологій: веб-платформа з підтримкою візуального відображення результатів.

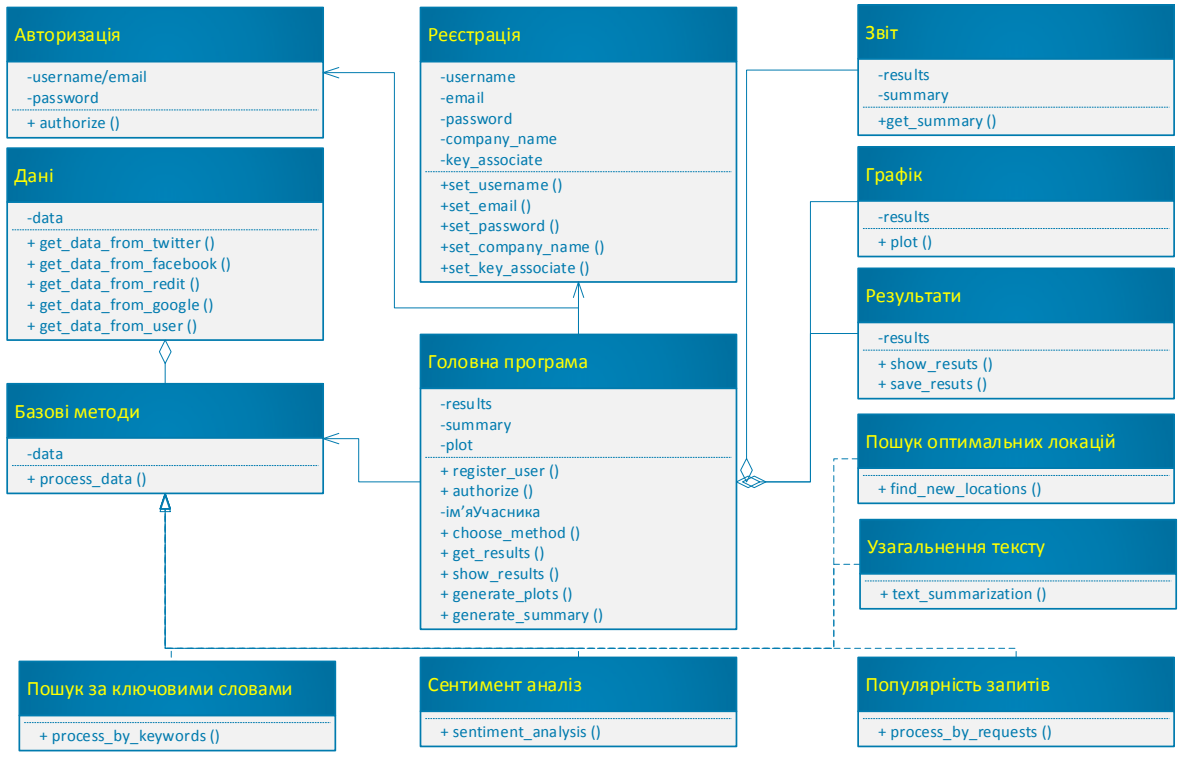


Рисунок 4 – Діаграма класів

DOI 10.15588/1607-3274-2023-3-11

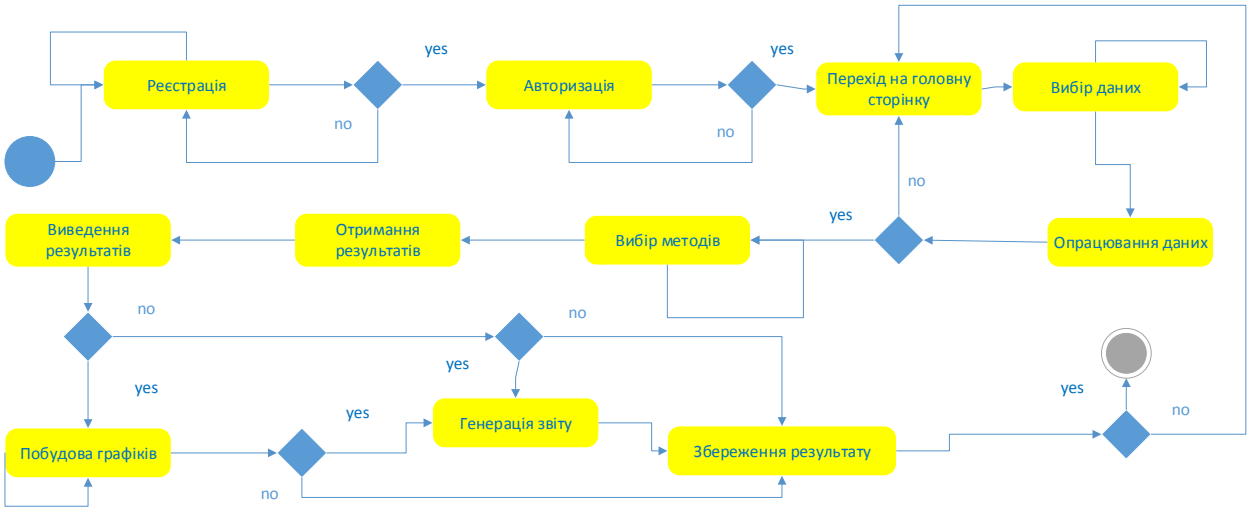


Рисунок 5 – Діаграма діяльності

На рис. 6 зображені компоненти, з яких складається розроблювана програма. Опишемо взаємодію між компонентами програми:

1. Main.py – даний компонент виконує роль керівника серед компонентів.

2. Authorization.py – компонент, який виконує роль авторизації, даний компонент ділиться на SignIn – логування в системі, та SignUp – реєстрація в системі;

3. DataGathering.py – компонент, який виконує роль збирання та опрацювання даних, включаючи:

– Get_data_from_twitter – дані з Twitter;
– Get_data_from_Reddit – дані з Reddit;
– Get_data_from_Google – дані з Google;
– Get_data_from_Facebook – дані з Facebook;
– Get_data_from_user – дані від користувача;

4. Methods.py – компонент, який містить різноманітні NLP методи та інші для аналізу даних:

– SentimentAnalysis – сентимент аналіз;
– Search_by_keywords – пошук за словом;
– Popularity_of_requests – популярність запиту;
– Text_summarization – узагальнення тексту;
– Look_for_new_locations – пошук нових локацій

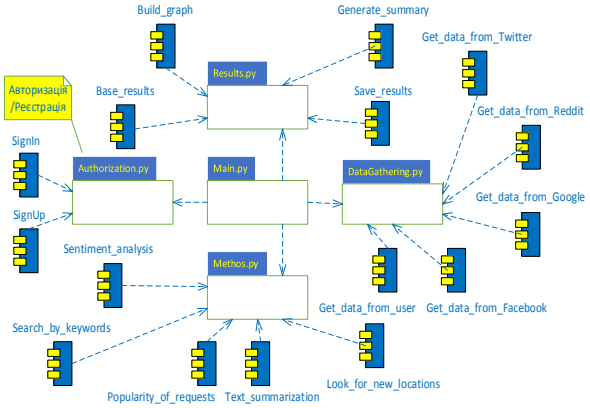


Рисунок 6 – Діаграма компонентів

5. Results.py – компонент, який відповідає за формування результатів, отриманих під час аналізу:

– Base_results – формує базові результати аналізу;
– Build_graph – побудова графіків результатів;
– Generate_summary – генерація результатів звіту.
– Save_results – збереження результатів.

5 РЕЗУЛЬТАТИ

При сентимент-аналізі користувач вводить ключ, згідно до якого він хоче отримати оцінку настроїв, наприклад назву компанії, товару, категорії товару тощо. Далі дані завантажуються з Twitter та опрацьовуються за допомогою regexp. Після чого ініціалізується об'єкт SentimentIntensityAnalyzer, також змінні для визначення кількості позитивних, негативних та нейтральних постів (рис. 7). Для кожного посту визначаємо оцінку настрою та за допомогою compound визначаємо до якої групи належить пост. Якщо ≤ -0.05, то негативний, а ≥ 0.05 – позитивний, інакше пост можна вважати нейтральним. Після чого формуємо відсотковий розподіл та відправляємо дані клієнту.

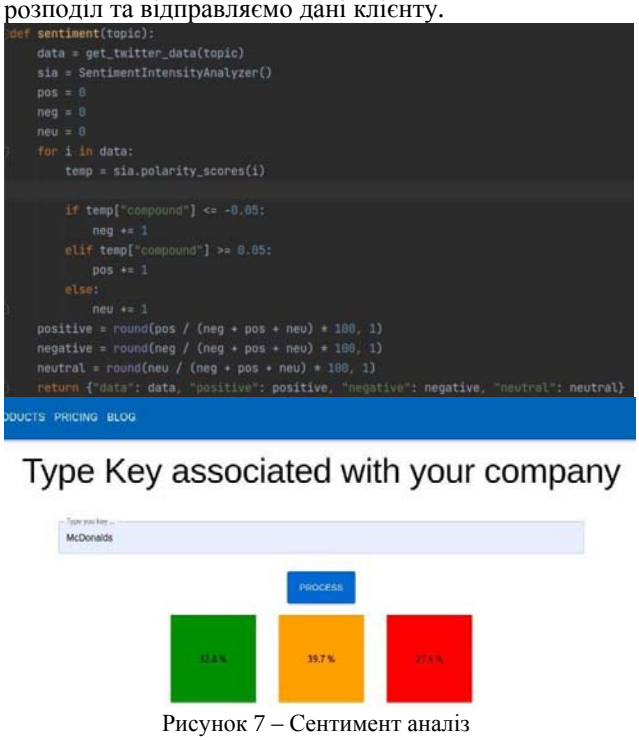


Рисунок 7 – Сентимент аналіз

DOI 10.15588/1607-3274-2023-3-11

При лематизації англомовного тексту застосовуємо RegexpTokenizer та WordNetLemmatizer (рис. 8). Для кожного посту реалізується токенізація, а для кожного токена застосовують Лемматайзер, після чого формуєтьмся множина результату лематизації.

Рисунок 8 – Лематизація

При стеммінгу ініціалізуються PorterStemmer, RegexpTokenizer та змінна results як масив (рис. 9). При узагальненні тексту постів їх об’єднують в однин масив, маркують стоп-слова датасету, ініціалізують та застосовюють RegexpTokenizer (рис. 10).

Рисунок 9 – Стеммінг

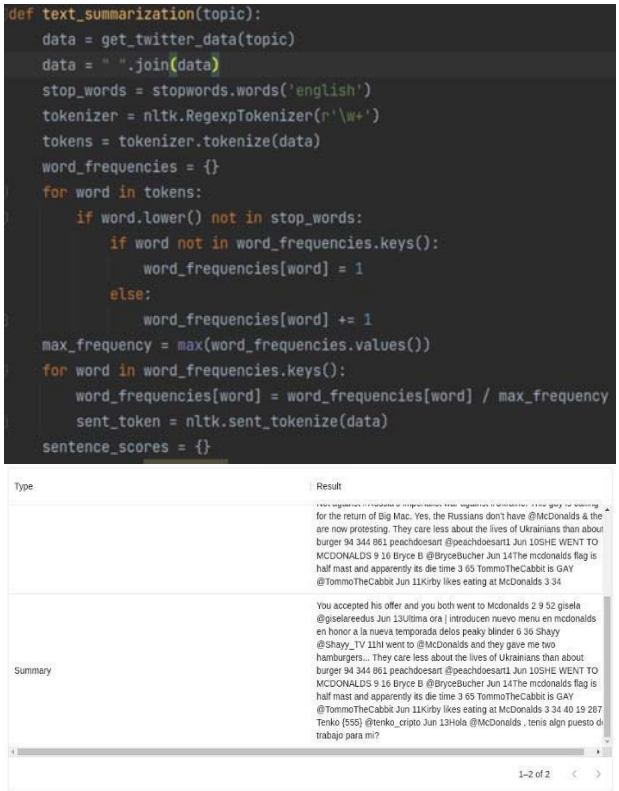


Рисунок 10 – Узагальнення тексту

Для всіх токенів, які не відносяться до стоп слів, створюємо частотний словник, а згодом нормалізуємо частоту на основі найбільшої знайденої частоти. Для кожного речення збираємо частоту появи слів в інших реченнях, після чого за допомогою nlargest алгоритму формуємо узагальнення та об’єднуємо в одне ціле.

При Pos Tagging слів ініціалізуємо RegexpTokenizer та змінну result як масив, застосовуємо алгоритм nltk.pos_tag (рис. 11).

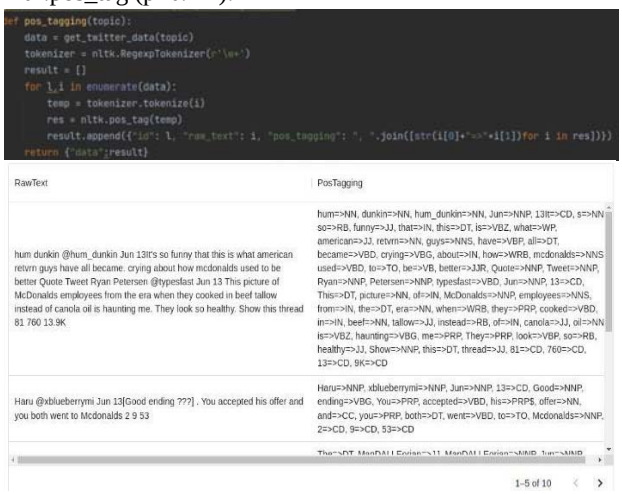


Рисунок 11 – Pos Tagging слів та Токенізація

DOI 10.15588/1607-3274-2023-3-11

При токенізації ініціалізуємо RegexpTokenizer та змінну result як масив (рис. 12).

```
def tokenization(topic):
    data = get_twitter_data(topic)
    tokenizer = nltk.RegexpTokenizer(r'\w+')
    result = []
    for l, i in enumerate(data):
        temp = tokenizer.tokenize(i)
        result.append({"id": l, "raw_text": i, "tokenization": ", ".join(temp)})
    return {"data": result}
```

RawText	Tokenization
hum dunkin @hum_dunkin Jun 13It's so funny that this is what american retvrn guys have all became. crying about how mcdonalds used to be better Quote Tweet Ryan Petersen @typesfast Jun 13 This picture of McDonalds employees from the era when they cooked in beef tallow instead of canola oil is haunting me. They look so healthy. Show this thread 81 761 13,9K	hum, dunkin, hum_dunkin, Jun, 13It, s, so, funny, that, this, is, what, american, retvrn, guys, have, all, became, crying, about, how, mcdonalds, used, to, be, better, Quote, Tweet, Ryan, Petersen, typesfast, Jun, 13, This picture, of, McDonalds, employees, from, the, era, when, they, cooked, in, beef, tallow, instead, of, canola, oil, is, haunting, me, They, look, so, healthy, Show, this, thread, 81, 761, 13, 9K
Haru @xblueberrymi Jun 13[Good ending ???] . You accepted his offer and you both went to Mcdonalds 2 9 53	Haru, xblueberrymi, Jun, 13, Good, ending, You, accepted, his, offer, and, you, both, went, to, Mcdonalds, 2, 9, 53
The ManDALLEorian @ManDALLEorian Jun 12Max Rebo working at McDonalds 15 132 1,013	The, ManDALLEorian, ManDALLEorian, Jun, 12Max, Rebo, working, at, McDonalds, 15, 132, 1, 013
Julia Davis @JuliaDavisNews Jun 12If only the Russians came out en masse to protest the war in Ukraine, but no, this is just the re-opening of McDonalds by another name. From 490 1,273 4,695	Julia, Davis, JuliaDavisNews, Jun, 12If, only, the, Russians, came, out, en, masse, to, protest, the, war, in, Ukraine, but, no, this, is, just, the, re, opening, of, McDonalds, by, another, name, From, 490, 1, 273, 4, 695
Inna Sovsun @InnaSovsun Jun 12The Russian protests. Do you know against what? Not against #Russia's imperialist war against #Ukraine. This	Inna, Sovsun, InnaSovsun, Jun, 12The, Russian, protests, Do, you, know, against, what, Not, against, Russia, s, imperialist, war, against, Ukraine,

1–5 of 10

Рисунок 12 – Реалізація та результат токенізації

Продемонструємо тепер роботу системи з україномовними текстами за допомогою бібліотеки Pandas (рис. 13). Датасет відгуків записаний у вигляді tsv-фалу (наявна пунктуація). Після завантаження тестових також потрібно створити масив стоп- слів (не несуть ніякого змісту або є надлишковим шумом).

```
data = pd.read_csv('data.tsv', delimiter = '\t', quoting =3)
slova = pd.read_csv('base.lst')
```

Рисунок 13 – Завантаження даних для навчання

Найпоширенішими службовими стоп-словами в українськомовних постах є ‘я’, ‘ти’, ‘там’, ‘де’, тощо (рис. 14). Також, стоп-словом є і ‘не’, проте виключимо це слово з масиву, оскільки воно достатньо сильно впливає на значення відгуку. Як частина класифікації система визначає найголовніше слово у відгуку на основі того, на скільки часто слово з’являється в українській мові.

```
stop_words = ['я',
 'мій',
 'та',
 'сам',
 'ми',
 'наш',
 'самі',
 'ти',
```

```
def most_import(review):
    d = {'words': review, 'freq': [0]*len(review)}
    df = pd.DataFrame(data = d)
    for word in review:

        if word not in set(stop_words) and word != 'не':
            for i in range(len(word)):

                for each in freq['word']:
                    if each.startswith(word[:i+1]) and len(each) <= len(word):
                        refr = each

                df.loc[df['words'] == refr, 'freq'] = freq['freq'][freq['word'] == refr].iloc[0]

    mst_word = df[df['freq'] == df['freq'].max()]['words'].iloc[0]

    return mst_word
```

Рисунок 14 – Масив стоп-слів української мови та визначення найголовнішого слова у відгуку

```
def Ukr_stem(review):
    stemmed = []

    for word in review:
        det = 0
        bord = 0
        word_len = len(word)
        found = False

        if word not in set(stop_words):

            if word_len <= 3 and not found:
                found = True
                stemmed.append(word)
            elif word_len == 4:
                for each in slova['word']:
                    if each == word:
                        found = True
                if found:
                    stemmed.append(word)
                else:
                    stemmed.append(word[:-1])

            else:

                root = word
                for i in range(len(word)):
                    for each in slova['word']:
                        if i != 0:
                            if each.startswith(word[:-i]) and i < (len(word)-3):
                                if len(word[:-i]) < len(root):
                                    root = word[:-i]

                stemmed.append(root)

    return stemmed
def ukr_stem2(review):
    stemmed = []
    review = [word for word in review if word not in stop_words]
    for word in review:
        root_len = len(word)-1 if word[-1:] in set(let_1) and len(word) > 2 else 0
        root_len = len(word)-2 if word[-2:] in set(let_2) and len(word) > 3 else root_len
        root_len = len(word)-3 if word[-3:] in set(let_3) and len(word) > 4 else root_len
        root_len = len(word)-4 if word[-4:] in set(let_4) and len(word) > 5 else root_len
        if root_len == 0:
            root = word
            for i in range(len(word)):
                for each in slova['word']:
                    if i != 0:
                        if each.startswith(word[:-i]) and i < (len(word)-3):
                            if len(word[:-i]) < len(root):
                                root = word[:-i]

            stemmed.append(root)
        else:
            root = word[:root_len]
            stemmed.append(root)
    return stemmed
```

Рисунок 15 – Функції стемінгу Ukr_stem та ukr_stem2

Функція перевіряє перші букви слів і знаходить таке слово, яке найближче підходить, ітеруючи стільки раз, скільки є букв у слові. З кожною ітерацією кількість перших букв збільшується і в кінці записується те слово зі словника, яке мало найбільшу кількість збігів. Далі, для оптимальної класифікації відгуків, необхідно їх підготувати, перед тим як навчати моделі на їх основі. Для цього потрібно провести ряд операцій: видалення пунктуації; переведення усіх літер в нижній регістр; токенізація; стемінг. Видалення та пониження регістру здійснюємо за допомогою імпортованої бібліотеки Re, яка призначена для роботи з регулярними висловами, та за допомогою функції lower() відповідно. Токенізацію здійснюємо за допомогою функції TreebankWordTokenizer. Стемінг – скорочення слів до найменшої можливої форми, коли зміст слова зберігається. Стемінг є ключовим у будь-яких NLP алгоритмах, оскільки грамотно проведене скорочення слів дозволяє оптимізувати роботу пізніших моделей. Розроблена функція стемінгу для української мови Ukr_stem (рис. 15).

Розроблена функція ukr_stem2 є другою ітерацією функції стемінгу українських слів. Функція Ukr_stem є повільною та неповорткою, але точною. Головна ідея Ukr_stem полягала у порівнянні кожного слова

DOI 10.15588/1607-3274-2023-3-11

відгуку з словами із словника, що займало багато часу. Функція ukr_stem2 є кращим варіантом для отримання швидко оперативних даних. Вона перевіряє закінчення слів та підбирає найкраще скорочення для нього. Для цього також створено масиви, що містять найпопулярніші закінчення слів в українській мові. За зразок взято дерево закінчень усіх можливих українських слів, розроблений на основі GNU Aspell Сеником Миколою (рис. 16) [51].

Рисунок 16 – Масиви закінчень слів української мови

```
def prepro(data):
    tokenizer = TreebankWordTokenizer()

    corpus = []
    most_import_word = []

    for rev in data:
        review = re.sub("[^а-яА-Я-іїґєЄ']", ' ', rev)
        review = review.lower()
        review = tokenizer.tokenize(review)
        most_import_word.append(most_import(review))
        review = ukr_stem2(review)
        review = ' '.join(review)
        corpus.append(review)
    df = pd.DataFrame(data = corpus)
    df['import_word'] = most_import_word
    return df
proc_reviews = prepro(data['review'])
proc_reviews.columns = ['review', 'import_word']
```

```
cv = CountVectorizer(max_features = 250)
x = cv.fit_transform(proc_reviews['review']).toarray()
y1 = data.iloc[:, 1].values
y2 = data.iloc[:, 2].values
```

Рисунок 17 – Функція prepro, створення Bag of Words

Як видно з рисунків, нова функція є набагато коротшою в плані коду, а також має набагато менше порівнянь, що скорочує час опрацювання відгуків, оскільки саме порівняння є найдовшою операцією в плані часу. Процес попереднього опрацювання відгуків подано функцією prepro (рис. 17).

Результати попереднього опрацювання текстів відгуків використаємо для навчання моделей та створюємо модель Bag of Words [52], на основі якої і навчатимуться моделі. У цій моделі текст (наприклад, речення або документ) подається як пакет (мультинабір) його слів, нехтуючи граматикою і навіть порядком слів, але зберігаючи множинність. Для класифікаторів наївного Байєса така модель підходить найкраще. Розбиваємо датасет на тренувальний та тестовий сети (рис. 18). Та проводимо навчання обох моделей. Тестувальний сет дорівнюватиме лише 2% від всього датасету для максимізації точності навчання моделей.

```
x_train, x_test, y_train1, y_test1 = train_test_split(x, y1, test_size = 0.02, random_state = 15)
x_train, x_test, y_train2, y_test2 = train_test_split(x, y2, test_size = 0.02, random_state = 15)
```

```
model1 = GaussianNB()
model1.fit(x_train, y_train1)

model2 = GaussianNB()
model2.fit(x_train, y_train2)
```

```
GaussianNB()
```

Рисунок 18 – Розподіл датасету, тренування моделей

Дослідимо точність роботи натренованих моделей на основі маркерів визначення сентименту (негативний/позитивний) відгука (рис. 19). Загальна точність сентиментальної моделі помилок для тестового сету є доволі задовільною (92,3%). Навчена модель добре класифікує позитивні відгуки, але має деякі проблеми з негативними (рис. 19а). Проблеми можуть бути через те, що негатив у відгуку людина, особливо українці, не передають на пряму, частіше нейтральними висловами, або сарказмом. Щодо другої моделі, то результати не настільки хороші при класифікації дії, з цим маємо проблеми (рис. 19б). Загальна точність моделі не є надто високою (61.5%). З матриці помилок випливає, що найбільше помиляється з тими відгуками, які не потребують дії.

```
cm1 = confusion_matrix(y_test1, predict1)
print(cm1)
accuracy_score(y_test1, predict1)*100
```

```
[[7 1]
 [0 5]]

92.3076923076923
```

а

```
cm2 = confusion_matrix(y_test2, predict2)
print(cm2)
(accuracy_score(y_test2, predict2))*100
```

```
[[5 4]
 [1 3]]

61.53846153846154
```

б

Рисунок 19 – Результат роботи сентимент- моделі та точність роботи сентимент-моделі

Проведемо тестування “живими” відгуками. Дамо класифікувати моделям щойно взятий відгук з інтернету та ще один, який написаний авторами. Обидва відгуки є позитивними і не потребують ніяких додаткових дій (рис. 20).

```
rev = ["Чудовий заклад! Приємний персонал, дуже смачна їжа, коктейлі і кальян) На 14 лютог
       "улюблений ресторан, кращого в світі не знайдеш"]
prot_test(rev)
```

```
Головне слово відгука: лютого;  Позитивне/негативне: 1; Потрібна допомога: 0
Головне слово відгука: світі;   Позитивне/негативне: 1; Потрібна допомога: 0
```

Рисунок 20 – Результат тестування «живих» відгуків

Також система визначила найголовніші слова відгуків. Для класифікації двох відгуків їй знадобилося 37 секунд, що означає приблизно 18.5 секунд на один відгук. Звісно, довжина відгуку сильно впливає на тривалість класифікації, оскільки найбільше часу все ще займає стемінг.

6 ОБГОВОРЕННЯ

Здійснено аналіз на основі методів машинного навчання (рис. 21): наївний Баєсів класифікатор (точність передбачення 71,13%), логістична регресія (точність передбачення 75,67%) та опорних векторів (точність передбачення 72,78%).

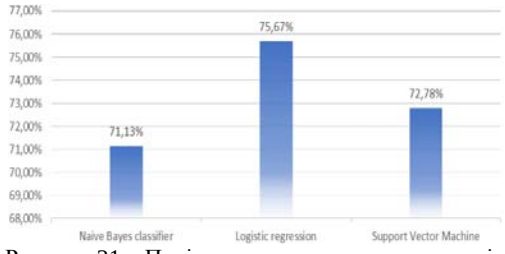


Рисунок 21 – Порівняння отриманих результатів

DOI 10.15588/1607-3274-2023-3-11

```
              precision    recall  f1-score   support

    negative       0.52      0.54      0.53       128
     neutral       0.76      0.87      0.81       575
    positive       0.66      0.46      0.54       267

    accuracy                           0.71       970
   macro avg       0.65      0.62      0.63       970
weighted avg       0.70      0.71      0.70       970

71.1340206185567
              precision    recall  f1-score   support

    negative       0.70      0.52      0.59       128
     neutral       0.78      0.88      0.83       575
    positive       0.70      0.60      0.65       267

    accuracy                           0.76       970
   macro avg       0.73      0.67      0.69       970
weighted avg       0.75      0.76      0.75       970

75.6701030927835
              precision    recall  f1-score   support

    negative       0.80      0.26      0.39       128
     neutral       0.72      0.97      0.82       575
    positive       0.77      0.42      0.55       267

    accuracy                           0.73       970
   macro avg       0.76      0.55      0.59       970
weighted avg       0.74      0.73      0.69       970

72.78350515463917
```

Рисунок 22 – Аналіз методом класифікатора Баєса, логістичної регресії та опорних векторів

На рис. 22 подані classifiaction report модулів для вимірювання якості прогнозів за алгоритмом класифікації (порівняння правдивих та хибних передбачень).Відповідно для прогнозування показників в classifiaction report використовують істинно позитивні TP, хибно позитивні FP, істинно негативні TN та хибно негативні FN показники передбачень [53–55]. Зокрема, коли випадок (подія) є при TN – негативним і прогнозовано негативним; TP – позитивним і прогнозовано позитивним; FN – позитивним, але передбачався негативним; FP – негативним, але передбачався позитивним.

ВИСНОВКИ

Описано застосування сентимент аналізу коментарів, відгуків, запитів та новин для підтримки та розвитку е-бізнесу. Проаналізовані аналоги дали можливість розробити інформаційну технологію для розв’язку NLP-задач е-бізнесу, адаптовану для української цільової аудиторії. Розроблена загальна типова структура інформаційної системи підтримки та розвитку е-комерції за рахунок аналізу зворотної реакції цільової аудиторії на основі технології машинного навчання та методів опрацювання природної мови. Серед методів реалізації основних функцій використані такі методи машинного навчання, як: наївний Баєсів класифікатор, логістична регресія та метод опорних векторів. Здійснено розробку програмного забезпечення та описано його структуру. Здійснено огляд звітів виконання методів машинного навчання. Це дало змогу краще переглянути та проаналізувати отримані результати. Опісля чого здійснено статистику виконання програми, описано її та проаналізовано отримані результати. А саме побудовано графік порівняння отриманих результатів. Також у ході роботи створена презентація про розроблений проект і написана стаття, у якій двома мовами, а саме українською та англійською, описано процес роботи над проектом. Найкраще з завданням аналізу впливу новини на фінансовий ринок впорався метод логістичної регресії, який показав точність 75,67%. Безперечно, це не є бажаним результатом, проте це найбільший показник із усіх розглянутих. Дещо гірше зі завданням впорався метод опорних векторів (SVM), який показав точність 72,78%, що є дещо гіршим результатом за той, який було отримано завдяки методу логістичної регресії. І найгірше зі завданням впорався метод наївного Баєсового класифікатора, який отримав точність 71,13%, що є меншою за отриману у двох попередніх методах. Звісно ж, що отримані результати далекі від ідеалу і демонструють точність у проміжку від 71 % до 76 %. Що означає те, що вони потребують удосконалення. На кінець хотілось би зазначити, що дана тема є неабияк популярною та актуальною, а аналогів на даний момент не існує.

ЛІТЕРАТУРА

1. Kuzminov M. Modern Development of Small Business in Ukraine / M. Kuzminov // Sciences of Europe. – 2022. – Vol. 107. – P. 29–31. DOI: 10.5281/zenodo.7479719
2. Definition of customer support. – Access mode: https://www.helpscout.com/helpu/definition-of-customer-support.
3. Edvardsson I. R. Strategic outsourcing in SMEs / I. R. Edvardsson, S. Durst, G. K. Oskarsson // Journal of small business and enterprise development. – 2020. – Vol. 27(1). – P. 73–84. DOI: 10.1108/JSBED-09-2019-0322
4. Sarkar D. Text analytics with Python: a practitioner’s guide to natural language processing / D. Sarkar. – Bangalore : Apress, 2019. – 674 p. DOI: 10.1007/978-1-4842-4354-1
5. Eisenstein J. Introduction to natural language processing / J. Eisenstein. – Cambridge : MIT press, 2019. – 536 p.
6. Goldberg Y. A Primer on Neural Network Models for Natural Language Processing / Y. Goldberg. – Access mode: https://jair.org/index.php/jair/article/view/11030/26198
7. Britannica dictionary. – Access mode: https://www.britannica.com/topic/outsourcing
8. Sykes. – Access mode: https://www.sykes.com
9. Sensee. – Access mode: https://www.sensee.co.uk/index.html
10. Serco. – Access mode: https://www.serco.com
11. Teleperformance. – Access mode: https://www.teleperformance.com/en-us
12. Repustate. Using NLP for business success. – Access mode: https://www.repustate.com/blog/using-nlp-for-business-success/
13. Repustate. How can sentiment analysis help you with Patient Voice? – Access mode: https://www.repustate.com/patient-voice/
14. SkywellSoftware. How does Siri work: technology and algorithm. – Access mode: https://skywell.software/blog/how-does-siri-worktechnology-and-algorithm/
15. Grammarly. How Grammarly uses Natural Language Processing and Machine Learning to identify the main points in a message. – Access mode: https://www.grammarly.com/blog/engineering/nlp-mlidentify-main-points/

DOI 10.15588/1607-3274-2023-3-11

16. Klevu. Smart Search Overview. – Access mode: https://www.klevu.com/smart-search/
17. IBM. Natural Language Processing (NLP). What is natural language processing? – Access mode: https://www.ibm.com/cloud/learn/natural-language-processing#tocwhat-is-na-jLju4DjE
18. SaS. Natural Language Processing (NLP). What it is and why it matters. – Access mode: https://www.sas.com/en_us/insights/analytics/what-is-natural-languageprocessing-nlp.html
19. MonkeyLearn, What is NLP. – Access mode: https://monkeylearn.com/blog/what-is-natural-language-processing/
20. Topic Analysis: The Ultimate Guide. – Access mode: https://monkeylearn.com/topic-analysis/
21. Sentiment Analysis Explained. – Access mode: https://www.lexalytics.com/technology/sentiment-analysis/
22. MonkeyLearn. Intent Classification: How to Identify What Customers Want. – Access mode: https://monkeylearn.com/blog/intentclassification/
23. MonkeyLearn. Keyword Extraction. – Access mode: https://monkeylearn.com/keyword-extraction/
24. Edia. What is Keyword Extraction? – Access mode: https://www.edia.nl/keyword-extraction
25. Stemming vs. Lemmatization in NLP. – Access mode: https://towardsdatascience.com/stemming-vslemmatization-in-nlp-dea008600a0
26. Analytics steps. What is Stemming and Lemmatization in NLP? – Access mode: https://www.analyticssteps.com/blogs/what-stemming-andlemmatization-nlp
27. What is Tokenization in NLP?. – Access mode: https://www.analyticsvidhya.com/blog/2020/05/what-is-tokenizationnlp/
28. Stanford. Machine Translation. – Access mode: https://nlp.stanford.edu/projects/mt.shtml
29. Data Science UA. Machine Translation. – Access mode: https://data-science-ua.com/wiki/natural-language-processingnlp/machine-translation/
30. Text Summarization in NLP. – Access mode: https://www.topcoder.com/thrive/articles/text-summarization-in-nlp
31. What Is Text Summarization in NLP? – Access mode: https://www.analyticssteps.com/blogs/what-text-summarization-nlp
32. Dict_uk Github repository. – Access mode: https://github.com/brown-uk/dict_uk/tree/master/data
33. Advantages and disadvantages of different classification models. Access mode: https://www.geeksforgeeks.org/advantages-and-disadvantages-of-different-cl assification-models/
34. Naive Bayes Classifier. – Access mode: https://www.upgrad.com/blog/naive-bayes-classifier/
35. Coursera. What Is Python Used For?. – Access mode: https://www.coursera.org/articles/what-is-python-used-fora-beginners-guide-to-using-python
36. Python.org. Executive Summary. – Access mode: https://www.python.org/doc/essays/blurb/
37. PymBook. Introduction to Flask. – Access mode: https://pymbook.readthedocs.io/en/latest/flask.html
38. FastApi. – Access mode: https://fastapi.tiangolo.com/
39. NLTK. Natural Language Toolkit. – Access mode: https://www.nltk.org/
40. AWS. What is Apache Kafka? – Access mode: https://aws.amazon.com/ru/msk/what-is-kafka/
41. Tutorialspoint. System Analysis and Design – Overview. – Access mode: https://www.tutorialspoint.com/system_analysis_and_design/system_ana lysis_and_design_overview.htm
42. Lonnie D. Bentley. System Analysis and Design for the Global Enterprise / D. Lonnie. – Columbus : McGraw-Hill Education Ltd, 2007. – 747 p.
43. WeyBackMachine. System Analysis. – Access mode: https://web.archive.org/web/20070822025602/http://pespmc1.vub.ac.be/ASC/SYSTEM_ANALY.html
44. Ritchey T. Analysis and synthesis: on scientific method-based on a study by Bernhard Riemann / T. Ritchey // Systems research. – 1991. – Vol. 8(4). – P. 21–41. DOI: 10.1002/sres.3850080402
45. Booch G. Unified Modeling Language User Guide / G. Booch, J. Rumbaugh, I. Jacobson. – Boston : AddisonWesley. 2005. – 391 p.
46. Iso.org. ISO/IEC 19501:2005 – Information technology – Open Distributed Processing – Unified Modeling Language (UML) Version 1.4.2. – Access mode: https://www.iso.org/standard/32620.html
47. Iso.org. ISO/IEC 19505–1:2012 – Information technology – Object Management Group Unified Modeling Language (OMG UML) – Part 1: Infrastructure. – Access mode: https://www.iso.org/standard/32624.html
48. WeyBackMachine. Basic UML. – Access mode: https://web.archive.org/web/20121214050605/http://ooad.asf.ru/Files/U ML.djvu.zip
49. Chen F. Linguistic tone and non-linguistic pitch imitation in children with autism spectrum disorders: A cross-linguistic investigation / F. Chen, C. C. H. Cheung, G. Peng // Journal of Autism and Developmental Disorders. – 2022. – Vol. 52(5). – P. 2325–2343. DOI: 10.1007/s10803-021-05123-4
50. Robertson A. Emoji skin tone modifiers: Analyzing variation in usage on social media / A. Robertson, W. Magdy, S. Goldwater // ACM Transactions on Social Computing. – 2020. – Vol. 3(2). – P. 1–25. DOI: 10.1145/3377479
51. Tree of endings in Ukrainian language. – Access mode: http://www.senyk.poltava.ua/projects/ukr_stemming/ukr_endings.html
52. Ishihara S. Score-based likelihood ratios for linguistic text evidence with a bag-of-words model / S. Ishihara // Forensic Science International. – 2021. – Vol. 327. – P. 110980. DOI: 10.1016/j.forsciint.2021.110980
53. Understanding the Classification report through sklearn. – Access mode: https://muthu.co/understanding-the-classification-report-in-sklearn/
54. Emotion recognition system project of English newspapers to regional E-business adaptation / [O. Markiv, V. Vysotska, L. Chyrun et al.] // Computer science and information technologies : IEEE 17th International conference, Lviv, Ukraine, 10–12 November 2022 : proceedings. – Lviv: IEEE, 2022. – P. 392–397. https://doi.org/10.1109/CSIT56902.2022.10000527
55. Sentiment analysis technology of English newspapers quotes based on neural network as public opinion influences identification tool / [V. Vysotska, O. Markiv, S. Voloshyn et al.] // Computer science and information technologies : IEEE 17th International conference, Lviv, Ukraine, 10–12 November 2022 : proceedings. – Lviv: IEEE, 2022. – P. 83–88. https://doi.org/10.1109/CSIT56902.2022.10000627

Стаття надійшла до редакції 05.05.2023.
Після доробки 15.08.2023.

DOI 10.15588/1607-3274-2023-3-11

UDC 004.9

SENTIMENT ANALYSIS TECHNOLOGY FOR USER FEEDBACK SUPPORT IN E-COMMERCE SYSTEMS BASED ON MACHINE LEARNING

Tchynetskyi S. – PhD student of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

Polishchuk B. – PhD student of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

Vysotska V. – PhD, Associate Professor of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

ABSTRACT

Context. The interaction between a company and its target audience has been studied for centuries. From the very beginning of commercial relations, the relationship between the service provider and the recipient has been valued almost above all else. Trade is built on trust and respect. The image of an entrepreneur is often more important than the product he sells. For hundreds of years, the relationship between the merchant and the buyer, the entrepreneur and the client has not lost its importance, and in the era of mass digitalization, the quality of the relationship between the company and the target audience of different sizes and professional feedback support with clients often start the success of e-business. To provide these additional tools and information technologies to help businessmen monitor e-business development opportunities in a specific location, as well as establish feedback with users through social networks and mass media. Obtaining such tools will significantly expand the vision of market opportunities for e-business, it will clarify which of them make sense to invest in, and which ones are not worth paying time for. Also see what idea has the future and what business model needs to be implemented/maintained/developed for the rapid development of territorial/interregional e-business. It will also help to understand which levers have the greatest effect for business changes: what not to touch, and what policies to change to ensure high speed in the implementation of the plan based on the analysis of relevant research results, for example, to receive: direct feedback from customers, the dynamics of changes in overall satisfaction or interest of the target audience and advantages/disadvantages from users using NLP analysis; support for the development of e-business in relation to the location of their enterprise and the best directions; – graphs of business development (improvement/deterioration) depending on the content of comments.

Objective of the study is to develop information technology to support the development of e-business by analyzing business locations, processing feedback from users, analyzing and classifying customer feedback in real time from social networks: Twitter, Reddit, Facebook and others using deep learning and Natural methods. Language Processing of Ukrainian-speaking and English-speaking texts.

Method. NLP-methods were used to analyze the opinions of users and customers. Among the methods of implementing the main functions of English-language news classification, the following machine learning methods are used: naive Bayesian classifier, logistic regression, and the method of support vectors. The Naive Bayes algorithm was used to classify Ukrainian-language user feedback, as it performs well on small amounts of data, is easy to train and operate, and works well with text data. Naive Bayes classifier is a very good option for our system and considering that the number of responses in the dataset is smaller compared to the averages.

Results. A machine learning model was developed for the analysis and classification of Ukrainian- and English-language reviews from users of e-commerce systems.

Conclusions. The created model shows excellent classification results on test data. The overall accuracy of the sentimental model for the analysis of Ukrainian-language content is quite satisfactory, 92.3%. The logistic regression method coped best with the task of analyzing the impact of English-language news on the financial market, which showed an accuracy of 75.67%. This is certainly not the desired result, but it is the largest indicator of all considered. The support vector method (SVM) coped somewhat worse with the task, which showed an accuracy of 72.78%, which is a slightly worse result than the one obtained thanks to the logistic regression method. And the naïve Bayesian classifier method did the worst with the task, which achieved an accuracy of 71.13%, which is less than the two previous methods.

KEYWORDS: NLP, text pre-processing, sentiment analysis, feedback, comment, e-commerce, e-business, machine learning, content analysis.

REFERENCES

1. Kuzminov M. Modern Development of Small Business in Ukraine, *Sciences of Europe*, 2022, Vol. 107, pp. 29–31. DOI: 10.5281/zenodo.7479719
2. Definition of customer support. Access mode: https://www.helpscout.com/helpu/definition-of-customer-support.
3. Edvardsson I. R., Durst S., Oskarsson G. K. Strategic outsourcing in SMEs, *Journal of small business and enterprise development*, 2020, Vol. 27(1), pp. 73–84. DOI: 10.1108/JSBED-09-2019-0322
4. Sarkar D. Text analytics with Python: a practitioner's guide to natural language processing. Bangalore, Apress, 2019, 674 p. DOI: 10.1007/978-1-4842-4354-1
5. Eisenstein J. Introduction to natural language processing. Cambridge, MIT press, 2019, 536 p.
6. Goldberg Y. A Primer on Neural Network Models for Natural Language Processing. Access mode: https://jair.org/index.php/jair/article/view/11030/26198
7. Britannica dictionary. Access mode: https://www.britannica.com/topic/outsourcing
8. Sykes. Access mode: https://www.sykes.com
9. Sensee. Access mode: https://www.sensee.co.uk/index.html
10. Serco. Access mode: https://www.serco.com
11. Teleperformance. Access mode: https://www.teleperformance.com/en-us
12. Repustate. Using NLP for business success. Access mode: https://www.repustate.com/blog/using-nlp-for-business-success/
13. Repustate. How can sentiment analysis help you with Patient Voice? Access mode: https://www.repustate.com/patient-voice/
14. SkywellSoftware. How does Siri work: technology and algorithm. Access mode: https://skywell.software/blog/how-does-siri-worktechnology-and-algorithm/
15. Grammarly. How Grammarly uses Natural Language Processing and Machine Learning to identify the main points in a message. Access mode:

DOI 10.15588/1607-3274-2023-3-11

https://www.grammarly.com/blog/engineering/nlp-mlidentify-main-points/
16. Klevu. Smart Search Overview. Access mode: https://www.klevu.com/smart-search/
17. IBM. Natural Language Processing (NLP). What is natural language processing? Access mode: https://www.ibm.com/cloud/learn/natural-language-processing#tocwhat-is-na-jLju4DjE
18. SaS. Natural Language Processing (NLP). What it is and why it matters. Access mode: https://www.sas.com/en_us/insights/analytics/what-is-natural-languageprocessing-nlp.html
19. MonkeyLearn, What is NLP. Access mode: https://monkeylearn.com/blog/what-is-natural-language-processing/
20. Topic Analysis: The Ultimate Guide. Access mode: https://monkeylearn.com/topic-analysis/
21. Sentiment Analysis Explained. Access mode: https://www.lexalytics.com/technology/sentiment-analysis/
22. MonkeyLearn. Intent Classification: How to Identify What Customers Want. Access mode: https://monkeylearn.com/blog/intentclassification/
23. MonkeyLearn. Keyword Extraction. Access mode: https://monkeylearn.com/keyword-extraction/
24. Edia. What is Keyword Extraction? Access mode: https://www.edia.nl/keyword-extraction
25. Stemming vs. Lemmatization in NLP. Access mode: https://towardsdatascience.com/stemming-vslemmatization-in-nlp-dea008600a0
26. Analytics steps. What is Stemming and Lemmatization in NLP?. Access mode: https://www.analyticssteps.com/blogs/what-stemming-andlemmatization-nlp
27. What is Tokenization in NLP? Access mode: https://www.analyticsvidhya.com/blog/2020/05/what-is-tokenizationnlp/
28. Stanford. Machine Translation. Access mode: https://nlp.stanford.edu/projects/mt.shtml
29. Data Science UA. Machine Translation. Access mode: https://data-science-ua.com/wiki/natural-language-processingnlp/machine-translation/
30. Text Summarization in NLP. Access mode: https://www.topcoder.com/thrive/articles/text-summarization-in-nlp
31. What Is Text Summarization in NLP? Access mode: https://www.analyticssteps.com/blogs/what-text-summarization-nlp
32. Dict_uk Github repository. Access mode: https://github.com/brown-uk/dict_uk/tree/master/data
33. Advantages and disadvantages of different classification models. Access mode: https://www.geeksforgeeks.org/advantages-and-disadvantages-of-different-cl assification-models/
34. Naive Bayes Classifier. Access mode: https://www.upgrad.com/blog/naive-bayes-classifier/
35. Coursera. What Is Python Used For? Access mode: https://www.coursera.org/articles/what-is-python-used-fora-beginners-guide-to-using-python
36. Python.org. Executive Summary. Access mode: https://www.python.org/doc/essays/blurb/
37. PymBook. Introduction to Flask. Access mode: https://pymbook.readthedocs.io/en/latest/flask.html
38. FastApi. – Access mode: https://fastapi.tiangolo.com/
39. NLTK. Natural Language Toolkit. Access mode: https://www.nltk.org/
40. AWS. What is Apache Kafka? Access mode: https://aws.amazon.com/ru/msk/what-is-kafka/
41. Tutorialspoint. System Analysis and Design – Overview. Access mode: https://www.tutorialspoint.com/system_analysis_and_design/system_ana lysis_and_design_overview.htm
42. Lonnie D. Bentley. System Analysis and Design for the Global Enterprise. Columbus, McGraw-Hill Education Ltd, 2007, 747 p.
43. WeyBackMachine. System Analysis. Access mode: https://web.archive.org/web/20070822025602/http://pespmc1.vub.ac.be/ASC/SYSTEM_ANALY.html
44. Ritchey T. Analysis and synthesis: on scientific method-based on a study by Bernhard Riemann, Systems research, 1991, Vol. 8(4), pp. 21–41. DOI: 10.1002/sres.3850080402
45. Booch G., Rumbaugh J., Jacobson I. Unified Modeling Language User Guide. Boston, AddisonWesley, 2005, 391 p.
46. Iso.org. ISO/IEC 19501:2005 – Information technology – Open Distributed Processing – Unified Modeling Language (UML) Version 1.4.2. Access mode: https://www.iso.org/standard/32620.html
47. Iso.org. ISO/IEC 19505-1:2012 – Information technology – Object Management Group Unified Modeling Language (OMG UML), Part 1, Infrastructure. Access mode: https://www.iso.org/standard/32624.html
48. WeyBackMachine. Basic UML. Access mode: https://web.archive.org/web/20121214050605/http://ooad.asf.ru/Files/U ML.djvu.zip
49. Chen F., Cheung C. C. H., Peng G. Linguistic tone and non-linguistic pitch imitation in children with autism spectrum disorders: A cross-linguistic investigation, *Journal of Autism and Developmental Disorders*, 2022, Vol. 52(5), pp. 2325–2343. DOI: 10.1007/s10803-021-05123-4
50. Robertson A., Magdy W., Goldwater S. Emoji skin tone modifiers: Analyzing variation in usage on social media, *ACM Transactions on Social Computing*, 2020, Vol. 3(2). – pp. 1–25. DOI: 10.1145/3377479
51. Tree of endings in Ukrainian language. Access mode: http://www.senyk.poltava.ua/projects/ukr_stemming/ukr_endings.html
52. Ishihara S. Score-based likelihood ratios for linguistic text evidence with a bag-of-words model, *Forensic Science International*, 2021, Vol. 327, P. 110980. DOI: 10.1016/j.forsciint.2021.110980
53. Understanding the Classification report through sklearn. – Access mode: https://muthu.co/understanding-the-classification-report-in-sklearn/
54. Markiv O., Vysotska V., Chyrun L., Voloshyn S., Dyyak I., Panasyuk V. Emotion recognition system project of English newspapers to regional E-business adaptation, *Computer science and information technologies, IEEE 17th International conference*, Lviv, Ukraine, 10–12 November 2022, proceedings. Lviv, IEEE, 2022, pp. 392–397. https://doi.org/10.1109/CSIT56902.2022.10000527
55. Vysotska V., Markiv O., Voloshyn S., Dyyak I., Budz I., Schuchmann V. Sentiment analysis technology of English newspapers quotes based on neural network as public opinion influences identification tool, *Computer science and information technologies, IEEE 17th International conference, Lviv, Ukraine, 10–12 November 2022, proceedings*. Lviv, IEEE, 2022, pp. 83–88. https://doi.org/10.1109/CSIT56902.2022.10000627

DOI 10.15588/1607-3274-2023-3-11

p-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2023. № 3
e-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2023. № 3

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

PROGRESSIVE INFORMATION TECHNOLOGIES

UDC 004.94 : 004.2

TEST GRAPH-SCHEMES OF THE ALGORITHMS OF FINITE STATE MACHINES WORK FOR ASSESSING THE EFFICIENCY OF AUTOMATED SYNTHESIS IN XILINX VIVADO CAD

Barkalov A. A. – Dr. Sc., Professor, Professor of Institute of Computer Science and Electronics, University of Zielona Gora, Zielona Gora, Poland;

Titarenko L. A. – Dr. Sc., Professor, Professor of Institute of Computer Science and Electronics, University of Zielona Gora, Zielona Gora, Poland;

Babakov R. M. – Dr. Sc., Associate Professor, Associate Professor of department of information technologies, Vasyl Stus Donetsk National University, Vinnytsia, Ukraine.

ABSTRACT

Context. The problem of evaluating the effectiveness of the automated design of a microprogram finite state machine with the operational transformation of state codes using Xilinx Vivado CAD is considered. The object of the research was graph-schemes of control algorithms implemented by finite state machine and able to prove the effectiveness of the principle of operational transformation of state codes in comparison with standard synthesis methods built into the CAD, in the context of hardware expenses optimization.

Objective. Development and research of graph-schemes of control algorithms in order to substantiate the effectiveness of the application of structure of the finite state machine with datapath of transitions in comparison with the built-in methods of synthesizing finite state machines in Xilinx Vivado CAD in the basis of programmable logic devices.

Method. The research is based on the hypothetical assumption that the Xilinx Vivado CAD has built-in methods of automated design of the circuit of a finite state machine, the effectiveness of which, according to the criterion of hardware expenses, exceeds other known methods of optimizing hardware expenses in the finite state machine circuit. In order to refute this hypothesis, it is proposed to prove that in some cases known methods of hardware expenses optimization in the finite state machine circuit are more effective in comparison with the methods built into CAD. In this work, as a well-known optimization method, the method of operational transformation of state codes, which corresponds to the structure of a finite state machine with datapath of transitions, is chosen. The effectiveness of this method is demonstrated on the example of several test graph-schemes of algorithms, the structure of which is abstract and artificially adapted to the chosen optimization method. The adaptation of the selected graph-schemes of the algorithms consists in the fact that a relatively small number of transition operations is required for their implementation with the help of a finite state machine with datapath of transitions. This contributes to the simplification of the circuit of the finite state machine and the reduction of hardware costs for its implementation. At the same time, the test graph-schemes of the algorithms have the possibility of scaling, which allows to automate the construction of VHDL models of the corresponding finite state machines for graph-schemes of different sizes and to evaluate the optimization of hardware expenses for finite state machines of different complexity.

Results. Using the example of several graph-schemes of algorithms, it is demonstrated that in some cases none of the finite state machine synthesis methods built into the Xilinx Vivado CAD is able to surpass the method of operational transformation of state codes according to the criterion of hardware expenses for the implementation of a finite state machine circuit. At the same time, a several-fold gain in hardware expenses can be achieved, which indicates the expediency of using this method under certain conditions. The formal definition of such conditions for the considered and other known optimization methods is a separate unsolved scientific problem.

Conclusions. The conducted experiments confirmed that in some cases, the known methods of synthesis of finite state machines allow to obtain circuits with lower hardware expenses than when using the methods of synthesis of finite state machines contained in Xilinx Vivado CAD. This testifies to the general expediency of using existing and developing new methods of hardware expenses optimization in the circuit of the finite state machines and the current relevance of the theory of the synthesis of digital automata as a scientific direction.

KEYWORDS: graph-scheme of algorithm, finite state machine, datapath of transitions, hardware expenses, Xilinx Vivado CAD.

ABBREVIATIONS

FSM is a finite state machine;
DT is a datapath of transitions;
GSA is a graph-scheme of algorithm;
CPLD is a complex programmable logic device;
LUT is a look-up table;
XST is a Xilinx Synthesis Tool.

DOI 10.15588/1607-3274-2023-3-12

NOMENCLATURE

M is a number of FSM states;
A is a set of FSM states $\{a_1, ..., a_M\}$;
L is a number of logic conditions;
X is a set of logic conditions $\{x_1, ..., x_L\}$;
N is a number of microoperations;
Y is a set of microoperations $\{y_1, ..., y_N\}$;
R is a digit capacity of state code;
T is a transition function of the FSM;
H^{DT} is a number of hardware expenses in the circuit of FSM with DT;
H^{XST} is a number of hardware expenses in the circuit of FSM, synthesized by XST;
E – the efficiency of the structure of the finite state machine according to the criterion of hardware expenses used for the implementation of its logic circuit.

INTRODUCTION

Modern human activity is closely related to the use of digital systems [1]. One of the main components of the digital system is the control unit [2, 3]. There are various ways of implementing control units, among which the finite state machine (FSM) model stands out [4, 5]. This model implements a given control algorithm in the form of a hardware circuit and is characterized by the maximum hardware expenses among other models of control units. At the same time, the model ensures the maximum speed of execution of the control algorithm due to the possibility of performing multidirectional microprogram transitions in one cycle of the device. The structure of the FSM can correspond to the Mealy machine model or the Moore machine model [2–5].

Large hardware expenses for the implementation of the FSM logic circuit have an impact on such characteristics of the finite state machine as power consumption, dimensions, cost, reliability, etc. [6]. Optimizing the characteristics of FSM circuit, in particular hardware expenses, is an important scientific and practical problem, the solution of which is devoted to many scientific works all over the world [1–7]. The structure of the finite state machine with datapath of transitions (FSM with DT), which is considered in this paper, is specifically aimed at minimizing hardware expenses by means of operational transformation of state codes [8].

In practice, the synthesis of FSM circuits is carried out with the help of specialized CAD, oriented to the use of the elemental basis of certain FPGA manufacturers. One of the leading manufacturers of FPGA-type chips is Xilinx, which is also the developer of Xilinx Vivado CAD. A component of this CAD is the Xilinx Synthesis Tool (XST), which implements, in particular, a number of methods for the synthesis of finite state machines [9]. At the same time, the question of how much these methods contribute to the optimization of hardware expenses in comparison with other known methods remains unexplored. This work solves the scientific and practical problem of comparing the efficiency of FSM synthesis methods built into the Xilinx Vivado CAD with the method of operational transformation of state codes. The solution to this problem is carried out by using graph-schemes of algorithms (GSA), adapted specifically for the structure of FSM with DT and able to demonstrate the advantages of this structure in comparison with the methods of synthesis of FSM in Xilinx Vivado.

The object of study is the process of synthesizing the logic circuit of a finite state machine in Xilinx Vivado CAD according to the VHDL model that corresponds to the given GSA.

This process can be carried out in automatic mode using the XST tool built into CAD according to the VHDL model recommended by Xilinx [9]. In the case of FSM with DT, a separate VHDL model must be developed, in the synthesis of which the capabilities of the XST tool are not used.

The subject of study is graph-schemes of control algorithms, which allow to prove the principle possibility of building a circuit with lower hardware expenes in the case of FSM with DT in comparison with circuits synthesized by Xilinx Vivado CAD in automatic mode.

The purpose of the work is the development and research of graph-schemes of control algorithms in order to substantiate the effectiveness of the application of the structure of FSM with DT in comparison with the built-in methods of synthesizing FSMs in Xilinx Vivado CAD in the basis of programmable logic devices.

1 PROBLEM STATEMENT

Let us assume that the finite state machine is given by the graph-scheme of the algorithm G and is characterized by the sets of states $A=\{a_1, ..., a_M\}$, input signals $X=\{x_1, ..., x_L\}$ and microoperations $Y=\{y_1, ..., y_N\}$. The synthesis of the FSM logic circuit involves the implementation of the transition function $T=T(X, T)$ and the output function $Y=Y(X, T)$ in the FPGA element base using Xilinx Vivado CAD. As a result of the synthesis of FSM according GSA G using the built-in XST tool, the circuit of the FSM is numerically characterized by hardware expenses H^{CAD}. As a result of the synthesis of FSM with DT in Xilinx Vivado without the use of XST, the circuit is characterized by hardware expenses H^{DT}.

The work solves the problem of finding several examples of graph-schemes of algorithms for which

$$H^{CAD} > H^{DT}, \quad (1)$$

which will prove the expediency of using (under certain conditions) the method of operational transformation of state codes instead of FSM synthesis methods built into Xilinx Vivado CAD.

2 REVIEW OF THE LITERATURE

Today, a wide range of methods for optimizing hardware expenses in the circuit of a finite state machine is known. These include, for example, the so-called methods of structural decomposition [7]. The essence of the meth-

DOI 10.15588/1607-3274-2023-3-12

ods lies in the multiple transformation of logic signals, which leads to corresponding changes in the structural scheme of the FSM.

In this article, the method of operational transformation of state codes is considered as a method of optimizing hardware expenses [8]. According to it, the transformation of state codes into FSM is not created using a system of canonical Boolean equations, but using a set of arithmetic and logical operations. Circuits that implement these operations are combined into the so-called datapath of transitions (DT). As a result, a structure of FSM with DT is formed, the synthesis of which is discussed, in particular, in [10].

The paper [11] substantiates the effectiveness of FSM with DT in comparison with the canonical structure of FSM based on the criterion of hardware expenses. However, today the canonical structure of FSM plays a more theoretical role, while the practical implementation of FSM circuits is carried out with the help of appropriate CAD, for example, Xilinx Vivado CAD. This is primarily due to the use of an elemental basis supported by CAD (as a rule, an FPGA basis).

Since FSM is often used as part of designed digital systems, support for its synthesis is implemented at the Xilinx Vivado CAD as part of the XST tool [9]. This tool supports several FSM synthesis methods aimed at optimizing various characteristics of the device circuit when implemented in the FPGA basis. Modeling the process of synthesizing the FSM circuit allows you to obtain the numerical values of the hardware expenses in the circuit of the device, expressed in the number of used LUT-elements.

During research conducted by the authors, the following hypothetical assumption was put forward. The use of FSM synthesis methods built into the Xilinx Vivado CAD will always allow you to obtain a machine circuit with lower hardware expenses than using other methods of optimizing the device circuit, which are not part of this CAD. This assumption is based on the fact that the Xilinx Vivado product has long been known in the world and contains developed technologies for the synthesis of specialized digital devices. In addition, CAD is focused on the use of its own elemental basis, which allows it to use the technological features of microchips to optimize circuits. The question of comparing the effectiveness of the synthesis of finite state machine by different methods in CAD Xilinx Vivado is not sufficiently considered today and does not allow to confirm or refute the hypothesis. This article is devoted to the solution of this issue on the example of a finite state machine with datapath of transitions.

3 MATERIALS AND METHODS

The XST synthesis tool, built into the Xilinx Vivado CAD, is able, under certain conditions, to find in the VHDL description of the device fragments of code that correspond to the description of the finite state machine (by state machine we mean a machine with undefined state codes). This process is called FSM extraction. For the found state machine, the XST tool performs the following actions:

– coding of states according to the chosen method;

– synthesis of the register circuit in accordance with the chosen method of states encoding;

– synthesis and optimization of the circuit for transition and output functions.

To ensure the possibility of automatic extraction of the FSM in its VHDL description, the following provisions should be observed:

1. The FSM states are specified in the form of a set of letters combined in an element of the Enumeration Type.

2. The memory register must be synchronous and have the possibility of being resetted to the initial state by a reset signal.

3. The implementation of the transition and output functions system is implemented using the case operator.

These requirements make it possible to specify an finite state machine in the VHDL language using one, two or three processes [9, 12–14]. Regardless of how many processes describe the FSM, the XST tool is capable of extracting the finite state machine from the VHDL code and coding the states according to the chosen coding method. For this purpose, the "fsm_extraction" parameter is provided in the synthesis process setting section, which can take on the following values [9]:

1. “One-hot”. A separate trigger is used to encode each state. The number of triggers is equal to the number of FSM states. At each point in time, only one trigger can have a ones value. To form the value of each trigger, a logical equation is used, in which the number of terms is equal to the number of transitions to the corresponding state.

2. “Sequential”. The XST tool finds in the FSM long sequences of states consisting of unconditional transitions, and encodes the states within them with consecutive binary codes of minimum sufficient digit capacity. As a result, the input signals of the FSM are not fed to the address inputs of the LUT elements, but only the code of the current state is fed, which usually has a small digit capacity compared to the number of input signals. Sequential encoding of states ensures more optimal filling of the static memory cells of LUT elements and reduces the number of unused cells.

3. “Johnson”. Coding of states is performed using the Johnson code. Each value of this code contains only one continuous sequence of ones, and any two adjacent values in the ordered sequence of values differ by only one digit. The Johnson code is a cyclic code with redundancy and allows you to reduce the number of electrical disturbances caused by the simultaneous switching of several bits of the register circuit.

4. “Gray”. Coding of states is performed using the Gray code, in which two adjacent values in an ordered sequence of values differ by the value of one binary bit, and the number of bits matches the number of bits in the case of sequential coding. It is advisable to use the Gray code for encoding chains of states, since each microprogram transition in such chain will be accompanied by a

DOI 10.15588/1607-3274-2023-3-12

change of only one digit in the memory register of the FSM.

5. “Auto”. The XST tool chooses one of the coding methods described above at its discretion based on the results of the analysis of the VHDL model of the FSM. The choice of the coding method also depends on other XST settings (for example, on the leading optimization strategy – hardware expenses or speed), but the generalized approach is as follows: if the FSM contains a small number of states, “One-hot” coding is used; with an average number of states, the Johnson code is used; with a large number of states, the Gray code is used.

6. “None”. The method of encoding states is not regulated. If in the five modes considered above, the coding of states is noted in the protocol of the synthesis process, then in this case, information about the coding of states is not provided. It is usually assumed that state codes correspond to the sequential number of state identifiers when they are listed in the VHDL description, starting from zero, but the XST tool does not officially define this and reserves the right to code states at its own discretion. This feature does not interfere with the synthesis of the correct FSM circuit, but it does not allow the application of additional optimization methods based on the known values of state codes.

The considered values of the “-fsm_extraction” parameter (except for the “Auto” and “None” parameters) correspond to different methods of FSM circuit synthesis built into Xilinx Vivado CAD. We will conduct a study of the effectiveness of the automatic FSM synthesis according to built methods in comparison with the synthesis of FSM with DT according to the criterion of hardware expenses. Efficiency will be determined by the following expression:

$$E^{DT} = \frac{H^{CAD}}{H^{DT}}, \qquad (2)$$

where H^{CAD} is the minimum possible hardware expenses for the implementation of the FSM circuit when using built-in CAD methods; H^{DT} – hardware expenes for the implementation of the FSM circuit with datapath of transitions. The unit of measurement of these parameters will be the number of used LUT elements of the selected FPGA chip. The value $E^{DT} > 1$ will mean that the circuit of the FSM with DT has lower hardware expenes compared to the FSM circuit synthesized by Xilinx Vivado built-in methods. Achieving the values $E^{DT} > 1$ will prove that, under certain conditions, the use of third-party methods for optimizing hardware expenses is more appropriate than the methods built into CAD.

In the context of the considered problem, the authors investigated five test GSAs G_1–G_5, which have the following common properties:

1. GSA reproduces only the transition function of the FSM and does not contain information about the output function (all operator nodes are empty). This makes it possible to determine the hardware expenses for the implementation of the transition function, although it cannot be considered an indicator of the efficiency of the FSM as a whole.

2. GSA has a regular structure, that is, it is a sequence of identical fragments. This approach, firstly, simplifies the scaling of the GSA by increasing the number of fragments, and secondly, allows you to automate the process of generating the VHDL code to describe the FSM.

Let’s consider GSAs investigated in this work.

GSA G_1

The GSA does not contain conditional nodes, has a completely linear structure and is marked by M states of the Moore machine (Fig. 1).

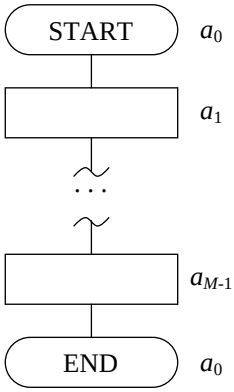


Figure 1 – GSA G_1

To implement sequential transitions in the corresponding FSM with DT, natural coding of states can be used, in which the binary code of the state coincides with its index in Fig. 1, and the state codes are transformed using a counter [3, 7].

GSA G_2

GSA corresponds to a Moore machine and has the following structure (Fig. 2):

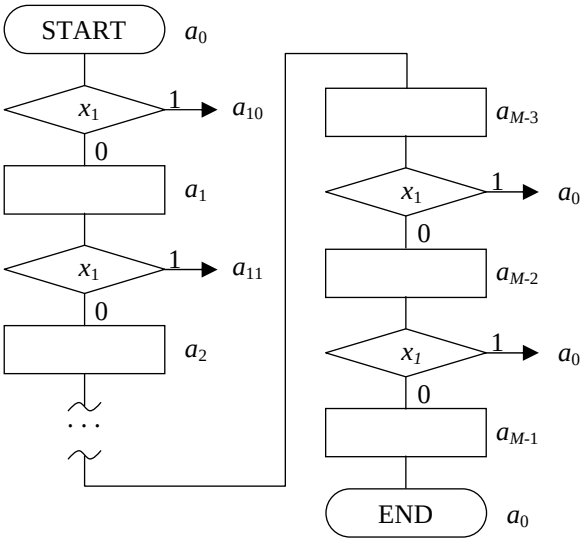


Figure 2 – GSA G_2

DOI 10.15588/1607-3274-2023-3-12

– each operator node is followed by a conditional node;

– the same logical condition x_1 is checked in all conditional nodes;

– when $x_1 = 0$ the transition leads to the next FSM state, when $x_1 = 1$ the transition leads ten states forward;

– for the last ten states, under the condition $x_1 = 1$ the transition leads to the state a_0.

In this case, the natural sequence of state coding is obvious, which makes it possible to implement the operational transformation of state codes based on the counter.

GSA G_3

The structure of the GSA is similar to the G_2 GSA, but the number of the state to which the transition is made under the condition $x_1 = 1$, is generated in a pseudo-random way within the limits of $[0; M-1]$. The general appearance of the GSA G_3, marked by the states of the Moore machine, is shown in Fig. 3.

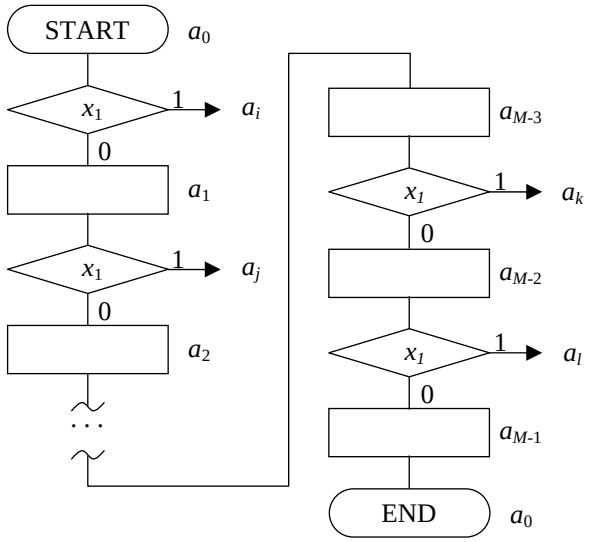


Figure 3 – ГСА G_3

In Fig. 3 states a_i, a_j, a_k, a_l mean some different states within the GSA. As in GSA G_2, states are coded in a natural order. Implementation of pseudo-random transitions is carried out with the help of a shift register with feedback based on the XOR operation. Such a register allows you to generate a sequence of $(2^R - 1)$ unique bit vectors that can be used to encode the states of the FSM. For this, the feedback must be organized according to a special primitive polynomial of length R, where R is the bit capacity of the state code of the FSM.

GSA G_4

GSA is similar in structure to G_2, but in each conditional node a different logical condition is checked, from x_1 to x_{M-1}. The general appearance of the GSA G_4, marked by the states of the Moore machine, is shown in Fig. 4.

GSA G_5

GSA is similar in structure to G_3, but in each conditional node a different logical condition is checked, from x_1 to x_{M-1}. The general appearance of the GSA G_5, marked by the states of the Moore machine, is shown in Fig. 5.

For each of the considered GSAs two VHDL models were built. The first model describes an FSM in the form of two processes and is intended for automated synthesis by Vivado XST. The second model represents the RTL description of the FSM with DT and is designed for circuit synthesis in Xilinx Vivado without using the "FSM Extraction" option.

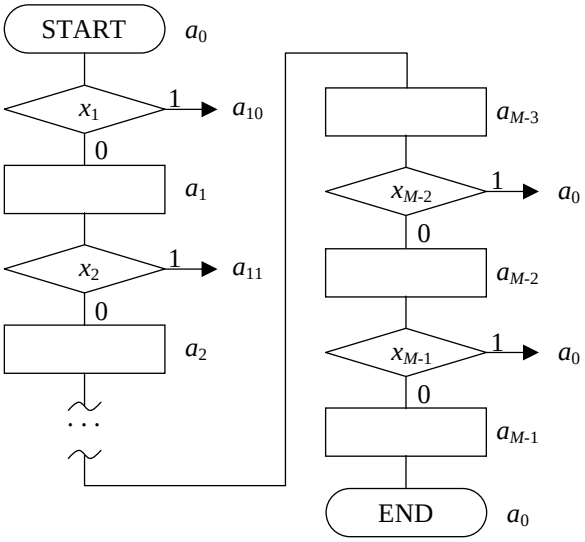


Figure 4 – GSA G_4

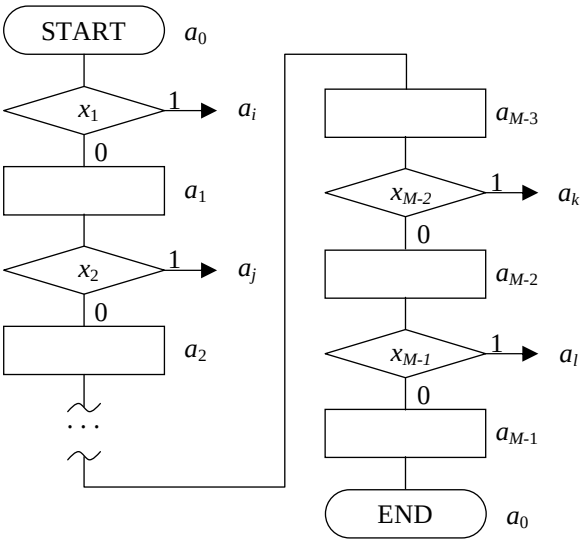


Figure 5 – GSA G_5

As an example, consider VHDL-models corresponding to GSA G_2. In Fig. 6 shows an example of the synthesized part of the FSM VHDL model intended for synthesis using XST. This model describes a machine with only

DOI 10.15588/1607-3274-2023-3-12

five states, but can easily be scaled to a machine with an arbitrary number of states. The model corresponds to the structure of the canonical FSM [2, 3, 9], but does not contain a part corresponding to the output function of the FSM. This is due to the fact that in the FSM with DT, the optimization of hardware expenses is carried out only in the transitions formation circuit, while the circuit of forming microoperations remains the same as in the canonical FSM. In order to emphasize the saving of hardware expenses precisely in the transition formation circuit, the used VHDL-models (as well as GSA) do not contain parts that correspond to the FSM output function).

```
package types is
    type state_type is (a0, a1, a2, a3, a4);
end package types;

library IEEE;
use IEEE.STD_LOGIC_1164.ALL;
use ieee.numeric_std.all;
use work.types.all;

entity FSM is                    -- Canonical FSM
    port (x1: in std_logic;
          C: in std_logic;      -- Clock
          Reset: in std_logic; -- Reset
          O: out state_type);  -- Current state
end FSM;

architecture FSM_A of FSM is
signal state, next_state: state_type;
begin

        process(C)          -- Memory Register
        begin
             if rising_edge(C) then
               if Reset = '1' then
                    state <= a0;
                else
                   state <= next_state;
               end if;
             end if;
        end process;

       process (state, x1) -- Trans. circuit
       begin
         case state is
           when a0 => if x1 = '0' then
                         next_state <= a1;
                      else
                         next_state <= a2;
                      end if;
           when a1 => if x1 = '0' then
                         next_state <= a2;
                      else
                         next_state <= a3;
                      end if;
           when a2 => if x1 = '0' then
                         next_state <= a3;
                      else
                         next_state <= a4;
                      end if;
           when a3 => if x1 = '0' then
                         next_state <= a4;
                      else
                         next_state <= a0;
                      end if;
           when a4 => next_state <= a0;
```

```
         end case;
       end process;

       O <= state;
end FSM_A;
```

Figure 6 – Synthesizable part of the VHDL model of canonical FSM for GSA G_2

We emphasize that in the model in Fig. 6 FSM states are declared by enumerating the literals “a0”, “a1”, etc. Specific state codes are not defined in this model. It is this approach that makes it possible to use the automatic extraction of the finite state machine by XST tool.

In Fig. 7 shows the VHDL model of the FSM with DT corresponding to GSA G_2 for the case of M=100 states.

На рис. 7 наведена VHDL-модель МПА з ОАП, що відповідає ГСА G_2 для випадку M=100 станів.

```
library IEEE;
use IEEE.STD_LOGIC_1164.ALL;
use ieee.numeric_std.all;

entity FSM is
    -- Bit capacity of state code:
    generic (R: integer := 7);

    port (C: in std_logic;
          Reset: in std_logic;
          D: out unsigned (R-1 downto 0));
end entity FSM;

architecture OAP_A of OAP is
signal state,
next_state: unsigned (R-1 downto 0);
begin

    process (C)   -- Memory Register
    begin
        if rising_edge(C) then
           if Reset = '1' then
               state <= "0000000";
           else
               state <= next_state;
           end if;
        end if;
    end process;

    process (state, x1)--Datapath of transitions
    begin
      -- If state is in range from a0 to A(M-10)
      if state < "1011010" then
          if x1 = '0' then -- If x1=0
             next_state <= state + 1;  --  +1
          else             -- If x1=1
             -- Transition to next state:
             next_state <= state + 10;
          end if;
       -- If it is state a(M-1)
       elsif state = "1100011" then
          next_state <= "0000000";-- To state a0

       -- If it is one of other ten states
       else
          if x1 = '0' then -- If x1=0
             -- Transition to next state:
             next_state <= state + 1;
          else  -- If x1=1
             -- Transition to initial state a0
             next_state <= "0000000";
          end if;
       end if;
```


DOI 10.15588/1607-3274-2023-3-12

```
    end process;

    D <= state;
end architecture OAP_A;
```

Figure 7 – Synthesizable part of the VHDL model of FSM with DT for GSA G_2

It should be noted that for the model in Fig. 6, the VHDL code length will increase along with the increase in the number of states of the machine. This is primarily due to an increase in the number of branches of the "case" operator. Instead, in the case of the equivalent FSM with DT (Fig. 7), a change in the number of states will only lead to a change in the bit rate and values of the constants in the model code. However, at the same time, this VHDL model maintains a "hard" binding to the GSA G_2 structure, which corresponds to the concept of FSM as a machine with "hardware" logic.

4 EXPERIMENTS

On the basis of the developed test GSAs, studies of the effectiveness of FSM with DT in comparison with the methods of synthesis of finite state machines built into the Xilinx Vivado CAD were carried out. The criterion of efficiency is the value of hardware expenses in the circuit of the FSM when it is implemented in the FPGA basis. The essence of the experiments was as follows:

1. VHDL models of canonical FSM and FSM with DT containing 100, 200, 500, 1000, 2000 states were built for each of GSA G_1–G_5.

2. For each model of the canonical FSM, the synthesis of the circuit was performed in the mode of automatic extraction of the finite state machine. The synthesis is performed for each of the following values of the "fsm-extraction" parameter: "one-hot", "sequential", "johnson"", "gray", "auto". According to the results of each synthesis, the numerical value of hardware expenses H^{CAD} was obtained, expressed in the number of used LUT-elements of the FPGA. The synthesis was carried out for the microchip xc7s6cpga196-2 of the Spartan-7 series. All Xilinx Vivado settings except the parameter "fsm_extraction" are selected by default.

3. For each GSA G_1–G_5, such a value of the "fsm-extraction" parameter is defined, at which the FSM circuit has the lowest hardware expenses in comparison with other values of this parameter. The values of hardware expenses obtained in this case are considered to be the minimum possible, which can be obtained when using FSM synthesis methods built into CAD.

4. For each model of FSM with DT, synthesis of the circuit was performed without using the mode of automatic extraction of the finite state machine. According to the results of each synthesis, the numerical value of the hardware expenses H^{DT} was obtained, expressed in the number of used LUT-elements of the FPGA. The synthesis conditions are the same as in the synthesis of models of canonical FSM.

5. For each model, the efficiency was calculated according to expression (2). As values H^{CAD}, the minimum values of hardware expenses are taken in accordance with clause 3.

It should be emphasized that for all models, the operability of the logic circuit of the FSM in Xilinx Vivado was verified using an additional VHDL model of the behavioral type. The function of this model was the generation of input signals, reset and synchronization signals.

5 RESULTS

The results of experimental studies are given in tables –5. The row of the table containing the minimum values according to clause 3 is marked with a gray background. The symbol "–" in the cells of the tables means that the numerical values of the hardware expenses for the corresponding parameters were not obtained due to the extremely high complexity of the resulting circuit.

Table 1 – Results of GSA G_1 studies

M		100	200	500	1000	2000
H^{CAD}, LUT	one-hot	60	116	292	570	1181
	sequential	9	15	18	23	27
	johnson	273	457	979	–	–
	gray	18	20	29	38	40
	auto	60	116	292	570	1181
H^{DT}, LUT		6	8	11	12	13
E		**1.5**	**1.88**	**1.63**	**1.92**	**2.08**

Table 2 – Results of GSA G_2 studies

M		100	200	500	1000	2000
H^{CAD}, LUT	one-hot	109	215	539	1489	3091
	sequential	18	18	23	27	22
	johnson	267	759	3047	–	–
	gray	26	38	56	68	66
	auto	109	215	539	1489	3091
H^{DT}, LUT		13	15	19	20	21
E		**1.38**	**1.20**	**1.21**	**1.35**	**1.05**

Table 3 – Results of GSA G_3 studies

M		100	200	500	1000	2000
H^{CAD}, LUT	one-hot	112	250	668	1454	3334
	sequential	44	94	239	511	1177
	johnson	419	1642	–	–	–
	gray	43	92	239	510	1166
	auto	112	250	668	1454	3334
H^{DT}, LUT		21	41	99	190	407
E		**2.05**	**2.24**	**2.41**	**2.68**	**2.86**

Table 4 – Results of GSA G_4 studies

M		100	200	500	1000	2000
H^{CAD}, LUT	one-hot	155	311	799	1958	4051
	sequential	130	269	656	1368	2767
	johnson	412	1219	3978	–	–
	gray	127	254	631	1240	2498
	auto	155	311	799	1958	4051
H^{DT}, LUT		51	72	155	318	586
E		**2.49**	**3.54**	**4.07**	**3.90**	**4.26**

DOI 10.15588/1607-3274-2023-3-12

Table 5 – Results of GSA G_5 studies

M		100	200	500	1000	2000
H^{CAD}, LUT	one-hot	145	298	831	1782	4050
	sequential	166	357	1048	2286	5123
	johnson	589	1967	–	–	–
	gray	155	348	995	2281	4922
	auto	145	298	831	1782	4050
H^{DT}, LUT		58	117	225	454	931
E		**2.50**	**2.54**	**3.69**	**3.92**	**4.35**

The content of the last row of each table is calculated as the result of dividing the values in the row of the table, marked with a gray background, by the value in the row H^{DT} of the table. For example, in the table 1, the value of 1.5 in the last row of the table is calculated as a fraction of the division of 9 by 6, where 9 is the minimum possible value of hardware expenses obtained using the XST tool, 6 is the value of hardware expenses for the implementation of circuit of the equivalent FSM with DT.

6 DISCUSSION

A finite state machine with datapath of transitions differs from a canonical finite state machine in that it uses a set of arithmetic and logic operations to transform state codes, which form a datapath of transitions. The complexity of the DT circuit depends on the number of operations used by datapath to transform state codes. If, for a given GSA, the complexity of the DT turns out to be less than the complexity of the transitions formation circuit of the canonical FSM, the use of the structure of FSM with DT is preferable compared to the canonical FSM.

The synthesis results shown in tables 1–5 demonstrate a clear gain (sometimes several times) in hardware expenses when implementing the FSM transition function using the method of operational transformation of state codes and the structure of FSM with DT. This allows us to draw the following conclusions.

1. Finite state machine synthesis methods built into Xilinx Vivado CAD do not always give the best result in terms of hardware expenses in comparison with third-party FSM optimization methods.

2. The use of finite state machine synthesis methods built into CAD does not allow choosing specific values of state codes. This makes it impossible to simultaneously use other known methods of optimization of the FSM circuit, in particular, optimization of the output function circuit. Instead, the use of FSM with DT is based on preselected values of state codes, which potentially allows combining other optimization methods with the method of operational transformation of state codes.

3. The theory of synthesis and optimization of circuits of finite state machine remains relevant today, provided that modern CAD digital systems and elemental basis are used. A promising scientific and practical direction is the formalization and algorithmization of well-known methods of FSM circuit optimization.

Certain limitations of the conducted research should also be taken into account when analyzing the obtained results.

1. The comparative analysis of hardware expenses was carried out only for the circuit of transition function without taking into account expenses in the circuit that implements the output function. If, for a given FSM, the hardware expenses for the implementation of the output function significantly exceed expenses for the implementation of the transition function, the effect of using an FSM with DT instead of a canonical FSM will be much smaller. However, from the point of view of Xilinx Vivado CAD, the use of FSM with DT is not the only third-party approach to reducing hardware expenses. Therefore, the potential possibility of surpassing the methods built into CAD remains for other well-known synthesis and optimization methods.

2. The structure of studied GSAs is artificially adapted in such a way that the implementation of microprogram transitions in the corresponding FSM with DT takes place with the help of a smaller number of operations. This made it possible to obtain efficiency values greater than 1. This fact indicates the need to optimize the FSM with DT circuit, which is possible under the condition of development and application of special methods of synthesis of this FSM class. At that time, the use of methods built into CAD allows for synthesis in automatic mode, following only the rules of building VHDL models to ensure automatic extraction of the finite state machine by means of XST.

However, these limitations do not negate (and in some ways emphasize) the general possibility and expediency of using known optimization methods, as opposed to the methods built into CAD.

CONCLUSIONS

The article proposes a solution to the scientific problem of researching the effectiveness of the method of operational transformation of state codes in comparison with the methods of synthesis of finite state machines built into Xilinx Vivado CAD, according to the criterion of hardware expenses.

The scientific novelty of the work consists in the experimental confirmation of the advantage (under certain conditions) of the use of well-known methods of synthesis of finite state machines (in particular, the method of operational transformation of state codes) compared to the methods of synthesis of finite state machines built into Xilinx Vivado CAD. This confirmation is provided by the efficiency values obtained during research (Tables 1–5). Thus, the hypothesis that finite state machine synthesis methods built into Xilinx Vivado are always able to generate FSM circuits with lower hardware expenses compared to other hardware expenses optimization methods is disproved.

The practical use of the obtained results is possible in the development of methods for evaluating the effectiveness of the structure of a finite state machine with datapath of transitions, as well as other structures and methods aimed at optimizing the characteristics of the circuit of a finite state machine.

DOI 10.15588/1607-3274-2023-3-12

Prospects for further research consist in solving a range of scientific and practical problems related to the development, implementation and evaluation of the effectiveness of structures and methods of synthesis of finite state machines with optimized hardware expenses.

ACKNOWLEDGEMENTS

The work is supported by the state budget scientific research project of Vasyl' Stus Donetsk National University "Methods, algorithms and tools of computer-aided design of control units of computing systems" (state registration number 0122U200085).

REFERENCES

1. Bailliul J., Samad T. Encyclopedia of Systems and Control. Springer, London, UK, 2015, 1554 p.
2. Sklyarov V., Sklyarova I., Barkalov A., Titarenko L. Synthesis and Optimization of FPGA-Based Systems; Volume 294 of Lecture Notes in Electrical Engineering. Springer, Berlin, Germany, 2014, 432 p.
3. Baranov, S. Logic and System Design of Digital Systems. Tallin, TUTPress, 2008, 267 p.
4. Micheli G. D. Synthesis and Optimization of Digital Circuits. McGraw-Hill, Cambridge, MA, USA, 1994, 579 p.
5. Minns P., Elliot I. FSM-Based Digital Design Using Verilog HDL, JohnWiley and Sons, Hoboken, NJ, USA, 2008, 408 p.
6. Grout I. Digital Systems Design with FPGAs and CPLDs. Elsevier Science, Amsterdam, The Netherlands, 2011, 784 p.
7. Baranov S. Logic Synthesis for Control Automata. Dordrecht, Kluwer Academic Publishers, 1994, 312 p.
8. Barkalov A. A., Babakov R. M. Operational formation of state codes in microprogram automata, *Cybernetics and Systems Analysis*, 2011, Volume 47 (2), pp. 193–197.
9. Xilinx. XST UserGuide. V.11.3. Available online: https://www.xilinx.com/support/documentation/sw_manuals/xilinx11/xst.pdf (accessed on 12 April 2023).
10. Barkalov A. A., Titarenko L. A., Babakov R. M. Synthesis of Finite State Machine with Datapath of Transitions According to the Operational Table of Transitions, *Radio Electronics, Computer Science, Control*, 2022, Volume 3 (62), pp. 109–119.
11. Barkalov A. A., Babakov R. M. Determining the Area of Efficient Application of a Microprogrammed Finite-State Machine with Datapath of Transitions, *Cybernetics and Systems Analysis*, 2019, Volume 54 (3), pp. 366–375.
12. Czerwinski R., Kania D. Finite State Machine Logic Synthesis for Complex Programmable Logic Devices. Berlin, Springer, 2013, 172 p.
13. Mano M. Digital design (4th Edition). New Jersey, Prentice Hall, 2006, 624 p.
14. Zwolinski M. Digital System Design with VHDL. Boston, Addison-Wesley Longman Publishing Co., Inc. 2000, 416 p.

Received 19.06.2023.
Accepted 11.08.2023.

УДК 004.94 : 004.2

ТЕСТОВІ ГРАФ-СХЕМИ АЛГОРИТМІВ РОБОТИ МІКРОПРОГРАМНИХ АВТОМАТІВ ДЛЯ ОЦІНКИ ЕФЕКТИВНОСТІ АВТОМАТИЗОВАНОГО СИНТЕЗУ В САПР XILINX VIVADO

Баркалов О. О. – д-р техн. наук, професор, професор Інституту комп'ютерних наук та електроніки університету Зеленогурського, м. Зельона Гура, Польща.

Тітаренко Л. О. – д-р техн. наук, професор, професор Інституту комп'ютерних наук та електроніки університету Зеленогурського, м. Зельона Гура, Польща.

Бабаков Р. М. – д-р техн. наук, доцент, доцент кафедри інформаційних технологій Донецького національного університету імені Василя Стуса, м. Вінниця, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто задачу оцінки ефективності автоматизованого проєктування мікропрограмного автомата з операційним перетворенням кодів станів із використанням САПР Xilinx Vivado. Об'єктом дослідження були граф-схеми алгоритмів керування, що імплементуються мікропрограмним автоматом та здатні довести ефективність принципу операційного перетворення кодів станів у порівнянні зі стандартними методами синтезу, вбудованими в САПР, в контексті оптимізації апаратурних витрат.

Мета. Розробка і дослідження граф-схем алгоритмів керування з метою обґрунтування ефективності застосування структури мікропрограмного автомата з операційним автоматом переходів у порівнянні із вбудованими методами синтезу автоматів в САПР Xilinx Vivado в базисі програмувальних логічних пристроїв.

Метод. В основу дослідження покладено гіпотетичне припущення про те, що САПР Xilinx Vivado має вбудовані методи автоматизованого проектування схеми мікропрограмного автомата, ефективність яких за критерієм апаратурних витрат перевершує інші відомих методи оптимізації апаратурних витрат в схемі автомата. З метою спростування даної гіпотези запропоновано довести, що в окремих випадках відомі методи оптимізації апаратурних витрат в схемі автомата є більш ефективними у порівняні з методами, вбудованими в САПР. В даній роботі в якості відомого методу оптимізації обраний метод операційного перетворення кодів станів, що породжує структуру мікропрограмного автомата з операційним автоматом переходів. Ефективність цього методу доводиться на прикладі кількох тестових граф-схем алгоритмів, структура яких є абстрактною і штучно адаптована до обраного методу оптимізації. Адаптація обраних граф-схем алгоритмів полягає в тому, що для їх реалізації за допомогою мікропрограмного автомата з операційним автоматом переходів потрібна відносно мала кількість операцій переходів. Це сприяє спрощенню схеми автомата і зменшенню апаратурних витрат на її реалізацію. Разом з тим тестові граф-схеми алгоритмів мають можливість масштабування, що дозволяє автоматизувати побудову VHDL-моделей відповідного автомата для граф-схем різного розміру і оцінити оптимізацію апаратурних витрат для автоматів різної складності.

DOI 10.15588/1607-3274-2023-3-12

Результати. На прикладі декількох граф-схем алгоритмів продемонстровано, що в окремих випадках жоден із методів синтезу кінцевих автоматів, вбудованих в САПР Xilinx Vivado, не здатен перевершити метод операційного перетворення кодів станів за критерієм апаратурних витрат на реалізацію схеми мікропрограмного автомата. При цьому може досягатись кількаразовий виграш у витратах апаратури, що свідчить про доцільність використання даного методу за певних умов. Формальне визначення таких умов для розглянутого та інших відомих методів оптимізації є окремою невирішеною науковою проблемою.

Висновки. Проведені експерименти підтвердили, що в окремих випадках відомі методи синтезу мікропрограмних автоматів дозволяють отримати схеми автоматів із меншими витратами апаратури, ніж при використанні методів синтезу автоматів, вбудованих в САПР Xilinx Vivado. Це свідчить про загальну доцільність використання існуючих і розробки нових методів оптимізації апаратурних витрат в схемі автомата та про сьогоденну актуальність теорії синтезу цифрових автоматів як наукового напрямку.

КЛЮЧОВІ СЛОВА: граф-схема алгоритму, мікропрограмний автомат, операційний автомат переходів, апаратурні витрати, САПР Xilinx Vivado.

ЛІТЕРАТУРА

1. Bailliul J. Encyclopedia of Systems and Control / J. Bailliul, T. Samad. – Springer : London, UK, 2015. – 1554 p.
2. Sklyarov V. Synthesis and Optimization of FPGA-Based Systems; Volume 294 of Lecture Notes in Electrical Engineering / V. Sklyarov, I. Sklyarova, A. Barkalov, L. Titarenko. – Springer : Berlin, Germany, 2014. – 432 p.
3. Baranov S. Logic and System Design of Digital Systems / S. Baranov. – Tallin : TUTPress, 2008. – 267 p.
4. Micheli G. D. Synthesis and Optimization of Digital Circuits / G. D. Micheli. – McGraw-Hill : Cambridge, MA, USA, 1994. – 579 p.
5. Minns P. FSM-Based Digital Design Using Verilog HDL / P. Minns, I. Elliot. – JohnWiley and Sons: Hoboken, NJ, USA, 2008. – 408 p.
6. Grout I. Digital Systems Design with FPGAs and CPLDs / I. Grout. – Elsevier Science: Amsterdam, The Netherlands, 2011. – 784 p.
7. Baranov S. Logic Synthesis for Control Automata / S. Baranov. – Dordrecht : Kluwer Academic Publishers, 1994. – 312 p.
8. Barkalov A. A. Operational formation of state codes in microprogram automata / A. A. Barkalov, R. M. Babakov // Cybernetics and Systems Analysis. – 2011. –Volume 47 (2). – P. 193–197.
9. Xilinx. XST UserGuide. V.11.3. Available online: https://www.xilinx.com/support/documentation/sw_manuals/xilinx11/xst.pdf (accessed on 12 April 2023).
10. Barkalov A. A. Synthesis of Finite State Machine with Datapath of Transitions According to the Operational Table of Transitions / A. A. Barkalov, L. A. Titarenko, R. M. Babakov // Radio Electronics, Computer Science, Control. – 2022. – Volume 3 (62). – P. 109–119.
11. Barkalov A. A. Determining the Area of Efficient Application of a Microprogrammed Finite-State Machine with Datapath of Transitions / A. A. Barkalov, R. M. Babakov // Cybernetics and Systems Analysis. – 2019. – Volume 54. (3). – P. 366–375.
12. Czerwinski R. Finite State Machine Logic Synthesis for Complex Programmable Logic Devices / R. Czerwinski, D. Kania. – Berlin : Springer, 2013. – 172 p.
13. Mano M. Digital design (4th Edition) / M. Mano. – New Jersey : Prentice Hall, 2006. – 624 p.
14. Zwolinski M. Digital System Design with VHDL / M. Zwolinski. – Boston : Addison-Wesley Longman Publishing Co., Inc., 2000. – 416 p.

DOI 10.15588/1607-3274-2023-3-12

p-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2023. № 3
e-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2023. № 3

UDC 711.57:004.9

INFORMATION TECHNOLOGY OF FORMING THE EDUCATIONAL NETWORK OF THE TERRITORIAL COMMUNITY

Danshyna S. Yu. – Dr. Sc., Professor of Dept. of Geo-information Technologies and Space Monitoring of the Earth, National Aerospace University "KhAI", Kharkiv, Ukraine.

Nechausov A. S. – PhD, Associate Professor of the Dept. of Geo-information Technologies and Space Monitoring of the Earth, National Aerospace University "KhAI", Kharkiv, Ukraine.

ABSTRACT

Context. Local government organizations have significant decision-making power in the field of education. This requires the development of understandable tools that help form a network of educational institutions that provide high-quality educational services. The object of the study was the process of forming an educational network (ENW) of the community in the implementation of territorial development projects as part of the reform of local self-government.

Objective. The goal of the work is to increase objectivity of decisions made in the formation of the ENW territorial community, summarizing demographic, infrastructural, personnel, economic factors when choosing the location of ENW facilities.

Method. The study of the classical placement problem and its further adaptation to real problems arising from the implementation of the education reform made it possible to represent ENW territorial communities as a set of independent complete bipartite graphs. In this case, to solve the problem of choosing the location of an educational institution on the network, an information technology (IT) was developed to form an ENW territorial community. Based on the adapted *p*-median model and methods of geospatial analysis, generalizing the requirements of the current legislation, a set of input and output parameters of IT and a set of its operations are formed. The representation of the IT structure in the form of an IDEF0-model clearly explains how a combination of various factors is processed and generalized when making decisions while creating ENW and looking for ways to improve it.

Results. The developed IT was investigated in solving the problem of forming a network of lyceums in the Kharkiv region using geospatial information, open statistical data and data on lyceums. The proposed accommodation options make it possible to achieve a general level of accessibility of specialized secondary education of almost 94%. IT has additional tools for solving the problems of transporting students to the place of study and home.

Conclusions. The experiments carried out confirmed the operability of the proposed IT. The generalization of the results obtained makes it possible to recommend it for practical use in solving the problems of analyzing the current state of ENW, finding ways to improve it and possible directions for development, as well as evaluating solutions related to the spatial planning of ENW.

KEYWORDS: *p*-median model, geospatial analysis, model of information flows of the process, IDEF0-model, accessibility assessments.

ABBREVIATIONS

ENW is an educational network;
IT is an information technology;
GIS is a geoinformation systems;
VGI is voluntary geographic information.

NOMENCLATURE

Λ is the admissible value of the spatial accessibility of the educational institution relative to the points of demand;

μ_j is the weight of the vertex u_j, depending on the number of students at the demand point D_i;

φ is an update function;

ψ is an output function, that generates an output data;

A is a set of operations that implement the ENW forming process;

D_i is a demand point (i=1, …, I);

E is a set of graph edges $K_{1,D_{I+1}}$;

I_Pr is a model of information flows of the ENW forming process;

$K_{1,D_{I+1}}$ is a complete bipartite graph as a graphical model "center – points of demand";

$L(\cdot)$ is a distance function between educational institution (center) and demand points;

N is the maximum number of students assigned to an educational institution;

N_{ENS} is an estimated number of students;

N_{PA} is a population in the planning area;

O is a set of output data the ENW forming process;

R is a location of the educational institution (center);

U is a set of graph vertices $K_{1,D_{I+1}}$;

V is a set of input data, incoming into an input of the ENW forming process;

(x_j, y_j) are graph vertex coordinates $K_{1,D_{I+1}}$;

Z is a set of documents, regulating the ENW forming process.

INTRODUCTION

Education is a vital "driving force" for the development of countries, a necessary proviso for achieving social, economic, political, and cultural advantages. Globally, it affects the level of employment of the country's population, the growth of its well-being, individual income, the distribution of wages, the efficiency of resources distribution, the state of health [1], becomes a key incentive and factor that allows countries to introduce new technologies, increase labor productivity, change production systems for development of society. Experts from the World Economic Forum emphasize that among

DOI 10.15588/1607-3274-2023-3-13

the key indicators that affect the level of economic and industrial development of the country, education takes the most important place (its importance is comparable to the development of Technology & Innovation and Global Trade & Investment) [2]. At the same time, according to the estimates of the economies of countries of the nascent archetypes, such as Cyprus, Saudi Arabia, Latvia, Ukraine, Bulgaria et al., a strong positive correlation (with a correlation coefficient) is recorded between the level of development of the drivers of production of the country and education outcomes score (Fig. 1).

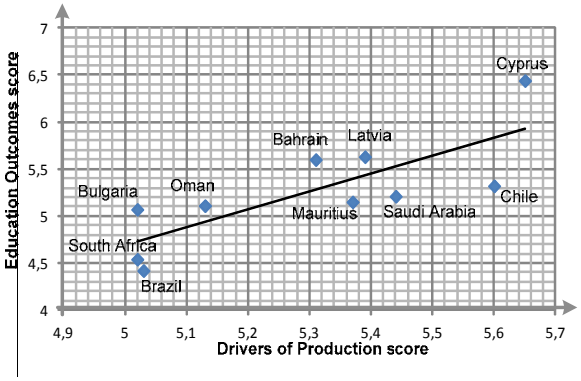


Figure 1 – The scatterplot (for the top-countries of the nascent archetypes) that illustrates dependence between indicators of drivers of production and education outcomes (According to the results of the analysis of work [2])

Therefore, education becomes a necessary condition for the effective transformation of production forces in the world of rapidly developing technologies, characterizes: the country's ability to respond to shifts in the labor market that are triggered by the Fourth Industrial Revolution, current labor force capabilities to adopt and use emerging technologies in production systems, the ability in the future to cultivate the right skills and talent of workforce [2].

The reform of local self-government, carried out in Ukraine since 2014, gives the authorities significant powers and functions to independently decide on the methods and resources for providing quality and affordable basic education in the respective territory. But the problems that are most often fixed at the local level (unequal occupancy of schools and classes, the lack of access to education for all children, the existence of a cumbersome school infrastructure with low occupancy in the presence of overcrowded schools, etc.) require the improvement of methodological approaches to the integrated development of territories and the development of effective tools for creating an educational network (ENW), which infrastructurally, organizationally, personnel, and financially will ensure the provision of quality education [3].

The object of study is the process of forming the educational network of the territorial community in the implementation of territorial development projects as part of the reform of local self-government.

To create conditions that ensure the provision of quality education, it is necessary to systematically study the existing ENW of the territorial community, identifying shortcomings and developing possible directions for its improvement in terms of changes that have occurred in connection with the education reform in the context of constantly changing legal norms, while there are no building requirements for the placement of educational institutions, establishments. Geospatial analysis methods are considered to be one of the most powerful and promising tools for solving such problems [4–6]. Therefore, in order to increase the objectivity of decisions made when forming the ENW of the community, it is necessary to expand the use of these methods, introducing approaches that, in the current economic conditions, will take into account the demographic characteristics of the settlements of the community, as well as the availability and prospects for the development of the existing ENW within the planning area.

The subject of study is the information technology for the formation of a territorial community educational network that uses geospatial analysis to determine the location of its infrastructure and find ways to improve it.

Therefore, the known sampling methods [4 – 14], which, unfortunately, are focused on assessing the availability of educational services, have excellent prospects for using information technology (IT) to form an educational community network.

The purpose of the work is to increase the objectivity of the decisions made when forming the educational network of the territorial community, generalizing various factors (demographic, infrastructural, personnel, economic) when choosing the location of its facilities.

1 PROBLEM STATEMENT

Consider the ENW facility as a center designed to meet the demand for educational services of students living in the planning area. In accordance with the J. Von Thunen model, the choice of the location of this center is subject to the provision of a minimum of transport costs, i.e.:

$$\arg\min_{R} F(R) = \arg\min_{R} \sum_{i=1}^{I} L(R, D_i)\,. \qquad (1)$$

Let us assume that the demand for educational services is unchanged, and the decision on which educational institutions (centers) to satisfy it is made centrally with the assignment of students from district to the corresponding territorial areas. In this case, there is a need to develop IT, focused on the use of geoinformation systems (GIS) in the processes of forming an ENW territorial community, which allows an educational institution to find R, given that

$$\sum_{i=1}^{I} L(R, D_i) \to \min \text{ for } \forall D_i \ (i = \overline{1, I}).$$

DOI 10.15588/1607-3274-2023-3-13

2 REVIEW OF THE LITERATURE

The modern concept of sustainable territorial planning of Professor C. Moreno is aimed at creating self-sufficient neighborhoods with the necessary functions for life, health care, education, etc. by decentralizing functions and services. In this regard, a territorial community is a complex socio-economic system for which the availability and accessibility of social infrastructure are of fundamental importance in ensuring the necessary quality of life for the population [4, 7]. For example, in the works of Ch. Jing et al. [4] or M. Ogryzek et al. [8] noted that the growth of cities increases the demand for education and other social services, leading to both positive outcomes and significant threats, affecting resilience. The spatial arrangement of social facilities, in particular educational institutions, plays an important role in organizing a number of economic processes, affects the attractiveness of places of residence, and leads to the movement of labor resources that are looking for better living conditions in their habitat. Similar studies of the urban ENW record uneven accessibility to social infrastructure and a trend towards a constant decrease in this accessibility [4, 8, 9]. However, a relatively small area of study (a city, its district or street) helps to solve a separate problem of placing an educational institution within a small area, but the results obtained are difficult to transfer to larger territorial counterparts (ENW of a district or region), take into account the influence of geographical factors when identifying inequality compulsory education. At the same time, in the generalizing studies of M. Kompil et al. [5] and A. Milbert et al. [9], the hypothesis about the uneven availability of services (including educational ones) in European countries is confirmed. The availability of services differs in urban areas and differs significantly between urban and rural areas, significantly increasing the average distance to services and inequality in education between rural and urban residents [5, 9].

Y. Gao et al. [10] analyze the inequality of compulsory education in terms of unbalanced spatial distribution, confirming the presence of educationally excluded areas – areas where the location of schools does not correspond to the real needs of residents. As noted in the work of J. Rao and J. Ye [11], attempts to reduce the gap between cities and rural areas by adjusting financial investments in basic education in rural areas and adapting the school map require significant efforts, affect rural areas and exacerbate the problem of educational disparity, in fact widening the gap between urban and rural areas, and between the rich and the poor people.

In this regard, the social infrastructure of Ukrainian regions also functions and develops very unevenly, leading to a significant imbalance in the field of social services provided to the population, which manifests itself in a massive decline in housing and communal services, consumer services, education, sports and medicine [12]. At the same time, it is noted that the governing bodies need an effective system that can help in analyzing the current state of the educational infrastructure, education itself and its progress, as well as in making decisions and formulating a policy for providing quality and affordable basic education in the relevant territory [6, 8, 10].

S. Agrawal and R. D. Gupta [6] suggest using GIS to analyze the current state of education in the region, as well as to identify possible problem areas in the study area. Focusing on gender inequality in education, on the demographic, social and economic situation in this area, the authors illustrate the advantage of geospatial analysis as an innovative way to critically study a variety of social and economic problems. These findings support D. Rodriguez-Segura et al. [13] and O. Mustapha et al. [14]. It is noted in their works that recent advances in remote sensing are making detailed geospatial population data available and can be used to pinpoint educationally excluded areas with unprecedented scale, detail, and cost-effectiveness [13], locating schools based on population size. population and their proximity, ensuring the sustainable development of territories [14].

Thus, regardless of the considered aspects of the problem of ENW territorial community formation, the literature review results confirm the following. The effectiveness of its formation is reduced due to the large size of the planning area, restrictions on the number of specialists involved in the analysis, and the reduction in funding allocated to local governments [3, 11, 13]. At the same time, studies of the potential of geospatial analysis methods confirm the possible reduction in socio-economic costs when the ENW facility is located in the context of the service area, taking into account a number of basic factors [6, 12, 14]:

– demographic, taking into account the demographic characteristics of settlements in the planning area;

– infrastructural, taking into account the availability and prospects for the development of the relevant educational infrastructure;

– economic, taking into account the correspondence of the volume of financing to the real needs and possibilities of the local budget, etc.

Therefore, in order to increase the objectivity of decisions made when forming an ENW territorial community based on the analysis of independent diverse data, it is necessary to introduce IT that combines spatial analysis technology with minimum distance methods, ensuring spatial accessibility (taking into account the real distance of students to the ENW object) and the availability of opportunities (taking into account the possibility of obtaining educational services in the existing ENW).

3 MATERIALS AND METHODS

The Law of Ukraine “On Education” guarantees equal rights to quality education, defining accessibility as “... a set of conditions that help meet the educational needs of individuals and provide each of them with equal opportunities to receive education ...” [15]. Local self-government bodies are responsible for the implementation of state policy in this area, plan and ensure the development of a network of educational institutions, guaranteeing their accessibility to all citizens living in the respective territory [3, 15]. Based on this, in model (1), the

DOI 10.15588/1607-3274-2023-3-13

places of residence of students can be considered as points of demand for educational services – D_i ($i=\overline{1,I}$), but, based on the idea of self-sufficient neighborhoods [7], consider the ENW territorial community model as a collection of complete bipartite graphs of the type $K_{1,D_{I+1}}$. From the condition of centralized assignment of students to the corresponding educational institution, each aggregate graph $K_{1,D_{I+1}}$ we will consider it an independent structure that has little influence on other columns of the ENW territorial community. Then, in the general case, while $i=1,...,I$ structure $K_{1,D_{I+1}}$ defines in the space a set of vertices $U=\{u_j \mid j=\overline{1,(i+1)}\}$ with coordinates (x_j, y_j) and a set of edges (without loops) $E=\{e_k \mid k=\overline{1,i}\}$.

Let's assume that at the top 1 of the graph $K_{1,D_{I+1}}$ an educational facility is located. Every vertex u_j ($j\neq 1$) match the weight $\mu_j > 0$, the value of which depends on the number of students living at the point of demand D_i ($j=i+1, \forall D_i$ while $i=\overline{1,I}$). In doing so

$$\sum_j \mu_j \le N, \qquad (2)$$

which determines the possibility of obtaining educational services by the *N*-th number of students assigned to this educational institution.

The value of *N* is selected taking into account the differences and different possibilities for filling schools in rural and urban areas, which leads to significant fluctuations in its value [3, 16]. This is due to the fact that, for example, the number of students in a class in rural areas should be at least 5, ideally, it is necessary to provide an average occupancy rate for elementary schools of 13 students, for a gymnasium – 17 students and for a lyceum – 21 students. The average occupancy of classes in urban areas for all three levels is 25 students or more (ideally, 30 students per class) [16]. Thus, the value of *N* affects the maximum number of demand points and, as a result, the number of vertices in the graph $K_{1,D_{I+1}}$.

Each edge of the graph e_k corresponds to its length $l_k > 0$ (for $e_k=(u_1,u_j)$ while $j=\overline{2,(i+1)}$ introduced designation ($l_{1,j}=l_k$). In this case, for any pair of vertices (1,*j*) ($j\neq 1$) it is possible to find the function of spatial (territorial) accessibility as a function of distance (or length) $L((x_1,y_1),(x_j,y_j))$ between vertex 1 with coordinates (x_1,y_1) and vertex j with coordinates (x_j,y_j). Then formula (1) will take the form:

$$\arg\min_{(x_1,y_1)} F(x_1,y_1) = \arg\min_{(x_1,y_1)} \sum_{j=2}^{I+1} \mu_j L((x_1,y_1),(x_j,y_j)). \qquad (3)$$

In the Law of Ukraine «On Complete General Secondary Education» it is mentioned, that «…every child has the right to receive education in the most accessible and closest institution to his place of residence" [16]. Therefore,

$$L((x_1,y_1),(x_j,y_j)) \le \Lambda \qquad (4)$$

for $\forall j=\overline{2,(i+1)}$ while $i=\overline{1,I}$.

Inequality (4) limits the real distance from students to the ENW object, directly affecting the location of the educational institution (x_1,y_1).

Note that for all the importance (even at the legislative level) of the concept of spatial accessibility, its unambiguous definition does not exist. Most often (for example, in [9, 12, 17]), accessibility is discussed in the context of social services, considering it from different points of view.

By spatial accessibility, we mean "... a set of indicators for quantitative assessment, reflecting the ease with which a person (the population of a city, a separate municipality, etc.) can reach a certain object from their place of residence using different modes of movement ... [18]". This provides convenient and fast connections with all functional areas of the settlement, leads to an increase in the quality of life and social activity, reducing socio-spatial inequality [17, 18]. In this context, better accessibility means the shortest travel time, the limit of which, depending on the number of inhabitants of the settlement, is determined by the Building Code 2.07.01–89 "Urban planning. Planning and development of urban and rural settlements".

Thus, the solution of the placement problem and the search for the coordinates of the educational institution (x_1,y_1) are possible by multiobjective optimization while minimizing the function (3) taking into account the restrictions on the number of students (2) (in the context of the availability of opportunities) and the distance from students to the educational institution (4) (in the context of spatial availability).

Function (3) corresponds to the *p*-median placement model, for which the analytical finding of coordinates is possible by solving a system of differential equations of the form [19, 20]:

$$\begin{cases} \dfrac{\partial F((x_1,y_1))}{\partial x_1} = 0; \\ \dfrac{\partial F((x_1,y_1))}{\partial y_1} = 0. \end{cases}$$

Note that the presence of a significant number of demand points D_i while simultaneously observing constraints (2) and (4) leads to certain computational difficulties and requires the use of special algorithms and methods for finding coordinates (x_1,y_1), for example, using the ideas of geospatial analysis.

DOI 10.15588/1607-3274-2023-3-13

The search for a solution to the formed placement problem (3) in the presence of constraints (2) and (4) is implemented as part of the information process of the ENW territorial community formation. Based on the requirements of the current legislation, we systematize the information flows of this process and, using the author's research methodology [21, 22], conceptually represent it in the form of a set-theoretic model of information flows:

$$I\_\Pr = (V, O, A, \psi, Z, \varphi), \quad (5)$$

the internal content of the sets of which is shown in Fig. 2.

In model (5), the set $V = \{v_1, v_2, v_3\}$ is converted to a set $O = \{o_1, o_2, o_3, o_4\}$ by successive implementation of set operations $A = \{a_1, a_2, a_3\}$, defined by a sequence [12, 13]:

Step 1. Assess the demographic situation and its trends in the planning area.

Step 2. Analyze the community's existing ENW.

Step 3. Choose the location of the educational institution.

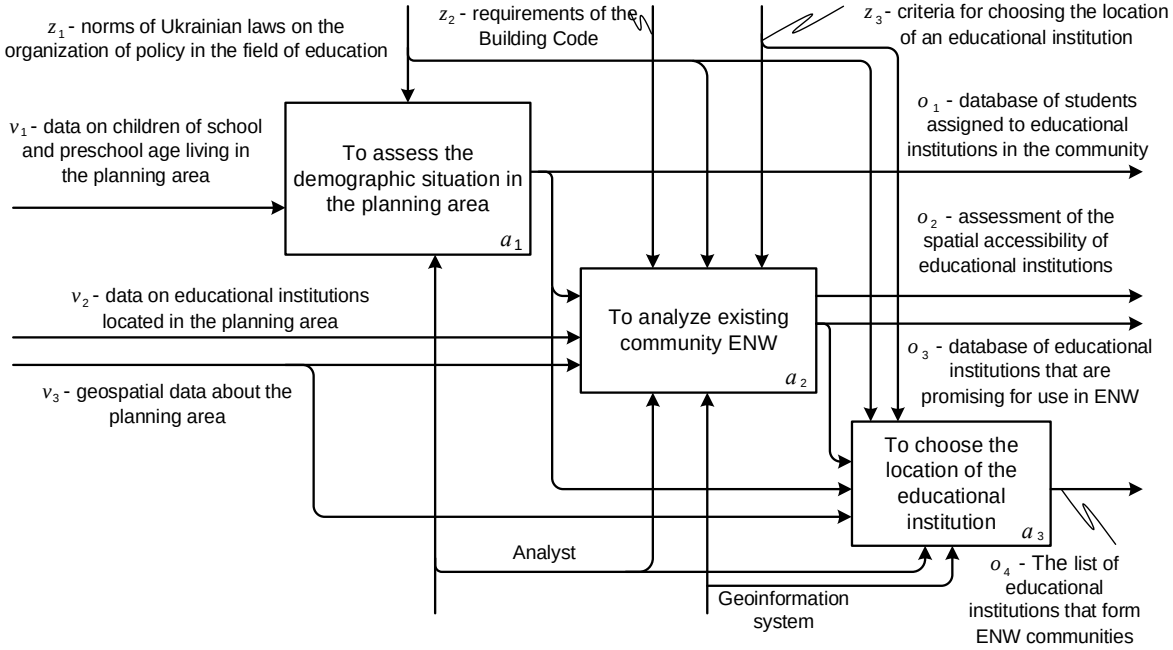


Figure 2 – IDEF0-model as a structure of information technology of forming the ENW territorial community

The set transformation is defined by the output function ψ – displaying the view [21, 22]:

$$\psi : A \times V \to O, \quad (6)$$

the implementation rules of which are disclosed in Table 1, where each row corresponds to one operation of the formation of an ENW territorial community a_n $(n = 1,...,3)$, and the column is one valid input element of the set v_m $(m = 1,...,3)$. The cell at the intersection of a row and a column contains an operation a_n, the execution of which occurs when an element arrives at the input v_m, and the output element o_q $(q = 1,...,4)$, which will appear when the operation is performed a_n [22].

In Table 1, for example, upon receipt of an input v_3 to get the output o_4 during operation a_3 is required. In this case, is a necessary output o_1, obtained as a result of the operation a_1, output o_2, obtained as a result of the operation a_2 by summarizing input data v_2 and v_3, and output o_3, obtained as a result of the operation a_2.

Table 1 – Exit Function Tabular View (6)

Operations of the set A	Input elements of the set V		
	v_1	v_2	v_3
a_1	$a_1 \vert o_1$	$a_2 \vert o_1$	$a_3 \vert o_1$
a_2	–	$a_2 \vert o_2$	$a_2 \vert o_2$
a_3	–	$a_2 \vert o_3$	$a_3 \vert o_4$

To normalize the internal content of the information flows of the model (5) when converting input data into output data, a set of regulatory documents is required. They are determined by a set $Z = \{z_1, z_2, z_3\}$. The normalization rules are set by the update function φ [21, 22]:

$$\varphi : V \times Z \to V, \quad (7)$$

which for each element of the set V specifies the implementation $v_m = \varphi(v_m, z_k)$ $\forall v_m \in V$, $\forall z_k \in Z$ while $m, k = 1,...,3$. For example, for operation a_2 function execution (7) allows you to create, supplement and

DOI 10.15588/1607-3274-2023-3-13

expand the database template of establishments that are promising for using in ENW. So, the implementation $\varphi(v_2, z_1)$ fulfills the requirement on the need to take into account information about the material base of the educational institution, the qualifications of its teaching staff, etc., the implementation of $\varphi(v_2, z_2)$ expands this template, supplementing it with information about the design capacity of the educational institution, age, etc., the implementation of $\varphi(v_3, z_2)$ requires taking into account the geospatial data of educational institutions, etc., in the database.

Summarizing the information flows, operations and functions of the model (5), in order to facilitate their perception while maintaining the rigor and formality of the representation, we depict it as an IDEF0-model (Fig. 2) [22, 23]. Based on the provisions of the functional modeling methodology, IDEF0-model determines the structure of IT of ENW territorial community formation. The implementation of proposed IT is possible by using the ArcGIS software suite by Esri.

Thus, the proposed IT, based on a consistent analysis of demographic, personnel, and infrastructural factors, will allow, when forming the ENW territorial community, to find a solution to problem (3), taking into account constraints (2) and (4), focusing on the real volumes of funding for the activities of educational institutions. The results obtained with the help of IT will make it possible to make informed decisions when choosing directions for optimization, for example, in conditions of limited financial support, when significant costs for maintaining schools with low occupancy can be reduced by reorganizing or reprofiling them [3, 15].

4 EXPERIMENTS

The possibility of using the proposed IT in the formation of a network of lyceums – institutions (structural divisions) that provide specialized secondary education (III level education) – is considered on the example of the communities of the Kharkiv region. At the same time, data on the city of Kharkiv and its educational institutions were excluded and were not considered when constructing ENW, which is associated with the specifics of the tasks of a large metropolis, which are not typical for other settlements [4, 8] of the region.

The purpose of the experiment was to assess the possibility of using spatial data to justify the choice of the location of an educational institution on the territory of the community in the formation of ENW. In particular, attention was focused on assessing the impact on the structure of the network element of the set Z – z_3 – criterion for choosing the location of the educational institution. In this case, such possible formulations were used z_3:

1. Lyceums are located in settlements with a population of at least 10,000 inhabitants;

2. Lyceums are located in the settlements of the districts of the territorial community, provided that the total population of the district is not less than 10,000 inhabitants.

Thus, according to the wording z_3 and taking into account the requirements of the Building Code 2.07.01–89, in expression (4) admissible value Λ for spatial accessibility was assumed to be 30 min. Due to the specificity of the information and the lack of accurate login data v_1 assessment of the demographic situation at the operation a_1 carried out using voluntary geographic information (VGI), the effectiveness of which for such tasks is confirmed in the works [12, 13, 23]. At the same time, the estimated number of lyceum students in the planning area, found by the formula [3, 12]:

$$N_{ENS} = N_{PA} \cdot \frac{97402}{2583141}, \quad (8)$$

is considered as a part of the number of residents in the planning area, taking into account the fact that the total population of the Kharkiv region as of January 1, 2022 is 2.583.141 people, including 15–18 years old (the age of applicants for specialized education) – 97.402 people.

In accordance with the requirements of The Law of Ukraine "On Complete General Secondary Education", lyceums must ensure the functioning of at least two classes in three areas of study at the level of specialized secondary education. Taking into account the cycles of the profile educational process and restrictions on the number of students in the class [1, 16], in expression (2) the value of N is taken equal to 500 people.

The following data were used as initial data:

– statistical data of the Main Department of Statistics of the Kharkiv region (http://kh.ukrstat.gov.ua/);

– data on educational institutions from the Education Management Information System of the Kharkiv region (https://kh.isuo.org/);

– VGI about the planning area from the OpenStreetMap (https://www.openstreetmap.org/) and Detsentralizatsiia (https://decentralization.gov.ua/) services.

As a result of the experiment, conclusions were drawn about the possibility of using geospatial analysis methods, as well as their effectiveness in the context of the reform of local self-government and the creation of new territorial communities, in particular, when solving problems:

– search for ways to improve ENW and possible directions of its development;

– analysis of the current state of ENW and assessment of decisions made related to spatial planning and the implementation of accessibility standards for specific places.

5 RESULTS

There are 19 settlements in the Kharkiv region (excluding Kharkiv) with a population of more than 10,000 people. Using VGI, according to formula (8), the estimated number of students was found and, taking into account the accepted value of N, the required number of lyceums was determined as 28. Their spatial distribution

DOI 10.15588/1607-3274-2023-3-13

on the territory of the Kharkiv region is shown in Fig. 3, while in the Bohodukhiv city there is 1 lyceum, in the Balakliia city – 2 lyceums, in the Izium city – 3 lyceums, in the Krasnohrad city – 1 lyceum, in the Slobozhanske urban-type settlement – 1 lyceum, in the Kivsharivka urban-type settlement – 1 lyceum, in the Kupiansk city – 2 lyceums, in the Lozova city – 4 lyceums, in the Pervomaiskyi city – 2 lyceum, in the Vysokyi urban-type settlement– 1 lyceum, in the Vovchansk city – 1 lyceum, in the Derhachi city – 1 lyceum, in the Liubotyn city – 1 lyceum, in the Merefa city – 1 lyceum, in the Nova Vodolaha urban-type settlement – 1 lyceum, in the Pisochyn urban-type settlement – 1 lyceum, in the Solonytsivka urban-type settlement – 1 lyceum, in the Zmiiv city – 1 lyceum, in the Chuhuiv city – 2 lyceums.

The spatial distribution of lyceums (Fig. 3) was found under the condition of uniform distribution of students throughout the territory of settlements. In case of their uneven settlement, the location of the lyceum is chosen in the places of maximum location of demand points D_i.

Figure 3 – A fragment of the distribution of lyceums across the territory of the Kharkiv region in settlements with a population of more than 10,000 people

In accordance with expression (3), for each lyceum, the distance to the expected demand points was found and the accessibility zones were determined (walking accessibility zone – 3 km, transport accessibility zone – 15 km and 30 min.) The results of the experiment are presented in Table 2.

Table 2 – Comparative results of lyceum accessibility assessments obtained using different placement criteria

Assessment parameter	Placement of lyceums in settlements with population		Comparison result, times
	More than 10.000 people	More and less 10.000 people	
Level of accessibility (%), including: walking distance (3 km) using transport distance (15 km) using transport time (30 minutes)	 0.95 16.46 47.6	 1.76 38.69 93.96	 ↑ 1.85 ↑ 2.35 ↑ 1.97
The area of the transport accessibility zone (30 min.), km^2	14 968.36	29434.36	↑ 1.97
Estimated number of lyceums in ENW	28	84	↑ 3.0

Taking into account the total number of inhabitants of the district (the second version of the wording) leads to the need to locate lyceums in settlements with a population of less than 10,000 people. In this case, the number of lyceums increases to 84. For example, in the Bohodukhiv district, instead of one lyceum in Bohodukhiv territorial community, 9 lyceums are created and operate, providing specialized secondary education in Bohodukhiv, Valky, Zolochiv, Krasnokutsk territorial communities. This allows you to increase the total area of the area of transport accessibility of the lyceums at the region to 29434,36 km^2, reaching the level of accessibility of specialized secondary education to almost 94 % (Table 2).

The final geospatial assessment of the accessibility zones of lyceums, obtained using the developed information technology for different formulations z_3, is presented in Fig. 4.

6 DISCUSSION

Analysis of statistical data on the distribution of the population of the Kharkiv region showed that the number of urban residents is more than 5 times higher than the population of rural areas. This confirms the correctness of the idea of locating educational institutions in urban areas of the region. However, the focus on placing lyceums only in settlements with a population of at least 10,000 people (and in the first edition of the law [16] – at least 50,000 people) leads to a significant violation of the requirement for the availability of specialized education. The zone of regulated transport accessibility covers only 47.6% of the Kharkiv region. (Table 2), which can lead to the problem of the lack of places in the created lyceums for all students (including those who are living in neighboring settlements) due to exceeding their design capacity (violation of constraint (2)).

Due to the significant distance between the recommended settlements (Fig. 5), areas with difficult access to specialized education are formed, which is especially noticeable in areas located closer to the borders of the region

DOI 10.15588/1607-3274-2023-3-13

(violation of the constraint (4)). In this case, the resulting economic effect from the reduction in the number of lyceums is leveled by the need to create and maintain boarding schools for students from remote areas. These problems explain the public outcry and negative attitude towards the first edition of the law “On Complete General Secondary Education” and the subsequent withdrawal of the discriminatory norm “on the placement of lyceums” in its new edition.

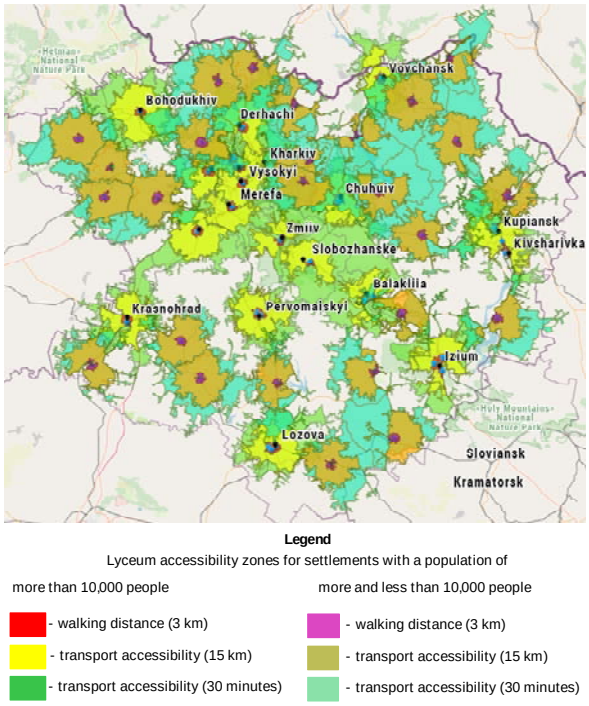


Figure 4 – An example of a geospatial assessment of the accessibility of lyceums obtained as a result of the application of the developed information technology

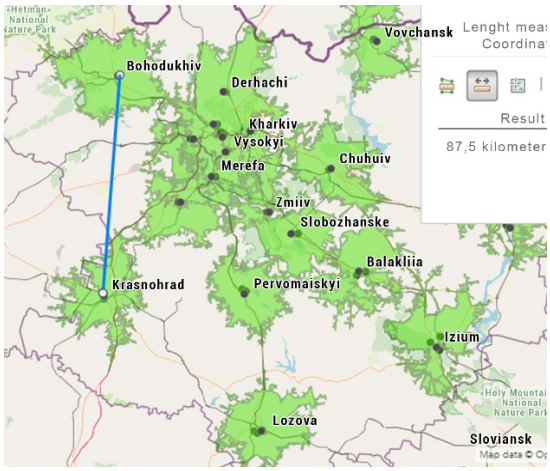


Figure 5 – Estimation of the distance between two neighboring lyceums in Bohodukhiv and Krasnohrad, Kharkov region

On the other hand, in cities located closer to Kharkiv, access zones overlap each other (Fig. 5). As a result, the problem of small-group classes arises, which, in addition to the emergence of financial support problems, is the cause of a decrease in the quality of education, a decrease in the number of education profiles, etc.

The placement of lyceums in the settlements of the districts of the territorial community in accordance with the second formulation of the criterion allows, on average, to increase the total accessibility zone by 2 times (Table 2), reduce the burden on cities, provide children with comfortable conditions and reduce the distance between lyceums so that students can choose the right lyceum which is closer to where you live. At the same time, if the idea of locating educational institutions in large settlements of the region’s districts is observed, the recommended number of lyceums is increased by 3 times (Table 2).

CONCLUSIONS

The urgent problem of developing scientific and methodological support of information support for the process of forming the educational network of the community in order to find ways to improve it in the implementation of community development projects has been solved.

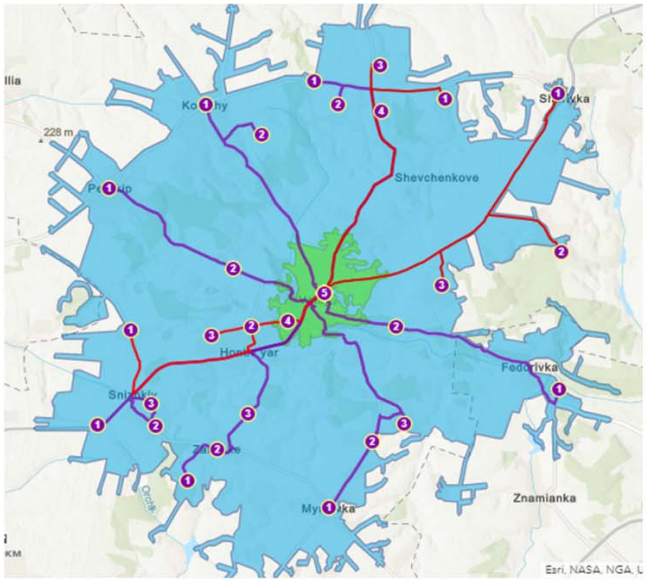
The scientific novelty of obtained results is that the adaptation of the *p*-median model to solve the problem of creating an ENW in the presence of constraints made it possible to combine demographic and infrastructural factors in making decisions about choosing the location of educational institutions. As a tool for solving the problem of choosing the location of educational institutions in ENW, the methodology for studying information processes was further developed by clarifying the set-theoretic model of information flows of the process. As a result of the clarification, an information technology for the formation of an ENW community was developed, which, based on an adapted *p*-median model and geospatial analysis methods, explains how a combination of various factors (demographic, infrastructural, personnel, economic) are processed and presented to support decision-making when formation of the educational network of the community and search for ways to improve it.

The practical significance of the obtained results is that the representation of the IT structure based on the IDEF functional modeling standard allows one to proceed to the creation of information systems for the formation of ENW based on spatial data. The conducted experiment on studying the capabilities of the developed IT showed its effectiveness in solving classical problems of accommodation, taking into account the accepted restrictions on the capacity of the educational institution and its spatial accessibility. At the same time, additional tools for solving the problems of ensuring the accessibility of education appear, in particular, the task of transporting students to the place of study and home. Using data about students, supplemented by geospatial information about their places of residence, and geospatial data about the locations of lyceums, the process implementation mechanism were proposed in the IDEF0-model – GIS – makes it possible to form student transportation routes, focusing on the al-

DOI 10.15588/1607-3274-2023-3-13

lowable value of Λ (Fig. 6). At the same time, all the necessary information for solving this problem is in the databases of outputs o_1 and o_3. The results of the experiment make it possible to recommend the proposed IT for the formation of a competitive educational network in accordance with the Laws of Ukraine, as well as to determine the effective conditions for its application.

Prospects for further research are in adjusting the data obtained in accordance with the actual occupancy of the classes of lyceums located in the planning territories. The use of actual data in the proposed IT will allow optimizing the network of lyceums, taking into account the possibility of organizing the functioning of parallel classes, transporting students from other settlements, etc., while the procedure for choosing the location of lyceums will remain unchanged.



N route	Start of the route (town)	End of the route (lyceum location)	Number of stops	Travel time, minutes	Route length, km
1	Perekip	Valky, Lyceum named after O. Maselsky	3	26	11,75
2	Hryntsiv Rih		4	34	16,19
3	Khvorostove		4	46	17,26
4	Koviahy		3	29	14,86
5	Sharivka		4	44	25,05
6	Fedorivka		3	24	13,12
7	Mynakivka		4	30	9,26
8	Buhaivka		4	30	13,19
9	Yaseneve		3	31	14,62
10	Kantakuzivka		4	46	17,36

Note: Routes are marked red, the implementation of which does not allow to achieve the value of spatial accessibility at the accepted (recommended) level without applying the additional organizational measures.

Figure 6 – An example of the routes for transporting students to the Valky lyceum of O. Maselsky obtained with the help of developed information technology

Author Contributions: Review and analysis of references, S. Danshyna, A. Nechausov; development of conceptual provisions and methodology of research, S. Danshyna; validation, A. Nechausov; analysis of research results, S. Danshyna, A. Nechausov; writing–original draft, S. Danshyna; writing–review and editing, A. Nechausov; project administration, S. Danshyna. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS

The work is supported by the state budget scientific research project of National Aerospace University "KhAI" "Methodological bases of distributed systems for monitoring environmental objects creating" (state registration number 0122U002298) and with the support of the Regional Center of Space Monitoring of the Earth "Slobozhanshchina".

Authors express my special gratitude to Iryna Buhai for experimental confirmation of the effectiveness of the developed scientific-methodological basis.

REFERENCES

1. Kremen V. H., Liashenko O. I. and Lokshyna O. I. Zahalna serednia osvita Ukrainy v konteksti osvity krainy Yevropy: tryvalist i struktura, *Bulletin of the National Academy of Pedagogical Sciences of Ukraine*, 2020, No. 2 (2), pp. 1 – 10. DOI: https://doi.org/10.37472/2707-305X-2020-2-2-14-1 (In Ukrainian).
2. Readiness for the Future of Production Report 2018; ed. A. T. Kearney. Switzerland, 2018.
3. Bilyk I., Diatlenko S., Kohut I., Oliinyk V., Putsova A. and Ruda O. Kryterii formuvannia spromozhnoi osvitnoi merezhi v OTH. Kyiv, 2019, 68 p. (In Ukrainian).
4. Jing Ch., Zhou W. and Qian Yu. Spatial disparities of social and ecological infrastructures and their coupled relationships in cities, *Sustainable Cities and Society*, 2022, Vol. 86, article No. 104117. DOI: https://doi.org/10.1016/j.scs.2022.104117.
5. Kompil M., Jacobs-Crisioni C., Dijkstra L. and Lavalle C. Mapping accessibility to generic services in Europe: A market-potential based approach, *Sustainable Cities and Society*, 2019, Vol. 47, article no. 101372. DOI: https://doi.org/10.1016/j.scs.2018.11.047.
6. Agrawal S. and Gupta R. D. School mapping and geospatial analysis of the schools in Jasra development block of India, in *Proc. of XXIII Int. Congress : International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2016, Vol. 41, pp. 145–150. DOI: 10.5194/isprsarchives-XLI-B2-145-2016.
7. Khavarian-Garmsir A. R., Sharifi A. and Sadeghi A. The 15-minute city: Urban planning and design efforts toward creating sustainable neighborhoods, *Cities*, 2023, Vol. 132, article No. 104101. DOI: https://doi.org/10.1016/j.cities.2023.104101.

DOI 10.15588/1607-3274-2023-3-13

8. Ogryzek M., Podawka K. and Cienciala A. Geospatial tools in the analyzes of land use in the perspective of the accessibility of selected educational services in Poland, *Land Use Policy*, 2022, Vol. 122, article No. 106373. DOI: https://doi.org/10.1016/j.landusepol.2022.106373.
9. Milbert A., Breuer I.-M., Rosik P., Stepniak M., and Velasco X. Accessibility of services of general interest in Europe, *Romanian Journal Regional Science*, 2013, Vol. 7, pp. 37 – 65.
10. Gao Y., He Q., Liu Y., Zhang L., Wang H. and Cai E., Imbalance in spatial accessibility to primary and secondary schools in China: Guidance for education sustainability, *Sustainability*, 2016, Vol. 8, Issue 12. Available at: https://www.mdpi.com/2071-1050/8/12/1236. DOI: https://doi.org/10.3390/su8121236.
11. Rao J. and Ye J. From a virtuous cycle of rural-urban education to urban-oriented rural basic education in China: An explanation of the failure of China's Rural School Mapping Adjustment policy, *Journal of Rural Studies*, 2016, Vol. 47, part B, pp. 601–611. DOI https://doi.org/10.1016/j.jrurstud.2016.07.005.
12. Danshyna S. Yu. and Nechausov A. S. Solution of the problem of placing medical facilities in city development projects, *Radio Electronics, Computer Science, Control*, 2020, No. 3 (54). pp. 138–149. DOI: https://doi.org/10.15588/1607-3274-2020-3-12.
13. Rodriguez-Segura D. and Kim B. H. The last mile in school access: Mapping education deserts in developing countries, *Development Engineering*, 2021, Vol. 6, article no. 100064. DOI: https://doi.org/10.1016/j.deveng.2021.100064.
14. Mustapha O.O., Akintunde O. S., Alaga A. T., Badru R. A., Ogbole J. O., Samuel P. O. and Samuel S. A. Spatial distribution of primary schools in Ilorin west local government area, Kwara State, Nigeria, *Journal of Scientific Research & Reports*, 2016, Vol. 9(6), Issue 2, pp. 1–10. DOI: https://doi.org/10.9734/JSRR/2016/22128.
15. *On Education*: Law of Ukraine of September 05, 2017, No. 2145-VIII, Bulletin of the Verkhovna Rada, 2017, No. 38–39, Article. 380, P. 5 (In Ukrainian).
16. *On Complete General Secondary Education*: Law of Ukraine of January 16, 2020, no. 463-IX, Bulletin of the Verkhovna Rada, 2020, No. 31, article 226, pp. 5 (In Ukrainian).
17. Curtis C. and Scheurer J. Planning for sustainable accessibility: The implementation challenge, *Progress in Planning*, 2010, Vol. 74, Issue 2, pp 53–106. DOI: https://doi.org/10.1016/j.progress.2010.05.001.
18. Chapman S. and Weir D. Accessibility planning methods. New Zealand, Booz and Comp. Ltd, 2008.
19. Levanova T. V. and Gnusarev A. Yu. Algoritmyi s chereduyuschimisya okrestnostyami dlya konkurentnoy zadachi razmescheniya predpriyatiy s gibkim sprosom, *Journal of Applied and Industrial Mathematics*, 2020, Vol. 27, Issue 4, pp. 80–103. DOI: https://doi.org/10.33048/daio.2020.27.575 (In Russian).
20. Drezner T. and Drezner Z. The gravity p-median model, *European Journal of Operational Research*, 2007, Vol. 179, Issue 3, pp. 1239–1251. DOI: https://doi.org/10.1016/j.ejor.2005.04.054.
21. Danshyna S. Yu., Nechausov A. S. and Andrieiev S. M. Information technology of transport infrastructure monitoring based on remote sensing data, *Radio Electronics, Computer Science, Control*, 2022, No. 4 (63), pp. 86–97. DOI: https://doi.org/10.15588/1607-3274-2022-4-7.
22. Danshyna S. and Cheranovskiy V. Formalizing the land inventory process for information support of land projects management, *Radioelectronic and Computer Systems*, 2022, No. 3, pp. 97–107. DOI: https://doi.org/10.32620/reks.2022.3.03.
23. Goodchild M. F., and Li L. Assuring the quality of volunteered geographic information, *Spatial Statistics*, 2012, Vol. 1, pp. 110–120. DOI: https://doi.org/10.1016/j.spasta.2012.03.002.

Received 06.03.2023.
Accepted 15.08.2023.

УДК 711.57:004.9

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ФОРМУВАННЯ ОСВІТНЬОЇ МЕРЕЖІ ТЕРИТОРІАЛЬНОЇ ГРОМАДИ

Даншина С. Ю. – д-р техн. наук, професор кафедри геоінформаційних технологій та космічного моніторингу Землі, Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна.

Нечаусов А. С. – канд. техн. наук, доцент кафедри геоінформаційних технологій та космічного моніторингу Землі, Національний аерокосмічний університет «Харківський авіаційний інститут», Харків, Україна.

АНОТАЦІЯ

Актуальність. Органам місцевого самоврядування надано значні повноваження щодо прийняття рішень у сфері освіти. Це потребує розроблення чітких і зрозумілих інструментів, які допоможуть сформувати мережу освітніх закладів для забезпечення якісного надання освітніх послуг. Об'єктом дослідження є процес формування освітньої мережі громади при реалізації проєктів розвитку під час здійснення реформи місцевого самоврядування. Метою роботи є підвищення об'єктивності рішень, що приймаються при формуванні освітньої мережі територіальної громади, узагальнюючи демографічні, кадрові, інфраструктурні, економічні чинники при виборі місця розміщення її об'єктів.

Метод. Дослідження класичної задачі про розміщення та її подальша адаптація до реальних завдань, що виникають при реалізації освітньої реформи, дозволили подати освітню мережу територіальної громади як сукупність незалежних повних дводольних графів. У цьому випадку для вирішення завдання вибору місця розміщення учбового закладу на мережі розроблено інформаційну технологію формування освітньої мережі громади. Грунтуючись на адаптованій *p*-медіанній моделі та методах геопросторового аналізу, узагальнюючи вимоги чинного законодавства, сформовано множину вхідних і вихідних параметрів інформаційної технології та множину її операцій. Подання структури ІТ у вигляді IDEF0-моделі наочно пояснює, як обробляють та узагальнюють сукупність різних факторів під час формування рішень при створенні освітньої мережі та пошуку шляхів її вдосконалення.

Результати. Розроблену інформаційну технологію досліджено під час вирішення завдання формування мережі ліцеїв у Харківській області з використанням геопросторової інформації, відкритих статистичних даних і даних про ліцеї. Запропо-

DOI 10.15588/1607-3274-2023-3-13

новані варіанти розміщення ліцеїв дають змогу досягти загального рівня доступності профільної середньої освіти майже 94%. У запропонованій ІТ є додаткові інструменти вирішення завдання підвезення учнів до місця навчання та додому.

Висновки. Проведені експерименти підтвердили працездатність запропонованої ІТ. Узагальнення отриманих результатів дає змогу рекомендувати її до використання на практиці при вирішенні завдань аналізу поточного стану освітньої мережі, пошуку шляхів її вдосконалення та можливих напрямів розвитку, а також оцінювання рішень, пов'язаних із просторовим плануванням мережі.

КЛЮЧОВІ СЛОВА: *p*-медіана модель, геопросторовий аналіз, модель інформаційних потоків процесу, IDEF0-модель, оцінки доступності.

ЛІТЕРАТУРА

1. Кремень В. Г. Загальна середня освіта України в контексті освіти країни Європи: тривалість і структура / В. Г. Кремень, О. І. Ляшенко, О. І. Локшина // Вісник НАПН України. – 2020. – № 2(2). – С. 1–10. DOI: https://doi.org/10.37472/2707-305X-2020-2-2-14-1.
2. Readiness for the Future of Production Report 2018 / ed. A. T. Kearney. – Switzerland, 2018. – 266 p.
3. Критерії формування спроможної освітньої мережі в ОТГ / [І. Білик, С. Дятленко, І. Когут та ін.]. – Київ, 2019. – 68 с.
4. Jing Ch. Spatial disparities of social and ecological infrastructures and their coupled relationships in cities / Ch. Jing, W. Zhou, Yu. Qian // Sustainable Cities and Society. – 2022. – Vol. 86. – Article No. 104117. DOI: https://doi.org/10.1016/j.scs.2022.104117.
5. Mapping accessibility to generic services in Europe: A market-potential based approach / [M. Kompil, C. Jacobs-Crisioni, L. Dijkstra, C. Lavalle] // Sustainable Cities and Society. – 2019. – Vol. 47. – Article No. 101372. DOI: https://doi.org/10.1016/j.scs.2018.11.047.
6. Agrawal S. School mapping and geospatial analysis of the schools in Jasra development block of India / S. Agrawal, R. D. Gupta // International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences: proceeding. – 2016. – Vol. 41. – P. 145–150. DOI: 10.5194/isprsarchives-XLI-B2-145-2016.
7. Khavarian-Garmsir A. R. The 15-minute city: Urban planning and design efforts toward creating sustainable neighborhoods / A. R. Khavarian-Garmsir, A. Sharifi, A. Sadeghi // Cities. – 2022. – Vol. 132. – Article no. 104101. DOI: https://doi.org/10.1016/j.cities. 2022.104101.
8. Ogryzek M. Geospatial tools in the analyzes of land use in the perspective of the accessibility of selected educational services in Poland / M. Ogryzek, K. Podawka, A. Cienciala // Land Use Policy. – 2022. – Vol. 122. – Article no. 106373. DOI: https://doi.org/10.1016/j.landusepol.2022.106373.
9. Accessibility of services of general interest in Europe / [A. Milbert, I.-M. Breuer, P. Rosik et al.] // Romanian Journal Regional Science. – 2013. – Vol. 7. – P. 37–65.
10. Imbalance in spatial accessibility to primary and secondary schools in China: Guidance for education sustainability [Electronic resource] / [Y. Gao, Q. He, Y. Liu et al.] // Sustainability. – 2016. – Vol. 8, Issue 12. Access mode: https://www.mdpi.com/2071-1050/8/12/1236. DOI: https://doi.org/10.3390/su8121236.
11. Rao J. From a virtuous cycle of rural-urban education to urban-oriented rural basic education in China: An explanation of the failure of China's Rural School Mapping Adjustment policy / J. Rao, J. Ye // Journal of Rural Studies. – 2016. – Vol. 47, Part B. – P. 601–611. DOI: https://doi.org/10.1016/j.jrurstud.2016.07.005
12. Danshyna, S. Yu. Solution of the problem of placing medical facilities in city development projects / S. Yu. Danshyna, A. S. Nechausov // Radio Electronics, Computer Science, Control. – 2020. – № 3 (54). – P. 138–149. DOI: https://doi.org/10.15588/1607-3274-2020-3-12.
13. Rodriguez-Segura, D. The last mile in school access: Mapping education deserts in developing countries / D. Rodriguez-Segura, B. H. Kim // Development Engineering. – 2021. – Vol. 6. – Article No. 100064. DOI: https://doi.org/10.1016/j.deveng.2021.100064
14. Spatial distribution of primary schools in Ilorin west local government area, Kwara State, Nigeria / [O. O. Mustapha, O. S. Akintunde, A. T. Alaga et al.] // Journal of Scientific Research & Reports. – 2016. – Vol. 9(6), Issue 2. – P. 1–10. DOI: https://doi.org/10.9734/JSRR/2016/22128.
15. Про освіту: Закон України від 5 вересня 2017 р., № 2145-VIII // Відомості Верховної Ради України. – 2017. – № 38–39. – Ст. 380. – С. 5.
16. Про повну загальну середню освіту: Закон України від 16 січня 2020 р., № 463-IX // Відомості Верховної Ради України. – 2020. – № 31. – Ст. 226. – С. 5.
17. Curtis C. Planning for sustainable accessibility: The implementation challenge / C. Curtis, J. Scheurer // Progress in Planning. – 2010. – Vol. 74, Issue 2. – P. 53–106. DOI: https://doi.org/10.1016/j.progress.2010.05.001.
18. Chapman S. Accessibility planning methods / S. Chapman, D. Weir. – New Zealand: Booz and Comp. Ltd, 2008. – 110 p.
19. Леванова Т. В. Алгоритмы с чередующимися окрестностями для конкурентной задачи размещения предприятий с гибким спросом / Т. В. Леванова, А. Ю. Гнусарев // Дискретный анализ и исследование операций. – 2020. – Т. 27, Вып. 4. – С. 80–103. DOI: https://doi.org/10.33048/daio.2020.27.575.
20. Drezner T. The gravity p-median model / T. Drezner, Z. Drezner // European Journal of Operational Research. – 2007. – Vol. 179, Issue 3. – P. 1239–1251. DOI: https://doi.org/10.1016/j.ejor.2005.04.054.
21. Danshyna S. Yu. Information technology of transport infrastructure monitoring based on remote sensing data / S. Yu. Danshyna, A. S. Nechausov, S. M. Andrieiev // Radio Electronics, Computer Science, Control. – 2022. – № 4 (63). – P. 86–97. DOI: https://doi.org/10.15588/1607-3274-2022-4-7.
22. Danshyna, S. Formalizing the land inventory process for information support of land projects management / S. Danshyna, V. Cheranovskiy // Radioelectronic and computer systems. – 2022. – № 3. – P. 7–19. DOI: https://doi.org/10.32620/reks.2022.3.01
23. Goodchild, M. F. Assuring the quality of volunteered geographic information / M. F. Goodchild, L. Li // Spatial Statistics. – 2012. – Vol. 1. – P. 110–120. DOI: https://doi.org/10.1016/j.spasta.2012.03.002.

DOI 10.15588/1607-3274-2023-3-13

UDC 519.87:004.02

ON THE RECURSIVE ALGORITHM FOR SOLVING THE TRAVELING SALESMAN PROBLEM ON THE BASIS OF THE DATA FLOW OPTIMIZATION METHOD

Ivohin E. V. – Dr. Sc., Professor, Professor of the Department of System Analysis and Decision Support Theory, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.

Gavrylenko V. V. – Dr.Sc., Professor, Professor of the Department of Information Systems and Technologies, National Transport University, Kyiv, Ukraine.

Ivohina K. E. – PhD Student of the Department of Information Systems and Technologies, National Transport University, Kyiv, Ukraine.

ABSTRACT

Context. The article considers a technique for the sequential application of flow schemes for distributing a homogeneous resource for solving the traveling salesman problem, which is formulated as the problem of finding a route to visit a given number of cities without repetitions with a minimum duration of movement. The task of formalizing the algorithm for solving the traveling salesman problem by the method of streaming resource distribution using the backtracking scheme is posed. The use of Orlin's method to optimize the flow distribution on the graph is proposed.

Objective. The goal of the work is to develop an algorithm for solving the traveling salesman problem based on the implementation of the method of streaming resource distribution and the backtracking scheme with the minimum duration of movement along the route.

Method. This paper proposes a method for solving the traveling salesman problem by the method of streaming resource distribution with the backtracking scheme. A scheme for formalizing the procedure for solving the traveling salesman problem with the minimum duration of movement along the route is described. A variant of accelerating the speed of the developed algorithm is proposed, which consists in using a greedy technique in the procedure for selecting route sections: planning each subsequent stage of movement is determined based on the choice of the fastest direction of movement. The results of the proposed algorithm for calculating solutions to the traveling salesman problem with minimization of the duration of movement are presented, the obtained solutions are compared with the solutions found by other exact and heuristic methods.

Results. The method for solving the traveling salesman problem using the method of streaming resource allocation and using the backtracking scheme is developed. A variant of accelerating the speed of the developed algorithm is proposed, which consists in using a greedy technique in the procedure for selecting route sections: planning each subsequent stage of movement is determined based on the choice of the fastest direction of movement. The application of the greedy approach makes it possible to obtain a constructive scheme for solving the traveling salesman problem. The results of the proposed algorithm for calculating solutions to the traveling salesman problem with minimization of the duration of movement are presented, the obtained solutions are compared with the solutions found by other exact and heuristic methods.

Conclusions. The paper considers a method for formalizing the algorithm for solving the traveling salesman problem using the method of streaming resource allocation and the backtracking scheme. The use of Orlin's method to optimize the flow distribution on the graph is proposed. The scheme of formalization of the procedure for using the method with the implementation of the backtracking scheme for solving the traveling salesman problem with the minimum duration of movement along the route is briefly described. A variant of accelerating the speed of the developed algorithm is proposed.

KEYWORDS: traveling salesman problem, resource allocation method, recursive backtracking scheme, greedy approach.

ABBREVIATIONS

TSP is a traveling salesman problem.

NOMENCLATURE

t is cyclic permutation of numbers;

$j_1,..., j_n$ are different city numbers;

n is a number of cities;

c_{ij} are the travel time between all pairs of vertices;

C is a matrix of c_{ij}, $i, j=\overline{1,n}$;

i, j, k are the indexes;

I is a the set of vertex indices;

X is a binary matrix of transitions between vertices x_{ij}, $i, j=\overline{1,n}$;

x_{ij} are the elements of matrix X, which equal to 0 or 1;

$G(V,E)$ is a graph;

V is a a non-empty set of vertices;

E is a set of edges;

v_i is a vertex of graph, $i=\overline{1,N}$;

N is a number of vertices of the graph;

e_i is a edge of graph, $i=\overline{1,M}$;

M is a number edges of the graph;

V_s is a subset of initial nodes of the graph;

V_p is a subset of intermediate nodes of the graph;

V_e is a subset of final nodes;

DOI 10.15588/1607-3274-2023-3-14

$F(i)$ is a subset of edges of the graph coming out of the i-th vertex;

e'_i time distribution coefficients, $i=\overline{1,K}$;

K is a the power of subset vertices V_d;

V_d is a set that unites sets V_s and V_p;

$f(\cdot)$ is a criterion for optimization;

J_k is a power of $F(k)$;

N_1 is a power of V_s;

N_2 is a power of V_p;

N_3 is a power of V_e;

$t_k(\cdot)$ is a element of vector-function that determines the amount of time spent on moving to the *k*-th vertex of the graph, $k=\overline{1,N_3}$;

H^{in} is a matrix for the forward flows of dimension $N\times M$;

H^{out} is a matrix for backward flows of dimension $M\times N$;

S^r is a incidence matrix $N\times N$ for a path of multiplicity *r*;

r is a number of edges through which there is a path from vertex v_i to vertex v_j;

$w'_j(\cdot)$ is a element of vector-function of dimension *N* that determines the amount of the time resource formed by the initial vertices;

$\gamma_R^j(\cdot)$ is a element of the vector-function of dimension *N*;

$\gamma_1^m(\cdot)$ is a element of the vector-function of dimension *N*;

$\beta(\cdot)$ is a matrix-function;

Q is a vector of dimension N_3 whose elements determine the numbers vertices of the final points of the movement.

INTRODUCTION

Recently, most global companies have experienced disruptions in logistics caused by the pandemic and the war in Ukraine. Due to the sanctions and events related to the pandemic, managers of logistics companies have experienced serious disruptions in determining the routes and volumes of transportation, as the mentioned processes have exposed the weaknesses of the traditional existing supply chains in logistics.

A lack of vertical vision of manufacturing processes and connections, outdated demand management processes, insufficient resilience to changes in demand, and unexpected disruptions due to reliance on manual efforts in logistics operations have disrupted the supply chain.

Logistics companies are forced to analyze their logistics processes. It is clear that changes in customer behavior and expectations are unlikely to address these unexpected logistics challenges, as shoppers expected faster delivery and easier product tracking.

It is becoming clear that companies need to quickly optimize their logistics management. Depending on the task at hand, there are many different mathematical approaches to various logistical problems, such as linear programming, network optimization, decision analysis, genetic algorithms, and so on.

Logistics problems experience their own difficulties, some of which are solved thanks to the work of the management department, while others involve the analysis and optimization of logistics operations, including planning, coordination and control of the movement and storage of goods, services and information, optimization of network flows [1–3]. Simulation modeling methods and models allow you to create computer models of a logistics system and use them to test various scenarios and optimize system performance.

Attracting mathematical approaches to solving logistics problems is becoming widespread, the specific content of which depends on the nature of the problem and the available data. Sometimes it is possible to find atypical methods for solving known problems, one of which is the traveling salesman problem.

The object of study is the process of optimal route search for the traveling salesman problem with a minimum duration of movement.

The subject of study is the development of the efficient algorithm for solving for solving the traveling salesman problem by the method of streaming resource distribution using the backtracking scheme.

The purpose of the work is to develop an algorithm for solving the traveling salesman problem based on the implementation of the method of streaming resource distribution and the backtracking scheme with the minimum duration of movement along the route

1 PROBLEM STATEMENT

According to the content of the traveling salesman problem (TSP, Traveling Salesman Problem), it is necessary to create a route of movement within a given set of interconnected points (bridges) that form the transport network of a particular region [4]. A feature of the problem is that the route must contain all the points specified in the task, and each of the points must be visited no more than once. It is clear that such trips take a lot of time, so it is logical that it is necessary to plan the route in such a way that the distance to be covered, or the time to overcome it, is minimal (finding the path with the least cost can also be considered as a criterion).

The traveling salesman problem is a combinatorial problem that can be solved using mathematical programming methods. To reduce the problem to a general form, we number the cities by numbers (1, 2, 3, ..., *n*), and describe the traveling salesman's route by a cyclic permutation of numbers $t=(j_1, j_2,..., j_n, j_1)$, where all $j_1,..., j_n$ are different numbers. The number j_1, repeated from the very beginning and at the end, shows that the permutation is cyclic [5].

The set of cities can be considered as the vertices of some graph with given distances (or travel time) between

DOI 10.15588/1607-3274-2023-3-14

all pairs of vertices c_{ij} that form the matrix $C=(c_{ij})$, $i,j=\overline{1,n}$. We assume that the matrix is symmetric. The formal problem then is to find the shortest route (in time or length) t that goes through each city and ends at the starting point. In this formulation, the problem is called the closed traveling salesman problem, which is a well-known mathematical integer programming problem.

Let us formulate a mathematical model of the TSP problem. Let $I=\{1,...,n\}$ be the set of vertex indices of the problem graph. The objective function is the total distance or time of the route, including all the vertices of the task graph. The parameters of the problem are the elements of the matrix $C=(c_{ij})$, $i,j\in I$.

Shift tasks are elements of the binary matrix of transitions between vertices $X=\{x_{ij}\}$, $i,j\in I$, which are equal to 1 if there is an edge (v_i,v_j) in the constructed route for the task, 0 otherwise [6]. The shortest route in terms of distance or time is optimal:

$$\sum_{i\in I}\sum_{j\in I,\,j\neq i} c_{ij}x_{ij}\to\min \qquad (1)$$

with constraints

$$\sum_{j\in I,\,j\neq i} x_{ij}=1,\ i\in I,$$
$$\sum_{i\in I,\,j\neq i} x_{ij}=1,\ j\in I, \qquad (2)$$
$$v_i-v_j+nx_{ij}\le n-1,\ 1\le i\neq j\le n.$$

The last inequality ensures the connectivity of the vertex traversal route; it cannot consist of two or more unconnected parts.

2 REVIEW OF LITERATURE

Algorithms that allow solving the problem of finding the optimal route are divided into exact and heuristic. Exact methods guarantee finding the optimal solution to the problem in a certain time or taking into account certain resource constraints. In this case, the search for solutions is based on optimization methods such as linear programming, dynamic programming, or the branch and bound method [7]. However, it is expedient to use exact methods only for small-scale problems (for example, for the purpose of primary design of a small-sized transport network), since their implementation requires large computing power.

On the other hand, heuristic methods are algorithms that do not guarantee finding an optimal solution, but are aimed at quickly finding a locally optimal solution. Traditionally, "trial and error" approaches, such as random search or greedy algorithm, are used to quickly explore the solution space and find a promising solution [8]. Heuristics are more flexible and can be applied to larger problems, but the solution they offer may not be optimal. Among such heuristic methods, attention should also be paid to methods that imitate biological (ant colony algorithm and genetic algorithm [9, 10]) or physical processes (imitation annealing [11]).

3 MATERIALS AND METHODS

When forming a route, it is necessary to pay attention to the fact that each subsequent stage of movement can be chosen based on the consistent use of methods for optimizing the distribution of a homogeneous resource, one of the most effective among which is the Orlin method [12]. Then the problem of this study can be formulated by formalizing the algorithm for solving the traveling salesman problem using the method of streaming resource allocation and using a backtracking scheme.

Consider the application of the method for our problem. This method allows solving the problem of distribution of a homogeneous resource with intermediate points in the form of a directed graph without loops and parallel edges, given by a set of a non-empty set of vertices and a set of edges

$$E\subset\{v_i,v_j\}=\langle V,E\rangle,\ G(V,E)=\langle V,E\rangle,\ V\neq\varnothing,\ v_i,v_j\in V,i\neq j \qquad (3)$$

where $V=\{v_1,v_2,...,v_N\}$, $E=\{e_1,e_2,...,e_M\}$, N and M are the total number of vertices and edges of the graph, respectively. It is assumed that the set V of graph vertices $G(V,E)$ is represented by a set of non-intersecting subsets:

1. V_s – a subset of initial nodes (vertices) of the graph;

2. V_p – a subset of intermediate nodes (vertices) of the graph;

3. V_e – a subset of final nodes (vertices),

that is $V=V_s\cup V_p\cup V_e$, provided that $(V_s\cup V_p)\cap V_e=\varnothing$ and $|V_s|=N_1$, $|V_p|=N_2$, $|V_e|=N_3$, $N=N_1+N_2+N_3$, and the weight of the ribs is understood as the time to overcome the corresponding stage of the route.

Let's denote $V_d=V_s\cup V_p$. Then the weight of the edges from the set E emanating from the vertices of the subset V_d is determined by the value $E'=\{e'_1,e'_2,...,e'_K\}$ of the corresponding time distribution coefficients for the route, where K – the number of vertices of the graph $G(V,E)$ belonging to the subset V_d, i.e. $|V_d|=K=|V_s|+|V_p|=N_1+N_2$. Let be $F(i)\subset E$ – a subset of edges of the graph $G(V,E)$ coming out of the i-th vertex, with $E=\bigcup_{i=1}^{N}F(i)$ and $\bigcap_{i=1}^{N}F(i)=\varnothing$.

Then the problem of optimal distribution of a homogeneous resource is the problem of determining the weight of edges emanating from the vertices of the subset V_d, taking into account the criterion

DOI 10.15588/1607-3274-2023-3-14

$$f\left(T_1,..,T_{N_1},e'_1,..,e'_K\right)=(-1)\sum_{k=1}^{N_3}t_k\left(T_1,..,T_{N_1},e'_1,..,e'_K\right)\to\min_{\substack{e'_1,..,e'_K\\T_1,..,T_{N_1}}} \quad (4)$$

and the restriction on the distribution coefficients, which is given by the relation:

$$\sum_{j=1}^{J_k}e'^{k}_{j}=1, \quad (5)$$

where $e'^{k}_{j}\in F(k)$, $0\le e'^{k}_{j}\le 1$, $j=\overline{1,J_k}$, $J_k=|F(k)|$, $F(k)\subset E$, $k=\overline{1,K}$.

To find a solution of the optimization problem (4) in the form of a vector function $t(T_1,..,T_{N_1},e'_1,..,e'_K)=(t_1(T_1,..,T_{N_1},e'_1,..,e'_K),\ldots,t_{N_3}(T_1,..,T_{N_1},e'_1,..,e'_K))$, the k-th element of which characterizes the time spent to move to the k-th vertex ($k=\overline{1,N_3}$), we introduce the notation. The indicative graph $G(V,E)$ will be specified in the form of incidence matrices for the forward H^{in} and backward H^{out} flows of dimensions $N\times M$ and $M\times N$, respectively, whose elements are defined as:

$$H^{in}_{i,m}=\begin{cases}1, \text{ node } v_i \text{ is incident t o edge } e_m \text{ and is its end}\\ 0, \text{in opposite case}\end{cases}; \quad (6)$$

$$H^{out}_{m,i}=\begin{cases}1, \text{node } v_i \text{ is incident to edge } e_m \text{ and is its beginning}\\ 0, \text{in opposite case}\end{cases}; \quad (7)$$

$i=\overline{1,N}, m=\overline{1,M}$.

For a indicative graph $G(V,E)$, we define a matrix S^r of dimension $N\times N$, which is the incidence matrix for a path of multiplicity r (r specifies the number of edges through which there is a path from vertex v_i to vertex v_j). The matrix S^r is defined by the equality:

$$S^r=\left(H^{in}\left(-H^{out}\right)^T\right)^r. \quad (8)$$

Consider a vector function $w(T_1,..,T_{N_1},e'_1,..,e'_K)=(w_1(T_1,..,T_{N_1},e'_1,..,e'_K),\ldots,w_N(T_1,..,T_{N_1},e'_1,..,e'_K))$, the j-th element of which determines the amount of time spent on moving to the j-th vertex of the graph $G(V,E)$, $j=\overline{1,N}$:

$$w_j(T_1,..,T_{N_1},e'_1,..,e'_K)=w'_j(T_1,..,T_{N_1})+\left[\sum_{p=1}^{N}\left(H^{in}*diag\left(\gamma_R\left(T_1,..,T_{N_1},e'_1,..,e'_K\right)\right)*\left(-H^{out}\right)^T\right)\right]_j, \quad (9)$$

where the sum in the second term is taken over all N elements of the j-th row of the matrix

$$H^{in}*diag\left(\gamma_R\left(T_1,..,T_{N_1},e'_1,..,e'_K\right)\right)*\left(-H^{out}\right)^T,$$

the vector function $w'(T_1,..,T_{N_1})=(w'_1(T_1,..,T_{N_1}),\ldots,w'_N(T_1,..,T_{N_1}))$ of dimension N determines the amount of the time resource formed by the initial vertices of the graph $G(V,E)$, the vector function

$\gamma_R(T_1,..,T_{N_1},e'_1,..,e'_K)=(\gamma^1_R(T_1,..,T_{N_1},e'_1,..,e'_K),\ldots,\gamma^N_R(T_1,..,T_{N_1},e'_1,..,e'_K))$ of dimension N, the elements of which are calculated recursively by the formula

$$\gamma^m_R(T_1,..,T_{N_1},e'_1,..,e'_K)=\gamma^m_1\left(T_1,..,T_{N_1},e'_1,..,e'_K\right)+\left[\sum_{q=1}^{M}\left(\beta(e'_1,..,e'_K)H^{in}diag\left(\gamma_{r-1}\left(T_1,..,T_{N_1},e'_1,..,e'_K\right)\right)\right)\right]_m, \quad (10)$$

$r=\overline{2,R}$, $m=\overline{1,N}$, and the sum in the second term is taken over all M elements of the m-th row of the matrix $\beta(e'_1,..,e'_K)H^{in}diag\left(\gamma_{r-1}\left(T_1,..,T_{N_1},e'_1,..,e'_K\right)\right)$.

In the recursive expression (10), the initial values of the elements of the vector function $\gamma_1(T_1,..,T_{N_1},e'_1,..,e'_K)$ and the elements of the matrix function $\beta(e'_1,..,e'_K)$ are determined by relations:

$$\gamma^m_1\left(T_1,..,T_{N_1},e'_1,..,e'_K\right)=\left[\sum_{p=1}^{N}\left(\beta(e'_1,..,e'_K)*diag\left(v'(T_1,..,T_{N_1})\right)\right)\right]_m,$$

$$m=\overline{1,N};$$

$$\beta(e'_1,..,e'_K)=[\left(-H^{out}\right)*diag(E'(e'_1,..,e'_K))]^T. \quad (11)$$

Then the elements of the vector function $w(T_1,..,T_{N_1},e'_1,..,e'_K)$ determine the elements of the original vector function $t(T_1,..,T_{N_1},e'_1,..,e'_K)$, which is the solution of the optimization problem (4):

$$t_k(T_1,..,T_{N_1},e'_1,..,e'_K)=w_{Q_k}(T_1,..,T_{N_1},e'_1,..,e'_K), \quad (12)$$

where Q is a vector of dimension N_3 whose elements determine the numbers of the graph $G(V,E)$ vertices that make up the subset V_e of the final points of the movement, $k=\overline{1,K}$.

DOI 10.15588/1607-3274-2023-3-14

It is clear that such a search for a route involves the use of a solution technique with return (backtracking) [13]. The solution of the problem based on the use of backtracking is reduced to a consistent expansion of a particular solution. If the expansion fails at the next step, then a return to a shorter particular solution occurs and the search continues further in a new direction. This algorithm allows you to find all solutions to the problem, if any. It is known that the use of algorithms based on the backtracking scheme in solving practical problems is significantly limited by the low speed of operation and puts forward high requirements for computing resources. To speed up the work of the method, they try to organize calculations in such a way as to identify non-optimal options as early as possible, or use a selection scheme based on a greedy approach when constructing each step. This can significantly reduce the time to find a solution.

The greedy approach is formulated in accordance with the principle of choosing the optimal solution at every step, despite previous steps or those taken ahead. In other words, the greedy technique is based on a locally optimal choice with the hope that this choice will lead to a globally optimal solution.

It should be noted that there is no way to check the quality of the application of greedy algorithms in solving a specific applied problem, however, for problems in which the sequence of local optima goes to the global optimal solution, this approach is very promising.

The greedy method proposed by the authors assumes consideration at each stage of the formation of the route of the fastest in time section of the route of movement. A combined approach based on the method of resource allocation and greedy choice of the direction of movement made it possible to implement a constructive scheme for solving the traveling salesman problem, that can be formulated as the following *recursive algorithm* for a network of N nodes and a given travel time for each pair of vertices. :

Step 0. We form the initial information for the flow distribution method. The starting vertex of the traveling salesman route defines a subset of the initial nodes of the method, the set of directions from it defines a subset of intermediate nodes, and the graph vertices accessible from this subset defines the set of end nodes.

Based on the Orlin method of flow distribution, we determine the time to reach each of the end vertices on a subnet of initial, intermediate, and final vertices.

We select the shortest travel time and the corresponding stage of the route, mark the selected vertices and proceed to the formation of data for a new flow distribution problem. We pass to the next step of the algorithm.

Step s, s=1,2,… We construct new subsets of initial, intermediate, and final vertices, excluding from further consideration the previously noted vertices.

If at the current step it is impossible to determine new subsets (all vertices are marked), we return to the previous step, unmark the route stage, marking the dead end direction, and move on to the next possible one by choosing the fastest direction of movement.

We repeat this process until we reach the end point of the route, which coincides with the starting point.

If the route is built, but does not include all the vertices of the graph, we return to the previous levels and rebuild all the working subsets, choosing new directions of movement, taking into account the speed of movement.

Final step. As a result of the work, we finally obtain a cyclic permutation of the numbers of the vertices of the graph, which determines the sequence of stages of the traveling salesman's route.

4 EXPERIMENTS

To analyze the efficiency of the algorithm, computational experiments were carried out, in which various methods (complete search, greedy, annealing and the one proposed above) were used to solve the traveling salesman problem on a network of 11 points [14]. The graph of the network of movements with the given time costs is shown in Figure 1.

5 RESULTS

The results of the numerical experiments performed are shown in Table 1.

The optimal route in the considered problem was found by full search and is determined by the sequence of numbers 1, 2, 5, 9, 7, 4, 3, 6, 8, 11, 10, 1 or 1, 2, 5, 9, 7, 4, 3, 6, 10, 11, 8, 1. The proposed algorithm made it possible to quickly find a route for visiting all graph vertices (1, 4, 3, 6, 2, 5, 9, 7, 10, 11, 8, 1), but the time it took more to move along this route.

As a result of the computational experiments the efficiency of using the developed algorithm was established, the obtained solutions are compared with the solutions found by other exact and heuristic methods.

6 DISCUSSION

Several remarks should be noted. The search procedure is based on the use of the backtracking technique, according to which the solution of the problem is reduced to a sequential expansion of a particular solution. If at the next step the expansion fails, then a return to a shorter particular solution occurs and the search continues further. This algorithm allows you to find all solutions to the problem, if any. It is known that the use of algorithms based on the backtracking scheme in solving practical problems is significantly limited by the low speed and significant requirements for computing resources. To speed up the work of the method, calculations are organized in such a way as to identify non-optimal options as early as possible, or use selection schemes based on a greedy campaign. This can significantly reduce the time to find a solution.

The greedy technique is based on a locally optimal choice under the assumption that this choice will lead to a globally optimal solution. Unfortunately, there is no way to check the quality of using greedy algorithms in solving a specific applied problem, however, for problems in which a sequence of local optima goes to a global optimal solution, this approach is very promising.

DOI 10.15588/1607-3274-2023-3-14

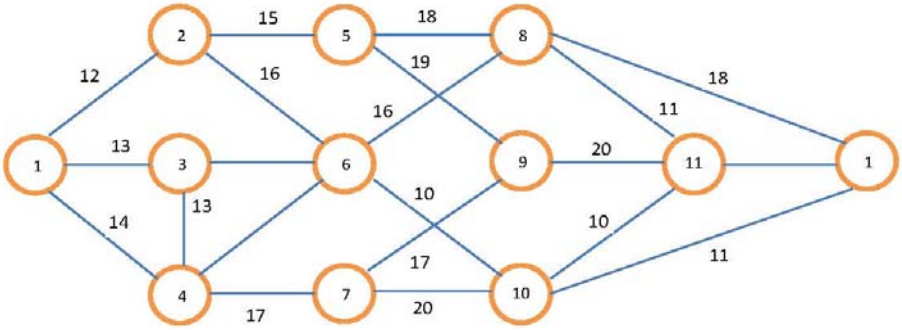


Figure 1 – The network of n=11 nodes and a given travel time for each pair of vertices for the traveling salesman problem

Table 1 – The comparison of search time and solutions of the traveling salesman problem for n=11

Calculation method	Operation time	Optimal solution	Solution characteristic
Complete search	30 sec	157 h	Exact
Greedy algorithm	21 sec	169 h	Approximate
Annealing method	23 sec	174 h	Approximate
The proposed algorithm	25 sec	169 h	Approximate

The greedy method proposed by the authors assumes consideration at each stage of the route formation of the fastest direction of movement in terms of time. A combined approach based on the method of resource allocation and greedy choice of the direction of movement made it possible to implement a constructive scheme for solving the traveling salesman problem.

CONCLUSIONS

The paper considers a method for formalizing the algorithm for solving the traveling salesman problem using the method of streaming resource allocation and using the backtracking scheme. The use of Orlin's method to optimize the flow distribution on the graph is proposed. The scheme of formalization of the procedure for using the method with the implementation of the backtracking scheme for solving the traveling salesman problem with the minimum duration of movement along the route is briefly described. A variant of accelerating the speed of the developed algorithm is proposed, which consists in using a greedy technique in the procedure for selecting route sections: the planning of each next stage of movement is determined based on the choice of the fastest direction of movement, which makes it possible to obtain a constructive scheme for solving the traveling salesman problem. The results of the proposed algorithm for calculating solutions to the traveling salesman problem with minimization of the duration of movement are presented, the obtained solutions are compared with the solutions found by known exact and heuristic methods. The influence of the greedy approach on the speed of the developed algorithm was analyzed. Conclusions are drawn, further development of the proposed methodology for solving traveling salesman problems based on the use of other principles of greedy choice of direction of movement and for solving fuzzy and dynamic traveling salesman problems is proposed.

ACKNOWLEDGMENTS

The work is the result of research carried out by the Faculty of Computer Science and Cybernetics of the Taras Shevchenko National University of Kyiv, the Department of Information Systems and Technologies of the National Transport University and the travel company Bon Voyage during the implementation of the project for building optimal multi-day tourist routes.

REFERENCES

1. Christopher M., Peck H. Building the Resilient Supply Chain, *The Intern. Journal of Logistics Management*, 2004, Vol. 15, No. 2, pp. 1–14. DOI:10.1108/09574090410700275
2. Alan Harrison, Remko van Hoek Logistics Management and Strategy. Financal Times Management, 3d edition, 2005, 308 p. ISBN: 978-0-273-71276-3
3. Ghiani G., Laporte G., Musmanno R. Introduction to Logistics Systems Planning and Control. John Wiley & Sons, Ltd, 2013, 377 p. DOI:10.1002/9781118492185
4. Vajda S. Linear programming. Dordrecht, Springer, 1981, 15 p. DOI: 10.1007/978-94-011-6924-0_1
5. Schrijver A. Theory of Linear and Integer Programming. Wiley, 1998, 484 p. DOI: 10.2307/253980
6. Vanderbei R. J. Linear programming: Foundations and extensions. Springer, 2014, 414 p. DOI: 10.1007/978-3-030-39415-8
7. Korte B., Vygen J. Combinatorial Optimization: Theory and Algorithms (Algorithms and Combinatorics). Berlin, Heidelberg, Springer, 2018, 455 p. DOI: 10.1007/978-3-662-56039-6
8. Moss L. T., Atre Shaku Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley Professional, 2003, 576 p. ISBN: 978-0-201-78420-6
9. Buontempo F. Genetic Algorithms and Machine Learning for Programmers. Pragmatic Bookshelf, 2019, 236 p. ISBN: 9781680506204
10. Kshemkalyani A. D., Singhal Mukesh Distributed Computing: Principles, Algorithms, and Systems. Cambridge University Press, 2008, 756 p. DOI: 10.1017/CBO9780511805318
11. Rai S., Ettam R. K. Simulation-based optimization using simulated annealing for optimal equipment selection within print production environments, *2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World, December, 2013: proceedings*, 2013, pp. 1097–1108. DOI: 10.1109/WSC.2013.6721499
12. Orlin J. B. A Faster strongly polynomial minimum cost flow algorithm, *Operations research*, 1993, V. 41, Iss.2, pp. 338–350. DOI: 10.1145/62212.62249
13. Watson D. A Practical Approach to Compiler Construction. Springer, 2017, 254 p. DOI: 10.1007/978-3-319-52789-5

DOI 10.15588/1607-3274-2023-3-14

14. Chvatal V., Cook W. J., Dantzig G. B., Dantzig G. B., Fulkerson D. R., Johnson S. M. Solution of a large-scale traveling-salesman problem, *In 50 Years of Integer Programming 1958–2008. From the Early Years to the State-of-the-Art*, 2010, pp. 7–28. DOI: 10.1007/978-3-540-68279-0_1

Received 26.06.2023.
Accepted 17.08.2023.

УДК 519.87:004.02

ЕФЕКТИВНИЙ МЕТОД РОЗВЯЗАННЯ ЗАДАЧІ РОЗПОДІЛУ ПОТУЖНОСТЕЙ КАНАЛІВ З УРАХУВАННЯМ НЕЧІТКИХ ОБМЕЖЕНЬ НА ОБСЯГИ СПОЖИВАННЯ

Івохін Є. В. – д-р фіз.-мат. наук, професор, професор кафедри системного аналізу та теорії прийняття рішень Київського національного університету імені Тараса Шевченка, Київ, Україна.

Гавриленко В. В. – д-р. фіз.-мат. наук, професор, професор кафедри інформаційних систем і технологій Національного транспортного університету, Київ, Україна.

Івохіна К.Є. – Аспірант кафедри інформаційних систем і технологій Національного транспортного університету,, Київ, Україна.

АНОТАЦІЯ

Актуальність. Важливою сучасною проблемою є швидке відновлення та оптимізація управління логістикою. В залежності від поставленої задачі існує багато різних математичних методів та підходів до вирішення різних логістичних задач, розв'язування яких набуває широкого практичного впровадження. Його конкретний зміст залежить від характеру проблеми та повноти наявних даних. Іноді для розв'язання відомих задач, однією з яких є задача комівояжера, вдається знайти нетипові методики на основі поєднання декількох обчислювальних схем та методів.

Ціль. Мета роботи – розробити алгоритм розв'язання задачі комівояжера на основі реалізації методу потокового розподілу ресурсів і схеми backtracking з мінімальною тривалістю руху за маршрутом.

Метод. У статті розглядається методика послідовного застосування потокових схем розподілу однорідного ресурсу для розв'язання задачі комівояжера, що формулюється як задача знаходження маршруту відвідування заданої кількості міст без повторень з мінімальною тривалістю руху. Поставлено та вирішено задачу формалізації алгоритму розв'язання проблеми комівояжера на основі методу розподілу ресурсів з використанням схеми backtracking. Запропоновано використання методу Орліна для оптимізації розподілу потоку на графі. Розроблено конструктивний алгоритм розв'язання задачі. Проведено обчислювальні експерименти.

Результати. Розроблено метод розв'язання задачі комівояжера з використанням методу потокового розподілу ресурсів і схеми пошуку з поверненням. Запропоновано варіант прискорення швидкості розробленого алгоритму, яке полягає в залученні жадібного способу в процедурі вибору ділянок маршруту: планування кожного наступного етапу переміщення визначається виходячи з відбору найбільш швидкого напряму руху. Застосування жадібного підходу дозволило отримати конструктивну схему розв'язання задачі комівояжера. Представлено результати розрахунків за допомогою запропонованого алгоритму в задачах комівояжера з мінімізацією тривалості руху, проведено порівняння отриманих розв'язків з розв'язками, знайденими іншими точними та евристичними методами.

Висновки. У статті розглянуто метод формалізації алгоритму розв'язання задачі комівояжера з використанням алгоритму потокового розподілу однорідного ресурсу та схеми backtracking. Запропоновано використання методу Орліна для оптимізації розподілу потоку на графі. Описано схему формалізації процедури використання методу з реалізацією схеми з поверненням для розв'язання задачі комівояжера з мінімізацією тривалості руху за маршрутом. Запропонований варіант прискорення роботи розробленого алгоритму.

КЛЮЧОВІ СЛОВА: задача комівояжера, метод розподілу ресурсів, рекурсивна схема пошуку з поверненням, жадібний підхід.

ЛІТЕРАТУРА

1. Christopher M. Building the Resilient Supply Chain, The Intern / M. Christopher, H. Peck // Journal of Logistics Management. – 2004. – Vol. 15, No. 2. – P. 1–14. DOI:10.1108/09574090410700275
2. Harrison Alan Logistics Management and Strategy / Alan Harrison, Hoek Remko van. – Financal Times Management, 2nd edition, 2005. – 308 p. ISBN: 978-0-273-71276-3
3. Ghiani G. Introduction to Logistics Systems Planning and Control / G. Ghiani, G. Laporte, R. Musmanno. – John Wiley & Sons, Ltd, 2004. – 377 p.
4. Vajda S. Linear programming / S. Vajda. – Dordrecht : Springer, 1981. – 15 p. DOI: 10.1007/978-94-011-6924-0_1
5. Schrijver A. Theory of Linear and Integer Programming / A. Schrijver. – Wiley, 1998. – 484 p. DOI: 10.2307/253980
6. Vanderbei R. J. Linear programming: Foundations and extensions / R. J. Vanderbei. – Springer, 2014. – 414 p.
7. Korte B. Combinatorial Optimization: Theory and Algorithms (Algorithms and Combinatorics) / B. Korte, J. Vygen. – Springer Berlin, Heidelberg, 2018. – 455 p.
8. Moss L. T. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications / L. T. Moss, Atre Shaku. – Addison-Wesley Professional, 2003. – 576 p.
9. Buontempo F. Genetic Algorithms and Machine Learning for Programmers / F. Buontempo. – Pragmatic Bookshelf, 2019. – 236 p.
10. Kshemkalyani A. D. Distributed Computing: Principles, Algorithms, and Systems / A. D. Kshemkalyani, Mukesh Singhal. – Cambridge University Press, 2011. – 756 p.
11. Rai S. Simulation-based optimization using simulated annealing for optimal equipment selection within print production environments / S. Rai, R. K. Ettam // Winter Simulations Conference (WSC). – 2013. – P. 1097–1108.
12. Orlin J. B. A Faster strongly polynomial minimum cost flow algorithm / J. B. Orlin // Operations research. – 1993. – V. 41, Iss.2. – P. 338–350. DOI: 10.1145/62212.62249
13. Watson D. A Practical Approach to Compiler Construction / D. Watson. – Springer, 2017. – 254 p. DOI: 10.1007/978-3-319-52789-5
14. Solution of a large-scale traveling-salesman problem / [Chvatal V., Cook W. J., Dantzig G. B. et al.] // In 50 Years of Integer Programming 1958–2008. – From the Early Years to the State-of-the-Art, 2010. – P. 7–28.

DOI 10.15588/1607-3274-2023-3-14

UDC 004.77

PERFORMANCE ANALYSIS OF WIRELESS COMPUTER NETWORKS IN CONDITIONS OF HIGH INTERFERENCE INTENSITY

Khandetskyi V. S. – Dr. Sc., Professor, Head of the Department of Electronic Computing Machinery, Oles Honchar Dnipro National University, Dnipro, Ukraine.

Gerasimov V. V. – PhD, Associate Professor of the Department of Electronic Computing Machinery, Oles Honchar Dnipro National University, Dnipro, Ukraine.

Karpenko N. V. – PhD, Associate Professor of the Department of Electronic Computing Machinery, Oles Honchar Dnipro National University, Dnipro, Ukraine.

ABSTRACT

Context. The decrease in the probability of successful frame transmission in the infrastructure domain of IEEE 802.11 DCF wireless network is caused both by the influence of the collision intensity and by the impact of external interference in the radio path. Using the Markov chain approach as a baseline, we explicitly expressed the dependence of the network throughput on the number of operating stations, bit error rate (BER), and the frame fragmentation factor.

Objective. The purpose of this article is to study the influence of interference intensity on the throughput of a wireless network domain in a wide range of the number of operating stations when transmitting frames of various lengths in the absence and with the use of the fragmentation mechanism.

Method. The performed mathematical modelling showed, that in the range of increased and high noise intensity (BER = 10^{-5} – 10^{-4}), a decrease in the length of the frame data field from the standard length of 12000 bits to 3000 bits is accompanied by a decrease in the throughput for all values of the number of competing stations. At the same time, it must be noted that as the amount of the frame data decreases, the throughput becomes less susceptible to an increase in the noise intensity. Qualitatively different results are obtained in the region of very high interference intensity (BER = $2\cdot10^{-4}$). A significant increase in the probability of frame transmission in this region observed with a decrease in the standard length of the frame data field by 2–3 times, made it possible to increase the throughput compared to the original one. This effect is especially pronounced when the length is halved.

Results. The study of the standard frame transmitting process, but with a fragmented data field, showed that if for BER = $5\cdot10^{-5}$ and less with an increase in fragmentation factor, the throughput values decrease, in the entire range of the number of stations due to the predominant increase in overhead costs, then in the region of high (BER = 10^{-4}) and very high noise intensity (BER = $2\cdot10^{-4}$) we have the opposite effect. To the greatest extent, the throughput increases when the frame data is transmitted in two equal fragments. We have made a comparison of the network throughput determined by simply reducing the length of the frame data field and using fragmentation of a standard frame. The comparison showed that the use of the fragmentation mechanism is more beneficial both when throughput is stabilized under conditions of increased noise intensity and when the throughput is increased under conditions of high and very high noise intensity.

Conclusions. In this article, a mathematical model has been modified for direct calculation of the wireless network throughput. Using this model, we studied the changes in throughput over a wide range of BER and a number of operation stations for various values of the transmitted frame fragmentation factor. The conditions for increasing the throughput are determined.

KEYWORDS: IEEE 802.11 wireless networks, DCF, throughput, infrastructure domain, BER, frame, fragmentation factor, collision.

ABBREVIATIONS

AP is an access point;

BER is a bit error rate;

CSMA/CA is a Carrier Sense Multiple Access with Collision Avoidance;

DCF is a Distributed Coordination Function;

FCC is a Federal Communication Commission;

MAC is a Medium Access Protocol;

MIMO is a Multiple Input Multiple Output;

STA is a station;

Wi-Fi is a Wireless Fidelity;

WLAN is a Wireless Local Area Network.

NOMENCLATURE

ACK is a frame acknowledgment;

DIFS is an interframe space;

SIFS is a short interframe space;

bE is a number of erroneous bits;

H is a frame header transmission time;

E_b is an amount of energy per bit;

FER is a frame error rate i.e. probability of frame distortion;

k is a fragmentation factor;

L is length of the frame data field in bits;

L_0 is an initial length of the frame data field in bits;

MAC_{hdr} is a transmission time of a frame channel layer header;

m is a number of window doubling allowed;

N_0 is a noise power spectral density;

n is a number of competing stations;

P_b is a bit error probability;

PHY_{hdr} is a transmission time of a frame physical layer header;

P_S is a probability of successful frame transmission;

p is the collision probability;

R is a data transfer rate;

S is a network throughput;

T_c is an average time the channel is sensed busy because of collision;

T_{sc} is an average time the channel is sensed busy because of successful transmission;

DOI 10.15588/1607-3274-2023-3-15

t is a transmission time;
W_0 is a minimum value of contention window;
α is a multiplicative constant coefficient;
β is a power constant coefficient;
δ is a propagation delay;
η is a number of empty slots;
σ is a duration of one slot;
T is a probability of successful frame transmission over the channel without errors.

INTRODUCTION

In the last few decades, we have witnessed an exponential growth of the demand for wireless networks that provide reliable communications and ensure ubiquitous coverage, high spectral efficiency, and low latency [1, 2]. From 2012 to 2017, mobile networks have been a seventeen-fold cumulative growth, registering an increment in 71% in data traffic from 2016 to 2017 alone. Recent studies also show, that 54% of the traffic, generated by devices that support cellular and Wi-Fi connectivity was offloaded via Wi-Fi in 2017 and is expected that this number increases up to 59% by 2022 [3, 4].

The 802.11 DCF WLANs (Wi-Fi networks) work on the basic of the well-known carrier-sense multiple access with collision avoidance (CSMA/CA) protocol. In distributed WLAN's environment a common wireless medium is shared by a number of associated stations without any centralized coordination. Whenever a given station has a frame to transmit, it waits until the channel becomes idle for a given amount of time (DIFS interval), and then accesses the channel following exponential backoff rules. If a successful reception occurs the access point responds after a SIFS – interval with an ACKnowledgment frame. The management of the common medium is specified by two aspects: (1) multiple access resolution, i.e. the rules that govern how a given station acquires the right to use the channel; and (2) channel transmission operations, i.e. the rules that govern how a station that wins a contention performs transmissions without losing control over the channel.

Despite significant progress in solving these and other WLANs problems, achieved in the development of next-generation networks such as 802.11ac and 802.11ax, the effective throughput increases quite slowly, especially in dense networks operating under conditions of high interference intensity. A high noise level is usually caused by the presence of both external interference and interference specific to a given data transmission technology, usually caused the need to increase the transmission rate [5].

As noted by US Federal Communications Commission (FCC), an urgent problem for technologies using channels with a width Δf = 160 MHz, operating in the range with a central frequency f = 5 GHz, is "clearing the frequency range". The effect of noise increases as the channel bandwidth expands. A similar effect is also observed with a decrease in the inter-symbol interval of transmitted data and with an increase in the number of subcarrier frequencies used in modern wireless technologies. Being closer to each other adjacent subcarriers are more sensitive to noise and mutual interference [6].

In the presence of a significant number of obstacles in the signal propagation area, multiple reflected signals lose their initial energy and arrive to wireless router with a certain delay. To struggle the negative influence of multi-path propagation, several antennas are used on the sender side and on the receiver side of the channel (MIMO scheme). This also allowed the formation of several parallel spatial data streams. In 802.11ac technology, which uses 8 antennas in the router, a directional signal formation mode (Beamforming) has been created. This mode is used, for example, between two routers in the backbone of the wireless network. At the same time, the concentration of several spatial streams in one region of the channel, even despite, for example, different polarization of signals transmitted in each stream, leads to an increase in the mutual influence of signals. This effect is further enhanced with an increase in the intensity of external noise, blurring the distinctive features of signals of different streams.

The object of study is the process of data transmission in wireless networks with heavy traffic at high intensity of external interference.

The subject of study is the mathematical models of IEEE.802.11 DCF networks operation under conditions of collisions and external interference combine influence.

The purpose of the paper is to study the possibility of increasing the throughput of a dense wireless network at high noise intensity due to the fragmentation of transmitted frames.

1 PROBLEM STATEMENT

In conditions of increased interference intensity, the probability of successful frame transmission from the station to the access point can be defined as [7, 8]

$$P_S = T \cdot (1 - FER). \quad (1)$$

The probability T is traditionally determined using Markov-chain models in the form of the function $F(p, W_0, m)$, where p is the collision probability, which in turn depends on T; $W_0 = CW_{\min}$ is the minimum contention window; m is the number of window doublings after every next collision [9–11].

Collisions coming from WLAN's nodes using the same MAC protocol, and interference coming from devices outside the network, waste valuable transmission time and radiated power, having a negative impact on the energy efficiency, throughput, and delay of the system [12]. Network nodes cannot distinguish one type of loss from another because the symptoms are the same – not receiving the acknowledgment from the access point. The increase in the level of interference leads to the increase in the loss of information frames during transmission, which in turn decreases the network throughput.

Rational for improving the reliability of WLANs is to avoid losing frames due to occurrence of channel induced errors, collisions etc. The STA needs to retransmit the

DOI 10.15588/1607-3274-2023-3-15

whole frame even if it contains only one bit error. When the channel error rate is significantly high to get the frame through would require a large number of retransmissions. To mitigate this, fragmentation was proposed whereby big frames are sent in small fragments which are individually acknowledged or retransmitted. Doing this in case the error the STA needs to retransmit only the error fragment which takes short time compared to retransmitting the whole initial frame. If the medium is significantly noisy, a fragment has a higher probability to get through without errors because it can be fitted between error bursts [13, 14]. By operating this way, the STA increases its chances of successful frame transmission in bad channel conditions.

At the same time, it should be considered that fragmentation increases the amount of service information needed to transfer a given amount of data, which leads to a decrease in network throughput.

The purpose of this work is to expand the mathematical model proposed by us in article [15], which in an explicit analytical form describes the impact of collisions and external noise on the operation of IEEE 802.11 networks, for the case of frames fragmentation, and to study the possibility of increasing network's throughput due to fragmentation under conditions of high intensity of external interference.

2 REVIEW OF THE LITERATURE

The basic for the theoretical analysis of the frame transmission probability in IEEE 802.11 DCF wireless networks for more than the last two decades are Markov chain models [9–11, 16–20]. In the widely known work published by Bianchi [9] which was further developed by Tinnirello, Bianchi, and Xiao [10] the authors use the chain model for ideal channel conditions. The probability that a station accessed a channel depends on when the channel was idle or busy in a previous time slot. These aspects were studied in [16]. In article [17] a Markovian agent model is used to represent the behavior of wireless nodes based on CSMA/CA access method.

In articles using the Markov model, it is often assumed that traffic is saturated [18–21]. In these conditions nodes can be modeled as being equally likely to send in any slot, and this assumption also holds in the first approximation for unsaturated traffic which nearly Poisson [22].

Work [23] is devoted to study of the optimization problem of retransmission number on transmission performance. The number of stations in the network is 20, 40, and 80. The author provides an analytical model on the performance of real-time applications transmission over WLAN. The analytical model evaluates the random-access performance of real-time services based on two-dimensional Markov-chain model by taking into account the impact of the maximum optimal retransmission number on the service time of the packet transmissions. Modelling and performance evaluation of the IEEE 802.11 DCF for real-time control is also carried out in article [24].

Electromagnetic interference that reduces the efficiency of wireless networks occurs both due to external sources and depends on the architecture and operating conditions of the network itself. Sources such as automobiles, aircrafts, ignition electric motors and switching gear, high voltage wires and fluorescent lamps cause industrial noise. Electromagnetic interference is a disturbance generated by these external sources that affects an electrical circuit of electromagnetic induction, electrostatic coupling, discharging process or conduction disorder.

The problem of interference arises when more APs of wireless network are placed near each other and the coverage area of these APs starts to overlap, which causes degradation of the bandwidth and the service received by the recipients. Another challenge in wireless networks is the handover, which is process of switching users from one AP to other [25].

Significant interference in the process of information transmission in some cases introduced the effect of multipath signal propagation. This effect is also known as multipath interference or multipath distortion. Notable consequences of this are envelope fading and inter-symbol interference [26]. If the propagation delay of the rays is small compared to the channel symbol duration, then only wave interference occurs, leading to the fading. Due to the time difference between the base signal and the multiple reflected copies, the access point may have problems demodulating the received signal. In this case, the serious problem is the overlapping of information bits on each other, as a result of which the data is damaged. This effect is called inter-symbol interference.

Interferences that have a various physical nature differ in their spectral composition. At the same time, it is important to study the general patterns of the interference influence on data transmitted in wireless networks over a radio channel. For this purpose, it is advisable to use Gaussian noise as the most general noise model that describes a wide range of noise sources and their superposition quite well [27].

An example of a simple channel model that is widely used in information theory is additive white Gaussian noise channel without fading [27, 28]. In [29], Tianji et al analyze the throughput performance of the Block ACK scheme over a noisy channel. In [30] authors present a new discrete time Markov chain model to estimate the packet transmission probability. They propose an enhancement of the IEEE 802.11 RTS/CTS scheme to recognize the reason of transmission failure (collision or noise errors).

One of the changes that modern digital communication systems have brought to radio engineering is the need to end-to-end performance evaluations. The measure of that performance is usually bit error rate (BER), which quantified the reliability of the radio system from "bits in" to "bits out" [31],

$$\text{BER} = \text{number of corrupted bits} / \text{total number of bits} = bE / (R \cdot t). \quad (2)$$

DOI 10.15588/1607-3274-2023-3-15

In a noisy channel, the BER is often expressed as a function of the normalized carrier-to-noise ratio denoted E_b / N_0 (energy per bit to noise power spectral density) [28]. The Gaussian approximation of the noise in determining the BER is used to estimate the number of iterations needed to the convergence the parity code decoder in function of the level of noise power [32]. Bit-error rate analysis of low-density parity-check codes using Gaussian approximation of a channel is considered in [33].

In general, errors at different locations of an information sequence of length L can occur with different probabilities. In this article, we use for transmission a time-discrete channel without memory with white Gaussian noise. A channel of this type is characterized by the fact that the bit errors in it are independent and equally distributed over the bits of the frame data [27].

3 MATERIALS AND METHODS

Improving the transmission reliability of a frame can be achieved by reducing the size of its data field [28, 30, 34]. At the same time, this leads to an increase in the relative contribution of the time spent on the transmission of the MAC protocol information of the 802.11 standard, which is necessary to ensure a successful transmission process. Let us study this process under conditions of high noise intensity.

In work [15], we have expressed in an explicit analytical form the dependence of the throughput of the IEEE 802.11 DCF wireless computer network on the number of stations operation in saturation mode and the value of BER, which is determined by the intensity of interference in the radio path. This dependence can be represented in the following form:

$$S = \frac{2nqL}{2nq(T_{SC} - T_C) + (Q + 1 - 2q)[T_C(\frac{Q+1}{Q+1-2q})^n + \eta\sigma]}. \quad (3)$$

In expression (3)

$$T_{SC} - T_C = SIFS + ACK + \delta, \quad (4)$$

$$T_C = PHY_{hdr} + MAC_{hdr} + \frac{L}{R} + DIFS + \delta, \quad (5)$$

$$Q = \frac{W_0 \cdot 2^{\beta(n-1)}}{1 + \alpha(n-1)}. \quad (6)$$

In accordance with the justification given in the previous section, bit errors occurring in the noisy channel are independent and equally distributed over the bits of the frame data field. Then the probability that a frame with a data field of length L will be transmitted undistorted is equal to

$$q = (1 - P_{b1})(1 - P_{b2})(1 - P_{b3})...(1 - P_{bL}).$$

And since the probabilities of distortion of individuals bits are the same, then

$$q = (1 - P_b)^L. \quad (7)$$

Using expressions (3)–(7), we calculated the dependences $S(n, P_b)$ for different values of L. The probability of an error of one bit in a frame was taken equal to $P_b = 10^{-5}$, $5 \cdot 10^{-5}$, 10^{-4} which corresponds to increased and high interference intensity. We took the initial length of the frame data field equal to $L_0 = 12000$ bits.

To calculate the dependences $S(n, L)$ for different values of P_b, we used the following data [15, 35, 36]:

$SIFS$ = 16 μs, $DIFS$ = 34 μs, δ = 0,33 μs (the distance between the station and AP was taken equal 100 m), σ = 9 μs, ACK = 38,66 μs, $H = PHY_{hdr} + MAC_{hdr}$ = 68 μs, R = 54 Mbps, α = 0,05, β = 0,2, W_0 = 16.

4 EXPERIMENTS

Tables 1, 2 and 3 for $P_b = 10^{-5}$, $5 \cdot 10^{-5}$ and 10^{-4} respectively show the values of the throughput S depending on the number of simultaneously operating stations n and the length of the frame data field L.

In [36] the authors carried out the numerical study of the well-known bi-dimensional Markovian mathematical model [9, 10] under ideal channel conditions using the ns-3 discrete-event network simulator. Comparison of the throughput values S obtained by them with our data determined during the transmission in a noisy channel with a relatively low noise level $P_b = 10^{-5}$ (Table 1) showed acceptable results. For example, for n = 10, the decrease in the value of S in our case is 27%, for n = 10 – 26%.

Analyzing the numerical dependences presented in Table 1, the following can be noted. With an increase in the number n of stations competing for access to the communication channel, the value of the throughput S monotonically decreases. This is due to the growth of the collision intensity of simultaneously operating stations. Reducing the length of the transmitted frames data field from 12000 bits to 6000 bits and further to 3000 bits allows you to increase the probability of the frame successful transmission q in expression (3). So, at $L = L_0$ q = 0,887, at $L = L_0/2$ q = 0,942, at $L = L_0/4$ q = 0,97. However, despite this, in Table 1 we observe a decrease in the value of S with a decrease in the length of transmitted frames. This is due to the fact that with a decrease in amount of transmitted data L, the relative part of the overhead costs, i.e., service information that ensures the process of the frames transmitting increases.

The probabilities of successful frame transmission for Table 2 are: at $L = L_0$ q = 0,549, at $L = L_0/2$ q = 0,741, at $L = L_0/4$ q = 0,86.

The probabilities of successful frame transmission for Table 3 are: at $L = L_0$ q = 0,30, at $L = L_0/2$ q = 0,549, at $L = L_0/4$ q = 0,74.

Using Tables 1, 2 and 3 we determined the relative decrease in the throughput value with a growth of interference intensity for various lengths of the frame data field.

DOI 10.15588/1607-3274-2023-3-15

Table 1 – Results of throughput S calculations for $P_b = 10^{-5}$

n		5	7	10	15	20	25	30	40	50	60	80
S, Mbps	$L = L_0$	19.440	19.300	19.220	18.690	18.100	16.130	13.410	7.430	3.210	1.180	0.130
	$L = L_0 / 2$	12.550	12.470	12.430	11.980	11.480	9.960	8.020	4.200	1.400	0.630	0.068
	$L = L_0 / 4$	7.350	7.310	7.290	6.980	6.640	5.640	4.450	2.240	0.910	0.330	0.035

Table 2 – Results of throughput S calculations for $P_b = 5 \cdot 10^{-5}$

n		5	7	10	15	20	25	30	40	50	60	80
S, Mbps	$L = L_0$	16.80	16.79	16.76	16.11	15.27	12.86	10.08	5.07	2.07	0.74	0.08
	$L = L_0 / 2$	11.57	11.55	11.53	11.05	10.47	8.81	6.90	3.45	1.40	0.50	0.05
	$L = L_0 / 4$	7.03	7.01	6.98	6.67	6.31	5.28	4.10	2.02	0.82	0.29	0.03

Table 3 – Results of throughput S calculations for $P_b = 10^{-4}$

n		5	7	10	15	20	25	30	40	50	60	80
S, Mbps	$L = L_0$	12.61	12.59	12.55	11.96	11.07	8.79	6.51	2.99	1.16	0.41	0.04
	$L = L_0 / 2$	10.16	10.15	10.12	9.66	9.04	7.37	5.59	2.66	1.05	0.37	0.04
	$L = L_0 / 4$	6.59	6.58	6.56	6.25	5.87	4.83	3.70	1.78	0.71	0.25	0.03

Let the number of simultaneously operating stations $n = 20$. Then for $L = L_0$ an increase in P_b from 10^{-5} to $5 \cdot 10^{-5}$ leads to a decrease in the value of S by 15,6%, and a further increase in P_b from $5 \cdot 10^{-5}$ to 10^{-4} by another 27,5% relative to the level at $P_b = 5 \cdot 10^{-5}$. Similar calculations for $n = 30$ give $\Delta S_{12} / S_1 = 24{,}8\%$ and $\Delta S_{23} / S_2 = 35{,}4\%$, where 1, 2 and 3 are the Tables numbers. For $n = 40$ $\Delta S_{12} / S_1 = 31{,}8\%$ and $\Delta S_{23} / S_2 = 41\%$.

When the length L_0 of the frame data field is halved, we get the following data: for $n = 20$ $\Delta S_{12} / S_1 = 8{,}8\%$ and $\Delta S_{23} / S_2 = 13\%$; for $n = 30$ $\Delta S_{12} / S_1 = 14\%$ and $\Delta S_{23} / S_2 = 19\%$; for $n = 40$ $\Delta S_{12} / S_1 = 17{,}9\%$ and $\Delta S_{23} / S_2 = 22{,}9\%$. By reducing the length L_0 of the frame data field by four times, we get the following data: for $n = 20$ $\Delta S_{12} / S_1 = 5\%$ and $\Delta S_{23} / S_2 = 7\%$; for $n = 30$ $\Delta S_{12} / S_1 = 7{,}9\%$ and $\Delta S_{23} / S_2 = 9{,}8\%$; for $n = 40$ $\Delta S_{12} / S_1 = 9{,}9\%$ and $\Delta S_{23} / S_2 = 11{,}9\%$.

The performed calculation shows that with a decrease in the length L of the transmitted frame data field from 12000 bits to 6000 bits and further to 3000 bits, the throughput S becomes less susceptible to an increase in the noise intensity. So, for example, for $n = 20$, an increase in the error probability P_b from 10^{-5} to $5 \cdot 10^{-5}$ leads for $L = 12000$ bits to a decrease in the value of S by 15,6%, for $L = 6000$ bits – by 8,8%, for $L = 3000$ bits – by 5%. In the zone of more intensive noise increase in P_b from $5 \cdot 10^{-5}$ to 10^{-4} leads for $L = 12000$ bits to a decrease in S by 27,5%, for $L = 6000$ bits – by 13%, for $L = 3000$ bits – by 7%. Similar results are observed for other values of n.

Qualitatively different results are obtained in the region of very high interference intensity, at $P_b = 2 \cdot 10^{-4}$. Corresponding dependences $S(n)$ at $L = L_0$, $L = L_0 / 2$ and $L = L_0 / 4$ are shown in Fig. 1.

As can be seen from the graphs, the dependence $S(n)$ at $L = L_0 / 2$ is located significantly higher than the similar dependence at $L = L_0$. Thus, in the region of high-intensity noise, a two-fold decrease in the length of the frame data field made it possible to significantly increase the throughput S compared to the original one. Reducing L to 3000 bits also allows, although to a lesser extent, to increase the throughput compared to the original.

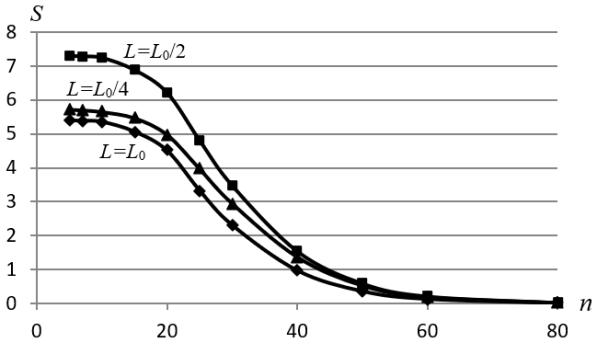


Figure 1 – Throughput S versus the number of competing stations n at different length of the frame data field L for BER = $2 \cdot 10^{-4}$, $L_0 = 12000$ bits

5 RESULTS

The increase in throughput within the framework of the fragmentation mechanism can be achieved by reducing the overhead that is spent on a frame transmission. In the overhead, we also include the time spent on the retransmissions of frames distorted by interferences.

Reducing the length of the data field L of the transmitted frames decreases the probability of its distortion by external interference and can, at the same time, increase throughput by reducing the number of retransmissions. It follows from the results of the previous section that this effect is more pronounced in the range of very high noise intensity. Let us study this process under the condition of a constant total length of the data field of the frame transmitted using the fragmentation mechanism.

We will divide the standard original frame with data field of length $L_0 = 12000$ bits into fragments so that the sum of the data fragments is L_0. In the basic DCF scheme, only first fragment in a transmitted frame contends for a channel access, the other fragments are transmitted after differing a SIFS interval and after each fragment an ACK is sent back by access point [14, 19, 34]. When creating a model for studying the throughput, we assumed that channel errors don't corrupt ACK frames. Since the ACK frames are usually transmitted at lower transmission rate than the data frames, this should be a reasonable assumption in many practical environments [37]. For all these conditions, and taking into account the fragmentation factor k, expression (3) is transformed to the following form:

DOI 10.15588/1607-3274-2023-3-15

$$S = \frac{2nL_0(1-P_b)^{\frac{L_0}{k}}}{2n(1-P_b)^{\frac{L_0}{k}}(T_{SC}-T_C)_k + [Q+1-2(1-P_b)^{\frac{L_0}{k}}]\{T_{Ck}[\frac{Q+1}{Q+1-2(1-P_b)^{\frac{L_0}{k}}}]^n + \eta\sigma\}}, \quad (8)$$

where

$$T_{Ck} = PHY_{hdr} + k \cdot MAC_{hdr} + \frac{L_0}{R} + DIFS + k\delta, \quad (9)$$

$$(T_{SC} - T_C)_k = k \cdot (SIFS + ACK + \delta). \quad (10)$$

Dependences $S(n)$ calculated in accordance with expressions (8)–(10) at different values of k for conditions of increased noise intensity (BER = $5 \cdot 10^{-5}$), high intensity (BER = 10^{-4}) and very high intensity (BER = $2 \cdot 10^{-4}$) are shown in Fig. 2, 3 and 4, respectively.

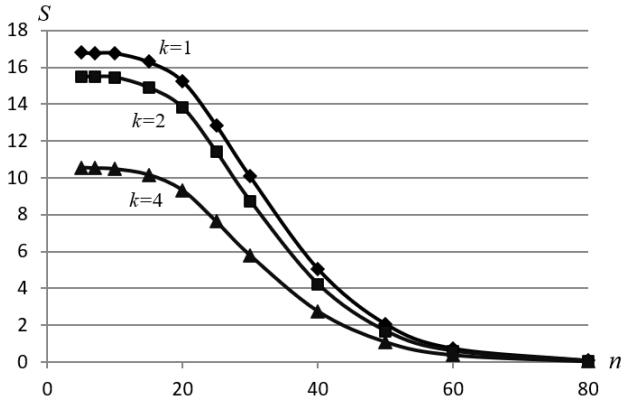


Figure 2 – Throughput S versus the number of competing stations n at different value of fragmentation factor k for BER = $5 \cdot 10^{-5}$

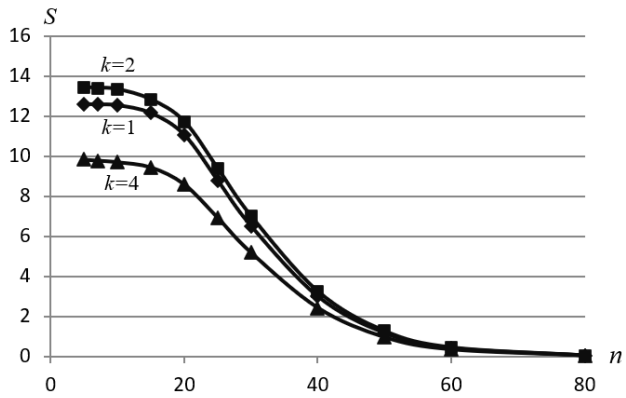


Figure 3 – Throughput S versus the number of competing stations n at different value of fragmentation factor k for BER = 10^{-4}

As follows from the graphs shown in Fig. 2–4, if for BER = $5 \cdot 10^{-5}$ with an increase in the fragmentation factor k, the values of S decrease for all values of n due to the predominant influence of an increase in overhead costs, then for BER = 10^{-4} and especially for BER = $2 \cdot 10^{-4}$ we observe the opposite effect. This is most characteristically for fragmentation with $k = 2$, in which the throughput values in relation to the initial frame increase most significantly for all values of n.

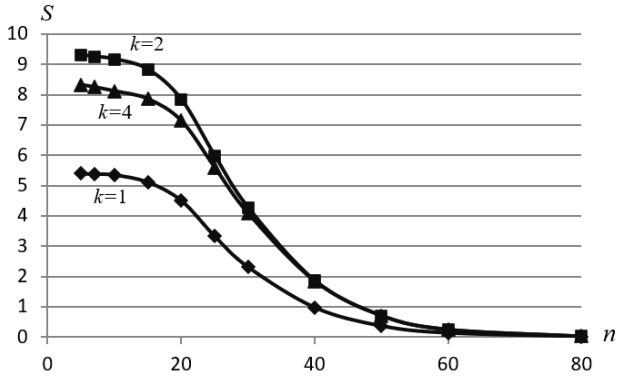


Figure 4 – Throughput S versus the number of competing stations n at different value of fragmentation factor k for BER = $2 \cdot 10^{-4}$

6 DISCUSSION

Based on the Markov chain model, which is used by many researchers as the basic for studying the behavior of wireless networks of IEEE 802.11 DCF standard, we expressed in an explicit analytical form the dependence of the throughput of the network infrastructure domain on the number of operating stations, the value of BER, which is determined by the interference intensity in radio path, and the size of the transmitted frames. This allowed us to investigate these dependencies for different lengths of the frame data field.

The initial length of the data field is taken equal to the standard – 12000 bits. The calculations were carried out with the error probabilities in one bit of data 10^{-5}, $5 \cdot 10^{-5}$, 10^{-4}, $2 \cdot 10^{-4}$. The first two values correspond to the probabilities 0.887 and 0.549 of standard-length frame successful transmission and correspond to increased noise intensity. The third and fourth values correspond to the probabilities 0.3 and 0.09 of successful transmission. We can assume that they correspond to high and very high noise levels, respectively. Reducing the length of the frame data field to 3000 bits made it possible to increase the probability of successful transmission to 0.97; 0.86; 0.74 and 0.55, respectively.

Analyzing the obtained dependencies, it should be noted that in the range of increased and high noise intensity, a decrease in the length of the frame data field from the standard length of 12000 bits to 3000 bits is accompanied by a decrease in the throughput for all values of the number of competing stations. This is because as the

DOI 10.15588/1607-3274-2023-3-15

amount of the transmitted data decreases, and the relative overhead increases. At the same time, it can be seen from the above tables that as the amount of the frame data decreases from 12000 bits to 6000 bits and further to 3000 bits, the throughput becomes less susceptible to an increase in the noise intensity. So, for example, for 20 stations, an increase in the error probability in one bit from $5 \cdot 10^{-5}$ to 10^{-4} leads to a decrease in the throughput by 27,5 % for 12000 bits, by 13% for 6000 bits, and by 7% for 3000 bits. Similar results are also observed for a different number of operating stations.

Qualitatively different results are obtained in the region of very high interference intensity. A significant increase in the probability of frame transmission in this region observed with a decrease in the standard length of the frame data field by 2 and 3 times, made it possible to increase the throughput compared to the original one. This effect is especially pronounced when the length is halved.

Consider the process of sending a frame with data of standard length but divided into a several fragments. In the basic DCF scheme, only the first fragment of the transmitted frame contends for access to the communication channel with the access point, each subsequent fragment is transmitted after the separating SIFS interval and is acknowledged by the ACK frame.

For these conditions, previously obtained by us expression for the throughput is transformed taking into account the number of transmitted fragments. As follows from the calculations, if for BER = $5 \cdot 10^{-5}$ with an increase in the fragmentation factor, the throughput values decrease over the entire range of the number of simultaneously operating stations, due to the predominant influence of the increase in the overhead costs, then already in the region of high (BER = 10^{-4}) and, accordingly, very high (BER = $2 \cdot 10^{-4}$) noise intensity we have the opposite effect. The throughput increases the most at a fragmentation factor of two.

It is of interest to compare the results obtained in sections 3 and 4 of this article at different noise levels.

For BER = $5 \cdot 10^{-5}$, a simple decrease in the frame data field from 12000 bits to 6000 bits and further to 3000 bits, for example, for 25 stations, leads to a decrease in the throughput by 1.46 and 2.44 times, respectively. In the case of using fragmentation, an increase in the fragmentation factor from $k = 1$ to $k = 2$ and further to $k = 4$ with the same number of the operating stations leads to a decrease in the throughput by 1.13 and 1.69 times, respectively, i.e., significantly less than in the previous case.

At BER = 10^{-4} for the same conditions, in the first case, a decrease in the throughput by 1,19 and 1,82 times is observed, and in the second case, for fragmentation, with an increase in k from 1 to 2, an increase in the throughput by 1,07 times is observed, and with an increase in k from 1 to 4 – a decrease in the throughput, but only by 1,27 times.

For BER = $2 \cdot 10^{-4}$ at 25 stations, reducing the frame data field from 12000 to 6000 bits leads to an increase in the throughput by 1,44 times, and when reducing from 12000 to 3000 bits to an increase in the throughput by 1.19 times. When fragmenting the data field of a standard length of 12000 bits, the transition from $k = 1$ to $k = 2$ gives an increase in the throughput by 1,8 times, and from $k = 1$ to $k = 4$ – by 1,68 times.

Thus, it can be stated that the fragmentation mechanism under conditions of increased and high noise intensity is more beneficial in terms of stabilization or increase in throughput than simply reducing the length of the frame data field.

CONCLUSIONS

The scientific novelty. Using an approach based on the Markov chains modeling of the operation of IEEE 802.11 DCF wireless networks, we first expressed in an explicit analytical form the dependence of the infrastructure domain throughput on the number of operating stations, the value of BER, which is determined by the level of interference in the domain space, and the fragmentation factor of the transmitted frame data field.

It is shown that despite the increase the noise immunity, a decrease in the length of the transmitted frame data field is accompanied by reduce in throughput in the range BER=$10^{-5} – 10^{-4}$. And only at a very high noise intensity, corresponding to BER = $2 \cdot 10^{-4}$, a decrease in the frame size leads to an increase in the throughput.

The study of the process of the standard frame transmitting, but with a fragmented data field, showed that if for BER = $5 \cdot 10^{-5}$ with an increase in fragmentation factor, the throughput values decrease in the entire range of the number of stations due to the predominant increase in overhead costs, then in the region of high (BER = 10^{-4}) and very high noise intensity (BER = $2 \cdot 10^{-4}$) we have the opposite effect. To the greatest extent, the throughput increases when the frame data is transmitted in two equal fragments.

For the first time, a comparison was made of the network throughput determined by simply reducing the length of the frame data field and using fragmentation of a standard frame. The comparison showed that the use of the fragmentation mechanism is more beneficial both when throughput is stabilized under conditions of increased noise intensity and when the throughput is increased under conditions of high and very high noise intensity.

The practical significance of this work lies in the fact that the results obtained in it make it possible to determine the optimal value of the fragmentation factor of the frame data field transmitted in the wireless network, depending on the number of operating in the domain stations and the intensity of electromagnetic interference.

One of the priorities of wireless networks is their use for the automation of production processes in a number of industries. A common factor that reduces their efficiency is the high electromagnetic interference level in the shops of industrial enterprises, due to the operation of technological equipment. Studying the possibilities of increasing the networks throughput in such conditions, with

DOI 10.15588/1607-3274-2023-3-15

BER = $10^{-5} – 2 \cdot 10^{-4}$, is of significant practical importance.

Prospects for further research are to study the reveal regularities of the joint influence of collisions and noise on the transmission efficiency of fragmented frames in the conditions of further development of modern wireless network technologies.

ACKNOWLEDGEMENTS

The work was carried out at the Department of Electronic Computing Machinery of the Oles Honchar Dnipro National University within the framework of research project “Improving the efficiency of information processing in the process of its formation and transmission using modern computer systems and networks”, No 0122U001400.

REFERENCES

1. Giordani M., Polese M., Mezzavilla M. et al. Toward 6G Networks: Use Cases and Technologies, *IEEE Communications Magazine*, 2022, Vol. 58, No. 3, pp. 55–61, DOI: 10.1109/MCOM.001.1900411.
2. Chien T. V., Ngo H. Q., Chatzinotas S. et al. Reconfigurable Intelligent Surface-Assisted Cell-Free Massive MIMO Systems Over Spatially-Correlated Channels, *IEEE Transactions on Wireless Communications*, 2022, Vol. 21, No. 7, pp. 5106–5128. DOI: 10.1109/TWC.2021.3136925.
3. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update : 2017–2022, 2019, pp. 1–33. Document ID 1486680503328360.
4. Dianne S. V. Medeiros, Helio N. Cunha Neto, Martin Andreoni Lopez et al.] A survey on data analysis on large scale wireless networks: online stream processing, trends and challenges, *Journal of Internet Services and Applications*, 2020, Vol. 11, No. 1. DOI: 10.1186/s13174-020-00127-2.
5. 802.11ac: The Fifth Generation of Wi-Fi. Technical White Paper [Electronic resource]. Access mode: https://www.cisco.com/c/dam/en/us/products/collateral/wireless/aironet-3600-series/white-paper-c11-713103.pdf.
6. IEEE 802.11ax: The Sixth Generation of Wi-Fi. White Paper [Electronic resource]. Access mode: https://www.cisco.com/c/en/us/products/collateral/wireless/white-paper-c11-740788.html.
7. Lee H., Tinnirello I., Yu J. et al. Throughput and Delay Analysis of IEEE 802.1le Block ACK with Channel Errors, *2nd International Conference on Communication Systems Software and Middleware, Bangalore, 7–12 January 2007 : proceeding*. IEEE, 2007, pp. 1–7. DOI: 10.1109/COMSWA.2007.382498.
8. Youngsoo K., Sunghyun Ch., Kyunghun J. et al.Throughput enhancement of IEEE 802.11 WLAN via frame aggregation, *IEEE 60th Vehicular Technology Conference : VTC2004-Fall, Los Angeles, 26–29 September 2004 : proceeding*, IEEE, 2004, Vol. 4, pp. 3030–3034. DOI: 10.1109/VETECF.2004.1400617.
9. Bianchi G. Performance analysis of the IEEE 802.11 distributed coordination function, *IEEE Journal on Selected Areas in Communications*, 2000, Vol. 18, No. 3, pp. 535–547. DOI: 10.1109/49.840210.
10. Tinnirello I., Bianchi G., Xiao Y. Refinements on IEEE 802.11 Distributed Coordination Function Modeling Approaches, *IEEE Transactions on Vehicular Technology*, 2010, Vol. 59, No. 3, pp. 1055–1067, DOI: 10.1109/TVT.2009.2029118.
11. Tay Y. C., Chua K. C. A capacity analysis for the IEEE 802.11 MAC protocol, *Wireless Networks*, 2001, Vol. 7, No. 2, pp. 159–171. DOI: 10.1023/A:1016637622896.
12. Vermeulen T., Reynders B. , Rosas F. E. et al. Performance analysis of in-band collision detection from dense wireless networks, *EURASIP Journal on Wireless Communications and Networking*, 2021. Article number: 87, DOI: 10.1186/s13638-021-01940-4.
13. Tourrilhes J. Fragment adaptive reduction: coping with various interferers in radio unlicensed bands, *IEEE international conference of communications (ICC’01)*. Helsinki, 11–14 June 2001, proceeding, IEEE, 2001, pp. 239–244.
14. Mafole P., Aritsugi M. Analysis and performance assessment of a fragment retransmission scheme for energy efficient IEEE 802.11 WLANs, *Springer Plus*, 2016, Vol. 5, No. 1. DOI: 10.1186/s40064-016-3023-6.
15. Khandetsky V. S., Karpenko N. V. Modeling of IEEE 802.11 Computer Networks at Increased Interference Intensity, *Radio Electronics, Computer Science, Control*, 2022, No. 2, pp. 132–139. DOI 10.15588/1607-3274-2022-2-13.
16. Foh C. H., Tantra J. W. Comments on IEEE 802.11 saturation throughput analysis with freezing of backoff counters, *IEEE Communications Letters*, 2005, Vol. 9, No. 2, pp. 130–132. DOI: 10.1109/LCOMM.2005.02008.
17. Scarpa M., Serrano S. A New Modelling Approach to Represent the DCF Mechanism of the CSMA/CA Protocol, *24th International Conference on Analytical and Stochastic Modelling Techniques and Applications, ASMTA 2017. Lecture Notes in Computer Science, Newcastle-upon-Tyne, 10–11 July 2017, proceeding.* Springer, Cham, 2017, Vol. 10378, pp. 181–195. DOI: 10.1007/978-3-319-61428-1_13.
18. Krishman M. N., Pollin S., Zakhor A. Local Estimation of Collision Probabilities in 802.11 WLANs with Hidden Terminals : Technical Report No. UCB/EECS-2009-2, January 2009, *Electrical Engineering and Computer Sciences University of California at Berkeley*, 11 p. [Electronic resource]. – Access mode: https://www2.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-2.pdf.
19. Sweedy A. M., Semeia A. I., Sayed S. Y. et al. The effect of frame length, fragmentation and RTS/CTS mechanism on IEEE 802.11 MAC performance, *10th International Conference on Intelligent Systems Design and Applications.* Cairo, 29.11.2010-1.12.2010 : proceeding, IEEE, 2010, pp. 1338–1344, DOI: 10.1109/ISDA.2010.5687095.
20. Liao R., Bellalta, Barcelo J. et al. Performance analysis of IEEE 802.11ac B. wireless backhaul networks in saturated conditions, *EURASIP Journal of Wireless Communications and Networking*, 2013. Article number 226, pp. 1–14. DOI: 10.1186/1687-1499-2013-226.
21. Ugwu G. O., Nwawelu U. N., Ahaneku M. A. et al. Effect of service differentiation on QoS IEEE 802.11e enhanced distributed channel access: a simulation approach, *Journal of Engineering and Applied Science*, 2022, Vol. 69, No. 1, pp. 1–18. DOI: 10.1186/s44147-021-00055-3.
22. Malone D., Duffy K., Leith D. Modeling the 802.11 Distributed Coordination Function in Nonsaturated Heterogeneous Conditions, *IEEE/ACM Transactions on Networking*, 2007. Vol. 15, No. 1, pp. 159–172. DOI: 10.1109/TNET.2006.890136.
23. Bozkurt A. Optimal delay analysis for real-time traffics over IEEE 802.11 wireless LANs, *EURASIP Journal of Wireless*

DOI 10.15588/1607-3274-2023-3-15

Communications and Networking, 2016, Article number 52, pp. 1–13. DOI: 10.1186/s13638-016-0545-0.
24. Tian G., Tia Y.-C. Modelling and performance evaluation of the IEEE 802.11 DCF for real-time control, *Computer Networks*, 2012, Vol. 56, No. 1, pp. 434–447. DOI: 10.1016/j.comnet.2011.10.001.
25. Aldbaibani O. A., Raschella A., Mohi-Ud-Din G. et al. A User Prioritization for Horizontal Handover in Dense WLAMs, *International Journal of Wireless Information Networks*, 2022, Vol. 29, pp. 130–142. DOI: 10.1007/s10776-021-00544-5.
26. Matthew Gast. 802.11 Wireless Networks: The Definitive Guide. O' Reilly Media, Inc, 2005, 630 p.
27. Cover T. M., Thomas J. A.. Elements of Information Theory. Second Edition. Hoboken, New Jersey, Wiley, 2006, 774 p.
28. Proakis J. G., Salehi M. Digital Communications. Fifth Edition. McGrow-Hill Education, 2007, 1150 p.
29. Li T., Ni Q., Turletti T. et al. Performance analysis of the IEEE 802.11e block ACK scheme in a noisy channel, *2nd International Conference on Broadband Networks, Boston, 7 October 2005 : proceeding, IEEE*, 2005, Vol. 1, pp. 511–517. DOI: 10.1109/ICBN.2005.1589655.
30. Yazid M., Aissani D., Bouallouche-Medijkoune L. et al. Modelling and enhancement of the IEEE 802.11 RTS/CTS scheme in an error-prone channel, *Formal Aspects of Computing*, 2015, Vol. 27, No. 1, pp. 33–52. DOI: 10.1007/s00165-014-0300-4.
31. Breed G. Bit error rate: fundamental concepts and measurement issues, *High Frequency Electronics*, 2003, Vol. 2, No. 1, pp. 46–48.
32. Ghani B., Launay F., Pousset Y. et al. Low complexity hybrid interference cancellation for sparse code multiple access, EURASIP Journal of Wireless Communications and Networking, 2022,Article number 95, pp. 1–26, DOI: 10.1186/s13638-022-02162-y.
33. Tan B. S., Li K. H., Teh K. C. Bit – error rate analysis of low-density parity-check codes with generalized selection combining over a Rayleigh-fading channel using Gaussian approximation, *IET Communication*, 2012, Vol. 6, No. 1, pp. 90–96. DOI: 10.1049/iet-com.2011.0271.
34. Yazid M., Bouallouche-Medikoune L., Aissani D. et al. Analytical analysis of applying packet fragmentation mechanism on IEEE 802.11b DCF network in non-ideal channel with infinite load conditions, Wireless Networks, 2014, Vol. 20, pp. 917–934. DOI: 10.1007/s11276-013-0653-2.
35. Chang Z., Alanen O., Huovinen T. et al. Performance Analysis of IEEE 802.11ac DCF with Hidden Nodes, *IEEE 75th Vehicular Technology Conference (VTC Spring), Yokohama, 6–9 May 2012 : proceeding*, IEEE, 2012, pp. 1–5. DOI: 10.1109/VETECS.2012.6240054.
36. Patidar R., Roy S., Henderson T. R. et al. Validation of Wi-Fi network simulation on ns-3 : Technical Report : August 7, 2017, University of Washington. Seattle, 2017, 19 p.
37. Lee H., Tinnirello I., Yu L. et al. A performance analysis of block ACK scheme for IEEE 802.11e networks, *Computer Networks*, 2010, Vol. 54, No. 14, pp. 2468–2481. DOI: 10.1016/j.comnet.2010.04.001.

Received 30.06.2023.
Accepted 27.08.2023.

УДК 004.77

АНАЛІЗ ФУНКЦІОНУВАННЯ БЕЗДРОТОВИХ КОМП'ЮТЕРНИХ МЕРЕЖ В УМОВАХ ВИСОКОЇ ІНТЕНСИВНОСТІ ЗАВАД

Хандецький В. С. – д-р техн. наук, професор, завідувач кафедри електронних обчислювальних машин Дніпровського національного університету імені Олеся Гончара, Дніпро, Україна.

Герасимов В. В. – канд. техн. наук, доцент кафедри електронних обчислювальних машин Дніпровського національного університету імені Олеся Гончара, Дніпро, Україна.

Карпенко Н. В. – канд. фіз.-мат. наук, доцент кафедри електронних обчислювальних машин Дніпровського національного університету імені Олеся Гончара, Дніпро, Україна.

АНОТАЦІЯ

Актуальність. Зниження імовірності успішної передачі фрейму в інфраструктурному домені бездротової мережі стандарту IEEE 802.11 DCF обумовлено як впливом інтенсивності колізій так і дією зовнішніх завад у радіоканалі. Використовуючи в якості базового підходу той, що, заснований на використанні Марківських ланцюгів, ми в явній аналітичній формі виразили залежність пропускної здатності мережі від кількості працюючих станцій, швидкості бітових помилок (BER) і коефіцієнта фрагментації фрейму.

Мета роботи. Метою даної статті є дослідження впливу інтенсивності шуму на величину пропускної здатності домену бездротової мережі в широкому діапазоні кількості працюючих станцій, під час передачі фреймів різної довжини за відсутності та із застосуванням механізму фрагментації.

Метод. Математичне моделювання показало, що в діапазоні підвищеної та високої інтенсивності шуму ($BER = 10^{-5} – 10^{-4}$), зменшення довжини поля даних фрейму, що передається, від стандартних 12000 біт до 3000 біт супроводжується зниженням пропускної здатності для будь-якої кількості конкуруючих станцій. Одночасно з цим слід відмітити, що пропускна здатність стає менш сприятливою до збільшення інтенсивності шуму. В області дуже високої інтенсивності завад ($BER = 2\cdot10^{-4}$) одержані результати, які якісно відрізняються. Значне збільшення імовірності передачі фрейму в цій області, яке спостерігається зі зменшенням стандартної довжини поля даних в 2–3 рази, дозволило підвищити пропускну здатність порівняно з початковою. Цей ефект є особливо вираженим у випадку, коли довжина зменшується вдвічі.

Результати. Дослідження процесу передачі фрейму стандартного розміру але з фрагментованим полем даних показало, що для $BER \le 5\cdot10^{-5}$ зі збільшенням коефіцієнту фрагментації значення пропускної здатності знижуються на всьому діапазоні кількості працюючих станцій переважно за рахунок впливу зростання накладних втрат. Однак в області високої ($BER = 10^{-4}$) і надвисокої ($BER = 2\cdot10^{-4}$) інтенсивності шуму ми маємо зворотний ефект. Найбільше зростання пропускної здатності спостерігається, коли дані фрейму передаються двома рівними фрагментами. Ми провели порівняння пропускної

DOI 10.15588/1607-3274-2023-3-15

здатності мережі, визначеної при простому зменшенні довжини поля даних фрейму і для передачі фрейму з даними стандартної довжини, розділеними на декілька фрагментів. Порівняння показало, що використання механізму фрагментації є більш вигідним як при стабілізації пропускної здатності в умовах підвищеної інтенсивності шуму, так і при збільшенні пропускної здатності в умовах високої і надвисокої інтенсивності шуму.

Висновки. У даній статті була модифікована математична модель, що дозволяє безпосередньо обчислювати пропускну здатність бездротової мережі. Використовуючи цю модель, ми дослідили зміну пропускної здатності в широкому діапазоні BER і кількості працюючих станцій, для різних значень коефіцієнта фрагментації фрейму, що передається. Визначено умови підвищення пропускної здатності.

КЛЮЧОВІ СЛОВА: бездротові мережі IEEE 802.11, DCF, пропускна здатність, інфраструктурний домен, BER, фрейм, коефіцієнт фрагментації, колізія.

ЛІТЕРАТУРА

1. Toward 6G Networks: Use Cases and Technologies / [M. Giordani, M. Polese, M. Mezzavilla et al.] // IEEE Communications Magazine. – 2022. – Vol. 58, No. 3. – P. 55–61. DOI: 10.1109/MCOM.001.1900411.
2. Reconfigurable Intelligent Surface-Assisted Cell-Free Massive MIMO Systems Over Spatially-Correlated Channels / [T. V. Chien, H. Q. Ngo, S. Chatzinotas et al.] // IEEE Transactions on Wireless Communications. – 2022. – Vol. 21, No. 7. – P. 5106–5128, DOI: 10.1109/TWC.2021.3136925.
3. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update : 2017–2022. – 2019. – P. 1–33. – Document ID 1486680503328360.
4. A survey on data analysis on large scale wireless networks: online stream processing, trends and challenges / [Dianne S. V. Medeiros, Helio N. Cunha Neto, Martin Andreoni Lopez et al.] // Journal of Internet Services and Applications. – 2020. – Vol. 11, No. 1. – DOI: 10.1186/s13174-020-00127-2.
5. 802.11ac: The Fifth Generation of Wi-Fi. Technical White Paper [Electronic resource]. – Access mode: https://www.cisco.com/c/dam/en/us/products/collateral/wireless/aironet-3600-series/white-paper-c11-713103.pdf.
6. IEEE 802.11ax: The Sixth Generation of Wi-Fi. White Paper [Electronic resource]. – Access mode: https://www.cisco.com/c/en/us/products/collateral/wireless/white-paper-c11-740788.html.
7. Throughput and Delay Analysis of IEEE 802.1le Block ACK with Channel Errors / [H. Lee, I. Tinnirello, J. Yu et al.] // 2nd International Conference on Communi-cation Systems Software and Middleware, Bangalore, 7–12 January 2007 : proceeding. – IEEE, 2007. – P. 1–7, DOI: 10.1109/COMSWA.2007.382498.
8. Throughput enhancement of IEEE 802.11 WLAN via frame aggregation / [K. Youngsoo, Ch. Sunghyun, J. Kyunghun et al.] // IEEE 60th Vehicular Technology Conference : VTC2004-Fall, Los Angeles, 26–29 Septem-ber 2004 : proceeding. – IEEE, 2004. – Vol. 4. – P. 3030–3034, DOI: 10.1109/VETECF.2004.1400617.
9. Bianchi G. Performance analysis of the IEEE 802.11 distributed coordination function / G. Bianchi // IEEE Journal on Selected Areas in Communications. – 2000. – Vol. 18, No. 3. – P. 535–547, DOI: 10.1109/49.840210.
10. Tinnirello I. Refinements on IEEE 802.11 Distrib-uted Coordination Function Modeling Approaches / I. Tinnirello, G. Bianchi, Y. Xiao // IEEE Transactions on Vehicular Technology. – 2010. – Vol. 59, No. 3. – P. 1055–1067, DOI: 10.1109/TVT.2009.2029118.
11. Tay Y. C. A capacity analysis for the IEEE 802.11 MAC protocol / Y. C. Tay, K. C. Chua // Wireless Net-works. – 2001. – Vol. 7, No. 2. – P. 159–171. DOI: 10.1023/A:1016637622896.
12. Performance analysis of in-band collision detec-tion from dense wireless networks / [T. Vermeulen, B. Reynders, F. E. Rosas et al.] // EURASIP Journal on Wireless Communications and Networking. – 2021. – Article number: 87. DOI: 10.1186/s13638-021-01940-4.
13. Tourrilhes J. Fragment adaptive reduction: coping with various interferers in radio unlicensed bands / J. Tourrilhes // IEEE international conference of commu-nications (ICC'01), Helsinki, 11–14 June 2001 : proceeding. – IEEE, 2001. – P. 239–244.
14. Mafole P. Analysis and performance assessment of a fragment retransmission scheme for energy efficient IEEE 802.11 WLANs / P. Mafole, M. Aritsugi // Springer Plus. – 2016. – Vol. 5, No. 1, DOI: 10.1186/s40064-016-3023-6.
15. Khandetsky V. S. Modeling of IEEE 802.11 Com-puter Networks at Increased Interference Intensity / V. S. Khandetsky, N. V. Karpenko // Radio Electronics, Computer Science, Control. – 2022. – No. 2. – P. 132–139. DOI 10.15588/1607-3274-2022-2-13.
16. Foh C. H. Comments on IEEE 802.11 saturation throughput analysis with freezing of backoff counters / C. H. Foh, J. W. Tantra // IEEE Communications Letters. – 2005. – Vol. 9, No. 2. – P. 130–132, DOI: 10.1109/LCOMM.2005.02008.
17. Scarpa M. A New Modelling Approach to Repre-sent the DCF Mechanism of the CSMA/CA Protocol / M. Scarpa, S. Serrano // 24th International Conference on Analytical and Stochastic Modelling Techniques and Applications, ASMTA 2017. Lecture Notes in Computer Science, Newcastle-upon-Tyne, 10–11 July 2017 : pro-ceeding. – Springer, Cham, 2017. – Vol. 10378. – P. 181–195. DOI: 10.1007/978-3-319-61428-1_13.
18. Krishman M. N. Local Estimation of Collision Probabilities in 802.11 WLANs with Hidden Terminals : Technical Report No. UCB/EECS-2009-2, January 2009 / M. N. Krishman, S. Pollin, A. Zakhor / Electrical Engineer-ing and Computer Sciences University of California at Berkeley. – 11 p. [Electronic resource]. – Access mode: https://www2.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-2.pdf.
19. The effect of frame length, fragmentation and RTS/CTS mechanism on IEEE 802.11 MAC perfor-mance / [A. M. Sweedy, A. I. Semeia, S. Y. Sayed et al.] // 10th International Conference on Intelligent Systems Design and Applications, Cairo, 29.11.2010-1.12.2010 : proceeding. – IEEE, 2010. – P. 1338–1344, DOI: 10.1109/ISDA.2010.5687095.
20. Performance analysis of IEEE 802.11ac wireless backhaul networks in saturated conditions / [R. Liao, B. Bellalta, J. Barcelo et al.] // EURASIP Journal of Wire-less Communications and Networking. – 2013. – Article number 226. – P. 1–14 p. DOI: 10.1186/1687-1499-2013-226.
21. Effect of service differentiation on QoS IEEE 802.11e enhanced distributed channel access: a simula-tion approach / [G. O. Ugwu, U. N. Nwawelu, M. A. Ahaneku et al.] //

DOI 10.15588/1607-3274-2023-3-15

Journal of Engineering and Applied Science. – 2022. – Vol. 69, No. 1. – P. 1–18. DOI: 10.1186/s44147-021-00055-3.
22. Malone D. Modeling the 802.11 Distributed Coor-dination Function in Nonsaturated Heterogeneous Condi-tions / D. Malone, K. Duffy, D. Leith // IEEE/ACM Transactions on Networking. – 2007. – Vol. 15, No. 1. – P. 159–172. DOI: 10.1109/TNET.2006.890136.
23. Bozkurt A. Optimal delay analysis for real-time traffics over IEEE 802.11 wireless LANs / A. Bozkurt // EURASIP Journal of Wireless Communications and Net-working. – 2016. – Article number 52. – P. 1–13. DOI: 10.1186/s13638-016-0545-0.
24. Tian G. Modelling and performance evaluation of the IEEE 802.11 DCF for real-time control / G. Tian, Y.-C. Tia // Computer Networks. – 2012. – Vol. 56, No. 1. – P. 434–447. DOI: 10.1016/j.comnet.2011.10.001.
25. A User Prioritization for Horizontal Handover in Dense WLAMs / [O. A. Aldbaibani, A. Raschella, G. Mohi-Ud-Din et al.] // International Journal of Wireless Information Networks. – 2022. – Vol. 29. – P. 130–142. DOI: 10.1007/s10776-021-00544-5.
26. Matthew Gast. 802.11 Wireless Networks: The Definitive Guide. – O' Reilly Media, Inc, 2005. – 630 p.
27. Cover T. M. Elements of Information Theory. Second Edition / T. M. Cover, J. A. Thomas. – Hoboken, New Jersey: Wiley, 2006. – 774 p.
28. Proakis J. G. Digital Communications. Fifth Edi-tion / J. G. Proakis, M. Salehi. – McGrow-Hill Education, 2007. – 1150 p.
29. Performance analysis of the IEEE 802.11e block ACK scheme in a noisy channel / [T. Li, Q. Ni, T. Turletti et al.] // 2nd International Conference on Broadband Networks, Boston, 7 October 2005 : proceeding. – IEEE, 2005. – Vol. 1. – P. 511–517. DOI: 10.1109/ICBN.2005.1589655.
30. Modelling and enhancement of the IEEE 802.11 RTS/CTS scheme in an error-prone channel / [M. Yazid, D. Aissani, L. Bouallouche-Medijkoune et al.] // Formal Aspects of Computing. – 2015. – Vol. 27, No. 1. – P. 33–52. DOI: 10.1007/s00165-014-0300-4.
31. Breed G. Bit error rate: fundamental concepts and measurement issues / G. Breed // High Frequency Elec-tronics. – 2003. – Vol. 2, No. 1. – P. 46–48.
32. Low complexity hybrid interference cancellation for sparse code multiple access / [B. Ghani, F. Launay, Y. Pousset et al.] // EURASIP Journal of Wireless Com-munications and Networking. – 2022. – Article number 95. – P. 1–26, DOI: 10.1186/s13638-022-02162-y.
33. Tan B. S. Bit – error rate analysis of low-density parity-check codes with generalized selection combining over a Rayleigh-fading channel using Gaussian approxi-mation / B. S. Tan, K. H. Li, K. C. Teh // IET Communi-cation. – 2012. – Vol. 6, No. 1. – P. 90–96. DOI: 10.1049/iet-com.2011.0271.
34. Analytical analysis of applying packet fragmenta-tion mechanism on IEEE 802.11b DCF network in non-ideal channel with infinite load conditions / [M. Yazid, L. Bouallouche-Medikoune, D. Aissani et al.] // Wireless Networks. – 2014. – Vol. 20. – P. 917–934. DOI: 10.1007/s11276-013-0653-2.
35. Performance Analysis of IEEE 802.11ac DCF with Hidden Nodes / [Z. Chang, O. Alanen, T. Huovinen et al.] // IEEE 75th Vehicular Technology Conference (VTC Spring), Yokohama, 6–9 May 2012 : proceeding. – IEEE, 2012. – P. 1–5. DOI: 10.1109/VETECS.2012.6240054.
36. Validation of Wi-Fi network simulation on ns-3 : Technical Report : August 7, 2017 / [R. Patidar, S. Roy, T. R. Henderson et al.] / University of Washington – Seattle, 2017. – 19 p.
37. A performance analysis of block ACK scheme for IEEE 802.11e networks / [H. Lee, I. Tinnirello, L. Yu et al.] // Computer Networks. – 2010. – Vol. 54, No. 14. – P. 2468–2481. DOI: 10.1016/j.comnet.2010.04.001.

DOI 10.15588/1607-3274-2023-3-15

UDC 519.85:004.42+004.02

MODIFIED GENETIC ALGORITHM APPROACH FOR SOLVING THE TWO-STAGE LOCATION PROBLEM

Serhieiev O. S. – Post-graduate student of the Department of System Analysis and Control, National TU Dnipro Polytechnic, Dnipro, Ukraine.

Us S. A. – PhD, Professor of the Department of System Analysis and Control, National TU Dnipro Polytechnic, Dnipro, Ukraine.

ABSTRACT

Context. Optimization of logistics processes is one of the important tasks of supply chain management in various fields, including medicine. Effective coordination in medical logistics is essential to ensure public health and prosperity. This is especially essential during global emergencies when the rapid and efficient distribution of medicines is critical. In addition, professional logistics management is critical to delivering humanitarian aid, where the timely transportation of medical supplies and resources can be life-saving. The most advanced technologies and algorithms are being used to improve medical logistics processes. This paper considers modifying the genetic algorithm for solving the two-stage location problem in supply chain management in the distribution of medicines and medical equipment.

Objective. The work aims to build a model and develop an algorithm for solving a two-stage location problem in the context of the medical logistics problem with further analysis of their applications and performance.

Method. We propose to use a genetic algorithm to solve a two-stage logistics problem. The peculiarities of this algorithm are the modification of evaluation procedures and the use of mixed mutation, which allows for solving the problem effectively, considering irregularities in the statement regarding the subject – the limits on the centers' location at several stages of the logistic process.

Results. The paper deals with a two-stage location problem with constraints on the maximum number of centers. Considering the specific requirements of medical logistics in the transportation context of medicines and medical equipment, a mathematical model and modification of the genetic algorithm are proposed. The developed algorithm is tested on model tasks and can produce effective solutions for problems ranging in size from 25 to 1000. The solution process takes longer for larger problems with dimensions from 1001 to 2035. Additionally, the influence of increasing the maximum generations number on the time of execution is investigated. When the maximum generation value increases from 50 to 100 and from 100 to 150 generations, the algorithm's execution time increases by 45.69% and 51.68%, respectively. 73% of the total execution time is dedicated to the evaluation procedure. The algorithm is applied to the medical logistics problem in the Dnipropetrovsk region (Ukraine). An efficient solution is obtained within an acceptable execution time.

Conclusions. A mathematical model for a two-stage location problem in the context of medical logistics is introduced. It considers the peculiarities of the medical field. A solution algorithm based on a genetic approach is developed and applied to the medical logistics problem. The algorithm has been tested on model tasks of varying sizes, with a comprehensive analysis conducted on the correlation between the problem size and the algorithm's running time. In addition, it is investigated how the maximum number of generations affects the algorithm's execution time. The role of each stage in the genetic algorithm research towards the overall effectiveness of the algorithm is researched. The obtained results indicate high efficiency and wide application possibilities of the proposed mathematical model and algorithm. The developed method demonstrates high performance and reliability.

KEYWORDS: two-stage location problem, genetic algorithm, priority-based encoding, medical logistics.

ABBREVIATIONS

RCs are regional centers;
SRCs are subregional centers;
RAM is a random access memory;
OSM is an Open Street Map service.

NOMENCLATURE

M_j are the operating expenses associated with the activation of a subregional center j ($j = \overline{1, J}$);

c_{ij}^1 are the transportation costs for volume weight unit of medicines and medical equipment from regional center i to subregional center j ($i = \overline{1, I}$, $j = \overline{1, J}$);

c_{jk}^2 are the transportation costs for volume weight unit of medicines and medical equipment from subregional center j to medical warehouse k ($j = \overline{1, J}$, $k = \overline{1, K}$);

d_k is the capacity of medical warehouse k ($k = \overline{1, K}$);

I is the number of regional centers;

J is the number of subregional centers;

K is the number of medical warehouses;

L is the limit value of subregional centers that can be located;

N is the limit value of regional centers that can be located;

P_i are the operating expenses associated with the activation of a regional center i ($i = \overline{1, I}$);

r_i are the medicines and medical equipment stocks in the regional center i ($i = \overline{1, I}$);

s_j is the capacity of subregional center j ($j = \overline{1, J}$);

v_i is a boolean variable, where $v_i = 1$, if regional center i is located, $v_i = 0$ otherwise;

DOI 10.15588/1607-3274-2023-3-16

x_{ij} are the volume weight units number of medicines and medical equipment transported from regional center *i* to subregional center *j*;

y_{jk} are the volume weight units number of medicines and medical equipment transported from subregional center *j* to medical warehouse *k*;

z_j is a boolean variable, where $z_j = 1$, if subregional center *j* is located, $z_j = 0$ otherwise.

INTRODUCTION

The optimization of the logistic processes is one of the critical tasks in managing supply chains across various sectors, including medicine. Effective management in medical logistics plays a pivotal role in ensuring healthcare and public welfare. It becomes particularly relevant during global crises when the quick and efficient distribution of medical supplies is crucial. Furthermore, the proper organization of logistics is vital for humanitarian aid when the timely delivery of medical resources can be life-saving. To enhance medical logistic processes, advanced technologies and algorithms are utilized. One approach gaining wide recognition is the use of genetic algorithms to solve two-stage location problems. Solving this problem allows running optimization of the distribution and location of medical facilities, regional centers, and resources. The healthcare industry can benefit greatly from the use of an application that aids providers and suppliers in making informed decisions. By doing so, the industry can experience higher levels of productivity, faster operations, and economic advantages in medical logistics.

The object of study is the process that involves locating centers of medical facilities in two stages while limiting their number.

The subject of study is the methods of modeling and solving two-stage location problems in the context of medical logistics.

The goal of study is to create a model and develop an algorithm to solve the two-stage location problem related to medical logistics issues. The next step is to analyze their application to model and practical tasks.

1 PROBLEM STATEMENT

In extremely critical situations in a region, there may arise the need for quick distribution of essential medical supplies (medicines and medical equipment) among the population. Each region has *I* regional centers (RCs) serving as primary distribution points. To facilitate the distribution process, the local government identifies a set of *J* potential subregional centers (SRCs) that can be used as intermediate nodes for transporting medicines and medical equipment. To optimize costs due to logistical difficulties and limited resources, only *N* out of *I* regional centers and *L* out of *J* potential subregional centers may be activated by the government. Activated subregional centers will receive medical goods from regional centers and then redistribute them to *K* medical warehouses distributed throughout the region. The task is to determine the most efficient combination of subregional centers and the optimal transportation plan for drugs and medical devices from subregional centers to facilities.

Let us move on to constructing the mathematical model. The objective of the problem is to minimize the total transportation costs of delivering drugs from regional centers to medical warehouses and to select optimal locations for regional and subregional centers:

$$\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}^{1} x_{ij} + \sum_{j=1}^{J}\sum_{k=1}^{K} c_{jk}^{2} y_{jk} + \sum_{i=1}^{I} P_i v_i + \sum_{j=1}^{J} M_j z_j, \quad (1)$$

under the following constraints:

$$\sum_{j=1}^{J} y_{jk} \le s_j z_j, \ \forall k = \overline{1,N}, \quad (2)$$

$$\sum_{j=1}^{J} x_{ij} \le r_i, \ \forall i = \overline{1,I}, \quad (3)$$

$$\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij} = \sum_{j=1}^{J}\sum_{k=1}^{K} y_{jk}, \quad (4)$$

$$\sum_{i=1}^{I} v_i \le N, \quad (5)$$

$$\sum_{j=1}^{J} z_j \le L, \quad (6)$$

$$\sum_{j=1}^{J} y_{jk} \ge d_k, \ \forall k = \overline{1,N}. \quad (7)$$

This model includes several constraints to ensure efficient medical logistics. Firstly, constraint (2) limits the capacity of subregional centers, meaning that the number of drugs transported from each subregional center should not exceed its capacity. Secondly, constraint (3) ensures that the number of drugs transported from each regional center does not exceed its stock. Constraint (4) is the balance equation between the first and second stages of medical logistics. Additionally, constraints (5) and (6) set a limit on the maximum number of regional and subregional centers. Lastly, condition (7) ensures that the number of medical supplies received by each medical warehouse does not exceed its capacity.

2 REVIEW OF THE LITERATURE

The facility allocation problem is a widespread problem in operations research and logistics. Its main goal is determining the optimal location of facilities (such as warehouses, factories and distribution centers) to serve a given set of demand points.

DOI 10.15588/1607-3274-2023-3-16

A genetic algorithm is utilized to study a two-stage transportation problem involving fixed route fees and goods transportation [1]. It is worth noting the generation of the first population considering different algorithms for building plans.

In [2], an iterative algorithm to solve a two-stage transportation problem is proposed, including a procedure for shuffling customers and efficiently removing duplicates. It is shown that the algorithm works efficiently with different input data.

Paper [3] is devoted to a multi-stage reverse logistics network using a genetic algorithm with priority encoding. Computational experiments confirm the effectiveness of the algorithm.

The task of improving the spatial planning of public health services through the development of location-allocation and accessibility models is considered in [4]. The work aims to determine the optimal location of hospitals and other medical facilities, considering factors such as population demand, accessibility, and distance to other healthcare facilities. The study is based on the example of the organization of health care in Lisbon (Portugal), where the application of the proposed methods has improved the quality and efficiency of its provision.

The article [5] deals with a two-stage transportation problem that aims to minimize logistics costs, considering the cost of placing distribution centers and transportation costs between facilities.

One potential drawback of discrete location problems is their limited scalability and, as a result, finding a timely and efficient solution for larger problems can be challenging. This is especially true for problems of significant dimensionality. By applying exact methods, the problem's dimensionality is limited even more (rather small-sized problems are usually being solved). Therefore, it is crucial to take into account continuum problems. For example, [6] proves the active requirements for more research on the case of continuous multi-stage problems and gives examples from the subject area. The authors develop a class of problems in which there are several groups of objects to be placed. Each group has its own set of possible locations (in some cases, these sets may overlap), and the relationships between the objects are regulated. The authors propose a model of a two-stage continuum location-allocation problem. Continuity is considered for each stage, sometimes combined with discrete cases. At the same time, research [7] discusses how to create new mathematical models for two-stage production allocation processes and develop methods for solving them. It emphasizes the importance of ensuring resource allocation continuity and managing multiple production stages. The article also highlights the significance of experimental studies and results analysis of model problem-solving. Additionally, the article provides examples of the practical use of these models in optimizing the two-stage distribution of material flow in fuel and energy enterprises.

In the paper [8], it is proposed to solve a two-stage transportation problem using tabu search and encoding procedures. The mathematical model corresponds to a discrete formulation of a two-stage transportation problem with a limit on the enterprises' number. The reasonable planning and optimization of shelter location is considered in [9]. The main goal of the study is to reduce losses from natural disasters and improve sustainable urban development. The authors propose a sequential approach to address a two-criteria problem by making decisions that aim to optimize economic sustainability and social utility step-by-step.

A two-stage transportation problem with a fixed fee for route usage is investigated in [10]. It is proposed to use a transition to a different form of the problem, similar to the two-stage transportation problem with the transportation unit cost. It can be represented as a conditional expression depending on whether the route is used. A genetic algorithm is utilized to solve the problem. The chromosome is encoded in the form of a matrix representation.

The problem of optimal warehouse location is solved in [11]. The authors propose a new mixed integer linear programming model to solve the warehouse location problem using Euclidean distance linearization.

Paper [12] proposes an original approach to the manufacturing enterprises' location using fuzzy logic and inference systems. This study provides decision-makers in the production industry with valuable insights on utilizing fuzzy logic and inference systems to effectively address complex issues related to the location of manufacturing facilities.

The problem of allocating production capacity with differentiated convex production costs, as described in reference [13], is a variation of the traditional allocation problem. The cost of production at each plant is modeled as a convex function of its production capacity. It is proposed a fast, accurate method based on the branch-and-price approach that takes advantage of the structure of the problem and the convexity of production costs.

3 MATERIALS AND METHODS

To solve the problem (1)–(7), an algorithm based on a genetic approach with priority encoding [5] is developed. The general scheme of it is as follows.

Step 1. The population $P(t)$ is initialized using priority-based encoding. In the current step, the first generation of possible solutions to the problem is obtained.

Step 2. The fitness of each chromosome in the population is calculated.

Step 3. Chromosomes are selected for reproduction using the roulette wheel selection method.

Step 4. The selected chromosomes are crossed to produce offspring.

Step 5. Some chromosomes are affected by mutation procedure to introduce variations.

Step 6. A new population $P(t+1)$ is formed from the offspring and some individuals from the current population. The new generation replaces the old one.

Step 7. The termination condition is checked. If satisfied, proceed to the next step. Otherwise, return to Step 2 with the population value $t+1$.

DOI 10.15588/1607-3274-2023-3-16

Step 8. Terminate the genetic algorithm. Decode the most effective chromosome, return the value of the objective function, and the transportation plan.

Let us move on to detail each of the procedures used in the genetic algorithm. A priority-based coding scheme can represent each transportation plan as a chromosome. In this case, each chromosome in the population consists of two parts: the first describes the transportation tree between regional and subregional centers, and the second describes the transportation tree between subregional centers and medical warehouses. Fig. 1 visualizes the transportation plan and the chromosome that corresponds to it and can be obtained using priority encoding.

The Prufer sequence is used to represent chromosomes in a population. Each chromosome is an array of integers, where each number represents a specific vertex in the graph. This encoding represents each tree in the population and has advantages, including compact representation and speed of modification during crossover and mutation processes.

Algorithm for encoding the transportation plan (using the example of the first stage).

Step 1. Priority allocation between nodes starts with the largest value for the total count of RCs (|*I*|) and SRCs (|*J*|).

Step 2. Select the node with the lowest transportation cost among the tree leaves (nodes with transportation volumes to/from equal to their stocks/capacities). The selected node is given the highest priority.

Step 3. Remove the corresponding arc from the transportation tree. It means that the selected node is no longer a leaf node.

Step 4. If all nodes have a priority, the algorithm terminates. Otherwise, select a new leaf node and repeat steps 2 and 3.

To evaluate each individual's fitness in the population, the Prufer sequence needs to be converted into a transportation plan using the decoding process. This process involves recovering a contiguity list or matrix, which can then be used to calculate different parameters for the evaluation function.

Algorithm for decoding the chromosome plan (using the example of the first stage).

Step 1. Select the SRC with the highest priority.

Step 2. Merge the selected SRC with the RC, considering the minimum transportation cost.

Step 3. Determine the transportation amount,

which is the minimum of the stock in the RC (a_i) and the capacity of the SRC (b_j). This can be expressed as $g_{ij} = \min\{a_i, b_j\}$.

Step 4. Update the values of the stock in the RC and the available capacity of the SRC using the following formulas: $a_i = a_i - g_{ij}$ and $b_j = b_j - g_{ij}$.

Step 5. If the available capacity on the SRC (b_j) equals zero, set the priority of the SRC to zero.

Step 6. Repeat steps 1–5 until all capacities of the SRC are filled.

Note that the encoding and decoding algorithms for the second stage are the same, given that the SRCs are replaced by medical warehouses and the RCs are used instead of SRCs.

Let us define the initialization algorithm.

Step 0. Assume that we are generating a population of size $U = |P|$.

Step 1. Generate random priorities for each part of the chromosome (for example, from 1 to (|*I*| + |*J*|) for the first part of the chromosome).

Step 2. Use the evaluation procedure to check whether the generated chromosome satisfies the problem's constraints. If so, insert chromosome to the population *P*. Otherwise, try to regenerate it with a second attempt. If 1000 unsuccessful attempts to generate a chromosome happen, the input data of the problem is incorrect. The genetic algorithm procedure is terminated with an error code.

Step 3. Continue to generate chromosomes until the population reaches size *U*.

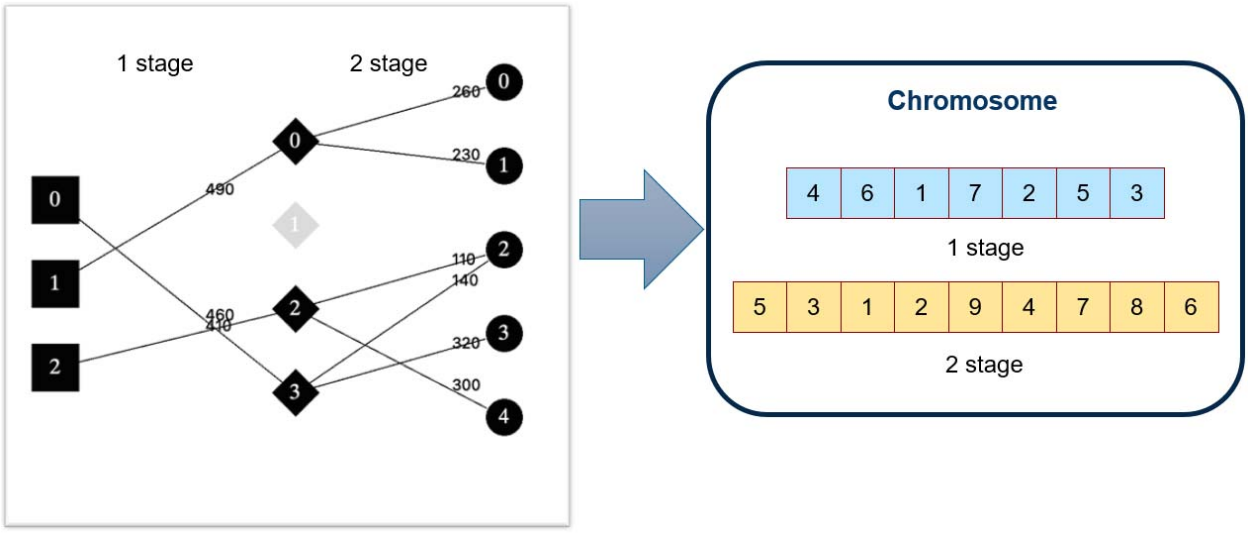


Figure 1 – Example of transportation plan encoding

DOI 10.15588/1607-3274-2023-3-16

The initial population is generated randomly. This diversity is important for a genetic algorithm because it allows a wide search in the solution space. The quality of the initial population can significantly affect the convergence rate and the quality of the solution.

The second step of initialization uses the evaluation function. The algorithm is the following.

Step 1. Apply the decoding procedure and determine the transportation plan.

Step 2. If the number of regional centers exceeds *N*, discard the chromosome, otherwise proceed to the next. If the transportation volume to a subregional center is zero, do not place it. Go to the next step.

Step 4. If the total number of placed subregional centers is greater than *L* from (6), then discard the chromosome, otherwise proceed to the next step.

Step 5. If constraints (2), (3), (4), (7) are not met with the specified transportation plan (transportation volume constraints), then discard the chromosome, otherwise, go to Step 6.

Step 6. Calculate the value of the objective function using (7).

The evaluation function (or fitness function) is a solvable measure. It is always specific to the problem and calculates a fitness score for each individual in the population, with the value of the objective function calculated according to (7). Higher fitness scores correspond to better solutions.

After the initialization procedure, it is necessary to introduce a selection procedure for each iteration. For this purpose, we use the roulette wheel selection.

Step 1. Calculate the value of *S* – the sum of all values of the chromosome objective functions for the population *P(t)*.

Step 2. Create an array of probabilities, where each element p_i is defined as:

$$p_i = \frac{fitness(chromosome_i)}{S}, \forall i \in |P|,$$

where $fitness(i)$ – the value of the evaluation function for some chromosome *i*.

Step 3. Randomly select the index corresponding to the chromosome from the probability array.

The roulette method is probabilistic – each individual's chance of being selected is in proportion to its fitness value. We revert the probabilities as the problem is about minimization. Thus, individuals with a lower objective function value are more likely to be selected for the next generation. This process mimics the survival of the fittest mechanism observed in natural evolution.

The weight mapping crossover is used. Here is its algorithm.

Input data: two parents v_1, v_2; chromosome length *n*.

Step 1. Randomly select a chromosome cut point $p = rand(1,n)$, where $rand(1,n)$ generates a random number from 1 to *n*.

Step 2. Determine the length of the chromosome segment after the cut point as $l = n - p$. Generate new chromosomes v_1' and v_2' by exchanging segments of parental chromosomes after the cut point:

$$v_1' = [v_1[1:p], v_2[p+1:n]],\ s_1[\cdot] = sort(v_1[p+1:n]),$$

$$v_2' = [v_2[1:p], v_1[p+1:n]],\ s_2[\cdot] = sort(v_2[p+1:n]),$$

Step 3. For each *i* from 1 to *l* and for each *j* from 1 to *l*, conditionally apply:

if $v_1'[p+i] = s_2[j]$, then $v_1'[p+i] = s_1[j]$ and

if $v_2'[p+i] = s_{1[j]}$, the $v_2'[p+i] = s_2[j]$.

This algorithm uses a random cut point to swap segments of the parent chromosomes and create new chromosomes. It then adapts the new chromosomes by replacing each value in the segment after the cut point with the corresponding value from the sorted segment of the original chromosome.

To solve the problem, we propose the use of mixed mutation: with a probability of 0.5, either a substitution or an insertion mutation is used. For the first type of mutation, two unique elements are selected from each part of the chromosomes and exchanged. For the insertion mutation, a random index is selected for each part of the chromosome, from which the element is moved to another random position.

4 EXPERIMENTS

We developed a software implementation for a two-stage location problem solution algorithm to run numerical experiments using the Python programming language and Qt5 as a user interface library. To evaluate the algorithm's performance in solving the two-stage location problem, we performed 456 tests using AMD Ryzen 7 5800X 3.8-4 GHz processor and 32 gigabytes of DD4-3200 RAM.

The input data is generated pseudorandomly within the range of values obtained by solving a practical problem. The number of regional centers ranged from 3 to 8; the number of subregional centers ranged from 5 to 13. The number of warehouses ranged from 10 to 500. Additional experiments are conducted for many warehouses – from 600 to 2000.

We take time benchmarks on the following parameters:

– the total execution of the genetic algorithm;
– one iteration execution;
– chromosomes generation;
– selection procedure;
– crossover procedure;
– mutation procedure;
– fitness procedure.

DOI 10.15588/1607-3274-2023-3-16

Table 1 – Expenses calculations for DAF XF (XF105) and Geely MC

Parameter	Calculating method	DAF XF (XF105)		Geely MC	
		Reference values	Result (UAH/km)	Reference values	Result (UAH/km)
Fuel cost (c_1)	Fuel consumption per km multiplied by fuel cost per liter	Fuel consumption – 0.28 l/km, fuel cost – 43.27 UAH/l	12.12	Fuel consumption – 0.07 l/km, fuel cost – 44.49 UAH/l	3.11
Driver's wage (c_2)	(Monthly salary / Average number of hours per month) / Average speed	Monthly salary – 30.000 UAH; 160 working hours. Average speed – 60 km/h	0.03	Monthly salary – 18.000 UAH 160 working hours. Average speed – 60 km/h	0.02
Depreciation (c_3)	(Vehicle cost / Average service life in km)	1.500.000 UAH / 1.000.000 km)	1.5	300.000 UAH / 200.000 km)	1.5
Maintenance expenses (c_4)	Average cost per km	0.20/km	0.20	0.10/km	0.1
Insurance cost (c_5)	cost per km	5000 UAH per year Expected mileage – 124.800 km/year	0.04	2500 UAH per year; Expected mileage – 124.800 km/year	0.02
		Total expenses per one km (c^1) = 13.89		Total expenses per one km (c^2) = 4.75	

Additionally, we introduce the value of estimate V – the very last iteration number in the genetic algorithm procedure when the improvement (decrease in the value of the objective function) occurs.

To determine the size of the problem, we use the size of the chromosome: $h_{size} = I + J + K$.

Let's consider the algorithm's application to the real medical logistic problem. The proposed mathematical model and algorithm modification can be used to solve the medical logistics problem in the Dnipropetrovsk region (Ukraine).

Before addressing the problem, we must determine the input parameters based on the model (1) – (7). To make this problem more factual, we consider a logistics scenario where the distances between different centers are crucial. To tackle this problem, we utilize the OpenStreetMap (OSM) service, which provides a comprehensive and publicly available geographic information database, including the road network. By leveraging OSM, we can obtain accurate data on the distance between medical facilities, allowing us to incorporate real-time and distance metrics into our proposed logistics optimization algorithms. Using real-world data from OSM ensures that our analysis and decision-making process is grounded on practical and reliable information.

Next, we determine the factors that affect the cost of transporting a unit. To accomplish this, we provide a breakdown of transportation costs for each kilometer of distance.

c_1 – fuel cost, which depends on the amount of fuel consumed per kilometer, the cost of fuel per liter, and the total distance;

c_2 – driver's wage, which can be calculated as the product of the driver's hourly rate and the time spent on transportation;

c_3 – vehicle depreciation, calculated as the product of the vehicle cost divided by the expected life of the vehicle in kilometers and the total distance;

c_4 – maintenance and repair costs, calculated as the product of the average maintenance and repair cost per kilometer divided by the total distance;

c_5 – insurance costs, expressed as a known coefficient.

The total cost of transportation (c) can be defined as:

$$c = c_1 + c_2 + c_3 + c_4 + c_5 .$$

When locating a regional or subregional center, it is important to factor in operating costs. These costs can be defined as follows:

$$M = M_1 + M_2 + M_3 + M_4 + M_5 + M_6 + M_7 ,$$

where M_1 – expenses on renting unloading equipment; M_2 – salaries of all employees, including taxes; M_3 – expenses on utility payments; M_4 – expenses on maintenance and repair of the building; M_5 – expenses on purchasing and maintaining the necessary equipment for the storage of medicines and medical equipment; M_6 – insurance costs for the building; M_7 – IT infrastructure costs.

The volumetric weight will be used to represent the transportation volume of medicines and medical equipment.

Next, we want to calculate the expense values considering the parameters specified in the problem statement for the subject area. The resulting calculations are available in Table 1. DAF XF (XF105) trucks with a load capacity of up to 20 tons are used for transportation from regional to subregional centers. Geely MC cars are used for transportation from subregional centers to medical warehouses.

DOI 10.15588/1607-3274-2023-3-16

Moving further, let us determine the cost of activating regional and subregional centers. According to the model, the cost of activating a center depends on the possible stock of drugs and capacity. The results of calculating the activation cost of regional and subregional centers are listed in Tables 2, 3. The minimum and maximum possible values are given for each parameter value.

Table 2 – Regional centers activation expenses

Expense type	Regional center (UAH/month)	
	Drugs storage: 180 – 225 m3 volumetric weight	
	min	max
Renting unloading equipment (M_1)	30000	50000
Salaries and wages and taxes (M_2)	420000	660000
Utility payments (M_3)	60000	80000
Maintenance and repairments (M_4)	15000	25000
Maintaining the necessary equipment (M_5)	100000	200000
Insurance (M_6)	10000	20000
IT infrastructure (M_7)	8350	8350
Total (M)	643350	1043350

Table 3 – Subregional centers activation expenses

Expense type	Subregional center (UAH/month)	
	Capacity: 110 – 150 m^3 volumetric weight	
	min	min
Renting unloading equipment (M_1)	15000	15000
Salaries and wages and taxes (M_2)	210000	210000
Utility payments (M_3)	20000	20000
Maintenance and repairments (M_4)	7500	7500
Maintaining the necessary equipment (M_5)	50000	50000
Insurance (M_6)	5000	5000
IT infrastructure (M_7)	4200	4200
Total (M)	311700	708700

To determine the problem's dimensionality, we use the Ministry of Health statistical data for 2021 presented in [14]. According to form N-47 "Report on the network and activities of medical institutions", we define the size of the problem as 4x7x65.

According to the mathematical model, we have I (number of regional centers) is equals 4; J (number of subregional centers) is equals 7; K (number of medical warehouses) is equals 65; N (the limit for regional centers) is equals 3; L (the limit for subregional centers) is equals 6.

The algorithm will be applied using the input data in Table 4.

The locations of regional, subregional centers and medical warehouses are shown in Fig. 2: red markers correspond to regional centers, blue markers indicate subregional centers and green dots indicate medical warehouses.

To solve this problem, we use the previously described genetic algorithm procedure with the following parameters:

– maximum population size – 50;

– maximum generations – 100;

– the probability of applying the mixed mutation procedure – 0.15.

Table 4 – Input data for the problem of medical logistics in the Dnipropetrovsk region

Input data	Values
Medicines and medical equipment stocks vector for regional centers	[213, 212, 215, 212]
Capacity vector for subregional center	[102, 102, 100, 101, 101, 102, 101]
Total demand for medical warehouses	586
Regional centers' activation cost vector	[846475, 906049, 937198, 669541]
Subregional centers activation cost vector	[322878, 517037, 425303, 422976, 528656, 610396, 384027]

The result of the solution is shown in Fig. 3. The first stage routes are marked in green, the second stage routes – in blue. The centers that were not activated are highlighted in gray. The following results are obtained:

– The objective function value for the most effective solution to the problem is 5057722,23 UAH;

– 3 regional centers were activated, except for the center in Pavlohrad;

– 6 subregional centers were activated, except for the center in Tomakivka;

– total execution time of the algorithm (with visualization) equals 47 seconds.

Fig. 4 (first stage only) and Fig. 5 (second stage only) show detailed maps for each stage of the medical logistics problem. It is shown that disabled centers have no inbound and outbound routes.

DOI 10.15588/1607-3274-2023-3-16

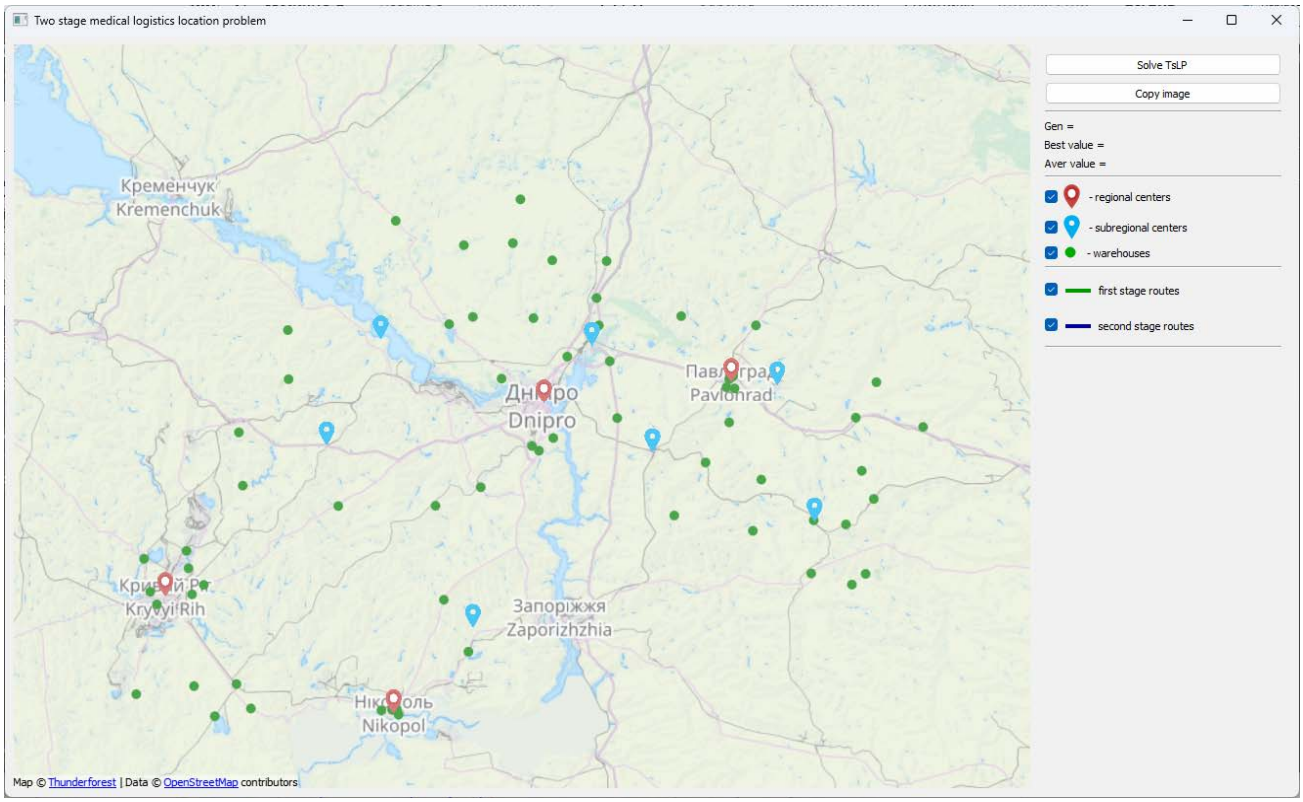


Figure 2 – Locations of the centers and medical warehouses

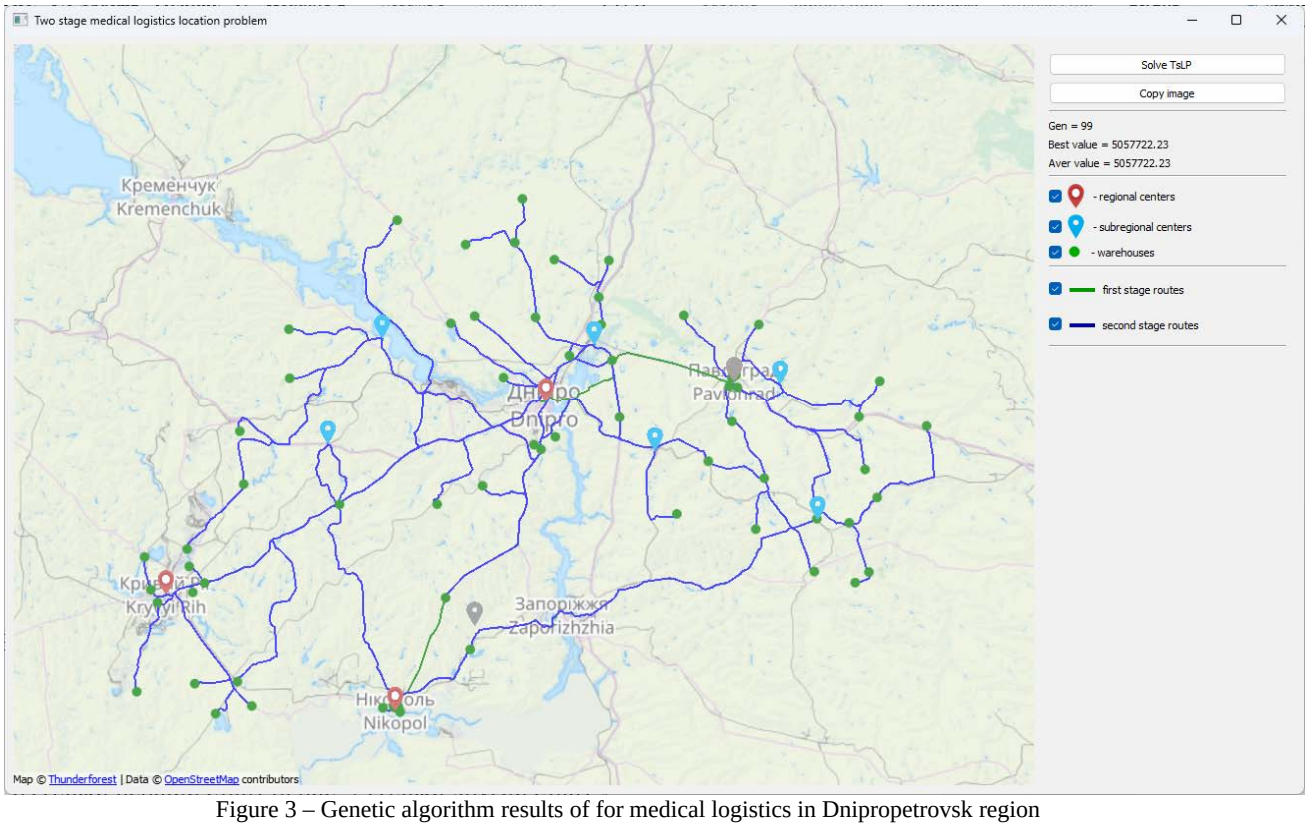


Figure 3 – Genetic algorithm results of for medical logistics in Dnipropetrovsk region

DOI 10.15588/1607-3274-2023-3-16

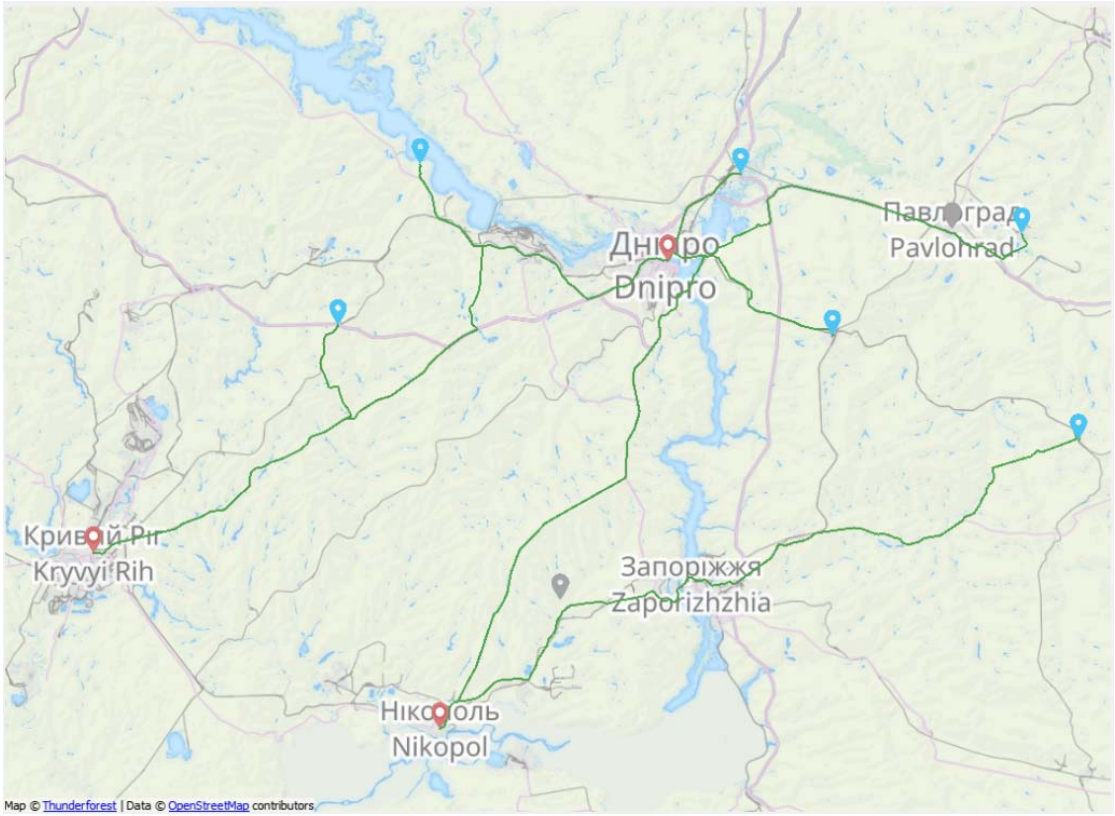


Figure 4 – Visualization of the first stage of the solution

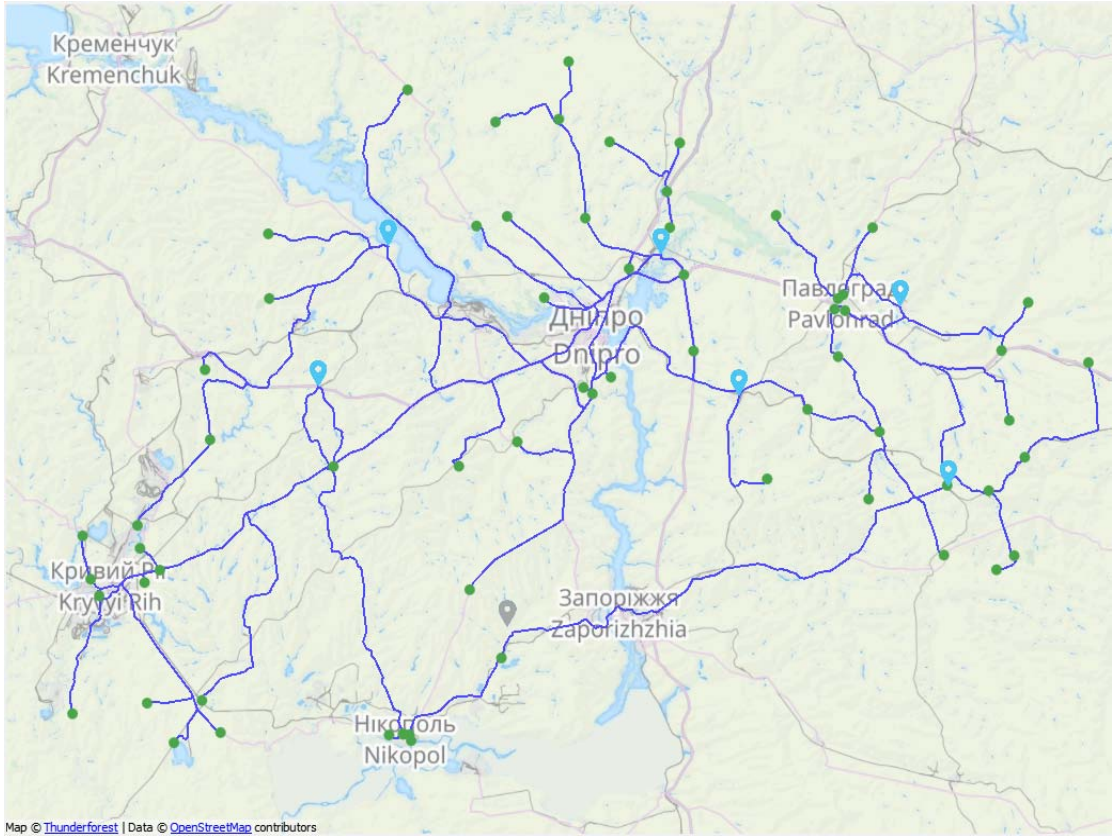


Figure 5 – Visualization of the second stage of the solution

DOI 10.15588/1607-3274-2023-3-16

5 RESULTS

Fig. 6 illustrates a graph of the dependence of the total execution time of the algorithm on the size of the problem. The results are compared for different limits for generations.

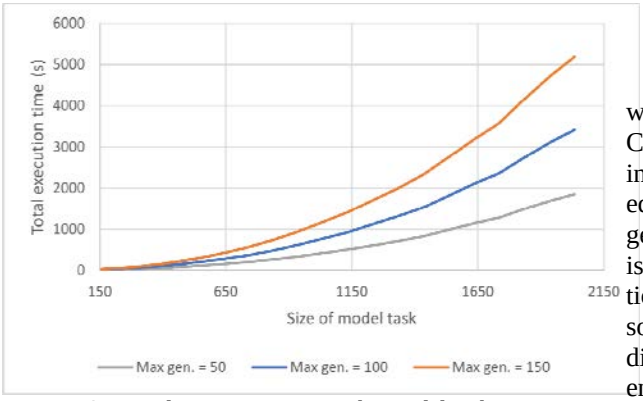


Figure 6 – Total execution time to the model task size

Fig. 7 shows a diagram containing the average time spent on each of the 456 tests. The value of the total time is taken as 100%. The benchmarks also contain the percentage of each procedure time execution of the algorithm.

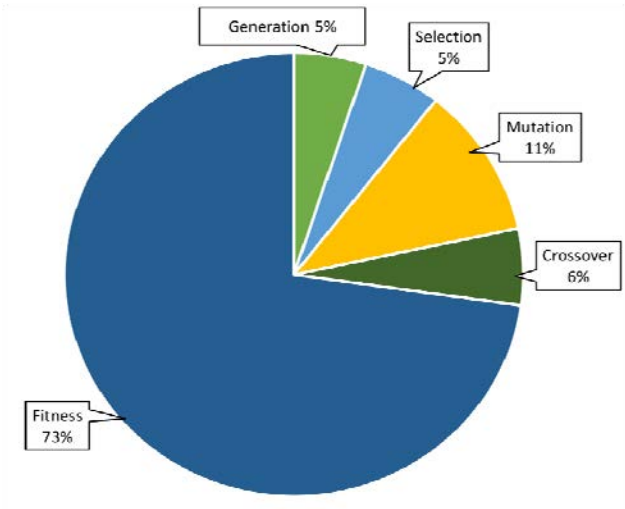


Figure 7 – Time execution of the genetic algorithm split by each procedure

6 DISCUSSION

The results from Fig. 6 demonstrate that, on average, the total execution time of the algorithm increases by 45.69% when the maximum number of generations is increased from 50 to 100. When increasing it from 100 to 150, the algorithm execution time increases by 51.68%. If the maximum generation limit is set at 150, it is improbable for value V to reach the highest number of iterations possible. Using higher values for the maximum number of generations will unlikely result in improvements. The algorithm's running time for problems of size 25 to 1000 is, at most, 17 minutes. For problems of size 1001 to 2035, the solution process takes much longer than those of smaller dimensions.

From the diagram in Fig. 7, we can conclude that most of the time is spent on the evaluation procedure, which includes decoding the transportation plan and checking all constraints.

CONCLUSIONS

The paper deals with a two-stage location problem with constraints on the maximum number of centers. Considering the specific requirements of medical logistics in the transportation context of medicines and medical equipment, a mathematical model and modification of the genetic algorithm are proposed. The developed algorithm is tested on model tasks and can produce effective solutions for problems ranging in size from 25 to 1000. The solution process takes more time for larger problems with dimensions from 1001 to 2035. Additionally, the influence of increasing the maximum generations number on the time of execution is investigated. When the maximum generation value increases from 50 to 100 and from 100 to 150 generations, the algorithm's execution time increases by 45.69% and 51.68%, respectively. 73% of the total execution time is spent on the evaluation procedure. The algorithm is applied to the medical logistics problem in the Dnipropetrovsk region (Ukraine). An efficient solution is obtained within an acceptable execution time.

The results of computational experiments demonstrate the proposed algorithm's correctness, high efficiency and wide application possibilities.

The scientific novelty of obtained results is that the mathematical model for a two-stage location problem with constraints on the maximum number of centers is proposed. An algorithm for solving this problem based on a genetic approach has been developed.

The practical significance of obtained results is that, based on research results, we recommend utilizing the developed algorithm and software implementation for medical logistics and other location problems in practical environments.

Prospects for further research are to study problems of higher dimensions and to extend the proposed approach to continuum two-stage location problems.

ACKNOWLEDGEMENTS

This publication was prepared as a part of the scientific theme 0123U100011 "Problems of analysis, modeling, and optimization of technological processes in complex systems of different nature", which is implemented on the System Analysis and Control Department at Dnipro University of Technology.

REFERENCES

1. Antony Arokia Durai Raj K., Rajendran Chandrasekharan A genetic algorithm for solving the fixed-charge transportation model: Two-stage problem, *Computers & Operations Research*, 2012, Vol. 39, No. 9, pp. 2016–2032. DOI: 10.1016/j.cor.2011.09.020

DOI 10.15588/1607-3274-2023-3-16

2. Cosma O., Pop P., Sabo C. An efficient solution approach for solving the two-stage supply chain problem with fixed costs associated to the routes, *Procedia Computer Science*, 2019, Vol. 162, pp. 900–907. DOI: 10.1016/j.procs.2019.12.066
3. Lee J.-E., Gen M., Rhee K.-G. Network model and optimization of reverse logistics by hybrid genetic algorithm, *Computers & Industrial Engineering*, 2009, Vol. 56, No. 3, pp. 951–964. DOI: 10.1016/j.cie.2008.09.021
4. Polo Gina [et al.] Location-Allocation and accessibility models for improving the spatial planning of public health services, *Plos One*, 2015, Vol. 10, No. 3, pp. 1–14. DOI: 10.1371/journal.pone.0119190
5. Gen M., Altiparmak F., Lin L. A genetic algorithm for two-stage transportation problem using priority-based encoding, *Or Spectrum*, 2006, Vol. 28, No. 3, pp. 337–354. DOI: 10.1007/s00291-005-0029-9
6. Us S. A., Koriashkina L. S., Stanina O. D. An optimal two-stage allocation of material flows in a transport-logistic system with continuously distributed resource, *Radio Electronics, Computer Science, Control*, 2019, No. 1, pp. 256–271. DOI: 10.15588/1607-3274-2019-1-24
7. Stanina O., Us S., Koriashkina L. Modeli ta metody rozviazannia zadach optymalnoho rozmishchennia dvoetapnoho vyrobnytstva z neperervno rozpodilenym resursom. Dnipro, Harant, 2021, 200 p.
8. Salman M. Using tabu search algorithm for solving two-stage transportation problem based on decoding procedure, *Sci.Int.*, 2018, Vol. 30, No. 4, pp. 643–651.
9. He L., Xie Z. Optimization of urban shelter locations using bi-level multi-objective location-allocation model, *International Journal of Environmental Research and Public Health*, 2022, Vol. 19, No. 7, pp. 4401. DOI: 10.3390/ijerph19074401
10. Calvete H. I. [et al.] // A matheuristic for the two-stage fixed-charge transportation problem, *Computers & Operations Research*, 2018, Vol. 95, pp. 113–122. DOI: 10.1016/j.cor.2018.03.007
11. You M. [et al.] Optimal mathematical programming for the warehouse location problem with Euclidean distance linearization, *Computers & Industrial Engineering*, 2019, Vol. 136, pp. 70–79. DOI: 10.1016/j.cie.2019.07.020
12. Paul S. K. [et al.] An advanced decision-making model for evaluating manufacturing plant locations using fuzzy inference system, *Expert Systems With Applications*, 2022, Vol. 191, pp. 116378. DOI: 10.1016/j.eswa.2021.116378
13. Christensen T. R. L., Klose A. A fast exact method for the capacitated facility location problem with differentiable convex production costs, *European Journal of Operational Research*, 2020. DOI: 10.1016/j.ejor.2020.11.048
14. Tsentr medstatystyky – Statystychni dani systemy MOZ [Elektronnyi resurs], *Tsentr medstatystyky*. Rezhym dostupu: http://medstat.gov.ua/ukr/statdanMMXIX.html (data zvernennia: 01.07.2023). – Nazva z ekrana.

Received 14.06.2023.
Accepted 23.08.2023.

УДК 519.85:004.42+004.02

МОДИФІКАЦІЯ ГЕНЕТИЧНОГО АЛГОРИТМУ ДЛЯ РОЗВ'ЯЗАННЯ ДВОЕТАПНОЇ ЗАДАЧІ РОЗМІЩЕННЯ

Сергєєв О. С. – аспірант кафедри системного аналізу і управління Національного ТУ «Дніпровська політехніка», Дніпро, Україна.

Ус С. А. – канд. фіз.-мат. наук, професор кафедри системного аналізу і управління Національного ТУ «Дніпровська політехніка», Дніпро, Україна.

АНОТАЦІЯ

Актуальність. Оптимізація логістичних процесів є одним із важливих завдань управління ланцюгами поставок у різних сферах діяльності, включаючи медицину. Ефективна координація у сфері медичної логістики має важливе значення для забезпечення громадського здоров'я та процвітання. Це стає особливо актуальним в умовах глобальних надзвичайних ситуацій, коли швидке та ефективне розповсюдження медикаментів має вирішальне значення. Крім того, вміле управління логістикою має першорядне значення для доставки гуманітарної допомоги, у випадку, коли своєчасне транспортування медичних препаратів і ресурсів може стати рятівним для життя. Задля вдосконалення процедур у медичній логістиці використовуються найсучасніші технології та алгоритми. У роботі розглянута модифікація генетичного алгоритму для розв'язання двоетапних задач розміщення у контексті управління ланцюгами поставок при дистрибуції ліків та виробів медичного призначення.

Мета. Метою роботи є побудова моделі та розробка алгоритму для розв'язання двоетапної задачі розміщення у контексті проблеми медичної логістики з подальшим аналізом їх застосування на модельних та прикладних задачах.

Метод. Запропоновано використання генетичного алгоритму для розв'язання двоетапної задачі логістики. Особливостями цього алгоритму є модифікація процедур оцінювання та використання змішаної мутації, що дозволяють ефективно розв'язувати проблему, враховуючи особливості предметної постановки – розміщення центрів на декількох етапах логістичного процесу.

Результати. Авторами запропоновано математичну модель і алгоритм, що враховують необхідність розміщення центрів на різних етапах з використанням модифікованих процедур мутації та оцінювання. Алгоритм протестовано на модельних задачах, досліджено вплив розмірності задачі на час його виконання. Розглянуто модельні задачі із розмірністю від 25 до 2035. Для задач розміром від 1001 до 2035 процес розв'язання займає значно більший час порівняно із задачами менших розмірностей. При збільшенні можливого обсягу популяції від 50 до 100 та від 100 до 150 поколінь, час виконання алгоритму збільшується на 45,69% та 51,68% відповідно. Найбільший час витрачається на процедуру оцінювання і становить 73% часу від загального часу розв'язання

Висновки. Введено математичну модель для двоетапної задачі розміщення у контексті медичної логістики з врахуванням особливостей галузі. Розроблено алгоритм розв'язання, що базується на генетичному підході та застосований для вирішення проблеми медичної логістики. Алгоритм досліджено на модельних задачах різних розмірів та проаналізовано

DOI 10.15588/1607-3274-2023-3-16

залежність часу роботи алгоритму від розміру задачі. Додатково розглянуто залежності часу роботи алгоритму від значення обсягу популяцій. Наведено аналіз частки участі кожного з етапів генетичного алгоритму у загальній роботі алгоритму. Отримані результати свідчать про високу ефективність та широкі можливості застосування запропонованої математичної моделі та алгоритму. Розроблений метод демонструє високу продуктивність та надійність, особливо для задач середньої та великої розмірності, підтверджуючи свою перспективність та універсальність.

КЛЮЧОВІ СЛОВА: двоетапна задача розміщення, генетичні алгоритми, кодування на основі пріоритетів, медична логістика.

ЛІТЕРАТУРА

1. Antony Arokia Durai Raj K. A genetic algorithm for solving the fixed-charge transportation model: Two-stage problem / K. Antony Arokia Durai Raj, Chandrasekharan Rajendran // Computers & Operations Research. – 2012. – Vol. 39, No. 9. – P. 2016–2032. DOI: 10.1016/j.cor.2011.09.020
2. Cosma O. An efficient solution approach for solving the two-stage supply chain problem with fixed costs associated to the routes / Ovidiu Cosma, Petrica Pop, Cosmin Sabo // Procedia Computer Science. – 2019. – Vol. 162. – P. 900–907. DOI: 10.1016/j.procs.2019.12.066
3. Lee J.-E. Network model and optimization of reverse logistics by hybrid genetic algorithm / Jeong-Eun Lee, Mitsuo Gen, Kyong-Gu Rhee // Computers & Industrial Engineering. – 2009. – Vol. 56, No. 3. – P. 951–964. DOI: 10.1016/j.cie.2008.09.021
4. Location-Allocation and accessibility models for improving the spatial planning of public health services / Gina Polo [et al.] // Plos One. – 2015. – Vol. 10, No. 3. – P. 1–14. DOI: 10.1371/journal.pone.0119190
5. Gen M. A genetic algorithm for two-stage transportation problem using priority-based encoding / Mitsuo Gen, Fulya Altiparmak, Lin Lin // Or Spectrum. – 2006. – Vol. 28, No. 3. – P. 337–354. DOI: 10.1007/s00291-005-0029-9
6. Us S. A. An optimal two-stage allocation of material flows in a transport-logistic system with continuously distributed resource / S. A. Us, L. S. Koriashkina, O. D. Stanina // Radio Electronics, Computer Science, Control. – 2019. – No. 1. – P. 256–271. DOI: 10.15588/1607-3274-2019-1-24
7. Станіна О. Моделі та методи розв’язання задач оптимального розміщення двоетапного виробництва з неперервно розподіленим ресурсом / О. Станіна, С. Ус, Л. Коряшкіна. – Дніпро : Гарант, 2021. – 200 с.
8. Salman M. Using tabu search algorithm for solving two-stage transportation problem based on decoding procedure / M. Salman // Sci.Int. – 2018. – Vol. 30, No. 4. – P. 643–651.
9. He L. Optimization of urban shelter locations using bi-level multi-objective location-allocation model / Lei He, Ziang Xie // International Journal of Environmental Research and Public Health. – 2022. – Vol. 19, No. 7. – P. 4401. DOI: 10.3390/ijerph19074401
10. A matheuristic for the two-stage fixed-charge transportation problem / Herminia I. Calvete [et al.] // Computers & Operations Research. – 2018. – Vol. 95. – P. 113–122. DOI: 10.1016/j.cor.2018.03.007
11. Optimal mathematical programming for the warehouse location problem with Euclidean distance linearization / Meng You [et al.] // Computers & Industrial Engineering. – 2019. – Vol. 136. – P. 70–79. DOI: 10.1016/j.cie.2019.07.020
12. An advanced decision-making model for evaluating manufacturing plant locations using fuzzy inference system / Sanjoy Kumar Paul [et al.] // Expert Systems With Applications. – 2022. – Vol. 191. – P. 116378. DOI: 10.1016/j.eswa.2021.116378
13. Christensen T. R. L. A fast exact method for the capacitated facility location problem with differentiable convex production costs / Tue Rauff Lind Christensen, Andreas Klose // European Journal of Operational Research. – 2020. DOI: 10.1016/j.ejor.2020.11.048
14. Центр медстатистики – Статистичні дані системи МОЗ [Електронний ресурс] // Центр медстатистики. – Режим доступу: http://medstat.gov.ua/ukr/statdanMMXIX.html (дата звернення: 01.07.2023). – Назва з екрана.

DOI 10.15588/1607-3274-2023-3-16

UDC 004.891

USER EVALUATION-DRIVEN RANKING CONCEPT

Zosimov V. V. – Dr. Sc., Professor of the Department of Applied Information Systems, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.

Bulgakova O. S. – PhD, Associate Professor of the Department of Applied Information Systems, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.

Perederyi V. I. – Dr. Sc., Professor of the Department of informatics and computer science, Kherson National Technical University, Kherson, Ukraine.

ABSTRACT

Context. The problem of personalizing search engine results, empowering users with search result management tools and developing new ranking models based on user's subjective information needs. The object of the study was to modeling information search results in the Internet based on user ratings.

Objective. The goal of the work is to form unique expert groups for each user, based on calculating the measure of agreement between the current user's opinions and potential experts.

Method. Introducing a novel method for ranking search results based on user ratings, which takes a subjective approach to the ranking process. This approach involves the formation of distinct expert groups tailored to individual users. Experts are selected based on the level of agreement between their opinions and the current user, determined by shared ratings on a specific set of web resources. User selection for the expert group is based on their weight relative to the current user, serving as a measure of agreement.

The proposed methodology offers a fresh approach to forming unique expert groups for each user, utilizing three different strategies depending on the presence of shared ratings on a particular set of web resources between the user and potential experts.

The developed ranking method ensures that each user receives a personalized list of web resources with a distinct order. This is accomplished by incorporating unique ratings from the expert group members associated with each user. Furthermore, each rating contributes to the ranking model of web resources with an individual weight, calculated based on an analysis of their past system activity.

Results. The developed methods have been implemented in software and investigated for complex web data operation in real time.

Conclusions. The conducted experiments have confirmed the effectiveness of the proposed software and recommend its practical use for solving complex web data operation in real time. Prospects for further research may include optimizing software implementations and conducting experimental investigations of the proposed methods on more complex practical tasks of various nature and dimensions

KEYWORDS: information search, ranking, search results, user ratings, expert groups, social profile, inductive algorithms, polynomial neural network, active neurons.

ABBREVIATIONS

ADE is a method of average differences of estimates, method of calculating the expert's weight;

CIDO is a complex Internet data operating system;

GMDH is a group method of data handling;

MIA is a multilayered iterative algorithm;

RIA is a relaxation iterative algorithm;

CIA is a combined iterative algorithm;

MICA is a multilayered iterative-combinational algorithm;

RICA is a relaxation iterative-combinational algorithm;

GIA is a generalized iterative algorithm;

DM is a dialogue mode;

IC is a iterative-combinatorial mode;

MR is a multilayered-relaxative mode;

AM is a arithmetic mean;

WAM is a weighted arithmetic mean;

HM is a harmonic mean;

WHM is a weighted harmonic mean.

NOMENCLATURE

U_0 is a current user, the user for whom the group of experts is formed and for whom the ranking of search results is carried out;

U_i is a first-level potential expert, the user who shares ratings with the current user, but whose level of agreement has not yet been calculated;

$\hat{U}_j$ is a second-level potential expert, the user who does not share ratings with the current user, but shares ratings with a first-level expert;

$\tilde{U}_k$ is a third-level potential expert, the user who does not share ratings with the current user or first-level experts;

$U_{0,\text{exp}}$ is a user's who do not have any common ratings with the current user or with the members of the expert group, as potential experts of the third level;

$X[U_0, \ldots, U_n]$ is a set of all users of the system;

$d(U_0, U_i)$ is a metric on the metric space (X, d);

W is a coefficient of concordance;

n is a number of indicators;

m is a number of experts;

r_{ij} is a rank of the i-th indicator determined by the j-th expert;

d_i is a sum of ranks of the i-th indicator by all experts;

T_i is a number of connections (types of repeated elements) in the evaluations of the i-th expert;

L_i is a number of links (types of repeating elements) in the evaluations of the i-th expert;

DOI 10.15588/1607-3274-2023-3-17

t_L is a number of elements in the L-th link for the i-th expert (the number of repeating elements);

χ^2 is a consistency criterion;

$F(x,\Theta)$ is a distribution law;

$x_{(0)}$ is a lower bound of the domain of definition of a random variable;

$x_{(k)}$ is an upper bound of the domain of definition of a random variable;

H_0 is a tested hypothesis;

$X_{r+1} = (y_1^r,...,y_F^r, x_1,\ldots,x_m)$ is the input matrix for a layer r+1 in GIA, where $x_1,\ldots,x_m$ are the initial arguments and $y_1^r,...,y_F^r$ are the intermediate ones of the layer r;

$d_k, k = 1,2,3$, $d_k = \{0,1\}$ are elements of the binary structural vector $d = (d_1\, d_2\, d_3)$ where values 1 or 0 mean inclusion or not a relevant argument;

CR is a selection criterion;

AR is a regularity criterion;

R^2 is a coefficient of determination;

K is a number of freedom degrees.

INTRODUCTION

At present, Internet advertising is the most effective way to promote a business. This has led search engines to no longer serve as information retrieval tools but rather transformed them into advertising platforms, where the pages displayed to users are not the most relevant to their informational needs but rather those that have invested more in promotion. Search engines benefit from artificially reducing the quality of organic search results, as contextual advertising appears more appropriate in comparison, even though it is often less relevant to the search query. The fact that search engines are not inclined towards high-quality organic search is evident from the introduction of the "Google SearchWiki" technology in 2009 by the world's most popular search engine, Google. It allowed users to customize the search results by sorting and removing them. Additionally, a global rating system for web resources was implemented. However, this technology was active for less than six months and was eventually discontinued due to low demand among users. Although it is evident that personalized search results, achieved through the accumulation of large amounts of statistical data and user ratings, will significantly enhance search efficiency in the long run, rendering many contemporary methods of artificial web resource promotion ineffective [1–2].

Search engine ranking algorithms take into account a large number of factors, but the main weight is given to the page rating, which is calculated based on the analysis of the number and quality of external links to the page [3]. Such assessment methods are objective, but they are easy to falsify in the presence of a certain advertising budget, due to the purchase of the necessary number of quality links from external sources [4]. It follows that they are focused on meeting the needs of advertisers, not users.

The development of methods for personalizing search engine results by providing users with search result management tools and the creation of new ranking models based on users' subjective information needs are therefore crucial tasks.

The object of study is the process of web data operation. The paper discusses an approach to personalize search results by utilizing a ranking model based on expert evaluations, which are considered authoritative for the user. The obtained ranking model is significantly more difficult to falsify as it is based on subjective factors. The ranking model will be unique for each user, making it even more challenging to falsify since it would require replicating each user's preferences, which is much more complex than acquiring links to one's website from authoritative sources.

The subject of study is the models and methods of personalizing the ranking of search results.

The ranking algorithms of search engines primarily rely on the page ranking, which is calculated based on the analysis of external links [3]. However, these methods can be easily manipulated through the acquisition of high-quality links [5], prioritizing the needs of advertisers over users.

The purpose of the work is to improve the quality of search output for current user by means of personalization methods in search engine systems, through providing users with search result management tools and developing new ranking models based on subjective user information needs.

1 PROBLEM STATEMENT

The paper considers an approach to personalization of search results through the use of a ranking model based on the assessments of experts whose opinion is authoritative for the user.

To rank data based on expert evaluations, such methods as Kemeny's median, Kendall's concordance coefficient, Bord's method, etc. are used [6–8]. The use of such methods requires the presence of a predetermined group of experts. However, in the real task of ranking search results, the input data is the evaluations of users for whom their status as an expert is not defined. It is obvious that it would not be correct to accept the opinion of all users who rated the web resource as expert. It is also obvious that the assessment and personal data specified during registration are not enough to uniquely identify the user as an objective expert in the subject area. However, these data are sufficient to determine subjective expert groups for each user based on the criterion of closeness of the user's ratings to the ratings of each of the experts.

The evaluation of service quality is highly subjective. It is also important to consider that the same service can be provided differently to different clients due to various subjective reasons of the provider. Therefore, the application of methods for assessing consensus among experts, such as the coefficient of concordance, may yield results lower than expected values.

DOI 10.15588/1607-3274-2023-3-17

The rankings of web resources in the search results for the current user's query are determined by calculating the weighted harmonic mean of ratings from the expert group. Unique expert groups are formed for each user in the background mode of the CIDO system [9] using three methods based on shared ratings for a specific set of web resources between the user and potential experts. Experts are categorized into three levels for clarity, corresponding to the method of calculating their weight. This categorization is purely logical, and all expert weights at each level hold equal significance during the ranking of search results without requiring additional coefficients.

Expert weight is a measure of agreement between the expert's opinions and the current user, calculated based on the similarity of their ratings for a certain set of web resources. The weight of expert U_i relative to the current user U_0 can be considered as the metric $d(U_0, U_i)$ in the metric space (X, d), where $X[U_0, ..., U_n]$ represents the set of all users in the system

The function d satisfies the identity, symmetry, and triangle axioms, but it cannot be defined for every pair of elements from the set X, as not all users have common ratings. Therefore, in the context of this problem, it is incorrect to use the term metric space. Instead, we will refer to the function $d(U_0, U_i)$ as an analogue of a metric for defining the weight of an expert.

The value of the expert's weight is a number in the interval $d(U_0, U_i) \in [0, \ldots, 0.9]$, therefore, not the value itself, but the result of the normalization function is used to determine the qualitative assessment:

$$W(d) = 1 - \left(1.1 \cdot \frac{d(U_0, U_i)}{10}\right).$$

The obtained value is used as an indicator of the expert's weight to further calculate the ranks of web resources.

It is necessary to find measures of consistency of the user's opinions with each of the potential experts, depending on the presence of joint evaluations for some set of web resources.

2 REVIEW OF THE LITERATURE

In [10], a learning-based ranking model is proposed to enhance recommendation systems using implicit user feedback. Adaptive learning is described in [11] to improve content-based recommendation systems. [12] introduces a hybrid ranking model for scientific articles, combining content-based and citation-based approaches. A neural network-based ranking model is presented in [13–14], which can handle incomplete data, making it versatile and user-friendly. However, understanding the principles of neural network operation can be complex. [15] explores the use of BERT (Bidirectional Encoder Representations from Transformers) for search engine ranking, demonstrating high accuracy compared to traditional methods but requiring significant computational resources. [16] proposes a novel ranking approach that employs reinforcement learning to aggregate diverse page ratings, leading to improved accuracy but with associated computational and implementation complexity.

Sentiment analysis methods are discussed in [17], covering rule-based and machine learning-based techniques and their applications in domains like social media monitoring and product review analysis. The integration of recommender systems and sentiment analysis is emphasized for more effective and personalized recommendations.

[18] presents a development aimed at enhancing search relevance within organizations by capturing employee knowledge and expertise. While this approach improves search results, its efficiency in utilizing contextual information may be limited. Considering context and adapting to evolving user requirements are crucial to avoid irrelevant or incomplete results.

Brytsov R. A. addresses this issue in his work [19], proposing a theoretical ranking model based on web resource visit statistics and document viewing time. However, it does not consider user opinions or agreement levels between users and experts. User rating-based methods commonly employed for product ranking in online stores rely solely on the number and values of ratings.

Recent studies demonstrate the potential of neural networks and reinforcement learning in improving search engine ranking accuracy, albeit at the cost of significant computational resources. Information retrieval, as defined in [20], involves searching for unstructured documentary information to satisfy individual and subjective user information needs. Accordingly, search result ranking algorithms should incorporate user-specific subjective factors.

3 MATERIALS AND METHODS

Let's consider the method of average differences of estimates. The direct calculation can be applied when the current user shares ratings with the set of potential experts X for a certain set of web resources. It allows for the calculation of the similarity of ratings separately for each pair of "current user – potential expert", denoted as $d(U_0, U_i)$. User ratings range from 1 to 10, where 10 represent the most acceptable option. The expert weight value is determined as the arithmetic mean of the absolute differences between each pair of user and potential expert ratings:

$$d(U_0, U_i) = \sum_{j=1}^{m} \frac{|U_{0j} - U_{ij}|}{m}. \quad (1)$$

A qualitative assessment of closeness is provided on the Chaddock scale. Candidates with high and very high connectivity strength are selected as experts. To determine the qualitative assessment on the Chaddock scale, the result of a normalization function $W(d)$ is used, rather than the actual value itself.

To justify the choice of this method for calculating the measure of agreement among experts, let's compare its

DOI 10.15588/1607-3274-2023-3-17

effectiveness with the results obtained from calculating Kendall's concordance coefficient for each pair of "current user – potential expert".

The concordance coefficient is a numerical value that serves as a measure of agreement among experts [21]:

$$W = \frac{S}{\frac{1}{12}m^2(n^3-n) - m\sum_{i=1}^{m} T_i}, \quad (2)$$

where

$$S = \sum_{i=1}^{m}\left(\sum_{s=1}^{d} r_{ij} - \bar{r}\right)^2, \quad (3)$$

$$T_i = \frac{1}{12}\sum_{i=1}^{L_i}(t_L^3 - t_L). \quad (4)$$

If there are no associated ranks, then T_i is zero. The significance assessment of the concordance coefficient is determined by the Pearson agreement criterion [22]. This consistency criterion is the most widely used criterion for testing the hypothesis that the studied sample $x_1, x_2, \dots x_n$, with volume n, belongs to some theoretical distribution law $F(x,\Theta)$.

The hypothesis testing procedure using χ^2-type criteria involves grouping observations. The domain of definition of the random variable is divided into k non-overlapping intervals with boundary points $x_{(0)}, x_{(1)}, x_{(k-1)}, x_k$.

According to the given partition, the number n_i of sample values falling into the i-th interval and the probabilities of falling into the interval $P_i(\theta) = F(x_{(i)},\theta) - F(x_{(i-1)},\theta)$ corresponding to the theoretical law with the distribution function $F(x,\theta)$ are calculated. In addition $n = \sum_{i=1}^{k} n_i$ and $n = \sum_{i=1}^{k} P_i(\theta) = 1$. When testing a simple hypothesis, both the form of the distribution law $F(x,\theta)$ and all its parameters (known scalar or vector parameter θ) are known. The statistics used in the conditions of goodness-of-fit type χ^2 are based on measuring deviations $\frac{n_i}{n}$ from $P_i(\theta)$. The Pearson's goodness-of-fit test statistic χ^2 is defined by the relationship:

$$X_n^2 = n\sum_{i=1}^{k}\frac{\left(\frac{n_i}{n} - P_i(\theta)\right)^2}{P_i(\theta)}. \quad (5)$$

In the case of testing a simple hypothesis within the limits as $n\to\infty$, this statistic follows a χ_r^2-square distribution with $r=k-1$ degrees of freedom if the tested hypothesis H_0 is true. The probability density function of the χ_r^2-square distribution, which is a specific case of the gamma distribution, is described by the formula:

$$g(s) = \frac{1}{2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}\right)} s^{\frac{r}{2}-1} e^{\frac{-s}{2}}. \quad (6)$$

The H_0 hypothesis is rejected at large statistics values, when the statistical value calculated from the sample X_n^{2*} is greater than the critical value $\chi^2_{r,\alpha}$, or the achieved significance level (p-value) is less than the specified significance level (given probability of error of the 1st kind) α:

$$P\left(X_n^2 > X_n^{2*}\right) = \frac{1}{2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}\right)} \int_{X_n^{2*}}^{\infty} s^{\frac{r}{2}-1} e^{\frac{-s}{2}} ds. \quad (7)$$

The calculated value χ^2 is compared with the table value for the number of degrees of freedom $K = n-1$ presented in Table 1.

Table 1 – The percentage points of distribution χ^2

α → / ↓ n	90%	70%	50%	30%
…	…	…	…	…
15	8.547	11.721	14.339	17.322
16	9.312	12.624	15.338	18.418
17	10.085	13.531	16.338	19.511
18	10.865	14.440	17.338	20.601
19	11.651	15.352	18.338	21.689
20	12.443	16.266	19.337	22.775
21	13.240	17.182	20.337	23.858
22	14.041	18.101	21.337	24.939
23	14.848	19.021	22.337	26.018
	10%	5%	1%	0,5%
…	…	…	…	…
15	22.307	24.996	30.578	32.801
16	23.542	26.296	32.000	34.267
17	24.769	27.587	33.409	35.718
18	25.989	28.869	34.805	37.156
19	27.204	**30.144**	36.191	38.582
20	28.412	31.410	37.566	39.997
21	29.615	32.671	38.932	41.401
22	30.813	33.924	40.289	42.796
23	32.007	35.172	41.638	44.181

Let's consider the generalized iterative algorithm.

The GIA approach is employed at the third level to determine the weight of users who lack any shared ratings.

The generalized iterative algorithm encompasses a collection of iterative and iterative-combinatorial algorithms, which are defined by three index sets: DM, IC, MR. Each iterative algorithm is considered a specific instance of the generalized GIA = {DM, IC, MR}. DM can assume three distinct values: 1 – standard automatic mode, 2 – planned automatic mode, 3 – interactive mode.

DOI 10.15588/1607-3274-2023-3-17

IC can be either 1 – iterative or 2 – iterative-combinatorial algorithms. MR has three potential values: 1 – classical multilayered, 2 – relaxative, 3 – combined algorithms [23]. In the case where DM is set to 1, we encounter three standard variations of iterative algorithms: MIA=GIA(1,1,1), RIA=GIA(1,1,2), CIA=GIA(1,1,3), as well as three iterative-combinatorial variants: MICA=GIA(1,2,1); RICA=GIA(1,2,2), CICA=GIA(1,2,3).

Formally, in general case, a layer of the GIA GMDH may be defined as follows, Figure 1 [23]:

1) the input matrix is $X_{r+1} = (y_1^r, ..., y_F^r, x_1, \ldots, x_m)$ for a layer r+1;

2) the operators of the kind:

$$\begin{aligned} y_l^{r+1} &= f(y_i^r, y_j^r),\ l = 1,2,\ldots, C_F^2,\quad i, j = \overline{1,F}, \\ y_l^{r+1} &= f(y_i^r, x_j),\ l = 1,2,\ldots, F_m,\ i = \overline{1,F},\ j = \overline{1,m}. \end{aligned} \quad (8)$$

may be applied on the layer r+1 to construct linear, bilinear and quadratic partial descriptions:

$$\begin{aligned} z &= f(u,v) = a_0 + a_1 u + a_2 v; \\ z &= f(u,v) = a_0 + a_1 u + a_2 v + a_3 uv; \\ z &= f(u,v) = a_0 + a_1 u + a_2 v + a_3 uv + a_4 u^2 + a_5 v^2. \end{aligned} \quad (9)$$

3) for any description, the optimal structure is searched by combinatorial optimization; e.g.:

$$f(u,v) = a_0 d_1 + a_1 d_2 u + a_2 d_3 v. \quad (10)$$

Then the best model will be described as $f(u,v,d_{opt})$, where

$$\begin{aligned} d_{opt} &= \arg\min_{l=\overline{1,q}} CR_l,\ q = 2^p - 1, \\ f_{opt}(u,v) &= f(u,v,d_{opt}). \end{aligned} \quad (11)$$

4) the algorithm stops when the condition $CR^r > CR^{r-1}$ is checked.

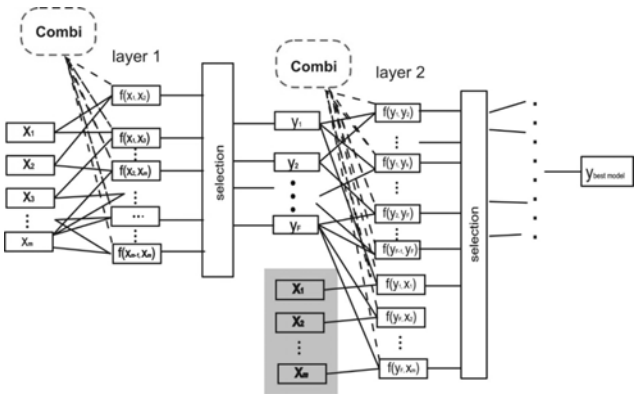


Figure 1 –The generalized architecture of GIA GMDH

4 EXPERIMENTS

To compare the results of expert agreement calculations using the described methods, a data sample (Table 2) was used, where 20 users evaluated 20 web resources based on the quality of presented information and ease of use. The evaluation scale ranged from 1 to 10, with 10 being the best rating. Expert №0 is the current user, and a group of experts is selected based on their ratings. To facilitate comprehension, the web resources will be denoted by capital Latin letters with a numerical index corresponding to the data sample number, and the users will be represented by numbers.

Table 2 – The fragment of the user evaluations data sample of web resources №1

web resource → / ↓ User	A_1	B_1	C_1	D_1	E_1	…	Q_1	R_1	S_1	T_1
0	10	10	8	5	7	…	8	8	3	5
1	10	10	9	6	8	…	8	8	4	4
2	8	8	9	5	6	…	7	7	3	4
3	9	9	9	5	7	…	7	6	3	3
4	10	10	9	7	7	…	9	9	4	3
5	2	7	1	6	7	…	7	2	7	9
6	9	10	7	6	8	…	7	6	4	4
7	7	7	3	9	10	…	5	2	9	9
8	8	8	6	4	7	…	9	7	3	4
9	7	9	8	5	6	…	9	7	4	3
10	7	9	7	4	8	…	9	8	3	4
11	9	10	8	5	7	…	7	9	4	4
12	8	8	9	6	7	…	8	8	3	4
13	8	10	7	6	8	…	3	4	9	10
14	7	7	5	4	8	…	8	9	2	4
15	9	3	9	1	7	…	5	3	7	10
16	4	2	10	3	1	…	4	6	9	7
17	5	2	9	3	2	…	4	4	9	10
18	7	7	10	6	4	…	5	8	6	7
19	7	9	9	8	5	…	6	7	7	9
20	8	7	9	7	4	…	5	8	8	4

Let's consider the calculation of the weight of first-level experts.

For each potential expert paired with user №0, the measure of agreement of opinions was calculated using two methods. The significance evaluation of the concordance coefficient was determined using the Pearson agreement criterion. The tabulated value of the Pearson criterion for k=20–1=19 degrees of freedom and a significance level of α=0.05 is equal to 30.144. If the calculated value χ^2 is greater than or equal to the tabulated value, it is considered that the value of the W criterion is not a random variable, and the obtained results are meaningful and can be used for further research. Otherwise, the value of the W criterion is considered a random variable. The calculation results are presented in Table 3.

The calculated values of the concordance coefficient passed the Pearson criterion test at the given significance level of α=0.05 for only 11 out of 20 potential experts. The values of the concordance coefficient for the remaining 9 users were very small (<30.144).

Additional research on artificially constructed datasets has shown that the coefficient of concordance starts to produce values that do not pass the Pearson criterion test with a specified significance level of α=0.05 when the

DOI 10.15588/1607-3274-2023-3-17

values according to the method of average differences of estimates (ADE) are less than 0.85. This indicates that the coefficient of concordance cannot accurately determine experts across the entire range of correspondence according to the Chaddock scale.

Table 3 – Measures of agreement between user №0 and each of the potential experts

User	Concordance coefficient		ADE method
	W	χ^2 | reliability (χ^2 >=30.144)	Weight value of the potential expert | Strength of the connection >0.7
1	W = 0.96	36.63 | +	0.9285 | +
2	W = 0.96	36.42 | +	0.8735 | +
3	W = 0.94	35.53 | +	0.9010 | +
4	W = 0.95	35.95 | +	0.9120 | +
5	W = 0.4	15.40 | –	0.6370 | –
6	W = 0.97	36.81 | +	0.9120 | +
7	W = 0.24	9.120 | –	0.5765 | –
8	W = 0.94	35.79 | +	0.8955 | +
9	W = 0.94	35.63 | +	0.8955 | +
10	W = 0.91	34.47 | +	0.9010 | +
11	W = 0.97	36.80 | +	0.9285 | +
12	W = 0.95	36.14 | +	0.9175 | +
13	W = 0.47	17.69 | –	0.6975 | –
14	W = 0.84	32.08 | +	0.8570 | +
15	W = 0.39	14.91 | –	0.6315 | –
16	W = 0.25	9.330 | –	0.5600 | –
17	W = 0.25	9.360 | –	0.5270 | –
18	W = 0.64	24.35 | +	0.7580 | +
19	W = 0.49	18.75 | –	0.7195 | +
20	W = 0.53	20.19 | –	0.7525 | +

These findings suggest that the coefficient of concordance and similar methods that take a list of rankings as input data cannot be used to solve the problem of expert selection for ranking web resources based on evaluations.

The obvious reason for the low percentage of correct results is the specific normalization of input data values in such a method. The rankings provided by experts for each object $S \in \{1..n\}$, where n is the number of objects, are used as input data for calculating the coefficient of concordance. In contrast, in the task of evaluating web resources, evaluations from 1 to 10 are used, regardless of the number of evaluated objects, which leads to a significant increase in the number of identical rankings for values of n far exceeding 10.

Therefore, the proposed method of average differences of estimates (ADE), is the most acceptable for calculating the weight of experts and will be used for further research on ranking methods. Its application requires a large amount of statistical evaluation data and can be effectively applied at the intermediate stages of system development.

Let's consider the calculation of the weight of second-level experts.

Forming expert groups only from users who have common ratings with the current user significantly narrows down the pool of potential experts. To address this issue, a method for calculating the weight of experts in the absence of common ratings with the current user has been proposed. It involves having shared ratings with first-level experts and determines the overall weight of a second-level potential expert relative to the current user $d\left(U_0,\hat{U}_i\right)$, taking into account their weight relative to the first-level expert $d\left(U_i,\hat{U}_j\right)$ and the weight of the first-level expert relative to the current user $d\left(U_i,\hat{U}_j\right)$.

To investigate the potential use of ratings from second-level experts, additional research was conducted. Table 4 presents a fragment of the data illustrating the calculation of the weight of all members in the expert group for dataset №1 in relation to each other.

Since the calculated weights of the experts are the same for both experts relative to each other, the weight table is symmetric about the diagonal. However, it is presented in its complete form for the sake of simplifying the illustration of further calculations. The analysis of the data presented in Table 4 allows the following observations:

– All experts except one have weight values >0.7, which, firstly, allows them to remain in the expert group, and secondly, demonstrates a high level of agreement within the expert group.

– It is evident that the weight of expert №14 relative to expert №19 shows a value below 0.7, because the weight of expert №14 relative to the current user has a value of 0.72 and is in close proximity to the lower boundary of the acceptable weight for participation in the expert group.

– The calculated weights of the experts relative to those who have a weight >0.9 relative to the current user have values close to their weights relative to the current user.

To develop a methodology for considering the weight of second-level experts in ranking, it is necessary to select criteria for evaluating the ratio of weights of second-level experts relative to the current user, expressed through the weight of first-level experts.

Table 5 presents the absolute differences in weights of second-level experts from Table 4 and their weights relative to the current user, which were calculated earlier (Table 3).

In Figure 2 are presented the obtained data.

Figure 3 shows a graph depicting the relationship between the average differences in the weights of second-level experts relative to the weights of first-level experts.

From the graph, it can be observed that the weight deviation significantly increases when the weight of the first-level expert is below 0.8.

Similar calculations based on expert groups from data sets 2 to 4 yield comparable results (Figures 4–9).

The weight of second-level experts relative to the current user is advisable to calculate as the product of the weight of the first-level expert relative to the potential second-level expert, who share common ratings, and the weight of the first-level expert relative to the current user.

$$W\left(d\left(U_0,\hat{U}_j\right)\right)=W\left(d\left(U_i,\hat{U}_j\right)\right)\cdot W\left(d\left(U_0,\hat{U}_i\right)\right).$$

The expert group selects potential second-level experts whose weight relative to the current user is $W\left(d\left(U_0,\hat{U}_j\right)\right)>0.7$ on the Chaddock scale.

DOI 10.15588/1607-3274-2023-3-17

Table 4 – The fragment of the sample: weights of experts relative to each other

Users	U_0	U_1	U_2	U_3	U_4	U_6	U_{11}	U_{12}	U_{14}	U_{18}	U_{19}	U_{20}
U_0	1	0.9285	0.8735	0.901	0.912	0.912	0.9285	0.9175	0.857	0.758	0.7195	0.7525
U_1	0.9285	1	0.868	0.9065	0.9285	0.8845	0.912	0.901	0.8625	0.7855	0.758	0.802
U_2	0.8735	0.868	1	0.8845	0.8405	0.8735	0.879	0.901	0.8295	0.7855	0.769	0.802
U_3	0.901	0.9065	0.8845	1	0.89	0.879	0.8955	0.8845	0.824	0.769	0.7525	0.7965
U_4	0.912	0.9285	0.8405	0.89	1	0.868	0.9065	0.8845	0.868	0.747	0.7305	0.7745
U_6	0.912	0.8845	0.8735	0.879	0.868	1	0.9285	0.8955	0.824	0.725	0.7085	0.7305
U_{11}	0.9285	0.912	0.879	0.8955	0.9065	0.9285	1	0.901	0.8515	0.7525	0.736	0.769
U_{12}	0.9175	0.901	0.901	0.8845	0.8845	0.8955	0.901	1	0.8735	0.7745	0.725	0.791
U_{14}	0.857	0.8625	0.8295	0.824	0.868	0.824	0.8515	0.8735	1	0.758	0.6865	0.7635
U_{18}	0.758	0.7855	0.7855	0.769	0.747	0.725	0.7525	0.7745	0.758	1	0.8075	0.8405
U_{19}	0.7195	0.758	0.769	0.7525	0.7305	0.7085	0.736	0.725	0.6865	0.8075	1	0.824
U_{20}	0.7525	0.802	0.802	0.7965	0.7745	0.7085	0.758	0.7525	0.769	0.791	0.736	1

Table 5 – The absolute difference values of the weights of second-level experts

Users	U_0	U_1	U_2	U_3	U_4	U_6	U_{11}	U_{12}	U_{14}	U_{18}	U_{19}	U_{20}
U_0	0	0.9285	0.8735	0.901	0.912	0.912	0.9285	0.9175	0.857	0.758	0.7195	0.7525
U_1	0.9285	0	0.0055	0.0055	0.0165	0.0275	0.0165	0.0165	0.0055	0.0275	0.0385	0.0495
U_2	0.8735	0.0605	0	0.0165	0.0715	0.0385	0.0495	0.0165	0.0275	0.0275	0.0495	0.0495
U_3	0.901	0.022	0.011	0	0.022	0.033	0.033	0.033	0.033	0.011	0.033	0.044
U_4	0.912	0	0.033	0.011	0	0.044	0.022	0.033	0.011	0.011	0.011	0.022
U_6	0.912	0.044	0	0.022	0.044	0	0	0.022	0.033	0.033	0.011	0.022
U_{11}	0.9285	0.0165	0.0055	0.0055	0.0055	0.0165	0	0.0165	0.0055	0.0055	0.0165	0.0165
U_{12}	0.9175	0.0275	0.0275	0.0165	0.0275	0.0165	0.0275	0	0.0165	0.0165	0.0055	0.0385
U_{14}	0.857	0.066	0.044	0.077	0.044	0.088	0.077	0.044	0	0	0.033	0.011
U_{18}	0.758	0.143	0.088	0.132	0.165	0.187	0.176	0.143	0.099	0	0.088	0.088
U_{19}	0.7195	0.1705	0.1045	0.1485	0.1815	0.2035	0.1925	0.1925	0.1705	0.0495	0	0.0715
U_{20}	0.7525	0.1265	0.0715	0.1045	0.1375	0.1815	0.1595	0.1265	0.0935	0.0825	0.1045	0

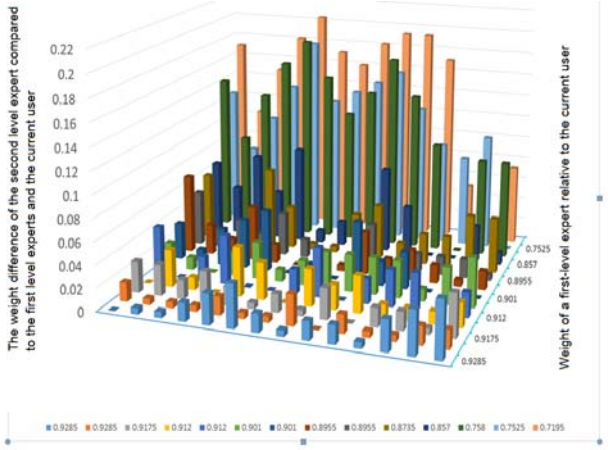


Figure 2 – The weight difference of second-level users

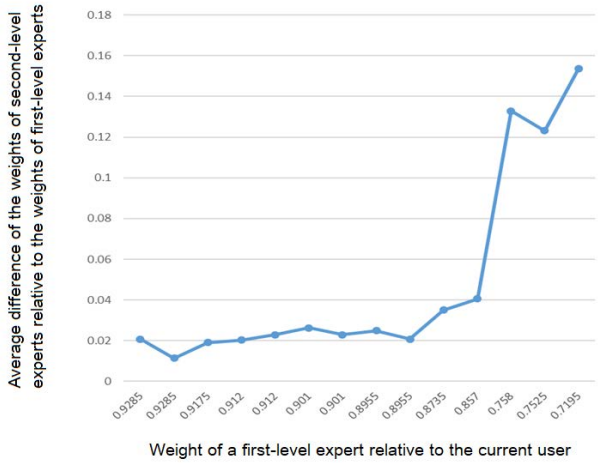


Figure 3 – The dependence of the average difference in weight of a second-level user on the weight of a first-level expert

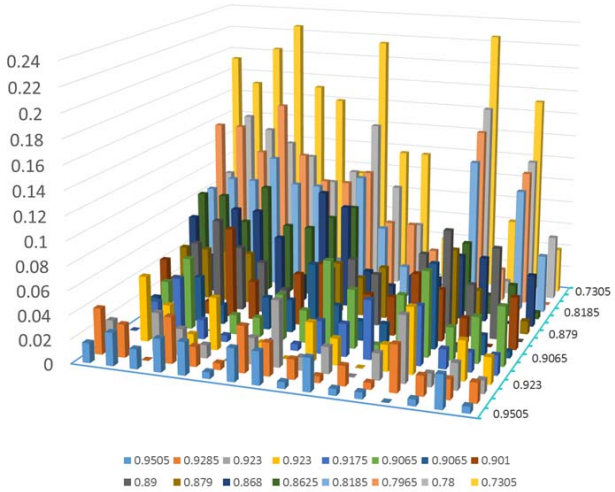


Figure 4 – The weight difference of second-level users for data set №2

Let's consider the calculation of the weight of users who do not have common ratings.

The indirect approach is suitable when there are limited or no shared ratings. It utilizes a model that calculates the weight of potential experts based on their individual social profiles. The most accurate results are obtained by analyzing the experts' previous activities. However, in the system's initial implementation phase and before accumulating a substantial rating database, situations may arise where there is insufficient data to apply this method. Accumulating a sufficient rating database means having a significant number of shared ratings to form expert groups for the majority of system users. Therefore, to ensure the system's proper functioning in its early stages, a method was developed to determine the weight of users with no shared ratings at all.

DOI 10.15588/1607-3274-2023-3-17

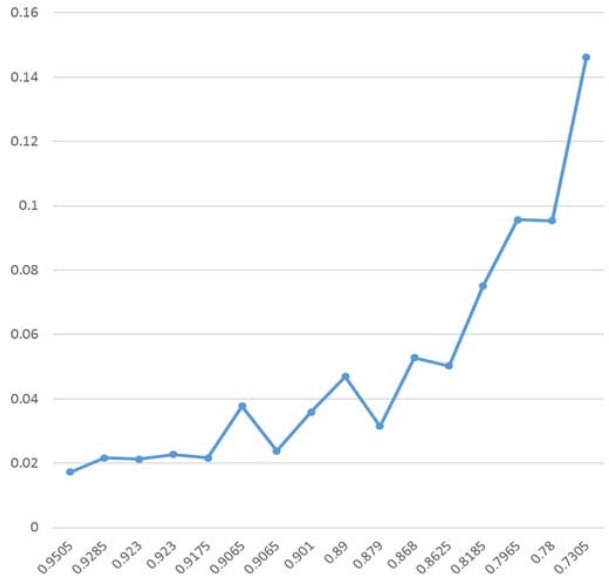


Figure 5 – The dependence between the average weight difference of second-level users and the weight of first-level experts in dataset №2

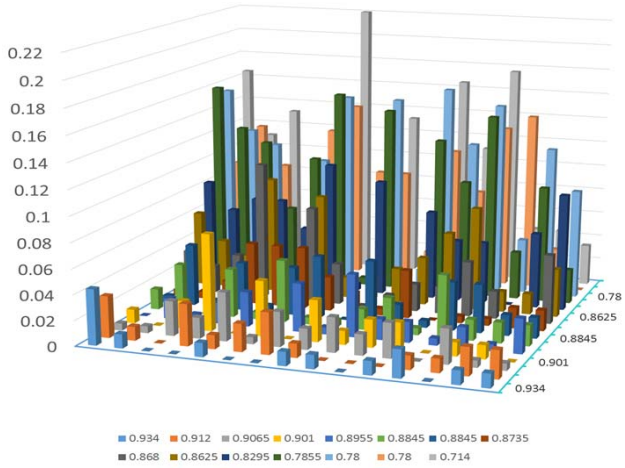


Figure 6 – The weight difference of second-level users for data set №3

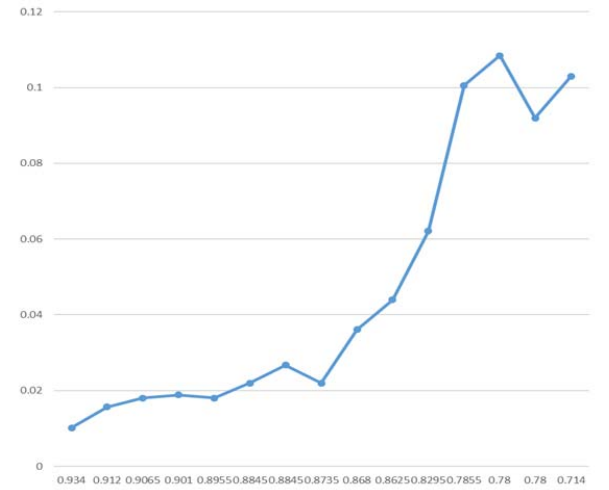


Figure 7 – The dependence between the average weight difference of second-level users and the weight of first-level experts in dataset №3

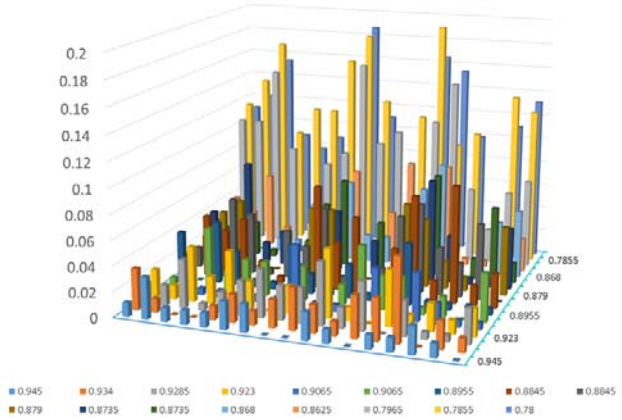


Figure 8 – The weight difference of second-level users for data set №4

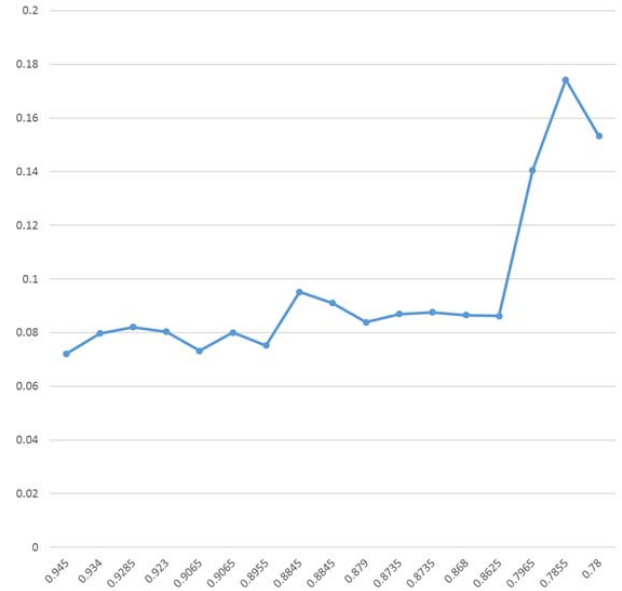


Figure 9 – The dependence between the average weight difference of second-level users and the weight of first-level experts in dataset №4

A user's social profile is formed based on the information they provide during system registration [24]. Taking into account global and domestic experience in conducting psychological research [25–27], a set of socio-personal factors that may influence the formation of a user's opinion was selected. Based on the research results, the factors found to be informative in building the model for determining the weight of potential experts will be included as mandatory fields in the system's registration form.

To build the model for calculating the weight of experts, a set of subjective features x_m was selected, which can directly or indirectly influence the visitor's rating. Social profiles of users (U_0 – U_{60}) who participated in previous experiments were used as input data. Since their weight relative to the user U_0 has already been calculated based on the data samples of ratings №1–4, the same expert groups presented in tables 2–5 were used to construct the model.

DOI 10.15588/1607-3274-2023-3-17

It is necessary to construct a model of the influence of socio-personal factors of Internet users on the degree of consensus between the potential expert of the third level $U_{0,exp}$ and the current user.

The sample contains $n = 20$ observation points and is divided into two parts: 2/3 of the points for the training sample A and 1/3 of the points for the validation sample B: $n_A = 14$, $n_B = 6$.

The accuracy of the obtained models was assessed using the coefficient of determination R^2.

The GIA algorithm is used at the third level to determine the weight of users who have no shared ratings at all [28–29].

The aggregated results of the modeling are presented in Table 6.

From Table 6, it can be observed that the best modeling results were obtained using the GIA. The GIA allows for the utilization of the mathematical model itself, which is more convenient for this particular task, as the obtained model serves as an intermediate step in solving the given problem within a limited time frame. Furthermore, as evident from the obtained dependency (GIA model), only 6 out of the 11 variables are significant. Detailed results for the GIA are presented in Table 7 and Figures 10–12.

Figure 10 show that the generalized algorithm reaches a minimum on the 7th layer.

Based on the above, a general conclusion can be drawn: the modeling results indicate that the best model was found using the generalized iterative algorithm, which incorporates all previous iterative structures.

Table 6 – Summarized results of modeling by GMDH algorithms

Algorithm	R^2, %	True monomials											Number of redundant monomials
		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	
Iterative algorithms with linear partial model description													
MIA	42.2	+	+	+		+	+	+	+	+			3
RIA	17.70			+									10
CIA	55.14	+	+	+		+	+	+	+	+	+		2
Iterative algorithms with quadratic partial model description													
MIA	56.25			+		+	+	+	+	+			5
RIA	54.25					+		+		+			8
CIA	77.17			+		+		+					8
MICA	71.80	+		+		+	+	+	+		+		4
RICA	54.25					+		+		+			8
GIA	80.27	+	+			+		+	+				6

Table 7 – GIA results

№	Data set	y (real)	$\hat{y}$ (model)	Error
1	*A (training)*	1	0.904	0.096
2		0.9285	0.895	0.033
3		0.8735	0.752	0.122
4		0.901	0.882	0.019
5		0.912	0.810	0.102
6		0.637	0.628	0.009
7		0.912	0.865	0.047
8		0.5765	0.637	–0.061
9		0.8955	0.818	0.077
10		0.8955	0.961	–0.066
11		0.901	0.911	–0.010
12		0.9285	0.886	0.042
13		0.9175	0.928	–0.010
14		0.6975	0.737	–0.040
15	*B (testing)*	0.857	0.856	0.001
16		0.6315	0.730	–0.098
17		0.56	0.680	–0.120
18		0.527	0.574	–0.047
19		0.758	0.788	–0.030
20		0.7195	0.814	–0.094
21		0.7525	0.750	0.003
		R^2, %	80.27%	

DOI 10.15588/1607-3274-2023-3-17

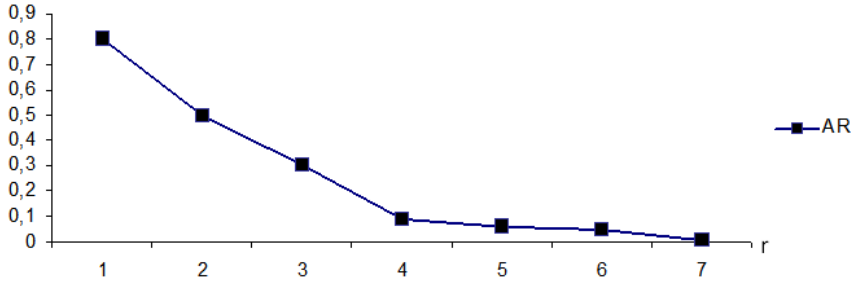


Figure 10 – The value of the evaluation criterion (*AR* criterion) by layers

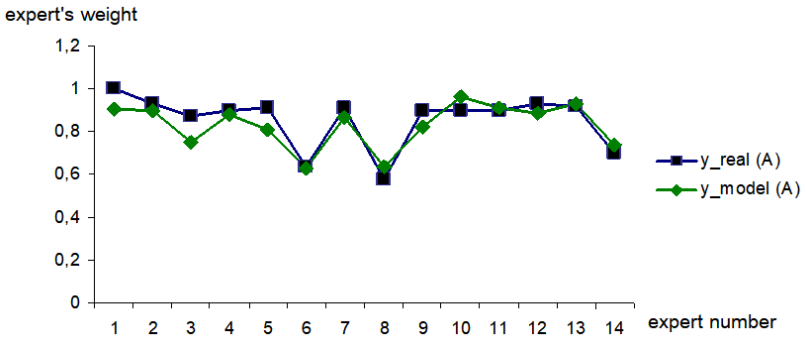


Figure 11 – The values of true and model results on sample *A*

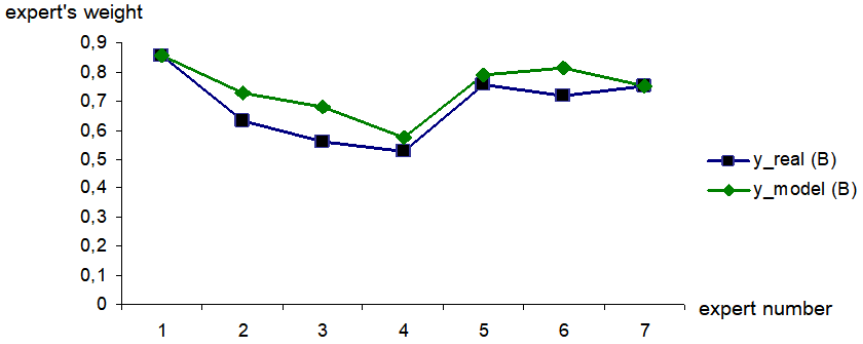


Figure 12 – The values of true and model results on sample *B*

5 RESULTS

The paper explores the challenge of ranking search engine results using google.com.ua as a case study. Initially, this task may seem formidable due to the vast number of results, often reaching hundreds of thousands or even millions. However, a closer examination of search engine algorithms reveals that the computational complexity is significantly lower [27]. In practice, the user's accessible search results are much fewer than what the system claims during query execution. Consequently, the ranking process effectively focuses on analyzing only a few hundred results, requiring minimal computational power.

To assess the effectiveness and accuracy of the aforementioned ranking methods based on user ratings, four experiments were conducted. Each experiment formed a distinct data sample, consisting of 20 users rating 20 web resources based on information quality and usability. A total of 60 users participated across the four samples, with some users rating web resources in multiple samples. Consistent numbering was employed for the same users across all experiments, ensuring User №1, for instance, remained consistent. Additionally, each experiment incorporated a combination of previous experts and new users, allowing for tracking the weight values of the same experts across different data samples.

Let's consider the determining the weight of experts by the ADE method.

The weight values of potential experts relative to user №0 for data sample №1 were calculated using the ADE method. The results are presented in Table 8.

Based on the results presented in the table, an expert group is formed for User №0. It includes users who have a calculated weight value greater than 0.7 according to the Cheddock scale.

Expert numbers: 1–4, 6, 8–12, 14, 18–20.

A new data sample is created for the ranking of web resources, which includes evaluations only from the experts (Table 9).

DOI 10.15588/1607-3274-2023-3-17

Table 8 – The weight values of potential experts relative to User №0 for data sample №1

User	User weight	Selected to the expert group
1	0.9285	+
2	0.8735	+
3	0.9010	+
4	0.9120	+
5	0.6370	–
6	0.9120	+
7	0.5765	–
8	0.8955	+
9	0.8955	+
10	0.9010	+
11	0.9285	+
12	0.9175	+
13	0.6975	–
14	0.8570	+
15	0.6315	–
16	0.5600	–
17	0.5270	–
18	0.7580	+
19	0.7195	+
20	0.7525	+

Table 9 – The fragment of the expert group evaluations

web resource → ↓ User	A_1	B_1	C_1	D_1	E_1	…	Q_1	R_1	S_1	T_1
0	10	10	8	5	7	…	8	8	3	5
1	10	10	9	6	8	…	8	8	4	4
2	8	8	9	5	6	…	7	7	3	4
3	9	9	9	5	7	…	7	6	3	3
4	10	10	9	7	7	…	9	9	4	3
6	9	10	7	6	8	…	7	6	4	4
8	8	8	6	4	7	…	9	7	3	4
9	7	9	8	5	6	…	9	7	4	3
10	7	9	7	4	8	…	9	8	3	4
11	9	10	8	5	7	…	7	9	4	4
12	8	8	9	6	7	…	8	8	3	4
14	7	7	5	4	8	…	8	9	2	4
18	7	7	10	6	4	…	5	8	6	7
19	7	9	9	8	5	…	6	7	7	9
20	8	7	9	7	4	…	5	8	8	4

Prior to commencing the ranking process using the selected methods, it is essential to acquire a reference ranking as a basis for comparison. Merely sorting the web resources based on User №0's ratings in descending order yields only approximate outcomes. This approach solely arranges clusters of web resources with identical ratings in descending order in Table 10. The precise order of web resources within each cluster remains unknown. Hence, to establish the reference ranking, User №0 manually assigned a rank ranging from 1 to 20 to each web resource, with 1 denoting the highest rank (Table 10). The reference ranking for the current user was created manually.

Let's consider the ranking of web resources by the method of average points.

Traditionally, the application of the average method involves the use of the arithmetic mean for value calculation. In this work, the following methods were used to calculate the averages: AM, WAM, HM, WHM.

Weighted modifications of the arithmetic mean and harmonic mean allow for the calculation of new rankings of web resources, taking into account the experts' weights. The calculation results of the web resource rankings using the average methods are presented in Table 11.

Table 10 – The reference ranking for the current user

The number of the web resource in the data set	Sort by decreasing rating	Manual ranking of the current user
1	A_1	B_1
2	B_1	A_1
3	I_1	K_1
4	J_1	J_1
5	K_1	I_1
6	C_1	C_1
7	H_1	R_1
8	Q_1	Q_1
9	R_1	H_1
10	E_1	E_1
11	L_1	L_1
12	G_1	G_1
13	M_1	N_1
14	N_1	M_1
15	D_1	D_1
16	T_1	T_1
17	F_1	F_1
18	O_1	O_1
19	P_1	S_1
20	S_1	P_1

Table 11 – The results of the web resource rankings using the average methods

Web resource	The rank of the web resource, calculated by methods:			
	AM	WAM	HM	WHM
A_1	8.142857	8.186422	8.158281	8.107372
B_1	8.642857	8.692747	8.699647	8.561141
C_1	8.142857	8.11233	8.101196	8.119847
D_1	5.571429	5.527025	5.614922	5.502869
E_1	6.571429	6.656235	6.639741	6.538092
F_1	5.214286	5.182274	5.272789	5.196464
G_1	5.571429	5.538319	5.598789	5.574416
H_1	7.428571	7.402497	7.459537	7.392929
I_1	7.857143	7.937367	7.836755	7.877952
J_1	8.071429	8.128224	8.07852	8.063776
K_1	8.142857	8.204493	8.148353	8.135257
L_1	6.428571	6.480204	6.459537	6.412008
M_1	6.857143	6.776984	6.810695	6.901067
N_1	6.071429	6.035609	5.976346	6.173482
O_1	3.928571	3.863192	3.928512	3.910907
P_1	4.357143	4.248973	4.402527	4.271915
Q_1	7.428571	7.505052	7.482288	7.391828
R_1	7.642857	7.643954	7.669863	7.644797
S_1	4.142857	4.045014	4.114847	4.126818
T_1	4.357143	4.268852	4.427347	4.288059

The ranking outcomes are displayed in Table 12. To evaluate the efficiency of the proposed techniques in computing average scores, the mean deviation from the reference ranking is utilized. This metric is obtained by averaging the differences in positions among the web resources.

DOI 10.15588/1607-3274-2023-3-17

Table 12 – The results of ranking

The reference ranking	ranking by methods				Position deviation by:			
	AM	WAM	HM	WHM	AM	WAM	HM	WHM
B_1	B_1	B_1	B_1	B_1	0	0	0	0
A_1	A_1	K_1	A_1	A_1	0	1	0	0
K_1	C_1	A_1	K_1	K_1	3	1	0	0
J_1	K_1	J_1	J_1	J_1	1	0	0	0
I_1	J_1	C_1	C_1	C_1	1	1	1	1
C_1	I_1	I_1	I_1	I_1	1	1	1	1
R_1	R_1	R_1	R_1	R_1	0	0	0	0
Q_1	H_1	Q_1	H_1	Q_1	1	0	1	0
H_1	Q_1	H_1	Q_1	H_1	1	0	1	0
E_1	M_1	M_1	M_1	M_1	4	4	4	4
L_1	E_1	E_1	E_1	E_1	1	1	1	1
G_1	L_1	L_1	L_1	L_1	1	1	1	1
N_1	N_1	N_1	N_1	N_1	0	0	0	0
M_1	D_1	G_1	D_1	D_1	1	2	1	1
D_1	G_1	D_1	G_1	G_1	3	0	3	3
T_1	F_1	F_1	F_1	F_1	1	1	1	1
F_1	P_1	T_1	T_1	T_1	3	1	1	1
O_1	T_1	P_1	P_1	P_1	2	2	2	2
S_1	S_1	S_1	S_1	S_1	0	0	0	0
P_1	O_1	O_1	O_1	O_1	2	2	2	2
The sum of deviations:					26	18	20	18
Average deviation value:					1.3	0,9	1	0,9

From the Table 12, it can be seen that the methods of WAM and WHM have the smallest error, indicating the relevance of considering the weight of experts when calculating the ranking of web resources.

To justify the feasibility and effectiveness of using ratings only from users who have a strong connection with the current user in the ranking process, additional calculations were performed to determine the rankings of web resources using ratings from all users in data set №1.

Below is a comparative table of the final rankings of web resources calculated based on the ratings from two groups of users, Table 13:

– Expert group.

– All users from Sample № 1.

Ranking based on the ratings of all users compared to ranking based on the ratings of a predefined expert group produces significantly worse results.

The results presented in Tables 11–13 prove the effectiveness of taking into account the weight of experts when calculating the ranks of web resources and justify the need to filter out users with low indicators of the degree of agreement of opinions relative to the current user. Experiments № 2, 3, 4 were conducted according to the same method. Therefore, their description is not given.

Table 13 – Comparative table of rankings of web resources calculated based on ratings from two groups of users

Reference ranking	Expert group				All users			
	AM	WAM	HM	WHM	AM	WAM	HM	WHM
B_1	B_1	B_1	B_1	B_1	C_1	J_1	I_1	I_1
A_1	A_1	K_1	A_1	A_1	I_1	I_1	M_1	J_1
K_1	C_1	A_1	K_1	K_1	B_1	C_1	J_1	K_1
J_1	K_1	J_1	J_1	J_1	K_1	K_1	K_1	M_1
I_1	J_1	C_1	C_1	C_1	A_1	A_1	A_1	A_1
C_1	I_1	I_1	I_1	I_1	M_1	M_1	Q_1	B_1
R_1	R_1	R_1	R_1	R_1	Q_1	Q_1	B_1	Q_1
Q_1	H_1	Q_1	H_1	Q_1	J_1	B_1	N_1	C_1
H_1	Q_1	H_1	Q_1	H_1	N_1	R_1	C_1	N_1
E_1	M_1	M_1	M_1	M_1	R_1	N_1	F_1	R_1
L_1	E_1	E_1	E_1	E_1	E_1	E_1	H_1	H_1
G_1	L_1	L_1	L_1	L_1	H_1	H_1	R_1	F_1
N_1	N_1	N_1	N_1	N_1	F_1	F_1	G_1	E_1
M_1	D_1	G_1	D_1	D_1	T_1	L_1	T_1	G_1
D_1	G_1	D_1	G_1	G_1	L_1	G_1	E_1	L_1
T_1	F_1	F_1	F_1	F_1	G_1	P_1	L_1	T_1
F_1	P_1	T_1	T_1	T_1	P_1	D_1	P_1	D_1
O_1	T_1	P_1	P_1	P_1	S_1	T_1	S_1	P_1
S_1	S_1	S_1	S_1	S_1	D_1	S_1	D_1	S_1
P_1	O_1	O_1	O_1	O_1	O_1	O_1	O_1	O_1
The sum of deviations	26	18	20	18	62	58	74	56
Average deviation value	1.3	0,9	1	0,9	3.1	2.9	3.7	2.8

DOI 10.15588/1607-3274-2023-3-17

Table 14 presents the summarized results of the conducted experiments. The last row of the table shows the count of instances where each method yielded the best ranking results. If more than one method achieved the best results during the experiment, all of them are considered the best.

Table 14 – The summary results of experiments

Experiment number	The method of calculating values by the method of average points			
	AM	WAM	HM	WHM
№1 The sum of deviations	26	18	20	18
Average value of deviation:	1.3	0,9	1	0,9
№2 The sum of deviations	32	28	32	24
Average value of deviation:	1.6	1.4	1.6	1.2
№3 The sum of deviations	32	30	24	22
Average value of deviation:	1.6	1.5	1.2	1.1
№4 The sum of deviations	22	22	20	20
Average value of deviation:	1.1	1.1	1	1
Best results	0	1	1	4

From the table 13, it can be seen that during the four experiments, the method of average scores calculated based on weighted harmonic mean showed the smallest deviation from the reference ranking. Based on these results, this approach will be used in the development of the meta-search engine.

6 DISCUSSION

The proposed methodology for forming unique expert groups for each user involves three approaches depending on the presence of shared ratings between the current user and potential experts:

1. When there are shared ratings between potential experts and the current user, the weight is calculated using the ADE method, which includes:

– calculating the average differences of estimates;

– applying a normalization function to scale the data from 0 to 0.99.

– selecting users who have a strong connection with the current user based on the Cheddock scale (values > 0.7).

Research has shown that the ADE method is more effective in solving research tasks and provides significantly better results compared to the Kendall's concordance method and similar approaches.

2. When there are no shared ratings with the current user but there are shared ratings between potential second-level experts and first-level experts, the weight of second-level experts relative to the current user is calculated as the product of the weight of the first-level expert relative to the potential second-level expert with shared ratings and the weight of the first-level expert relative to the current user.

3. When there are no shared ratings at all, the expert group is formed based on a model constructed from the user's social profile using inductive algorithms.

To calculate the rankings of web resources in search result ranking, the methods of average ratings were considered. However, the classical form of the average ratings method does not yield high results and does not take into account the weight of experts. The results of conducted experiments have shown that the best ranking results are achieved using the method of weighted average ratings, specifically using the weighted harmonic mean where the weights are based on the experts' expertise.

Comparative analysis of inductive modeling methods showed that GIA provides the most accurate results. The task involved building a model that captures the relationship between the measure of agreement of opinions and the socio-personal factors of users, which is why the neural network approach was not applied for comparison, as it does not allow obtaining such a model. The best model for calculating the weight of third-level potential experts relative to the current user was found using a generalized iterative algorithm that incorporates all previous iteration structures.

The described methodology of constructing a personalized model for ranking web resources based on user ratings has demonstrated high effectiveness, indicating the promising development of this direction.

CONCLUSIONS

The urgent problem of enhancing search efficiency, an approach to search result management based on user's subjective information needs is employed.

The scientific novelty of obtained results is introducing scientific novelty through a search result ranking method that generates a unique order of web resources for individual users. This is accomplished by leveraging ratings from user-specific expert groups and incorporating each rating with a distinct weight into the model for calculating final rankings. The weight is determined based on an analysis of the web resources' previous activities within the system.

The practical significance of obtained results is that the software that implements the proposed methods, along with conducting experiments to examine their properties. The experimental outcomes support the recommendation of the proposed methods for practical use, while also identifying effective conditions for their application.

Prospects for further research are to involve the exploration of building ranking models that incorporate a multitude of factors, similar to contemporary search systems.

REFERENCES

1. Gärdenfors P. How to make the Semantic Web more semantic, *Digests 3th Conf. Formal Ontology in Information Systems*, FOIS, 2004, pp. 17–34.
2. SearchWiki: https://ukraine.googleblog.com/2009/05/.
3. Zosimov V. The Significant Features for Search Engines Ranking Results, *Inductive Modelling: proceedings of the International Workshop*. Zhukyn IWIM, 18–20 July, 2016, Kyiv, 2016. P. 54–60.
4. Britsov R.A. Improvement of quality of the information search based on ranking rationalization and users' behavioristic characteristics, *T-Comm.*, 2016, Vol. 10, No. 2, pp. 63–66.
5. Dahake S., Thakre V. Search Engine Optimization Techniques – The Analysis, *International Journal of Advanced*

DOI 10.15588/1607-3274-2023-3-17

Research in Computer Science, 2014, Volume 5, No. 4(Special Issue), P. 163–167
6. Kaur G. Role and Importance of search engine optimization, *International journal of research – Granthaalayah*, 2017, Vol. 5, Issue 6, pp. 147–151.
7. Pham D. T., Dimov S. S., and Nguyen C. D. Selection of K in K-means clustering, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2005, Vol. 219 (1), pp. 103–119. DOI:10.1243/ 095440605x8298.
8. Hudry O. Complexity of computing median linear orders and variants, *Electronic Notes in Discrete Mathematics*, 2013, vol. 42, pp. 57–64. DOI:10.1016/j.endm. 2013.05.146.
9. Velychko O. M., Gordiyenko T. B., Kolomiets L. V. Methodologies of expert's competence evaluation and group expert evaluation, *Metallurgical and Mining Industry*, 2015, Issue 2, pp. 262–271
10. Zosimov V., Bulgakova O., Pozdeev V. In: Babichev, S., Lytvynenko, V., Wójcik, W., Vyshemyrskaya, S. (eds) Complex Internet Data Management System, *Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2020. Advances in Intelligent Systems and Computing.* Springer, Cham, 2021, Vol. 1246, pp. 639–652 DOI:10.1007/978-3-030-54215-3_41
11. Yu R., Liu Q., Ye Y., Cheng M., Chen E. and Ma J. Collaborative List-and-Pairwise Filtering From Implicit Feedback, *IEEE Transactions on Knowledge and Data Engineering*, 2022, Vol. 34, No. 6, pp. 2667–2680, DOI:10.1109/TKDE.2020.3016732.
12. Raj N. S. and R. V G. A Rule-Based Approach for Adaptive Content Recommendation in a Personalized Learning Environment: An Experimental Analysis, *IEEE Tenth International Conference on Technology for Education (T4E)*, 2019 pp. 138–141. DOI:10.1109/T4E.2019.00033.
13. Li Li, Yujie Zhang, Shuai Zhang, and Xinyu Li. A Hybrid Approach for Ranking Research Articles, *Information Processing & Management*, 2017, Vol. 53, No. 6, pp. 1407–1419.
14. Zhaopeng Meng, Jianxun Lian, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. Personalized Ranking Framework via Multilayer Perceptron Learning, *IEEE Transactions on Knowledge and Data Engineering*, 2017, Vol. 29, No. 7, pp. 1509–1521.
15. Shuai Shen, Zhiyuan Liu, Maosong Sun, and Yang Liu. Neural Ranking Models with Weak Supervision, *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2019, pp. 111–120. DOI:10.48550/arXiv.1704.08803
16. Makarov M., Tutubalina E., Braslavski P. BERT for Ad-hoc Retrieval: Baselines and Analysis, *Advances in Information Retrieval – 42nd European Conference on IR Research, ECIR 2020*, pp. 516–530.
17. Yongqiang Zhang, Min Zhang, Yiqun Liu, Shaoping Ma, and Jintao Li. A Reinforcement Learning Approach for Rank Aggregation in Information Retrieval, *IEEE Access*, 2021, Vol. 9, pp. 316–327.
18. Mohammed S., AL-Ghuribi Shahrul, Mohd Noah A. A Comprehensive Overview of Recommender System and Sentiment Analysis. URL: https://arxiv.org/ftp/arxiv/papers/2109/2109.08794.pdf
19. Brave S., Bradshaw R., Jia J., Minson C.. Method and apparatus for identifying, extracting, capturing, and leveraging expertise and knowledge. URL: https://patents.google.com/patent/US7698270
20. Britsov R. Ranking of information based on user ratings and behavior, *T-Comm, Telecommunications and Transport*, 2016, Vol. 10. No. 1, P. 62.
21. Christopher D. Manning. Introduction to Information Retrieval. Cambridge University Press, 2008, 520 p. URL: https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf
22. Velychko O. M., Gordiyenko T. B., Kolomiets L. V. Hardware-software complexes for evaluation of competence level of experts, *Metallurgical and Mining Industry*, 2015, Issue 7, pp. 396–401
23. Stoica P., Selen Y. Model-order selection: a review of information criterion rules, *IEEE Signal Processing Mag.*, 2004, Vol. 21, No. 4, pp. 36–47. DOI: https://doi.org/10.1109/MSP.2004.1311138.
24. Bulgakova O., Stepashko V., Zosimov V. Numerical study of the generalized iterative algorithm GIA GMDH with active neurons, *Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT* 2017, pp. 496–500. DOI:10.1109/STC-CSIT.2017.8098836.
25. Zosimov V., Bulgakova O., Pozdeev V.. Semantic Profile of Corporate Web Resources, *CEUR Workshop Proceedings*, 2021, Vol. 3179, pp. 389–397. URL: https://ceur-ws.org/Vol-3179/Short_13.pdf
26. Vink P. Advances in Social and Organizational Factors/Advances in Human Factors and Ergonomics. T.12, AHFE Conference, 2014, 598 p.
27. Bicchieri C. The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press, 2005, 306 p.
28. Zosimov V., Stepashko V., Bulgakova O. Ed. by Markus Spies at al. Inductive Building of Search Results Ranking Models to Enhance the Relevance of Text Information Retrieval. Database and Expert Systems Applications. Los Alamitos, IEEE Computer Society, 2015, pp. 291–295.
29. Zosimov V., Bulgakova O. In: Lytvynenko, V., Babichev, S., Wójcik, W., Vynokurova, O., Vyshemyrskaya, S., Radetskaya, S. (eds) Calculation the Measure of Expert Opinions Consistency Based on Social Profile Using Inductive Algorithms, *Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2019, Advances in Intelligent Systems and Computing*. Vol 1020, Springer, Cham, pp. 622–636. DOI:10.1007/978-3-030-26474-1_43

Received 21.06.2023.
Accepted 11.07.2023.

УДК 004.891

РАНЖУВАННЯ ДАНИХ НА ОСНОВІ КОРИСТУВАЦЬКИХ РЕЙТИНГІВ

Зосімов В. В. – д-р техн. наук, професор кафедри прикладних інформаційних технологій Київського університету імені Тараса Шевченка, Київ, Україна.

Булгакова О. С. – канд. техн. наук, доцент кафедри прикладних інформаційних технологій Київського університету імені Тараса Шевченка, Київ, Україна.

DOI 10.15588/1607-3274-2023-3-17

Передерій В. І. – д-р техн. наук, професор кафедри інформатики та комп'ютерних наук Херсонського національного технічного університету, Херсон, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто проблема персоналізації результатів роботи пошукових систем шляхом надання користувачу інструментів управління пошуковою видачею, а також розробки нових моделей ранжування, заснованих на суб'єктивних інформаційних потребах користувача. Об'єктом дослідження є моделювання результатів пошуку інформації в мережі Інтернет на основі користувацьких оцінок.

Мета роботи – формування унікальних для кожного користувача експертних груп, заснованого на розрахунку міри узгодженості думок поточного користувача з потенційними експертами.

Метод. Запропонований новий підхід до ранжування результатів пошуку на основі оцінок користувачів. Основною відмінністю даного методу є суб'єктивний підхід до процесу ранжування. Такий ефект досягається за рахунок попереднього формування експертних груп, унікальних для кожного користувача. Експерти відбираються за мірою узгодженості думок з поточним користувачем, що розраховується на основі спільних оцінок для деякої множини веб-ресурсів. Відбір користувачів до експертної групи відбувається на основі його ваги відносно поточного користувача, що є мірою узгодженості їх думок.

Запропонована нова методика формування унікальних для кожного користувача експертних груп, що передбачає три підходи в залежності від наявності спільних оцінок для деякої множини веб-ресурсів між поточним користувачем та потенційними експертами.

Розроблений метод ранжування результатів пошуку видає для кожного користувача список веб-ресурсів, що має свій унікальний порядок елементів. Такий ефект досягається за рахунок використання оцінок членів експертної групи, що є унікальною для кожного користувача, а також за рахунок того, що кожна оцінка входить до моделі розрахунку кінцевих рангів веб-ресурсів зі своєю унікальною вагою, розрахованою на основі аналізу їх попередньої діяльності в системі.

Результати. Розроблені методи реалізовано в програмному забезпеченні та досліджено для вирішення задач оперування даними в мережі Інтернет.

Висновки. Проведені експерименти підтвердили працездатність запропонованого програмного забезпечення та дозволяють рекомендувати його для використання на практиці для вирішення задач оперування даними в мережі Інтернет. Перспективи подальших досліджень можуть включати оптимізацію програмних реалізацій, а також експериментальне дослідження запропонованих методів на більш складних практичних задачах різної природи та розмірності.

КЛЮЧОВІ СЛОВА: пошук інформації, ранжування, результати пошуку, оцінки користувачів, експертні групи, соціальний профіль, індуктивні алгоритми, поліноміальна нейронна мережа, активні нейрони.

ЛІТЕРАТУРА

1. Gärdenfors P. How to make the semantic web more semantic / P. Gärdenfors // Formal Ontology in Information Systems: Digests 3th conference, FOIS, 2004 : proceeding. – IOS Press, 2004. – P. 17–34.
2. SearchWiki [Electronic resource]. – Access mode: https://ukraine.googleblog.com/2009/05/
3. Zosimov V. The Significant features for search engines ranking results / V. Zosimov // Inductive Modelling : proceedings of the International Workshop, Zhukyn IWIM, 18–20 July 2016 : proceeding. – Kyiv, 2016. – P. 54–60.
4. Britsov R. A. Improvement of quality of the information search based on ranking rationalization and users' behavioristic characteristics / R. A. Britsov // T-Comm. – 2016. – vol. 10, no. 2. – P. 63–66.
5. Dahake S. Search Engine optimization techniques – The Analysis / S. Dahake, V. Thakre // International Journal of Advanced Research in Computer Science. – 2014. – Vol. 5, No. 4. – P. 163–167.
6. Kaur G. Role and Importance of search engine optimization / G. Kaur // International journal of research – Granthaalayah. – 2017. – Vol. 5, Issue 6. – P. 147–151.
7. Pham D.T. Selection of K in K-means clustering / D.T. Pham, S.S. Dimov, C.D. Nguyen // Journal of Mechanical Engineering Science, Part C. – 2005. – Vol. 219, Issue 1. – P. 103–119.
8. Hudry O. Complexity of computing median linear orders and variants / O. Hudry // Electronic Notes in Discrete Mathematics. – 2013. – Vol. 42. – P. 57–64. . DOI:10.1016/j.endm. 2013.05.146.
9. Velychko O. M. Methodologies of expert's competence evaluation and group expert evaluation / O. M. Velychko, T. B. Gordiyenko, L. V. Kolomiets // Metallurgical and Mining Industry. – 2015. – Issue 2. – P. 262–271.
10. Zosimov V. Complex Internet data management system / V. Zosimov, O. Bulgakova, V. Pozdeev // Lecture Notes in Computational Intelligence and Decision Making (Advances in Intelligent Systems and Computing). – 2021. – Vol 1246. – P. 639–652. DOI:10.1007/978-3-030-54215-3_41
11. Yu R. Collaborative list-and-pairwise filtering from implicit feedback / [R. Yu, Q. Liu, Y. Ye et al.] // Transactions on Knowledge and Data Engineering. – 2022. – Vol. 34, № 6. – P. 2667–2680. DOI:10.1109/TKDE.2020.3016732.
12. Raj N. S. A rule-based approach for adaptive content recommendation in a personalized learning environment: An Experimental Analysis / N. S. Raj and R. VG // Technology for Education (T4E) : Ten international conference : proceedings. Los Alamitos: IEEE, 2019. – P. 138–141. DOI:10.1109/T4E.2019.00033.
13. Li L. A Hybrid approach for ranking research articles / [L. Li, Y. Zhang, S. Zhang et al.] // Information Processing & Management. – 2017. – Vol. 53, № 6. – P. 1407–1419.
14. Meng Z. Personalized ranking framework via multilayer perceptron learning / [Z. Meng, J. Lian, Y. Lan et al] // Transactions on Knowledge and Data Engineering. – 2017. – Vol. 29, № 7. – P. 1509–1521.
15. Neural ranking models with weak supervision / [S. Shen, Z. Liu, M. Sun et al.] // Research and Development in Information Retrieval : 42nd international ACM SIGIR conference, 2019. – P. 111–120. DOI:10.48550/arXiv.1704.08803
16. Makarov M. BERT for ad-hoc retrieval: baselines and analysis / M. Makarov, E. Tutubalina, P. Braslavski //Advances in Information Retrieval : 42nd European conference on IR Research, ECIR, 2020. – P. 516–530.

DOI 10.15588/1607-3274-2023-3-17

17. A Reinforcement learning approach for rank aggregation in information retrieval / [Y. Zhang, M. Zhang, Y. Liu et al.] // IEEE Access. – 2021. – Vol. 9. – P. 316–327.
18. Mohammed S. A Comprehensive overview of recommender system and sentiment analysis [Electronic resource] / S. Mohammed AL-Ghuribi Shahrul, A. Mohd Noah. – Access mode: https://arxiv.org/ftp/arxiv/papers/2109/2109.08794.pdf
19. Brave S. Method and apparatus for identifying, extracting, capturing, and leveraging expertise and knowledge. [Electronic resource] / S. Brave, R. Bradshaw, J. Jia, C. Minson. – Access mode: https://patents.google.com/patent/US7698270
20. Britsov R. Ranking of information based on user ratings and behavior / R. Britsov // T-Comm, Telecommunications and Transport. – 2016. – Vol. 10, № 1. – P. 62.
21. Christopher D. Manning introduction to information retrieval / D. Christopher. – Cambridge University Press, 2008. – 520 p. https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf
22. Velychko O. M. Hardware-software complexes for evaluation of competence level of experts / O. M. Velychko, T. B. Gordiyenko, L. V. Kolomiets // Metallurgical and Mining Industry. – 2015. – Issue 7. – P. 396–401.
23. Stoica P. Model-order selection: a review of information criterion rules / P. Stoica, Y. Selen // IEEE Signal Processing Mag. – 2004. – Vol. 21, № 4. – P. 36–47. https://doi.org/10.1109/MSP.2004.1311138.
24. Bulgakova O. Numerical study of the generalized iterative algorithm GIA GMDH with active neurons / Bulgakova O., Stepashko V., Zosimov V. // Computer Sciences and Information Technologies, CSIT : 12th international scientific and technical conference : proceeding. – IEEE, 2017. –P. 496–500. DOI:10.1109/STC-CSIT.2017.8098836.
25. Zosimov V. Semantic Profile of corporate web resources / V. Zosimov, O. Bulgakova, V. Pozdeev // CEUR Workshop Proceedings, 2021. – Vol. 3179. – P. 389–397. URL: https://ceur-ws.org/Vol-3179/Short_13.pdf
26. Vink P. Advances in social and organizational factors / P. Vink. – T.12 : Advances in human factors and ergonomics. – AHFE, 2014. – 598 p.
27. Bicchieri C. The grammar of society: the nature and dynamics of social norms / C. Bicchieri. – Cambridge University Press, 2005. – 306 p.
28. Zosimov V. Inductive building of search results ranking models to enhance the relevance of text information retrieval / V. Zosimov, V. Stepashko, O. Bulgakova // Database and expert systems applications. – Los Alamitos: IEEE Computer Society, 2015. – P. 291–295.
29. Zosimov V. Calculation the measure of expert opinions consistency based on social profile using inductive algorithms / V. Zosimov, O. Bulgakova // Lecture notes in computational intelligence and decision making. – 2019. – Vol. 1020. – P. 622–636. DOI:10.1007/978-3-030-26474-1_43

DOI 10.15588/1607-3274-2023-3-17

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

UDC 004.93

IMPROVED MULTI-OBJECTIVE OPTIMIZATION IN BUSINESS PROCESS MANAGEMENT USING R-NSGA-II

Filatov V. O. – Dr. Sc., Professor, Head of the Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Yerokhin M. A. – Post-graduate student of the Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

ABSTRACT

Context. Business process management is a critical component in contemporary organizations for maintaining efficiency and achieving operational objectives. Optimization of these processes in terms of time and cost can lead to significant improvements in overall business performance. However, traditional optimization techniques often face challenges in handling multi-objective problems with a known time-cost trade-off, necessitating more effective solutions. The integration of a business process model and notation for a stochastic process simulation provides a robust foundation for analyzing these business processes and complies with state-of-the-art business process management. In prior studies, we applied several heuristic algorithms, including the evolutionary NSGA-II, to find a Pareto-optimal set of solutions. We defined a solution as a pair of cost and time associated with a specific resource allocation. For one of the selected processes, the performance of NSGA-II was subpar compared to other techniques.

Objective. The goal of this study is to improve upon the NSGA-II's performance and, in turn, enhance the efficiency of multi-objective business process optimization. Specifically, we aim to incorporate reference points into NSGA-II. Our goal is to identify an optimized set of solutions that represent a trade-off between process execution time and the associated cost. We expect this set to have a higher spread and other quality metrics, compared to the prior outputs.

Method. To accomplish our objective, we adopted a two-step approach. Firstly, we modified the original genetic algorithm by selecting and integrating the reference points that served to guide the search towards the Pareto-optimal front. This integration was designed to enhance the exploration and exploitation capabilities of the algorithm. Secondly, we employed the improved algorithm, namely R-NSGA-II, in the stochastic simulations of the business processes. The BPMN provided the input for these simulations, wherein we altered the resource allocation to observe the impact on process time and cost.

Results. Our experimental results demonstrated that the R-NSGA-II significantly outperformed the original NSGA-II algorithm for the given process model, derived from the event log. The modified algorithm was able to identify a wider and more diverse Pareto-optimal front, thus providing a more comprehensive set of optimal solutions concerning cost and time.

Conclusions. The study confirmed and underscored the potential of integrating the reference points into NSGA-II for optimizing business processes. The improved performance of R-NSGA-II, evident from the better Pareto-optimal front it identified, highlights its efficacy in multi-objective optimization problems, as well as the simplicity of the reference-based approaches in the scope of BPM. Our research poses the direction for the further exploration of the heuristics to improve the outcomes of the optimization techniques or their execution performance.

KEYWORDS: business process, genetic algorithm, reference points, multi-objective optimization, spacing.

ABBREVIATIONS

BPM is a business process management;
BPMN is a business process model and notation;
BPS is a business process simulation;
MOEA is a multi-objective evolutionary algorithm;
MOOP is a multi-objective optimization problem;
MORAP is a multi-objective resource allocation problem;
NSGA-II is a non-sorting genetic algorithm;
R-NSGA-II is a reference non-sorting genetic algorithm;
CT is the cycle time;
TC is the total cost;
TCT is the theoretical cycle time;
OMG is Object Management Group;
HA is hyperarea;
SP is spacing;
IGD is inverted generational distance.

NOMENCLATURE

M is the number of the conflicting objective functions;
R^M is the objective function space;
$f()$ is a process simulation function;
S is a constrained search space;
a is a lower bound for the allocation;
b is an upper bound for the allocation;
x^* is a Pareto-optimal solution;

DOI 10.15588/1607-3274-2023-3-18

P_{ref} is a Pareto front containing Pareto-dominating points from the different applied algorithms;
P_{ref}' is P_{ref} including R-NSGA-II;
x is an allocation parameter used in Tabu Search;
x' is an allocation parameter used in R-NSGA-II;
H is a set of heuristics used;
C_i is the hourly cost for the resource;
t_i is the number of hours spent by the resource;
$p()$ is a purity metric function;
p_i is the probability of the outgoing sequence flow;
T_i is the average task execution time;
p_{rew} is the probability of necessity to rework the task;
r_i is the resource pool;
$cTime$ is an average CT of the process;
N is the number of resources assigned;
TC_r is the total cost for the resource pool;
R_p is a sequence of resource pools;
$rtCost$ is the function that retrieves TC for r_i;
$\in$ is a threshold parameter;
RP_c is a cost reference point;
RP_t is a time reference point;
RP_h is a reference point to optimize HA;
P_{approx} is a Pareto front obtained by a specific algorithm;
μ is the population size;
λ is the size of the offspring;
rp is the number of reference points.

INTRODUCTION

Resource optimization in business processes is a critical aspect of the organizational efficacy and profitability of enterprises. The process of resource management and assigning resources to the work roles, as well as simulating their behaviour in a business process has been widely addressed using various Petri net models [1, 2]. Although the Petri nets are generally compliant and transformable with the newer industry standard of BPMN [3], they possess a small toolset for modelling complex business processes that involve multiple objectives and high-level business concepts [4]. The descriptive possibilities of the core Petri nets and their' extensions (e.g. stochastic Petri nets or differential Petri nets) are of limited applicability to the modern well-defined business flows, and not coupled to the industry processes requirements. It leads to the necessity for the specific enterprise to develop a model with a custom architecture, which might require manual changes to be compliant with other models.

From a practical standpoint, the classical Petri nets typically aimed for a precise simulation of the basic mechanisms present in the system and clearly communicating its state. If the system becomes more complicated, considering privacy requirements, unequal resources, and other constraints, the specification complexity of this Petri net increases. Since the Places, Transitions, and Tokens are not self-descriptive, the built model has to carry all the rules and assumptions for each element. While not all of these constraints might be necessary for the simulation, they are required to explicitly define an AS-IS process and make it possible to interpret the results by management representatives. Moreover, BPMN provides a number of time-related primitives out-of-the-box, which are often critical to describe the simulation scenarios, whereas the notion of time is not natively defined in classical Petri nets [5].

There are a number of extended notations of Petri nets, such as stochastic, hierarchical, and differential Petri nets, that were attempting to address multi-level processes, multi-agent systems, stochastic transitions, and others. The diversity of non-standardized solutions could be a ground that led the industry to create a conventional standard, which includes the necessary primitives to exhaustively describe the business process and its scenarios, without the necessity to introduce the common definitions for each model.

BPMN 2.0 has become a standard, which expresses the control logic such as choice, sequences, parallel execution, and iteration and introduces the respective common concepts such as Task, Event, and Resource Pool instead of operating abstract places, transitions, and tokens. The more advanced elements such as timer event, lane, or data object allow to cover typical use cases with less architectural effort. BPMN has been approved by OMG, it describes and formalizes the manifold of use cases of the enterprise systems. Nevertheless, BPMN is two-way compatible with the majority kinds of Petri nets and other industrial frameworks, which are used by different simulation engines by building the mapping between BPMN and Petri nets [6, 7].

One of the classical problems of enterprise management concerns choosing the number of people or, more generally, resources to assign to a specific unit of work. Increasing the number of resources typically leads to lower execution times of the work units, which implies a faster business process cycle. In its turn, the amount of money (cost) required to spend on these resources also increases. The improvement of the cost or time efficacy of the business process is usually achieved by qualitative and quantitative analysis. In this research we prioritize the quantitative analysis, employing and assessing various optimization methods in the scope of business process modelling. However, enterprises usually do not seek to minimize only the cost of their business process or to make execution time the lowest possible. Instead, both cost and time are objectives, and the goal is to find a suitable trade-off from a set of existing allocations. This is a so-called MORAP.

In [8] the solution was developed to employ three algorithms for finding the Pareto front containing resource allocations for the given BPMN, so that there is not any single allocation in the set with a better cost and time simultaneously, compared to another allocation from the set. The Pareto front represents the trade-offs between multiple objectives and allows one to select a suitable solution depending on the management priority. The performance metrics of the algorithms were compared across multiple business processes. In particular, the output Pareto front for the business process of the call center, acquired by NSGA-II, diverged significantly with the

DOI 10.15588/1607-3274-2023-3-18

reference front accumulated from all of the three approaches.

In this paper, we plan to set up a more advanced variant of the NSGA-II evolutionary algorithm, namely R-NSGA-II, to assess the possibility to narrow the gap between accumulated outcomes of the regression descent algorithms and NSGA-II. We will select the reference points that fit the common sense bounds of the selected process, run the adjusted algorithm, and record the metrics of the updated Pareto front.

The object of study is the Pareto fronts of the dominating resource allocations in the business processes, their quality metrics, and the input data features.

The subject of study is the methods for identifying the Pareto-optimal set of resource allocations in a business process derived from the call center event log.

The purpose of the work is to improve the metrics of the outputting Pareto front for the selected BPMN model, by adding the reference points to the NSGA-II experimental run.

1 PROBLEM STATEMENT

Suppose given a classical MOOP, to find a single solution we can use the formalism as follows:

$$\begin{aligned} &minimize\ f(x) = (f_1(x), ..., f_m(x))^T, \\ &x \in S \subseteq R^M, \\ &m \in \{1, 2, ..., M\}, \end{aligned} \tag{1}$$

where f: $S \to R^M$. S is limited by the boundaries:

$$\begin{aligned} &x = (x_1, ..., x_M), \\ &a_i \le x_i \le b_i, \end{aligned} \tag{2}$$

considering the i-th resource. We say that the allocation x_{i_1} dominates x_{i_2} only if $f_i(x_{i_1}) \le f_i(x_{i_2})$ for all $i \in \{1, ..., M\}$ and $f_i(x_{i_1}) < f_i(x_{i_2})$ for at least one sequence entry, therefore we denote $x_{i_1} \prec x_{i_2}$. x^* encompasses a Pareto-optimal solution if $\nexists x \in S$ such that $x \prec x^*$. The aggregated set of solutions using H is a Pareto front

$$P_{ref} = \{f(x_1^*), ..., f(x_M^*)\} \in S \tag{3}$$

that contains global dominating points.

Then the overall goal of MOOP in this research is finding a set of well-spread non-dominated solutions using H_j wherein $p(P_{ref}') \to p(P_{ref})$.

2 REVIEW OF THE LITERATURE

In this section, we start with the state-of-the-art research and applications for MOOP and MORAP related to the BPM and proceed with the guided search methods.

In [9], the authors investigate the parcel delivery being procured by Employees and Drones. The resource-to-process assignment is assessed in two ways. Firstly, they selected a static allocation for both resources and measured how do those resources handle the variable workload, considering the execution time and the resource utilization. Secondly, they fixated on the number of work units and tracked how resource utilization and time are affected depending on the lower or higher number of drones and employees involved. The authors explored all possible allocations since the possible pool sizes were limited; the search space consisted of less than a hundred variants. This approach is not well-applicable to the larger search spaces and also does not consider the stochastic nature of the real-life business process.

Other than resource utilization, the measure of cost is typical and natural for human-related tasks. In [10] authors proposed a modification of the ant-colony algorithm for finding the global optimum of the multi-objective function. In their experimental setup, the global Pareto-optimal solutions are stored in the form of the Pareto front. The algorithm is aimed at the resource allocation problem but is also applicable to grouping and scheduling problems. The advantage of the algorithm is the ability to continue the exploration of the richer areas after reaching the local optimum, which reminds the Tabu Search in this regard. In the experiment, authors used cost and profit as the optimization objectives, but it is also possible to work with higher dimensionality. Authors claim a better computational time in comparison with genetic algorithms. On the other hand, the authors do not assess the quality of the resulting Pareto front.

Genetic algorithms have become a baseline to approach multi-objective problems. Specifically, NSGA-II is among the most popular and widely used algorithms because of the simplicity of tuning. In [11] authors developed the hybrid algorithm to minimize the number of simulations for multi-objective optimization. The paper describes the custom algorithm based on the evolutionary approach mixed with the predictions model. The authors used the hypervolume metric to compare and assess the Pareto front quality. The results are compared with the conventional genetic algorithms such as NSGA-2 and SPEA2 and claimed to be more efficient in the simulations usage.

A common drawback of the evolutionary search algorithms is a fixed sample size. While it is a convenient parameter to specify the degree of reduction of the search space, it may imply some uneven distribution of the solutions [12]. As a result, we might see some poorly explored areas, some parts of the Pareto set being distant from the reference figure or not present at all. In [12-15] the different methods proposed to guide the evolutionary algorithms towards the areas of interest. They do not focus on the BPM field, however, demonstrate a clear improvement in the convergence, diversity, and quality of the outputting Pareto front. In our paper, we decided to follow the approach described in [16], since it proved to work well on the two-objective problem, allows us to specify

DOI 10.15588/1607-3274-2023-3-18

more than one reference point, and is tuned to expand toward particular areas of interest.

Finally, we refer to [8] as a baseline for our experiment. The research introduces a comprehensive setup with the multiple BPMN models, concerns the different facets of MORAP in BPM, such as building the simulation model out of the event log, the accuracy of the simulation model, and proposes metrics to compare the outputting Pareto front from the applied algorithms. The authors selected the well-known Hill Climbing, Tabu Search, and NSGA-II to run. For one of the case studies, namely a business process of a call center, there is a gap that we try to cover in this paper.

3 MATERIALS AND METHODS

To estimate the business process performance, there are two traditional measures: average cycle time and cost of the process execution. In our case study, we assume that the cost is a static measure meaning it has a predefined value counted for a unit of time for each resource, therefore the total cost value is described as:

$$TC = \sum_{i=1}^{n} C_i t_i . \quad (4)$$

Calculating the cycle time for a flat single-lane business process is straightforward: we calculate the average execution time for each task and sum it up. However, there are additional constraints that add up to additional calculations:

– Alternative paths with the probability of choosing one sequence flow over another; in this case cycle time $\mathrm{CT}_{\mathrm{alternative}} = \sum_{i=1}^{n} p_i T_i$, considering the tasks and their respective sequence flows;

– Parallel $\mathrm{CT}_{\mathrm{parallel}} = \mathrm{Max}\{T_1, T_2 \ldots T_m\}$;

– Rework $\mathrm{CT}_{\mathrm{rework}} = \mathrm{T}/(1 - \mathrm{p}_{\mathrm{rew}})$.

Combining these three rules, we can calculate TCT, although it does not include waiting time, handover time, or other non-value-adding activities [17]. Queueing theory can be used to address those real-life parameters, however, it has its own drawbacks. To calculate factual CT, there is a BPS approach, which allows the tuning of the resources' waiting and handover times to calculate factual CT in a versatile way, given that the BPMN model can undergo frequent changes. Due to the high level of output noise, it is common to run multiple simulations, hence in our research we repeat the simulation 15 times for each resource allocation and assume *cTime* is an average CT of the process.

The accuracy of a BPS, and hence the usefulness of the outcomes, to a large extent relies on how accurately the process model and simulation parameters capture the observed reality. In general, process models are manually designed by enterprise analysts for the sake of management convenience. Typically, process models do not capture all the details and mechanisms of how the process is actually carried out. If there is a significant variation in service times the actual cycle time of the business process can diverge significantly from the predicted BPS metrics using flow analysis. The simulation parameters for BPS are commonly estimated based on the process manager's expertise and manual fitting, which does not always comply with the real-life process execution [18].

As an input for the simulation, we denote a resource allocation as a sequence of resource pools in the business process $R_p = \langle r_1, \ldots, r_n \rangle$, each corresponding to a subset of tasks in a business process. The function *rtCost:* $R_p \rightarrow TC_{r_i}$ retrieves the total cost of the selected resource pool r_i, $r_i \in \{N, T, Ci\}$.

R-NSGA-II is a multi-objective optimization algorithm. It is an improved version of the original NSGA-II algorithm, which aims to solve multi-objective optimization problems. While both of them keep the population of the Pareto-optimal points, R-NSGA-2 incorporates a reference point-based approach for selecting the individuals from the offspring. By using the reference points, R-NSGA-2 can effectively explore diverse and evenly-distributed solutions along the Pareto front, allowing decision-makers to make informed choices when dealing with multiple conflicting objectives.

The algorithm ranks the current Pareto-optimal points by Euclidean distance to each reference point in ascending order. The solution closest to the reference point obtains the rank of one. The next step ranks the solutions by crowding distance, meaning the solution should be closest to a set of reference points. Then the solutions are grouped by the sum of normalized distances between them using $\in$ threshold parameter. The farther groups are discouraged from being promoted to the next generation. The higher value of $\in$ increases the range of explored solutions.

By employing the aforementioned selection algorithm, it becomes possible to allocate equal attention to solutions that are in close proximity to each reference point. This enables the identification of multiple regions of interest concurrently.

4 EXPERIMENTS

In this research, we outline the three stages of multi-objective business process optimization: Process Discovery, Optimization, and Evaluation, which are denoted in Figure 1 in BPMN format. We aim to improve and focus on the latter two stages.

In the Process Discovery stage, we obtain the XES event log as input for the business process. This is a common data source for existing enterprises since building the BPMN model requires certain expertise and effort from the management resources. The process mining technique is used to build the BPMN simulation model. Specifically, we use the Simod tool to obtain the BPMN model approximation with the necessary simulation parameters described in [19]. These parameters include the initial resource allocation: resource pools, the number of resources for each of the resource pools, resource cost per

DOI 10.15588/1607-3274-2023-3-18

hour, resource pool linked to each of the BPMN tasks, and the distribution of the processing time for each task.

In the Optimization stage, we try to change the initial resource allocation, which means overriding the number of resources in the BPMN pool of the simulation model. To compare two resource allocations, we run the simulation using the BIMP tool. For a given resource allocation, it produces an output in CSV format. Among all of the simulation outcomes, we are interested in the average cycle time *cTime* of the simulation and the TC. In general, better allocation means both lower cycle time and lower total cost, making it Pareto-dominating. However, if only one parameter is better in one allocation than another, we cannot prefer a single solution. We have to keep a set of non-dominating solutions as a P_{ref}'.

Due to the computationally intensive nature of simulating the BPMN model, employing a brute-force approach to explore the vast search space of all potential resource allocations is not feasible. Therefore, we adopt established strategies to navigate the search space of solutions and iterate through various allocation possibilities.

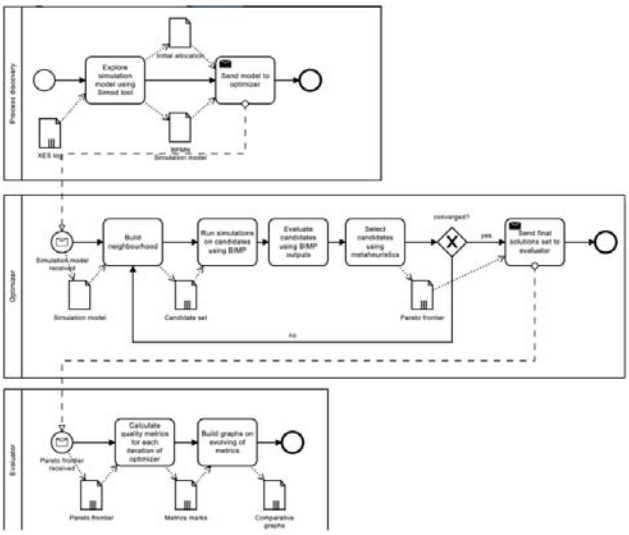


Figure 1 – The experiment outline in a BPMN format

During the Evaluation stage, we compare the quality of the output Pareto fronts obtained by different algorithms and assess the selected quality metrics proposed in [20]. To apply the reference point-based approach, we extended the experimental setup introduced in [8] using the R-NSGA-II from the Pymoo toolset [21]. During the experiment run, we want to measure and compare the efficacy of R-NSGA-II with a conventional regression descent algorithm – Tabu Search. In the prior study, Tabu Search outperformed the original NSGA-II evolutionary algorithm on one of the business processes describing a call center enterprise.

The resulting Pareto front significantly diverged in all of the selected quality metrics including the Hyperarea Ratio, IGD, Purity, and Spacing.

The research question to evaluate is as follows: how good is the Pareto front obtained by R-NSGA-II compared with the regression descent with respect to convergence, spread, and distribution, and does it perform better than pure NSGA-II?

Each experimental run constructs two outputs:

– P_{ref}' contains the Pareto front points from all of the applied approaches, in our case study it consists of the points, independently acquired by R-NSGA-II and Tabu Search. Namely, P_{ref}' represents the cross-dominating solutions from both of the algorithms;

– P_{approx} contains the Pareto front points obtained by a specific algorithm.

To set up R-NSGA-II, we selected three reference points, each one defining the desired direction of expansion:

– RP_c has to attract the Pareto front towards the *Resource Allocation Cost* axis and force the algorithm to explore more solutions with smaller *Cycle Time*. Therefore, we try to improve spread by *X*-axis;

– RP_t has to attract the Pareto front towards to the *Cycle Time* and force the algorithm to explore more solutions with smaller *Resource Allocation Cost*. Therefore, we try to improve spread by *Y* axis;

– RP_h aims to extend the Pareto front towards the corner, therefore improving the HA of the resulting Pareto front P_{ref}. This point stands for the wittingly unreachable solution, with both low cost and time.

According to the guidance in [16, 22], we selected the following parameters to reach the balanced and feasible advancement of a genetic algorithm, considering the modification with the reference points approach:

– The population size μ is set to 40;

– The size of the offspring λ is set to 20;

– The number of reference points rp is set to 3, and they are all infeasible as defined in [16];

– The $\in$ threshold for the sum of the normalized distances is set to 0.001.

In our experiment, we extend the quality evaluation with two metrics commonly used in MOOP:

– SP is a straightforward measure to assess the spread and distribution. Despite its known issue to process Pareto fronts with clearly distinct groups of points, this downside is not applicable to our output. It is calculated as follows:

$$\mathrm{SP(S)} = \sqrt{\frac{1}{|S|-1}\sum_{i=1}^{|S|}(\bar{d}-d_i)^2}, \qquad (5)$$

where $d_i = \min(s_i, s_j) \in S, s_i \neq s_j \, \| F(s_i) - F(s_j) \|_1$.

Higher value stands for better spread and diversity, in case Pareto fronts are similarly dispersed.

– IGD is a classical convergence metric, which ranks one Pareto front better than another if and only if the given Pareto front is always preferred according to the Pareto optimality rules:

$$\mathrm{IGD(S,P)} = \frac{1}{|P|}\left(\sum_{i=1}^{|P|} d_i^M\right)^{\frac{1}{M}}, \qquad (6)$$

where $d_i = \min x \in S \, \| F(x) - F(i) \|$. Lower value stands for better P_{approx}.

DOI 10.15588/1607-3274-2023-3-18

5 RESULTS

Figure 2 comprises the Pareto front P_{approx} discovered by R-NSGA-II and P_{ref}' containing all Pareto-optimal solutions from all selected algorithms. The filled markers in black stand for the solutions in both P_{approx} and P_{ref}', meaning these Pareto-optimal points were successfully identified by the algorithms. The hollow markers in blue designate the solutions in P_{ref}' but not in P_{approx}, which implies the points from the reference Pareto set were not identified by the current algorithm. The ones in red are the points in P_{approx} but not in P_{ref}', they were selected by the current algorithm, however, the more effective Pareto-dominating solution exists in the reference set. Nevertheless, the solutions marked with red might be useful in terms of improving the spread and diversity. Figure 3 denotes the same output for the previously winning TS algorithm. We can observe that R-NSGA-II found 29 of the 45 non-dominated points in P_{ref}', while TS found a different set containing only 23 of the points in P_{ref}'. This is a good entry indicator of the improved performance of R-NSGA-II.

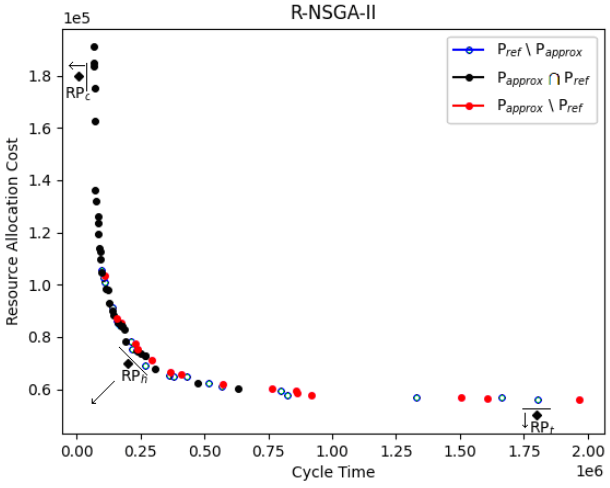


Figure 2 – Pareto front for the R-NSGA-II

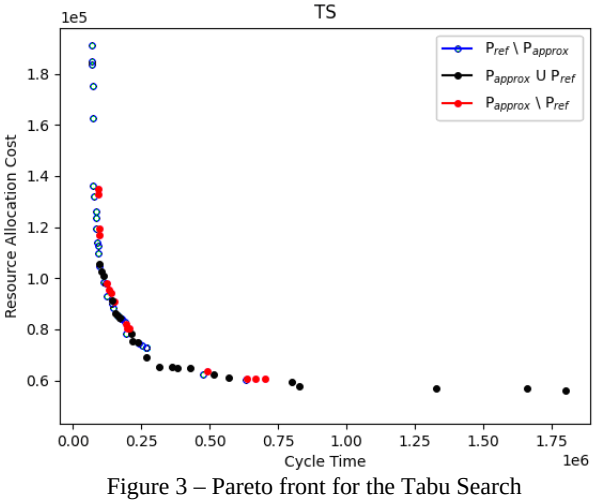


Figure 3 – Pareto front for the Tabu Search

Table 1 compares the quality metrics of the obtained Pareto fronts, and also the metrics for previously used pure NSGA-II. The values in bold designate the best metric across three algorithms. As we can see, adding the reference points significantly improved original NSGA-II efficacy, and overall demonstrated R-NSGA-II more performant than TS in our case study. The HA Ratio and Purity metrics in R-NSGA-II have overcome other algorithms, although the advantage is not very noticeable. The IGD dominance of TS over R-NSGA-II can be explained by the former visually more spread by X-axis and the latter more spread by Y-axis, while the X-axis possesses a higher order of scale. Although TS kept its performance dominance, the reference approach essentially improved this metric of NSGA-II. Regarding the spread and distribution metrics, namely Spacing and Delta, the results are arguable. Although Delta is generally considered a more sustainable and future-proof metric for Pareto fronts, the Spacing metric still might be more applicable in our research since the reference front follows the Gaussian distribution. This means that higher Spacing covers the solutions closer to the extreme points.

Table 1 – Comparative quality metrics of the selected algorithms

	TabuSearch	NSGA-II	R-NSGA-II
HA Ratio	0.998923	0.975051	**0.999989**
IGD	**16820.4**	426892.2	32366.9
Purity	0.60	0.0625	**0.65**
Spacing	88597.2	80713.4	**91556.6**
Delta	**1.07**	1.17	1.14

In general, we can observe that adding the reference points to the original genetic algorithm can significantly improve the quality of the output Pareto front, and also outperform the competitor such as Tabu Search in a number of metrics.

6 DISCUSSION

This paper presented an approach to involve a genetic algorithm for computing a set of Pareto-optimal resource allocations for a given business process. In particular, a prior case study evaluated the performance of the NSGA-II algorithm on a given set of business processes. The overall experimental setup remained unchanged and is based on the simulation model to evaluate the noisy value of the objective function. However, we selected a specific business process of a call center to optimize, since the genetic algorithm performed the worst in that example. The output significantly diverged from the Tabu Search output and yielded a less spread Pareto front with a significantly higher HA. We employed a more supervised variant of the algorithm, namely R-NSGA-II, based on the chosen reference points. They provide a clue for a regression run and can improve the convergence and exploration of the new solutions for the Pareto front. The evaluation found that providing three reference points made the Pareto front significantly closer to the one from the Tabu Search, but also explores more of a search field. Overall, providing the reference points can improve the output of NSGA-II, however, requires prior knowledge of the

DOI 10.15588/1607-3274-2023-3-18

Pareto front approximation. The further work direction includes comparing other multi-objective algorithms with conventional regression. We also attempt to revisit the actuality of the Spacing metric to assess the spread and the diversity of the Pareto fronts.

CONCLUSIONS

The resource allocation problem is common in BPM. Although the narrower field of MORAP has a series of research solutions, there are still some methods that are not well-represented in the scope of BPM. If we look at the specific case studies, there arises an even broader spectre of research questions.

The scientific novelty of the obtained results is that the method of populating the MOEA with the reference points in the scope of the resource allocation for BPM has been proposed. It characterizes the areas of interest for the management purpose, in a scenario when another evolutionary method did not output the adequate and desired set of the time-cost trade-offs.

The practical significance of the obtained results is that the applied reference points approach has improved the existing metrics in a specific scenario. While the experiment shows the potential of reference-based add-ons to explore previously unsearched areas of interest, it also asserts the extensibility of the existing framework to work with the different MOOP methods and BPMN derived from various sources.

Prospects for further research are to apply more algorithms to the MORAP. Considering the evolutionary approach, it is possible to extend the experiment with SPEA2; also neural networks are a more profound way to approximate the optimal Pareto front. Since the simulation model captures the stochastic nature of the simulated processes, it enables us to estimate if Bayesian optimization is an applicable strategy.

REFERENCES

1. Qin J., Zhao N., Xie Z. et al. Business Process Modelling based on Petri nets, MATEC Web of Conference, 2017, Vol. 139, № 1, pp. 105–113.
2. Shoylekova K., Grigorova K. Methods for Business Process Simulation Based on Petri Nets, *International Journal of Industrial and Systems Engineering*, 2015, Vol. 9, Issue 12, pp. 4148–4153.
3. Umair M. M., Barkaoui K., Li Z. et al. Transformation of Business Process Model and Notation models onto Petri nets and their analysis, *Advances in Mechanical Engineering*, 2018, Vol. 10, № 12, pp. 168–188.
4. De Backer M., Snoeck M. Deterministic Petri Net Languages as Business Process Specification Language [Electronic resource], *SSRN Electronic Journal*, 2006. Access mode: http://dx.doi.org/10.2139/ssrn.876906
5. Zeigler B. Praehofer H., Gon Kim T. Theory of Modeling and Simulation. Cambridge, Academic Press, 2018, Discrete Event & Iterative System Computational Foundations, pp. 567–599.
6. Raedts I., Petkovic M., Usenko Y. et al. Transformation of BPMN Models for Behaviour Analysis, *5th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems, MSVVEIS-2007*, Funchal, June 2007, proceedings, 2007, pp. 126–137.
7. Li Z., Ye Z. A Petri Nets Evolution Method that Supports BPMN Model Changes, *Scientific Programming*, 2021, Vol. 2021, Issue 3, pp. 1–16.
8. Pintado O. L., Dumas M., Yerokhin M. et al. Silhouetting the Cost-Time Front: Multi-objective Resource Optimization in Business Processes, *Business Process Management Forum, Rome, 6–10 September 2021: proceedings*. Berlin, Springer, 2021, pp. 92–108.
9. Durán F., Rocha C., Salaün G. Analysis of Resource Allocation of BPMN Processes, *17th International Conference on Service-Oriented Computing, Toulouse, 28–31 October 2019, proceedings*. Berlin, Springer, 2019, pp. 452–457.
10. Rizk-Allah R. M. A Novel Multi-Ant Colony Optimization For Multi-Objective Resource Allocation Problems, *International Journal of Mathematical Archive*, 2014, Vol. 5, № 9, pp. 183–192.
11. Couckuyt I., Deschrijver D., Dhaene T. Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization, *Journal of Global Optimization*, 2013, Vol. 60, №3, pp. 575–594.
12. Cheng R., Jin Y., Olhofer M. et al. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. IEEE transactions on neural networks, *IEEE Transactions on Evolutionary Computation*, 2016, Vol. 20, № 5, pp. 773–791.
13. Tong L., Du B. Neural architecture search via reference point based multi-objective evolutionary algorithm, *Pattern Recognition*, 2022, Vol. 132, № 11.
14. Lárraga G., Saini B. S., Miettinen K. Incorporating Preference Information Interactively in NSGA-III by the Adaptation of Reference Vectors, *Evolutionary Multi-Criterion Optimization, Hannover, 18–23 September 2023, proceedings*. Berlin, Springer, 2023, pp. 578–592.
15. Vargas D., Lemonge A. C. C., Barbosa H. et al. Solving multi-objective structural optimization problems using GDE3 and NSGA-II with reference points, *Engineering Structures*, 2021, Vol. 239, № 3.
16. Deb K., Jayavelmurugan S. Reference Point Based Multi-Objective Optimization Using Evolutionary Algorithms, *International Journal of Computational Intelligence Research*, 2006, Vol. 2, № 3, pp. 635–642.
17. Dumas M., La Rosa M., Mendling J. et al. Fundamentals of Business Process Management. Berlin, Springer, 2018, 527 p.
18. Estrada-Torres B., Camargo M., Dumas M. et al. Discovering business process simulation models in the presence of multitasking and availability constraints, *Data & Knowledge Engineering*, 2021, Vol. 134, № 1.
19. Camargo M., Dumas M., González-Rojas O. Automated discovery of business process simulation models from event logs, *Decision Support Systems*, 2020, Vol. 134, № 1.
20. Audet C., Bigeon J., Cartier D. Et al.Performance indicators in multiobjective optimization, *European Journal of Operational Research*, 2021, Vol. 292, № 2, pp. 397–422.
21. Blank J., Deb K. Pymoo: Multi-Objective Optimization in Python, *IEEE Access*, 2020, Vol. 8, № 1, pp. 89497–89509.
22. Tanabe R., Oyama A. The Impact of Population Size, Number of Children, and Number of Reference Points on the Performance of NSGA-III, *International Conference on Evolutionary Multi-Criterion Optimization, Münster, 19–22 March 2017: proceedings*. Berlin, Springer, 2017, pp. 606–621.

Received 27.06.2023.
Accepted 29.08.2023.

DOI 10.15588/1607-3274-2023-3-18

УДК 004.93

ПОКРАЩЕНА БАГАТОЦІЛЕВА ОПТИМІЗАЦІЯ В УПРАВЛІННІ БІЗНЕС-ПРОЦЕСАМИ ЗА ДОПОМОГОЮ R-NSGA-II

Філатов В. О. – д-р техн. наук, професор, завідувач кафедри штучного інтелекту Харківського національного університету радіоелектроніки, Харків, Україна.

Єрохін М. А. – аспірант кафедри штучного інтелекту Харківського національного університету радіоелектроніки, Харків, Україна.

АНОТАЦІЯ

Актуальність. Управління бізнес-процесами є критично важливим компонентом у сучасних організаціях для підтримки ефективності та досягнення операційних цілей. Оптимізація цих процесів з точки зору часу та витрат може призвести до значного покращення загальної ефективності бізнесу. Однак традиційні методи оптимізації часто стикаються з труднощами при вирішенні багатоцільової проблеми з відомим компромісом часу та вартості, що вимагає більш ефективних рішень. Використання моделі та нотації бізнес-процесів (BPMN) для стохастичного моделювання процесу забезпечує надійну основу для аналізу цих бізнес-процесів і відповідає найсучаснішому управлінню бізнес-процесами. У попередніх дослідженнях ми застосували кілька евристичних алгоритмів, включаючи еволюційний NSGA-II, щоб знайти оптимальний за Парето набір рішень. Ми визначили рішення як пару витрат і часу, пов'язаних із певним розподілом ресурсів. Для одного з вибраних процесів продуктивність NSGA-II була нижчою порівняно з іншими методами.

Мета роботи – покращення продуктивності NSGA-II і, у свою чергу, підвищення ефективності багатоцільової оптимізації бізнес-процесів. Зокрема, ми прагнемо включити контрольні точки в NSGA-II. Наша мета полягає в тому, щоб визначити оптимізований набір рішень, що представляє компроміс між часом виконання процесу та пов'язаними витратами. Ми очікуємо, що цей набір матиме вищий розкид та інші показники якості порівняно з попередніми результатами.

Метод. Щоб досягти нашої мети, ми застосували двоетапний підхід. По-перше, ми модифікували оригінальний генетичний алгоритм, вибравши та інтегрувавши опорні точки, які слугували для спрямування пошуку до оптимального за Парето фронту. Ця інтеграція була розроблена для покращення можливостей алгоритму для дослідження та використання. По-друге, ми застосували покращений алгоритм, а саме R-NSGA-II, для стохастичного моделювання бізнес-процесів. BPMN надав вхідні дані для цього моделювання, у якому ми змінили розподіл ресурсів, щоб спостерігати вплив на час і вартість процесу.

Результати. Наші експериментальні результати продемонстрували, що R-NSGA-II значно перевершив вихідний алгоритм NSGA-II для даної моделі процесу, отриманої з журналу подій. Модифікований алгоритм зміг ідентифікувати ширшу і більш різноманітну оптимальну за критерієм Парето криву, таким чином забезпечивши більш повний набір оптимальних рішень щодо вартості та часу.

Висновки. Дослідження підтвердило та підкреслило потенціал інтеграції опорних точок у NSGA-II для оптимізації бізнес-процесів. Покращена продуктивність R-NSGA-II, очевидна з кращою Парето-оптимальної кривої, яку алгоритм ідентифікував, підкреслює його ефективність у задачах багатоцільової оптимізації, а також простоту еталонних підходів у сфері BPM. Наше дослідження визначає напрямок для подальшого вивчення евристик для покращення результатів методів оптимізації або продуктивності їх виконання.

КЛЮЧОВІ СЛОВА: бізнес-процес, генетичний алгоритм, опорні точки, багатокритеріальна оптимізація, спейсінг.

ЛІТЕРАТУРА

1. Business Process Modelling based on Petri nets / [J. Qin, N. Zhao, Z. Xie et al.] // MATEC Web of Conference. – 2017. – Vol. 139, № 1. – P. 105–113.
2. Shoylekova K. Methods for Business Process Simulation Based on Petri Nets / K. Shoylekova, K. Grigorova // International Journal of Industrial and Systems Engineering. – 2015. – Vol. 9, Issue 12. – P. 4148–4153.
3. Transformation of Business Process Model and Notation models onto Petri nets and their analysis / [M. M. Umair, K. Barkaoui, Z. Li et al.] // Advances in Mechanical Engineering. – 2018. – Vol. 10, № 12. – P. 168–188.
4. De Backer M. Deterministic Petri Net Languages as Business Process Specification Language [Electronic resource] / M. De Backer, M. Snoeck // SSRN Electronic Journal. – 2006. – Access mode: http://dx.doi.org/10.2139/ssrn.876906
5. Zeigler B. Theory of Modeling and Simulation / B. Zeigler, H. Praehofer, T. Gon Kim. – Cambridge: Academic Press, 2018. – Discrete Event & Iterative System Computational Foundations. – P. 567–599.
6. Transformation of BPMN Models for Behaviour Analysis / [I. Raedts, M. Petkovic, Y. Usenko et al.] // 5th International Workshop on Modelling, Simulation, Verification and Validation of Enterprise Information Systems, MSVVEIS-2007, Funchal, June 2007: proceedings. – 2007. – P. 126–137.
7. Li Z. A Petri Nets Evolution Method that Supports BPMN Model Changes / Z. Li, Z. Ye // Scientific Programming. – 2021. – Vol. 2021, Issue 3. – P. 1–16.
8. Silhouetting the Cost-Time Front: Multi-objective Resource Optimization in Business Processes / [O. L. Pintado, M. Dumas, M. Yerokhin et al.] // Business Process Management Forum, Rome, 6–10 September 2021: proceedings. – Berlin: Springer, 2021. – P. 92–108.
9. Durán F. Analysis of Resource Allocation of BPMN Processes / F. Durán, C. Rocha, G. Salaün // 17th International Conference on Service-Oriented Computing, Toulouse, 28–31 October 2019: proceedings. – Berlin: Springer, 2019. – P. 452–457.
10. Rizk-Allah R. M. A Novel Multi-Ant Colony Optimization For Multi-Objective Resource Allocation Problems / R. M. Rizk-Allah // International Journal of Mathematical Archive. – 2014. – Vol. 5, № 9. – P. 183–192.
11. Couckuyt I. Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization / I. Couckuyt, D. Deschrijver, T. Dhaene // Journal of Global Optimization. – 2013. – Vol. 60, №3. – P. 575–594.

DOI 10.15588/1607-3274-2023-3-18

12. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. IEEE transactions on neural networks / [R. Cheng, Y. Jin, M. Olhofer et al.] // IEEE Transactions on Evolutionary Computation. – 2016. – Vol. 20, № 5 – P. 773–791.
13. Tong L. Neural architecture search via reference point based multi-objective evolutionary algorithm / L. Tong, B. Du // Pattern Recognition. – 2022. – Vol. 132, № 11.
14. Lárraga G. Incorporating Preference Information Interactively in NSGA-III by the Adaptation of Reference Vectors / G. Lárraga, B. S. Saini, K. Miettinen // Evolutionary Multi-Criterion Optimization, Hannover, 18–23 September 2023: proceedings. – Berlin: Springer, 2023. – P. 578–592.
15. Solving multi-objective structural optimization problems using GDE3 and NSGA-II with reference points / [D. Vargas, A. C. C. Lemonge, H. Barbosa et al.] // Engineering Structures. – 2021. – Vol. 239, № 3.
16. Deb K. Reference Point Based Multi-Objective Optimization Using Evolutionary Algorithms / K. Deb, S. Jayavelmurugan // International Journal of Computational Intelligence Research. – 2006. – Vol. 2, № 3. – P. 635–642.
17. Fundamentals of Business Process Management / [M. Dumas, M. La Rosa, J. Mendling et al.] – Berlin: Springer, 2018. – 527 p.
18. Discovering business process simulation models in the presence of multitasking and availability constraints / [B. Estrada-Torres, M. Camargo, M. Dumas et al.] // Data & Knowledge Engineering. – 2021. – Vol. 134, № 1.
19. Camargo, M. Automated discovery of business process simulation models from event logs / M. Camargo, M. Dumas, O. González-Rojas // Decision Support Systems. – 2020. – Vol. 134, №1.
20. Performance indicators in multiobjective optimization / [C. Audet, J. Bigeon, D. Cartier et al.] // European Journal of Operational Research. – 2021. – Vol. 292, № 2. – P. 397–422.
21. Blank J. Pymoo: Multi-Objective Optimization in Python / J. Blank, K. Deb // IEEE Access. – 2020. – Vol. 8, № 1. – P. 89497–89509.
22. Tanabe R. The Impact of Population Size, Number of Children, and Number of Reference Points on the Performance of NSGA-III / R. Tanabe, A. Oyama // International Conference on Evolutionary Multi-Criterion Optimization, Münster, 19–22 March 2017: proceedings. – Berlin : Springer, 2017. – P. 606–621.

DOI 10.15588/1607-3274-2023-3-18

UDC 616.1:67.05

APPLICATION OF BLOW-UP THEORY TO DETERMINE THE SERVICE LIFE OF SMALL-SERIES AND SINGLE ITEMS

Nahornyi V. V. – PhD, Associate Professor, Associate Professor of the Information Technologies Department, Sumy State University, Sumy, Ukraine.

ABSTRACT

Context. The actual task of developing a method for determining the service life of small-series and single items based on the blow-up modes theory has been solved.

Objective. Application of the blow-up theory in conditions where there are no statistical data on the dynamics of behaviour during the operation of small-series and single items.

Method. To determine the service life of a particular product manufactured in large series, information obtained for a set of similar products of the same type is used. This information is based on numerous experiments, mathematical statistics and probability theory. When operating small-series and single items, such information is not available. In this case, it is necessary to determine the individual resource of an individual product based on the results of an analysis of its behaviour in the past. The method presented in the article is based on the application for such an analysis of the method used when considering systems operating in blow-up mode. The essence of the technique is to extract the periodic component from the temporal realization of the control parameter. This component is modelled by a Fourier series consisting of log-periodic functions. The main coefficients of these functions are the time equal to the operating time of the product until the end of its service life.

Results. The method under consideration has been successfully tested in determining the service life of the transport-dumping bridge, related to products that are actually single items.

Conclusions. An analysis of the experimental data on the behaviour of the load-bearing elements of a transport-dump bridge confirms the assumption about the behaviour of the bridge structure as a system operating in a blow-up mode. This made it possible to determine in advance the service life of the power units of the bridge and obtain the result directly in units of time, without requiring information about the maximum permissible value of the controlled parameter to obtain this information.

For the first time, the possibility is shown to consider the behaviour of small-series and single items as dynamic systems operating in a blow-up mode.

Practical significance. A solution to the topical problem of determining the service life of small-series and single items is proposed.

KEYWORDS: life time, transport and dump bridge, small-series products, log-periodic component, direct and indirect control methods, blow-up theory.

NOMENCLATURE

T – service life of products;
t_f – moment of exacerbation;
a – actual exponent;
$\alpha + \beta i$ – complex exponent;
a_i – polynomial coefficients;
$F(\cdot)$ – some periodic function;
τ – size phase;
m, α, β – exponent;
C, η, γ – coefficients;
$f(t)$ – probability density function;
$F(t)$ – distribution function;
$R(t)$ – reliability function;
n – number of experiments;
B_{TR} – controlled signal trend;
A_{PER} – variable component of the controlled signal;
A_{SUM} – controlled signal value;
B_{MOD} – модель переменной составляющей;
A_0 – amplitude;
A_{EXT} – variable component extremum B_{VC}.
A_{EXTM} – extreme array model of the variable component B_{VC};
t_n, t_{n+1}, t_{n+2} – extreme time of the controlled signal variable component;
ρ – time parameter;
ω – log-periodic frequency;
φ – phase of log-periodic oscillations;
T_{RUL} – residual useful life ;
a_0, a_k, b_k – Fourier series coefficients;
κ – serial number of the Fourier series term;
t_0 – integration lower limit;
t_m – integration upper limit;
n – number of Fourier series terms.

INTRODUCTION

The service life of multi-series products is traditionally determined by the results of a mass, collective forecast based on numerous experimental data on the resource of products, mathematical statistics and probability theory. At the same time, the determination of the service life of a wide class of technical products is carried out, as a rule, indirectly upon the fact that the controlled parameter reaches the maximum permissible level according to the standards. This level is an average statistical value established by the results of a sufficiently long operation of a significant number of the same type of products. For small-series products, as well as products produced in single copies, this approach is not applicable. The solution to this problem is to determine for each product sample its

DOI 10.15588/1607-3274-2023-3-19

individual service life. In this case, it is desirable to determine the duration of operation not indirectly, but directly in units of time and to obtain the information of interest long before the end of the service life of the product.

These requirements are met by information obtained, for example, from the results of approximating the trend of the controlled parameter by a predictive model, which is a smooth monotonically changing function. The coefficients of the model determined in this case include a coefficient that coincides in magnitude and dimension with the end time of the product's operation. This coefficient can be determined long before the expiration of its service life each time when monitoring the condition of the product.

If the initial information does not meet these requirements, and fluctuations are superimposed on a smooth trend, the frequency of which increases with time, then the product can be considered as a dynamic system developing in a blow-up mode. A distinctive feature of the behaviour of these systems is that as the catastrophe approaches, which means the destruction of the system or a radical change in the law of its development, the frequency value reaches infinity.

The fact of frequency change is fixed long before the catastrophe, and the oscillation model allows you to directly determine the service life of the product, without requiring knowledge of the maximum permissible standard level of the controlled parameter.

The object of work is the process of determining the service life of small-scale and single items.

The subject of the work is a model that describes the change in the periodic component of the control parameter recorded during the operation of products.

The purpose of the work is to develop a method for determining the service life of small-series and single items.

1 PROBLEM STATEMENT

Periodic processes, apparently, are one of the foundations for constructing theories in various fields, including for determining the service life of small-scale and single products. Periodicity – the regular repetition of something in time – testifies to the cognoscibility of the world, in the causal conditionality of phenomena. Understanding the nature of periodicity makes it possible to predict events, and such predictions are the basis of a method for determining the life of a dynamic system under control.

In dynamic systems developing in blow-up mode, a periodic process is superimposed on the main trend of the controlled parameter. This process is described by a model, one of whose coefficients coincides in value and dimension with the moment of system destruction or a radical change in the law of its development [1].

Such modes are described by the following equation

$$\frac{dx}{dt} = x^{1+1/\alpha}. \qquad (1)$$

The equation solution increases without limit as we approach the peaking moment t_f:

$$x(t) \sim \left(t_f - t\right)^{-\alpha}. \qquad (2)$$

To obtain a solution acceptable for practice (2), we pass from the real indicator α to the complex one $\alpha + \beta i$, which allows us to obtain an equation of the following form:

$$x(t) = \mathrm{Re}\sum_k a_k \left(t_f - t\right)^{-\alpha+k\beta i} = \left(t_f - t\right)^{-\alpha} \cdot F\left(\ln\left(t_f - t\right)\right). \qquad (3)$$

The function $F(\cdot)$ is described by several multiple harmonics, characterizing in the general case the significant nonlinearity of systems developing in the blow-up mode. However, in practice [2], the function $F(\cdot)$ is limited to one first harmonic:

$$x(t) = \left(t_f - t\right)^{-\alpha} \cdot \left(a_0 + a_1 \cos\left(\beta \cdot \ln\frac{t_f - t}{\tau}\right)\right). \qquad (4)$$

This expression is a smooth trend, on which log-periodic fluctuations are superimposed, which serve as precursor of approaching the blow-up moment t_f. Taking $t \to t_f$ the oscillation frequency tends to infinity, which meets the dynamic law requirements followed by the blow-up mode. The continuous increase in the log-periodic oscillations frequency allows them to react sensitively to the course of catastrophically developing processes long before the blow-up moment.

If we consider the exhaustion moment of the tool life T as the blow-up moment t_f, then the materials cutting can be attributed to the blow-up modes. At the same time, to improve the quality of predicting tool life, it is necessary to isolate the sensitive log-periodic part of the recorded signal. In practice, this means that the total signal periodic component must be separated from the smooth trend and its behavior should be analyzed separately throughout the entire cutting process.

The periodic component model should be subjected to direct analysis, which fully describes the complex polyharmonic in structure of the actually recorded signal.

2 REVIEW OF THE LITERATURE

The service life of similar products produced in significant quantities is determined using a mass, statistical model, when the behavior of a set of products over time is observed. Based on the observation made over a certain time, a prediction is made of the behavior of one specific product in the future time interval. The service life of products in this case is determined, as a rule, in three ways [2]. The first method is based on the use of statistics on the cumulative probability function

DOI 10.15588/1607-3274-2023-3-19

P(t) of the normal distribution, showing the probability of failure of a given type of product depending on its service life (Fig. 1).

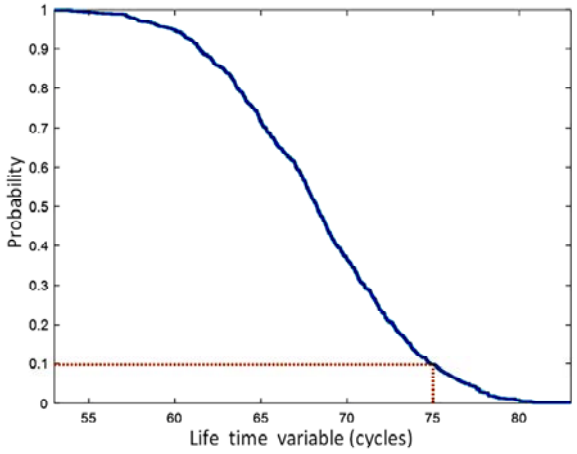


Figure 1 – Probability of product failure depending on from its service life [2]

The second method establishes the degree of similarity of current information about the trajectory of the controlled parameter that characterizes the behaviours of the product of interest (degradation profile), with statistical data on similar trajectories compiled from the results of operating similar products (Fig. 2).

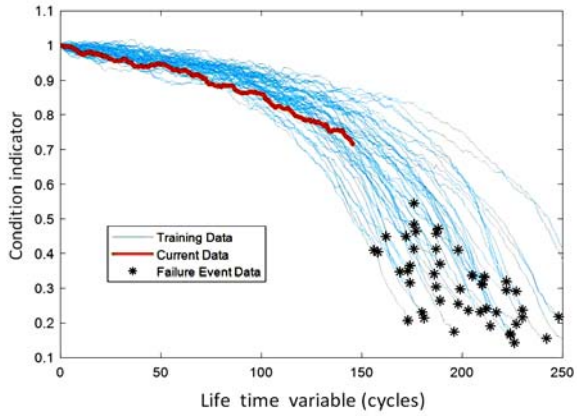


Figure 2 – Statistics of degradation profiles and the actual trajectory of the controlled parameter [2]

In Fig. 2, the statistical set of degradation profiles is highlighted in blue, the current trajectory of the controlled parameter is highlighted in red. In this case, based on the degree of closeness of the current curve and the set of blue curves, the residual life of the product is estimated by the authors at about 65 conditional cycles.

The third method predicts the moment of product failure by the value of the coordinate of the time axis (applicator) of the point of intersection of the skeletal curve of the trajectory of the controlled parameter with its threshold level (Fig. 3).

In practice, smooth monotonically varying functions are used as models of backbone curves [3]. In Fig. 4, for example, the behaviours of 4 types of backbone curves is shown, described by linear, parabolic, *S*-shaped (Gompertz curve) and exponential dependences.

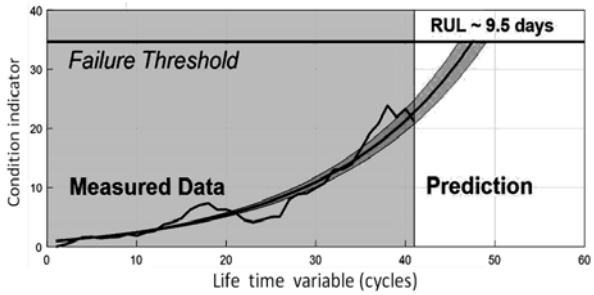


Figure 3 – Intersection of the skeletal curve, described by the exponent, with a threshold level of the controlled parameter [2]

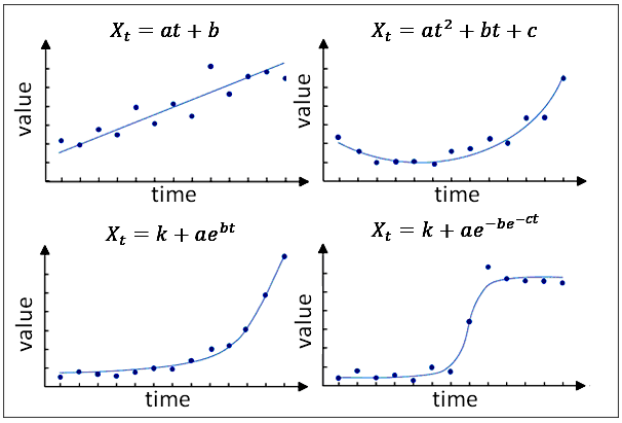


Figure 4 – Behaviours of 4 different analytical description of backbone curves

Forecasting the service life of a product using skeletal curves requires knowledge of the threshold level of the controlled parameter, which is not feasible for small-scale products, and, moreover, single products.

The controlled parameter of a given product sample changes along a single (individual) trajectory. Therefore, the threshold level, being in essence an average statistical value, refers to a specific sample of even the same type of products only with a certain degree of probability.

Thus, according to the most common normal distribution law for the uptime of a product, 50% of products fail before the threshold level is reached, and, accordingly, 50% work after it is crossed [4]. This serves as a serious, error-prone, problem of predicting the life of products in general, and not just small-scale or single ones.

In this case, it is legitimate to use an individual model, in which it is required to regularly monitor the technical condition of the product, comparing the recorded data with the skeletal curve. The analytical description of the skeletal curve model contains a coefficient that coincides in value and dimension with the operating time of the product until the end of its service life.

So, in [5], the model (5) is considered, which includes a similar coefficient *T*. The model was obtained on the basis of the expression for the fatigue curve.

DOI 10.15588/1607-3274-2023-3-19

$$\sigma = \sigma_{-1} \cdot \left(\frac{N_0}{N} \right)^{1/m} \Rightarrow A(t) = A_0 \cdot \left(\frac{T - t_0}{T - t_i} \right)^{\alpha}. \quad (5)$$

The coefficients of the model A(t), including the coordinate *T*, are determined numerically by minimizing the deviation from the backbone curve of the time series $A_{con}(t)$ compiled from the results of measuring the control parameters (6).

$$U(t) = \sum_{i=1}^{m} \left(A_{con}(t) - A(t) \right)^2. \quad (6)$$

A similar approach to determining the service life is proposed in [6]. Here, as a model of the backbone curve, the analytical expression (7) is used to describe the Weibull distribution.

$$A(t) = \left[-\ln\left(1 - \frac{t}{T}\right) \right]^{\lambda} + C. \quad (7)$$

When choosing models of backbone curves (5) and (7), following the phenomenological approach to modelling, a generalized scheme of the change in the controlled parameter during the life cycle of the product was reproduced (Fig. 5). Three characteristic sections are distinguished in this diagram: the initial section, where an increased flow of product failures is noted, due to the running in of its components and parts; stationary site – the main time of operation of the product; site of catastrophic failures of the product, leading to the termination of its operation. The generalized scheme of the change in the controlled parameter (Fig. 5), consisting of three sections, characterizes the change in the process of the life cycle of the product of gradual failures, the negative impact of which on the quality of the product's functioning gradually accumulates. For this reason, these failures in the literature are often referred to as "wear-out", and "wear-out" is understood in an extended sense [4]. Accordingly, the generalized scheme is called the "wear curve" [7].

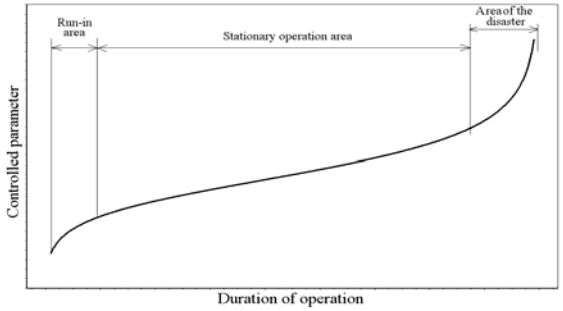


Figure 5 – Generalized scheme of changes in the product life cycle of a controlled parameter characterizing its technical Condition

Based on the results of gaining experience in operating the model of the skeletal curve (5), its modification was carried out [8]. The purpose of the modification was to develop a model (8) that describes all three sections of the product life cycle (Fig. 6). The graph of this model at the site of stationary operation of the product has a curvature of a different sign, which makes it possible to more flexibly describe the nature of the change in the technical condition of the supervised product during its operation.

$$A(t) = A_0 + \eta \cdot \left(\frac{t - t_0}{T - t} \right)^{\alpha} - \gamma \cdot \left(\frac{t - t_0}{T - t} \right)^{\beta}. \quad (8)$$

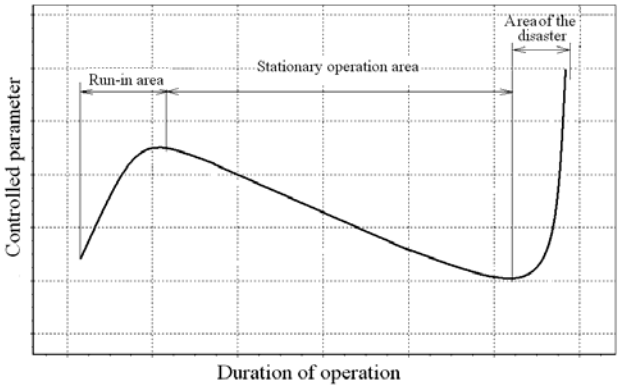


Figure 6 – Skeletal curve model (8)

The use of this approach to determine the service life of small-scale and single products in various fields of human activity is clearly shown in the monograph [9].

The introduction of digital measurement systems made it possible to refine the method for monitoring the state of the product. In particular, what earlier, with the analogy method of monitoring the state of the product, was considered as a measurement error, for example, fluctuations (fluctuations) of the measured value relative to some of its average value (skeletal curve) in practice turned out to be an informative component of the measured data [10].

The wear curve (Fig. 5, 6) ends with a catastrophe section, which characterizes a sharp change in the value of the controlled parameter, which is the result of the cumulative impact on the product of its gradual failures. This gives grounds to believe that the behaviour of the product during its operation can be interpreted as the behaviour of a system developing in a blow-up mode [1].

In this case, the trajectory of the control parameter should be considered as the sum of a smooth skeletal curve and periodic oscillations about it. Periodic oscillations obey the log-periodic law, according to which the frequency of oscillations increases as it approaches the moment of product failure.

Monitoring the fact of changing the frequency of log-periodic oscillations makes it possible to predict the moment of product failure individually for each of its samples, regardless of the stage of its operation and does not require knowledge of statistical data on the threshold level of the controlled parameter.

DOI 10.15588/1607-3274-2023-3-19

The latter is very important, since it allows solving the urgent problem of determining the service life of products manufactured in small batches or, in general, in single copies, which were the goal of the research, the results of which are presented in this article.

3 MATERIALS AND METHODS

The controlled parameter $A_{CON}(t)$ is considered as the sum of the smooth (trend) B_{TR} and the periodic component A_{PER}.

$$A_{CON}(t) = B_{TR} + A_{PER} \tag{9}$$

According to (4), at $T = t_f$, B_{TR} is determined from the following expression

$$B_{TR} = a_0 \cdot (T - t)^{-\alpha}. \tag{10}$$

The periodic component A_{PER} is extracted from the information (total) signal A_{SUM} by decomposing it into empirical modes [11].

$$A_{PER} = -0.25A_{SUM\,i-1} + 0.5A_{SUM\,i} - 0.25A_{SUM\,i+1}. \tag{11}$$

The periodic component A_{PER}, according to (4), is determined from the following expression

$$A_{PER} = a_1 \cdot \cos\left(\beta \cdot \ln\frac{T-t}{\tau}\right). \tag{12}$$

For the convenience of further research, expression (12) should be reduced to the classical form of the log-periodic function (13), considering it as a B_{MOD} model of the periodic component A_{PER}.

$$\begin{gathered} B_{MOD} = A_0 \cos(\omega \cdot \ln(T-t) - \varphi), \\ \text{where } A_0 = a_1 \cdot (T-t)^{-\alpha};\ \omega = \beta; \varphi = \beta \cdot \ln(\tau). \end{gathered} \tag{13}$$

Expression (13) contains fore unknown parameters: T, ω, φ, A_0. The first three parameters are determined by solving the system of two nonlinear equations (14).

$$\begin{cases} \ln(T-t_n) - \ln(T-t_{n+1}) = \dfrac{2\pi}{\omega}, \\ \ln(T-t_{n+1}) - \ln(T-t_{n+2}) = \dfrac{2\pi}{\omega}. \end{cases} \tag{14}$$

Equations (14) are based on the knowledge of the time t_n, which account for the extremes A_{EXT} of the periodic component A_{PER}.

To search for these extremes, the following algorithm is used

– at least three local extreme stand out in the periodic component A_{PER}. They are separated from each other in phase by an angle 2π, and there are consecutive and identical in sign (maximum or minimum);

–the time t is marked when extremes occur (t_n, t_{n+1}, t_{n+2});

– the parameter ρ is calculated that characterizes the relationship between the extremes occurrence time.

$$\rho = \frac{t_{n+1} - t_n}{t_{n+2} - t_{n+1}},\ \rho > 1. \tag{15}$$

Parameter ρ must exceed one. This indicates a decrease in the period of its oscillations, characteristic of the log-periodic function, over time.

A decrease in the period leads to an increase in the oscillation frequency of the log-periodic function in the limit to infinity. This function feature was the basis for choosing it as a model B_{MOD} (13) for describing systems operating in the blowup mode [1].

The set of extremes forms an array composed of discrete values of A_{EXT} extremes of the periodic component A_{PER}.

The solution of system (14) gives the following expressions for the first three unknowns of equation (13) [12]:

$$\begin{gathered} T = \frac{t_{n+1}^2 - t_{n+2}t_n}{2t_{n+1} - t_{n+2} - t_n}, \\ \omega = 2\pi / \ln(\rho), \\ \varphi = \pi - \omega \cdot \ln(T - t_{n+2}). \end{gathered} \tag{16}$$

To check the correctness of the obtained unknowns values (16) and, if necessary, to refine them, the difference between the components of the array of extreme values A_{EXT} and their model B_{MOD} (13) is minimized. In this case, the parameter A_0 is also determined.

$$\sum_i^m \left(A_{EXT\,i} - B_{MOD\,i}\right)^2 \Rightarrow \min. \tag{17}$$

In practice, the array of extreme values A_{EXT} contains a number of components, indicating the polyharmonic nature of the oscillations of the periodic component A_{PER}. Therefore, when refining the values of parameters (16), as a model A_{EXTM} (predictive model) describing an array of extreme values A_{EXT}, one should use a trigonometric polynomial composed of log-periodic functions (Fourier series).

$$A_{EXTM} = \frac{a_0}{2} + \sum_{k=1}^{n}\left[a_k \cos(k \cdot \omega \cdot \ln(T-t)) + b_k \sin(k \cdot \omega \cdot \ln(T-t))\right]. \tag{18}$$

The coefficients of the series a_0, a_k, b_k are determined from the following expressions

DOI 10.15588/1607-3274-2023-3-19

$$\begin{cases} a_0 = \dfrac{1}{t_0 - t_n}\int\limits_{t_0}^{t_m} A_{EXT}\cdot\dfrac{1}{T-t}dt, \\ a_k = \dfrac{2}{t_0 - t_n}\int\limits_{t_0}^{t_m} A_{EXT}\cdot\cos\left(k\cdot\dfrac{2\pi}{t_0 - t_m}\cdot\ln(T-t)\right)\cdot\dfrac{1}{T-t}dt, \\ b_k = \dfrac{2}{t_0 - t_n}\int\limits_{t_0}^{t_m} A_{EXT}\cdot\sin\left(k\cdot\dfrac{2\pi}{t_0 - t_m}\cdot\ln(T-t)\right)\cdot\dfrac{1}{T-t}dt. \end{cases} \quad (19)$$

The remaining useful life T_{RUL} is determined from the following expression.

$$T_{RUL} = T - t. \quad (20)$$

4 EXPERIMENTS

The purpose of the experiment was to test the effectiveness of determining the service life of small-series and single products, considering their behaviour during operation, as systems operating in the blow-up mode.

As an object of study, a hydro turbine, a representative of a small-series product [9] (Fig. 7), and a transport and dump bridge, related to single products [12] (Fig. 13), were considered.

5 RESULTS

The initial information for the analysis was the vibration of the turbine support, which was measured along the y-axis (Fig. 7).

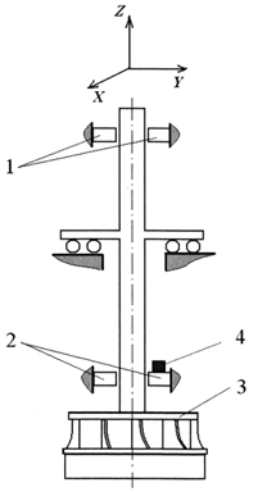


Figure 7 – Diagram of the hydro turbine indicating the point of its vibration monitoring: 1—generator bearing; 2—turbine bearing; 3 – turbine impeller;
4 – vibration sensor of the turbine bearing

The trajectory of the turbine vibration level during the observed period (120 days) is shown in Fig. 8.

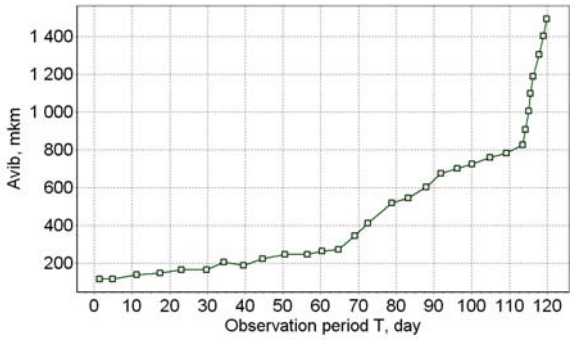


Figure 8 – Changing the vibration level of the turbine support

Fig. 9 shows the change over time in the frequency of logo-periodic oscillations superimposed on the trajectory of turbine vibrations

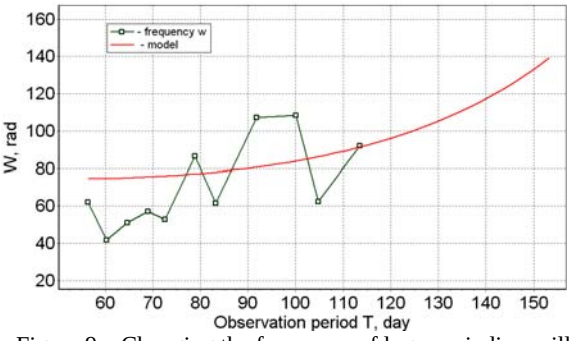


Figure 9 – Changing the frequency of logo-periodic oscillations, superimposed on the trajectory turbine vibration level changes

Fig. 10 shows the approximation by the log-periodic oscillations model B_{MOD} (13) of the periodic component of turbine vibration.

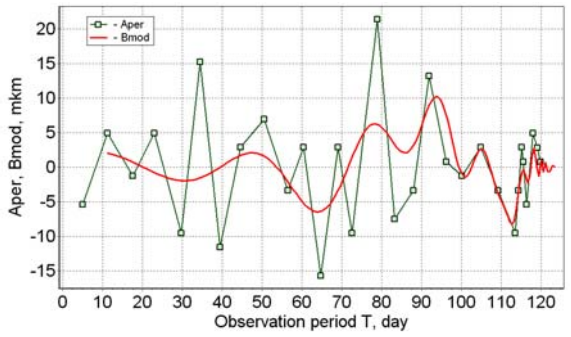


Figure 10 – Approximation by the model B_{MOD} (13) of the periodic component A_{PER} turbine vibration

The results of determining the service life of hydro turbine are shown in Fig. 11.

DOI 10.15588/1607-3274-2023-3-19

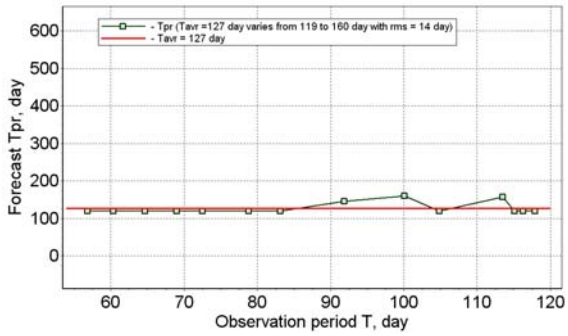


Figure 11 – Forecast of the service life of hydro turbine

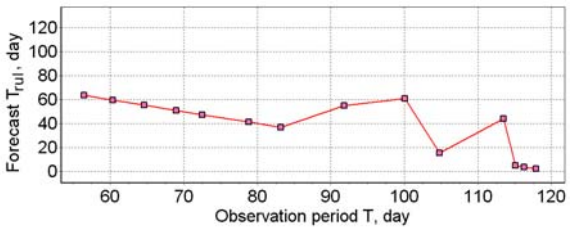


Figure 12 – Change in the forecast of the residual service life of the hydro turbine

Fig. 12 shows the change during the operation of the residual useful life of the T_{RUL} (20) of the hydro turbine.

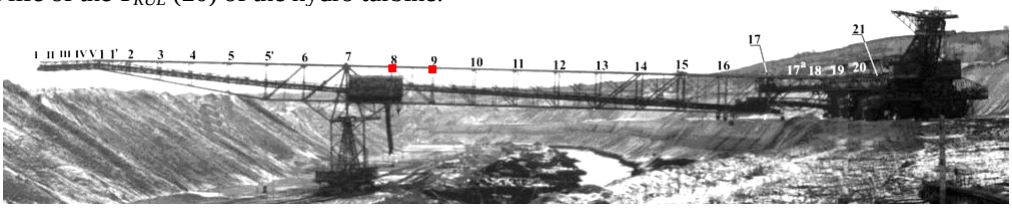


Figure 13 – Scheme for marking the nodes of the main truss of the transport and dump bridge [12]

Changes during the controlled period of the truss deflection trajectories in nodes No. 8 and No. 9 (marked in red in Fig. 13) are shown in Fig. 14.

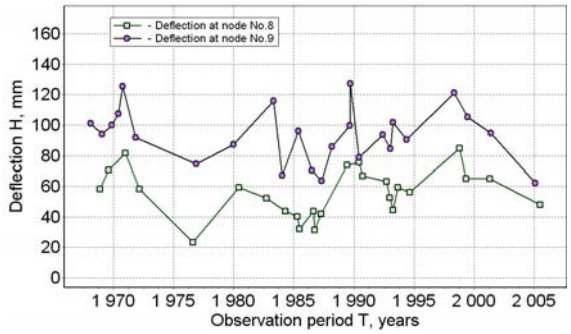


Figure 14 – The trajectory of deflections of the farm transport and dump bridge at its nodes No. 8 and 9 fixed during the period of control of his condition [12]

Fig. 15 and 16 show the change over time in the frequency of logo-periodic oscillations superimposed on the trajectory of the deflection of nodes No. 8 and No. 9 and the model approximating them.

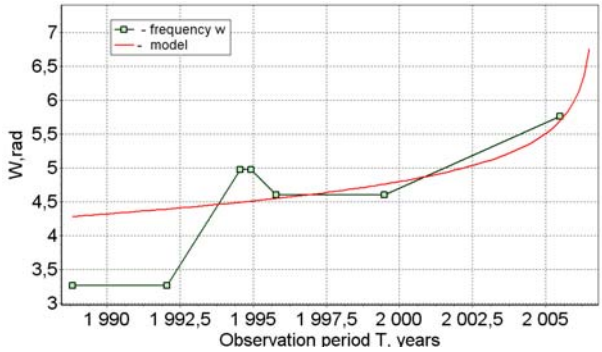


Figure 15 – Changing the frequency of logo-periodic oscillations, superimposed on the trajectory of the deflection of node No. 8

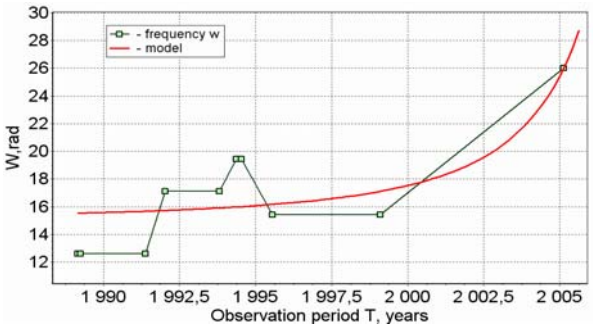


Figure 16 – Changing the frequency of logo-periodic oscillations, superimposed on the trajectory of the deflection of node No. 9

Figures 17 and 18 show the approximation by the log-periodic oscillations model B_{MOD} (13) of the periodic component of deflections A_{PER} recorded during the period of monitoring the state of the bridge, respectively, at nodes No. 8 and No. 9

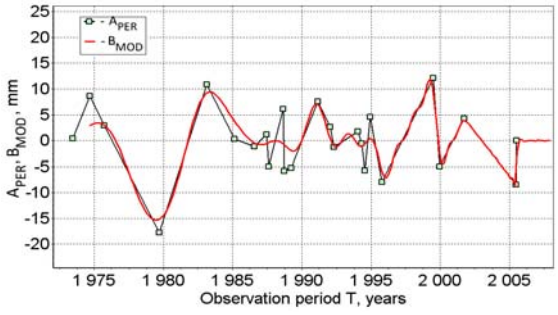


Figure 17 – Approximation by the model B_{MOD} (13) of the periodic component A_{PER} of the deflection of node No. 8

DOI 10.15588/1607-3274-2023-3-19

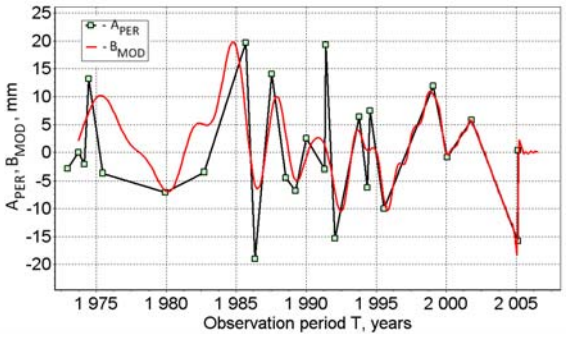


Figure 18 – Approximation by the model B_{MOD} (13) of the periodic component A_{PER} of the deflection of node No. 9

The results of predicting the service life of nodes No. 8 and 9 of the farm transport – dump bridge are shown in Fig. 19 and 20.

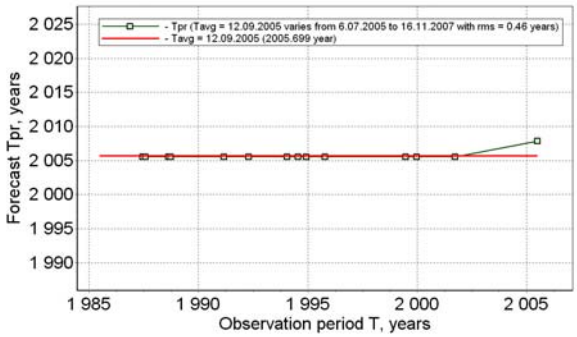


Figure 19 – Forecast of the service life of node No. 8.

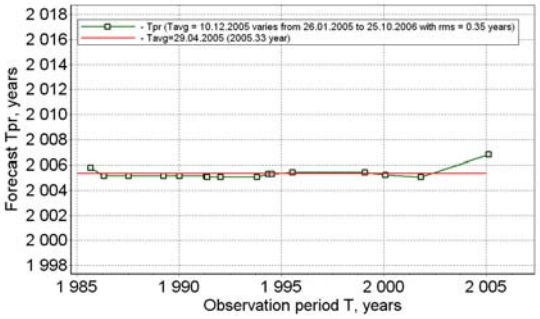


Figure 20 – Forecast of the service life of node No. 9

Fig. 21 shows the change during the operation of the residual useful life of the T_{RUL} (20) nodes No. 8 and 9 of the truss of the transport and dump bridge.

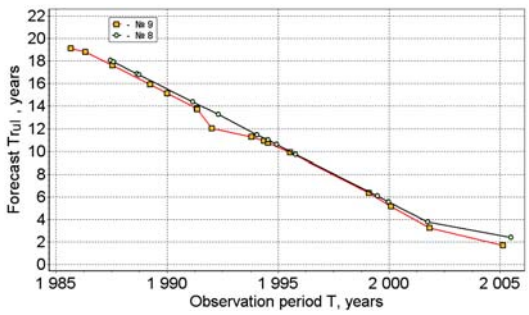


Figure 21 – Change in the forecast of the residual service life of nodes No. 8 and No. 9

6 DISCUSSION

Monitoring of the state of the turbine and the bridge confirmed the assumption that the trajectory of change of the control parameter (Fig. 8, 14) contains a variable component (Fig. 10, 17, 18). This component is described by the model of logo-periodic fluctuations. The frequency of these oscillations increases as the products approach and the expiration of their service life (Fig. 9, 15, 16), which corresponds to the behaviour of systems operating in the blow-up mode. The calculation of the coefficients of the logo-periodic oscillations model made it possible to determine the operating time of the turbine before the accident (127 days, fig. 11) and the calendar date of the resource exhaustion of nodes No. 8 and No. 9 (fig.19,20), which, respectively, was 09.12.2005 and 01.26.2005. The remaining service life of the turbine and bridge at the time of the last control of their condition was 2 days for the turbine, and 2.38 years for the bridge (node No. 8) and 1.69 years (node No. 9).

CONCLUSIONS

The scientific novelty of obtained results is that the analysis of experimental data on the behaviour of the hydro turbine and of the transport-dump bridge confirms the assumption about the behaviour of the turbine and bridge structure as a system operating in the blow-up mode.

The practical significance of obtained results is made it possible to determine in advance the moment when the service life of the turbine and power units of the bridge is exhausted, and to obtain a forecast directly in units of time, and without requiring information about the maximum permissible value of the controlled parameter (turbine vibration level and deflection of the truss unit) to obtain this information.

Prospects for further research are to study the possibility of using the developed technique to predict the service life of products of various purposes and designs.

ACKNOWLEDGEMENTS

The work is supported by the state budget scientific research project of Sumy State University “Models and methods of information technologies for the analysis and synthesis of structural, information and functional models of automated objects and processes” (state registration number 0120U103071).

REFERENCES

1. Sornette D. Why Stock Markets Crash: Critical Events in Complex Financial Systems. Princeton, N.J., Princeton U. Press, 2003, 421 p. DOI: 10.1063/1.1712506
2. Three Ways to Estimate Remaining Useful Life for Predictive Maintenance [Elektronnij resurs]. Rezhim dostupu: https://www.mathworks.com/company/newsletters/articles/three-ways-to-estimate-remaining-useful-life-for-predictive-maintenance.html
3. Klosterman R. E., Brooks K., Drucker J. et al. Planning support methods: Urban and regional analysis and projection. Lanham, Rowman & Littlefield Publishers, 2018, 298 p. ISBN 978-1-4422-2028-7.

DOI 10.15588/1607-3274-2023-3-19

4. Lad B. K., Shrivastava D., Kulkarni M. S. Machine Tool Reliability. H, John Wiley & Sons Inc, 2016, 336 p. ISBN 978-1-1190-3860-3.
5. Nahornyi V. V., Lavrov E. A., Chybiriak Y. I. Forecasting individual resource of technical systems, *Radio Electronics, Computer Science, Control*, 2019, № 1, pp. 48–56. DOI: 10.15588/1607-3274-2019-1-5
6. Cempel C., Natke H. G., Yao J. T. P. Symptom reliability and hazard for systems condition monitoring, *Mechanical Systems and Signal Processing*, 2000, Vol. 14, № 3, pp. 495–505. DOI: 10.1006/mssp.1999.1246
7. Zhou H., Farsi M., Harrison A., Parlikad A. K., Brintrup A. Civil aircraft engine operation life resilient monitoring via usage trajectory mapping on the reliability contour, *Reliability Engineering & System Safety*, 2023, Vol. 230, pp. 1–29. DOI: 10.1016/j.ress.2022.108878
8. Zaloga V. A., Nahornyi V. M., Nahornyi V. V. Control of the dynamic state of a metalworking technological system and forecasting its resource : monograph. Sumy, SSU, 2016, 242 p. ISBN 978-966-657-604-3.
9. Panda A., Nahornyi V. Forecasting catastrophic events in technology, nature and medicine : monograph. Berlin, Springer, 2021, 97 p. ISBN 978-3-030-65327-9. DOI: 10.1007/978-3-030-65328-6
10. Ding D. M. Zhang, X. Pan, M. Yang Modelling Extreme Events in Time Series Prediction, *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* New York, Association for Computing Machinery, 2019, pp. 1114–1122. DOI: 10.1145/3292500.3330896
11. Myasnikova N. V., Beresten' M. P., Priymak A. A. A "fast" extreme filtering algorithm, *University proceedings. Volga region. Engineering sciences*, 2020, Vol. 53, № 1, pp. 70–77. DOI: 10.21685/2072-3059-2020-1
12. Sornette D. Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth, *PNAS*, 2002, Vol. 230, №1, pp. 2522–2529. DOI: 10.1073/pnas.022581999
13. Sholomitsky A. A., Nahornyi V. V. Observation of deformations, assessment and forecasting of the technical condition of lifting and transport equipment, *Metallurgical processes and equipment*, 2009,Vol. 17, № 3, pp. 31–36

Received 31.05.2023.
Accepted 30.08.2023.

УДК 616.1:67.05

ЗАСТОСУВАННЯ ТЕОРІЇ РЕЖИМІВ ІЗ ЗАГОСТРЕННЯМ ДЛЯ ВИЗНАЧЕННЯ ТЕРМІНУ ЕКСПЛУАТАЦІЇ МАЛОСЕРІЙНИХ ТА ОДИНИЧНИХ ВИРОБІВ

Нагорний В. В. – канд. техн. наук, доцент, доцент кафедри інформаційних технологій Сумського державного університету, м. Суми, Україна.

АНОТАЦІЯ

Актуальність. Вирішено актуальне завдання розробки на основі теорії режимів із загостренням методики визначення терміну служби малосерійних виробів та виробів, виготовлених у одиничному екземплярі.

Ціль. Застосування теорії режимів із загостренням за умов, коли відсутні статистичні дані про динаміку поведінки в процесі експлуатації малосерійних виробів та виробів, виготовлених у одиничному екземплярі.

Метод. Для визначення терміну служби конкретного виробу, який виготовляють великими серіями, використовується інформація, отримана для сукупності подібних однотипних виробів. Ця інформація складена на основі численних експериментів, математичної статистики та теорії ймовірностей. При експлуатації малосерійних виробів та виробів, виготовлених у одиничному екземплярі, такої інформації немає. І тут необхідно визначати індивідуальний ресурс «виробу-індивіда» за результатами аналізу його поведінки у минулому. Поданий у статті метод заснований на застосуванні для такого аналізу методики, яка використовується при розгляді систем, що працюють у режимі із загостренням. Суть методики полягає у виділенні з часової реалізації контрольного параметра періодичної компоненти. Ця компонента моделюється рядом Фур'є, що складається з лого-періодичних функцій. Основним коефіцієнтом цих функцій є час, що дорівнює напрацюванню виробу до закінчення терміну його служби.

Результати. Метод, що розглядається, успішно апробований при визначенні терміну служби транспортно-відвального мосту, що відноситься до виробів, які виготовляються фактично в одиничних екземплярах.

Висновки. Аналіз експериментальних даних про поведінку несучих елементів транспортно-відвального мосту підтверджує припущення щодо поведінки конструкції моста як системи, що працює в режимі з загостренням. Це дозволило заздалегідь визначити термін служби силових агрегатів моста і отримати результат безпосередньо в одиницях часу, не вимагаючи для отримання інформації про гранично допустиме значення контрольованого параметра.

Вперше показано можливість розглядати поведінку малосерійних виробів та виробів, що виготовляються в одиничному екземплярі, як динамічних систем, що працюють у режимі із загостренням.

Практична значимість. Запропоновано вирішення актуального завдання щодо визначення терміну служби малосерійних виробів та виробів, що виготовляються в одиничних екземплярах

КЛЮЧОВІ СЛОВА: ресурс, транспортно-відвальний міст, дрібносерійна продукція, логперіодична складова, прямі та непрямі методи керування, теорія режимів із загостренням.

ЛІТЕРАТУРА

1. Sornette D. Why Stock Markets Crash: Critical Events in Complex Financial Systems / D. Sornette. – Princeton, N.J. : Princeton U. Press, 2003. – 421 p. DOI: 10.1063/1.1712506
2. Three Ways to Estimate Remaining Useful Life for Predictive Maintenance [Електронний ресурс]. – Режим доступу: https://www.mathworks.com/company/newsletters/articles/three-ways-to-estimate-remaining-useful-life-for-predictive-maintenance.html
3. Planning support methods: Urban and regional analysis and projection / [R. E.Klosterman, K. Brooks, J. Drucker et al.]

DOI 10.15588/1607-3274-2023-3-19

– Lanham : Rowman & Littlefield Publishers, 2018. – 298 p. ISBN 978-1-4422-2028-7.
4. Lad B. K. Machine Tool Reliability / B. K. Lad, D. Shrivastava, M. S. Kulkarni. – H. : John Wiley & Sons Inc, 2016. – 336 p. ISBN 978-1-1190-3860-3.
5. Nahornyi V. V. Forecasting individual resource of technical systems / V. V. Nahornyi, E. A. Lavrov, Y. I. Chybiriak // Radio Electronics, Computer Science, Control. – 2019. – № 1. – P. 48–56. DOI: 10.15588/1607-3274-2019-1-5
6. Cempel C. Symptom reliability and hazard for systems condition monitoring / C. Cempel, H. G. Natke, J. T. P. Yao // Mechanical Systems and Signal Processing. – 2000. – Vol. 14, № 3. – P. 495–505. DOI: 10.1006/mssp.1999.1246
7. Civil aircraft engine operation life resilient monitoring via usage trajectory mapping on the reliability contour / [H. Zhou, M. Farsi, A. Harrison et al.] // Reliability Engineering & System Safety. – 2023. – Vol. 230. – P. 1–29. DOI: 10.1016/j.ress.2022.108878
8. Zaloga V. A. Control of the dynamic state of a metalworking technological system and forecasting its resource : monograph / V. A. Zaloga, V. M. Nahornyi, V. V. Nahornyi. – Sumy : SSU, 2016. – 242 p. ISBN 978-966-657-604-3.
9. Panda A. Forecasting catastrophic events in technology, nature and medicine : monograph / A. Panda, V. Nahornyi. – Berlin : Springer, 2021. – 97 p. ISBN 978-3-030-65327-9. DOI: 10.1007/978-3-030-65328-6
10. Modelling Extreme Events in Time Series Prediction / [D. Ding, M. Zhang, X. Pan, M. Yang] // KDD ‘19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. – New York : Association for Computing Machinery, 2019. – P. 1114–1122. DOI: 10.1145/3292500.3330896
11. Myasnikova N. V. A “fast” extreme filtering algorithm / N. V. Myasnikova, M. P. Beresten’, A. A. Priymak // University proceedings. Volga region. Engineering sciences, 2020. – Vol. 53, № 1. – P. 70–77. DOI: 10.21685/2072-3059-2020-1
12. Sornette D. Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth / D. Sornette // PNAS. – 2002. – Vol. 230, № 1. – P. 2522–2529. DOI: 10.1073/pnas.022581999
13. Sholomitsky A. A. Observation of deformations, assessment and forecasting of the technical condition of lifting and transport equipment / A. A. Sholomitsky, V. V. Nahornyi // Metallurgical processes and equipment. – 2009. – Vol. 17, № 3. – P. 31–36.

DOI 10.15588/1607-3274-2023-3-19

UDC 621.396.2

USING ESP32 MICROCONTROLLER FOR PHYSICAL SIMULATION OF THE WIRELESS REMOTE CONTROL MODEM

Vakaliuk T. A. – Dr. Sc., Professor, Professor of the Department of Software Engineering, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Andreiev O. V. – PhD, Associate Professor, Associate Professor of the Department of Computer Technologies in Medicine and Telecommunications, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Nikitchuk T. M. – PhD, Associate Professor, Dean of the Faculty of Information and Computer Technologies Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

Osadchyi V.V. – Dr. Sc., Professor, Dean of the Faculty of Economics and Management Borys Grinchenko Kyiv University, Ukraine.

Dubyna O. F. – PhD, Associate Professor, Associate Professor of the Department of Computer Technologies in Medicine and Telecommunications, Zhytomyr Polytechnic State University, Zhytomyr, Ukraine.

ABSTRACT

Context. Due to the need for practical implementation of the theoretical provisions of the proposed method of the wireless transmission of the commands for controlling a moving object in conditions of intentional interference.

Objective of the work is a practical verification of the possibility of using a phase-pulse modulation of the linear-frequency-modulated signal for transmission a control commands through physical modeling using microcontrollers.

Method. Analytical calculations of the change in time of the voltage at the input and output of the device for optimal processing of the linear frequency-modulated signal were carried out exclusively using the computing capabilities of the microcontroller. The graphs of changes in time of the relevant parameters were built with the help of Excel using the data output to the monitor of the serial port of the Arduino IDE software environment. A digital oscilloscope with a USB host was used to monitor the operation of a wireless modem to transmit control commands to a moving object.

Results. Analytical calculations and physical modeling using a modern microcontroller proved the operability of the low-frequency part the wireless remote control modem with using phase-pulse modulation of the linear frequency-modulated signal.

Conclusions. The possibility of using phase-pulse modulation of the linear-frequency modulated signal for the transmission of control commands is considered. This method of transmitting the information component, unlike the existing methods, does not require changing the parameters of the linear frequency modulated signal. The use in the receiver of optimal processing of a linear-frequency-modulated signal of sufficiently big base will allow of the wireless transmission of the commands for controlling a moving object in conditions of the intentional interference. The use of modern microcontrollers made it possible to conduct a practical test of the functionality of the low-frequency part the wireless remote control modem with phase-pulse modulation of the linear-frequency modulated signal through physical modeling.

KEYWORDS: wireless, remote control, linear frequency modulation, phase-pulse modulation, simulation.

ABBREVIATIONS

LFM is a linear frequency modulated;
CSS is a chirp spread spectrum;
FHSS is a frequency hopping spread spectrum;
DSSS is a direct-sequence spread spectrum;
TH is a time hopping;
CDMA is a code-division multiple access;
GFSK is a Gaussian frequency-shift keying;
DBPSK is a differential phase-shift keying;
MU-MIMO is a multi-user MIMO;
MIMO is a multiple-input and multiple-output;
LoRa is a “long range”;
PPM is a phase-pulse modulation;
DAC is a digital-to-analog converter;
ADC is an analog-to-digital converter;
IoT is an Internet of Things.

NOMENCLATURE

τ_c is an impulse duration;

$u(t)$ is a voltage of the LFM signal;

Um is an amplitude of the radio pulse;

f_c is an average frequency of the signal;

Δf is a signal frequency deviation;

$f(t)$ is a change of the frequency of the LFM signal;

$V(t)$ is a pulse characteristic of the matched filter of the signal with LFM;

$f_V(t)$ is a change the frequency of the pulse characteristic of the matched filter;

C is a constant value;

$K(f)$ is a frequency characteristic of the optimal filter of the LFM signal;

$g^*(f)$ is a complex conjugation spectrum of the LFM radio impulse;

$w(t)$ is a voltage of the signal component at the output of the optimal filter;

τ is a duration of the pulse at the filter output;

i is a command number;

$x_i(t)$ is a command structure;

B is a base of the signal.

DOI 10.15588/1607-3274-2023-3-20

INTRODUCTION

Modern conditions of operation of radio-electronic communication complexes place increased demands on their work. This is because qualitative changes in the means of radio-electronic warfare of the leading countries of the world in the last few decades allow them to create a significant impact on the operation of radio-electronic communication complexes. Radio communication in range of short waves has special place in military communication [1]. Most of the short wave and ultra-short wave communication means use narrowband signals that are transmitted at a fixed frequency of the communication channel to transmit speech. A transmitter of sufficient power tuned to the same frequency can suppress a narrowband signal. This lack of protection against conventional radio signal interference led to the need to expand the spectrum of the emitted signal. At the same time, a much wider frequency band than the spectrum width of the information signal is used for information transmission. This leads to a decrease in the spectral power density of the signal emitted by the radio communication device. At the same distance from the means of radio frequency monitoring, the probability of detecting the radiation of a broadband signal transmitter will be less than that of a narrowband one with the same power. Drone communication channels are also vulnerable to cyberattacks [2]. Therefore, the study of the possibility of increase the interference protection of radio communication channels is an urgent scientific task. One of the problematic issues in the creation of a test sample is the verification of the correctness of theoretical calculations by conducting physical modeling using the latest elemental base.

The object of research is the process of the wireless transmission of commands for controlling a moving object.

The subject of the study is the wireless remote control modem of the moving object with using phase-pulse modulation (PPM) of the linear frequency-modulated (LFM) signal.

The purpose of the work is to a physical simulation of the operation the low-frequency part of the wireless remote control modem using of the modern microcontrollers.

1 PROBLEM STATEMENT

As a moving object, consider a car with a camera, for control of which it is enough to send five commands. The duration of command transmission should not exceed one second. During this interval, each command with the number i=1,...,5 will be transmitted by a pulse packet of a certain structure:

$$x_i(t) = \sum_{i=1}^{5} u\left(t + (i-1)\tau_c\right),$$

$$u(t) = U_m \cos\left(2\pi\left(f_c t + \frac{\Delta f t^2}{2\tau_c}\right)\right). \quad (1)$$

Each command to provide for a change in the time position of the pulse of the packet by a certain amount corresponding to the command number. The minimum value corresponds to the command with the number one. As the command number increases, the time shift increases by an amount equal to the pulse duration τ_c. The change of the signal frequency during the duration of the pulse should be less than 25 kHz. In order to increase the interference protection of the radio communication channel, optimal signal processing should be carried out in the receiver. We will assume that frequency conversion cascade is used in the transceiver. Therefore, the formation and processing of the signal can be carried out at a frequency significantly lower than the operating frequency of the transmitter.

To carry out a physical simulation of the operation of the wireless modem to transmit control commands, it is necessary to choose a controller based on which the modem will be implemented. It is also advisable to conduct mathematical modeling to determine signal parameters for command transmission. Next, using the mathematical principles of forming and processing the signal develop the program code for the controller. After that, conduct a physical simulation of the operation of the wireless modem using the selected microcontroller to check the possibility of practical implementation of the proposed solutions.

2 REVIEW OF THE LITERATURE

An overview of the promising directions of increasing the secrecy of operation and interference protection of communication means is carried out in [3–5]. There are several methods of expanding the spectrum of signals, which differ among themselves by the principles of obtaining a wide frequency band: the direct sequence method – DSSS, the frequency hopping method – FHSS, the method of expanding the spectrum with linear frequency modulated signal – CSS, the time-hopping method – TH. For example, the CDMA mobile communication system uses DSSS technology with signal spectrum spreading by mutually orthogonal Walsh-Hadamard functions. The IEEE 802.11 standard provides for the use of various options for expanding the signal spectrum for the organization of a wireless data transmission. The width of the spectrum of the radio signal is much greater than the data transmission rate, and the correlation function is significantly narrower than the correlation function of a narrowband signal. In particular, the use of FHSS technology with GFSK modulation of the radio signal, or DSSS and DBPSK modulation technology with the provision of data transmission at a speed of no more than 1 Mbit/s over a radio channel in the 2.4 GHz frequency range. As a perspective for a wireless local network, the IEEE 802.11ac standard with a radio channel in the 5 GHz frequency range and the use of MU-MIMO spatial multiplexing technology with the formation of an adaptive directional pattern can be considered. When designing complex electronic circuits, it is important to simulate their operation using appropriate software. The experience and results of

DOI 10.15588/1607-3274-2023-3-20

mathematical modeling in the MATLAB software environment to study the problem of synchronization in the receiver based on the technology of multiplexing with orthogonal frequency distribution – OFDM is given in [6]. The CSS spectrum expansion technology was used in the creation of a LoRa physical radio interface [7–9]. The frequency redundancy of the broadband radio signal determines its high interference protection. The CSS frequency of the radio signal can both increase and decrease. However, the LoRa technology, optimal processing of the CCS signal is not carried out, since the information is transmitted by the signal frequency hopping. The necessary range of the information transmission is provided by the change duration of the signal.

The most technologies of the signal spectrum expansion are aimed at use in the ultra-high frequency range. At the same time, the question arises about the possibility of implementing technologies for increase the interference protection of short-wave and ultra-short-wave communication means and to ensure their operation in the conditions of creating deliberate obstacles. A variant of expanding the signal spectrum using Walsh functions with a frequency separation of channels and frequency manipulation of the carrier is considered in [10]. It is proposed to use Walsh functions with a duration of an elementary symbol of 0.5 μs to expand the signal spectrum. At the same time, the width of the signal spectrum much exceeds the frequency band allocated in this range for one channel. The frequency band of the radio channel of the narrowband means of communication does not exceed 25 kHz. Therefore, when implementing the proposed method, it will be difficult to ensure the necessary value of the frequency band with the existing restrictions [11]. Therefore, it is necessary to offer the wireless method of transmitting control commands with increased interference protection within the frequency band allocated for the operation of one channel.

One of the possible options is the transmission of control commands using the method, which based on the LoRa technology, but instead of the signal frequency hopping will changing the pulse repetition frequency of the LFM pulses in accordance with the command being transmitted. So, it method of transmitting the information component, unlike the existing methods, does not require changing the parameters of the LFM signal and we will named is a PPM of the LFM signal. The frequency deviation of the LFM signal of a certain duration should not exceed 25 kHz. The optimal LFM signal processing in the receiver will provide increase the interference protection of the radio communication channel. Therefore, it is necessary to carry out a practical check of the possibility of creating blocks for the formation and processing of signal of the wireless modem with PPM of the LFM signal using the microcontrollers.

3 MATERIALS AND METHODS

To conduct research, we define the following parameters of the LFM signal: an average frequency and frequency deviation are ten kHz, an impulse duration is 0.1 sec. To carry out a physical simulation of the operation of the wireless remote control modem with PPM of the LFM signal, it is necessary to choose a controller based on which the modem will be implemented. For determinate conditions, it is advisable to consider of use of the microcontroller ESP-WROOM-32. It is based on the popular two-core 32-bit Xtensa LX6 processor with a variable clock frequency from 80 MHz to 240 MHz. ESP-WROOM-32 has a RAM of 520 kilobytes and a rich peripheral that includes such inter-faces as UART, SPI, I2C, I2S, a twelve-bit ADC, and an eight-bit DAC [12]. Choosing this controller is also advisable in case of use on the moving object the latest small size camera module ESP32-CAM. Based on this module, it is also possible to build a demodulator of the receiving device.

ESP32-CAM can be widely used in various the Internet of Things (IoT) applications, suitable for home smart devices, industrial wireless control, wireless monitoring, wireless positioning system signals and ets. Using low-power dual-core 32-bit CPU, can be used as an application processor. Main frequency up to 240MHz, built-in 520 KB SRAM, external 8MB PSRAM. Support UART/SPI/I2C/PWM/ADC/DAC and other interfaces and OV2640 and OV7670 cameras with picture WiFi upload [13].

In addition, by choosing these controllers, we can provide a connection to a Wi-Fi access point, or transfer data via the Bluetooth interface. Moreover, the use of the combination of the IoT and edge computing will allow control of network security and access to the data transmission channel [14]. The presence of the SPI interface allows you to connect a transceiver module that uses “LoRa” technology or an NRF24L01 radio module to the microcontroller. Based on the NRF24L01 radio module, it is possible to organize a mini-network in which up to 6 transmitters and 1 receiver can work on one frequency. Line-of-sight communication ranges up to 100 m is provided [15]. ESP32 microcontrollers also use ESP-NOW technology – a simplified WiFi communication protocol with the transfer of short packets between devices. The ESP-NOW technology provides communication distance near 190m in conditions of non-existence of local internet network [16].

The process of formation and optimal processing of the LFM signal is not much different from the methods used in radiolocation [17]. To use the digital method of forming and processing the LFM signal, consider the mathematical representation of the signal. Analytically, a single rectangular radio pulse with LFM duration τ_c is described by the expression (1).

The law of change of the frequency of the LFM signal can be obtained by differentiating the phase of the LFM signal (1):

$$f(t) = f_c + \frac{\Delta f}{\tau_c} \cdot t \,. \qquad (2)$$

DOI 10.15588/1607-3274-2023-3-20

The impulse characteristic of the matched filter of the signal with LFM, within the limits of the signal duration τ_c has the form

$$V(t) = CU_m \times \times \cos\left(2\pi\left(f_c(t_0 - t) + \frac{\Delta f}{2\tau_c}(t_0 - t)^2\right)\right). \quad (3)$$

The law of changing the frequency of the pulse characteristic of the matched filter of the signal with LFM at $t_0 = \tau_c/2$ will have the form

$$f_V(t) = f_c + \frac{\Delta f}{2} - \frac{\Delta f}{\tau_c}t. \quad (4)$$

and within the duration of the signal will decrease linearly in time, in contrast to the frequency of the signal with LFM, which increases linearly with time.

The synthesis of the optimal filter of the LFM signal can be carried out both according to its frequency and impulse characteristics. As is known, the frequency characteristic of the optimal filter is determined by the expression

$$\dot{K}(f) = C\dot{g}^*(f)\exp\{-j2\pi f t_0\}. \quad (5)$$

The frequency spectrum of the LFM-radio pulse is given by the expression

$$\dot{g}^*(f) = \frac{U_m\sqrt{\tau_c\Delta f}}{2\Delta f} \times \times \exp\left\{-j2\pi\frac{(f - f_c)^2}{2\Delta f}\tau_c\right\}, \quad |f - f_c| \le \frac{\Delta f}{2}. \quad (6)$$

Then the frequency characteristic of the optimal filter matched to the LFM signal will be represented by the expression:

$$\dot{K}(f) = \frac{CU_m\sqrt{\tau_c\Delta f}}{2\Delta f} \times \times \exp\left\{j2\pi\left[\frac{(f - f_c)^2}{2\Delta f}\tau_c - ft_0\right]\right\}, \quad |f - f_c| \le \frac{\Delta f}{2}. \quad (7)$$

The voltage of the signal component at the output of the optimal filter at time t can be obtained using the inverse Fourier transform of the frequency spectrum of the signal at the filter output:

$$w(t) = C\int_{-\infty}^{\infty}\dot{g}(f)\dot{g}^*(f)\exp\{j2\pi f(t - t_0)\}df = = C\frac{\sin(\pi\Delta f(t - t_0))}{\pi\Delta f(t - t_0)}\cos(2\pi f_c(t - t_0)). \quad (8)$$

The voltage at the input and output of the optimal filter of the LFM signal is shown in Fig. 1.

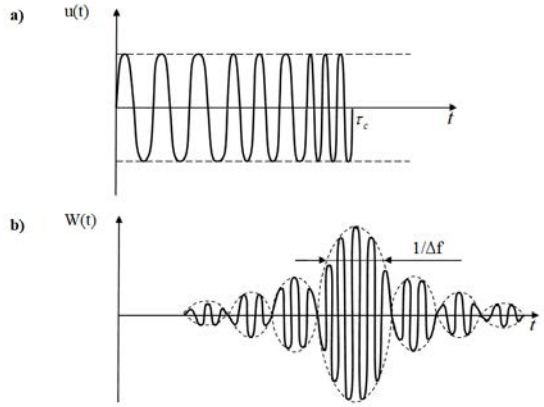


Figure 1 – Compression of the LFM signal in the optimal filter

An LFM signal is applied to the input of the filter (Fig. 1, a), and the output signal of the optimal filter (Fig. 1, b) is a harmonic oscillation $\cos(2\pi f_c t)$ with the envelope signal of the form $\sin(x/x)$. The maximum value of the output signal is reached at the moment ending of the input signal τ_c. The duration of the output pulse τ is inversely proportional to the width of the signal spectrum $1/\Delta f$. Since the base of the LFM signal $B = \tau_c\Delta f$, then $\tau = \frac{1}{\Delta f} = \frac{\tau_c}{B}$. Thus, the duration of the signal at the output of the optimal filter is less in base times than at its input. At the same time, the signal-to-noise ratio by the power at the output of the optimal filter will be greater than the signal-to-noise ratio by the power at its input.

4 RESULTS

A simplified structural diagram of the transceiver, which uses the PPM of the LFM signal, is shown in Fig. 2. The transmitter includes a modulator with phase-pulse modulation and the generator LFM. Since the phase-pulse modulation involves a change in the time position of the LFM pulse by a certain amount corresponding to the command number, then the modulator contains a “command-time” converter and a pulse shaper for starting the LFM generator. Since the phase-pulse modulation involves a change in the time position of the LFM pulse by a certain amount corresponding to the command number, then the modulator contains a “command-time” converter and a pulse shaper for starting the LFM generator. If a digital method of signal formation is used in the LFM generator, then the transmitter additionally includes a digital-to-analog converter (DAC), a fre-

DOI 10.15588/1607-3274-2023-3-20

quency converter, a high-frequency amplifier, and a band-pass filter. The receiver consists of a high-frequency amplifier, a frequency converter, an analog-to-digital converter (ADC), a matched LFM signal filter, and a demodulator, which contains a converter "time-command". At the same time, the PPM modulator and the LFM signal generator can be implemented on one ESP32 microcontroller. The LFM signal can be generated at an intermediate frequency and with a defined deviation value. The frequency converter will be carried out the generated signal to the frequency of the channel selected for operation. In the receiver, the selection of the required frequency channel will be carried out by changing the frequency of the local oscillator. From the output of the intermediate frequency amplifier, the signal is sent to the ADC, where it is converted to a digital form. An ADC, a matched LFM filter and a demodulator can be implemented on an ESP32CAM microcontroller. This controller also functions as a decoder for control commands coming to the flight controller.

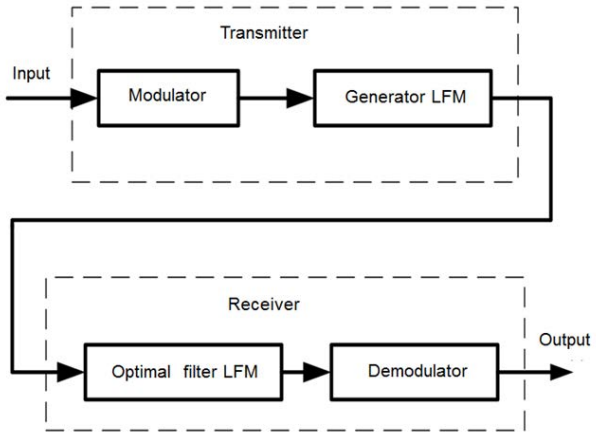


Figure 1 – Structural diagram of the receiver-transmitter

The results of mathematical modeling of the laws of change in time of the frequency of the LFM signal with a deviation of 10 kHz and the impulse characteristic of the filter matched to this signal using Excel are shown in Fig. 3. As can be seen, the LFM signal with a duration of $\tau_c = 0.1$ sec relative to the average signal frequency of 10 kHz occupies the frequency range Δf=10 kHz. Therefore, the width of the spectrum occupied by the LFM signal is determined by the frequency deviation Δf and does not depend on the duration of the signal τ_c.

The impulse characteristic of the optimal filter of the LFM signal is a mirror image of the input signal, which is evidenced by the nature of its frequency change.

Expression (1) was implemented programmatically based on the ESP-32 microcontroller. For a better visualization of the law of frequency change over time, the value of the frequency of the LFM signal variable from zero to one kHz. The voltage values of the signal, which output to the serial port of the Arduino-IDE software environment from the DAC output of the microcontroller, are shown in Fig. 4. As can be seen from the given data, the change in voltage over time at the analog output of the microcontroller is a signal with a linearly increasing frequency value within the duration of the signal, which corresponds to the analytically calculated value according to expression (1).

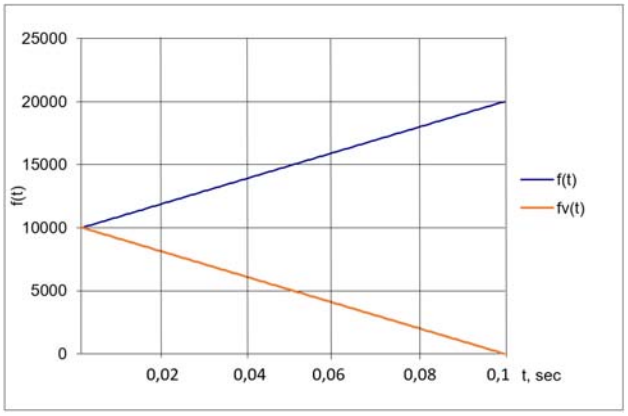


Figure 2 – Changing the frequency of the LFM signal and the impulse characteristic of the matched filter

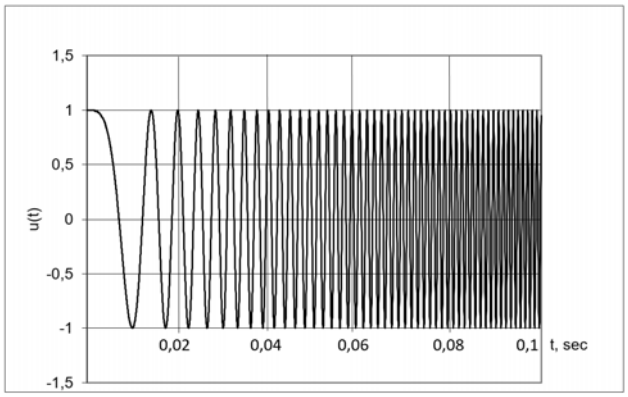


Figure 3 – Changing in time of the LFM signal

The result of the optimal processing of the LFM signal for command number two from the output of the serial port of the Arduino-IDE software environment is shown in Fig. 5. The value of the voltage at the output of the optimal processing block of the LFM signal for command number two is shown in Fig. 6. As can be seen from the given data, indeed, the duration of the signal at the output of the optimal processing block of the LFM signal is shorter than at its input. The time interval between the maximum values of the output signal of the optimal processing block is equal to 0.2 sec, which corresponds to command number two and proves the correct operation of the PPM modulator.

Fig. 7 show oscillograms that can be used to determine the operation of the remote control modem for the case of command number two. At the same time, it can be seen from the upper (violet) oscillogram that the pulse repetition interval of the LFM pulses is equal is 0.2 seconds, which corresponds to the command number two.

DOI 10.15588/1607-3274-2023-3-20

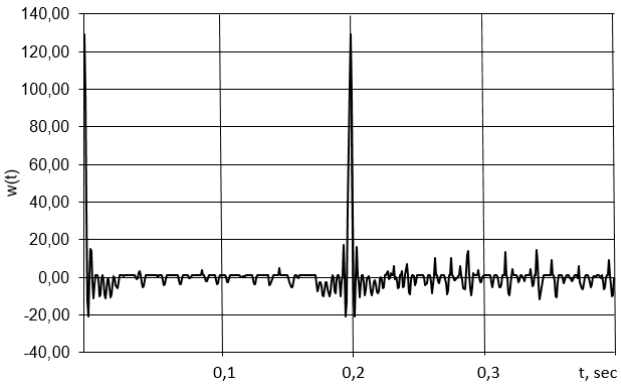


Figure 5 – The result of the optimal processing of the LFM signal for command number two

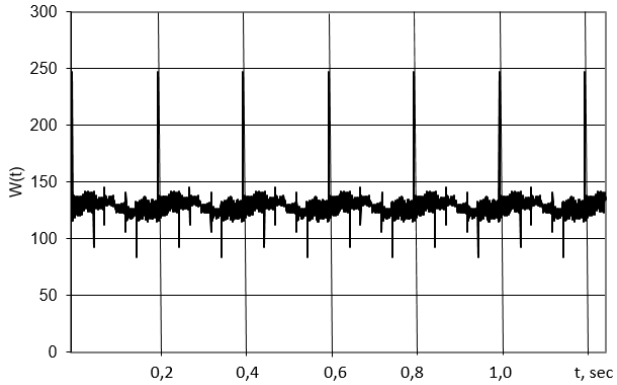


Figure 6 – The value of the voltage at the output of the optimal processing block of the LFM signal for command number two

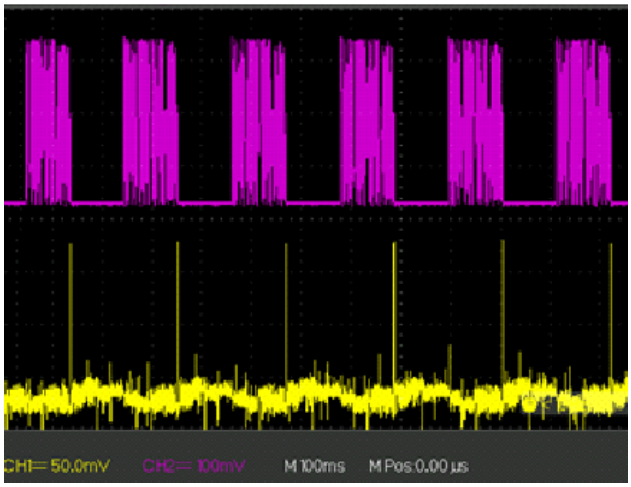


Figure 7 – The oscillograms of the voltage at the input and output of the optimal processing block of the LFM signal for command number two

The voltage at output of the optimal processing block of the LFM signal is showed to the lower (yellow) channel of the digital oscilloscope with use the second DAC of the ESP-32 microcontroller. The maximum value of the output signal is corresponds to the end of the input LFM signal. At the same time, the duration of the compressed pulse is much shorter than the duration of the LFM signal, which was equal to 0.1 sec. This corresponds to the graphs of the voltage at the output of the optimal processing block of the LFM signal shown in Fig. 6.

Testing of the modem operation was also carried out for the number one command with the minimum pulse repetition interval of the LFM signal, and the oscillograms corresponding to the conditions of transmission of the command number one are shown in Fig.8. The results of testing the low-frequency part of the wireless remote control modem for the conditions of transmission of the command number three are shown in Fig. 9.

Analysis of the given results confirms the possibility of using the proposed method of command transmission to control a moving object using a PPM of the LFM signal.

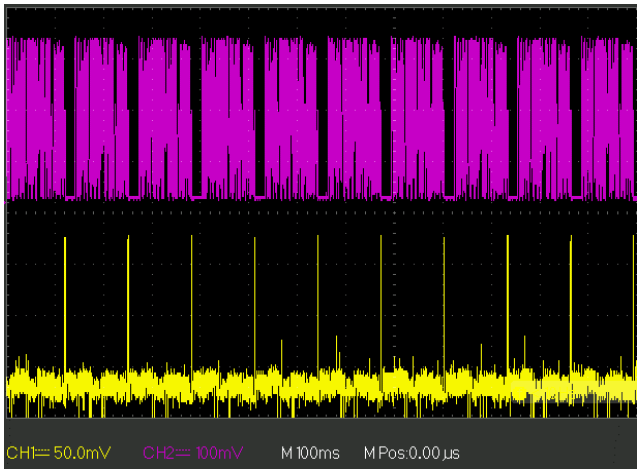


Figure 8 – The oscillograms of the voltage at the input and output of the optimal processing block of the LFM signal for command number one

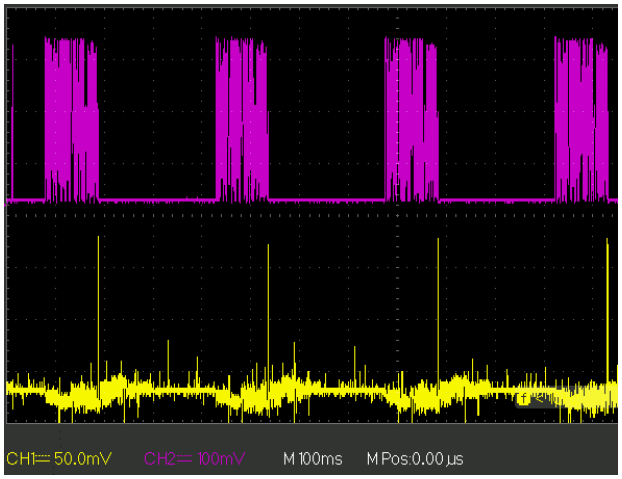


Figure 9 – The oscillograms of the voltage at the input and output of the optimal processing block of the LFM signal for command number three

Therefore, the possibility of conducting a practical test of a low-frequency part of the wireless remote control modem using an ESP32 microcontroller has been proven.

DOI 10.15588/1607-3274-2023-3-20

6 DISCUSSION

The paper investigates the possibility of using a PPM of the LFM signal to transmit commands for controlling a moving object. The use of digital methods makes it possible to form and process the LFM signal with defined parameters at a reduced frequency. Transfer to the operating frequency and power amplification will provide high-frequency cascades. Thus, the wireless remote control modem of the moving object consists of the low-frequency and high-frequency parts. The low-frequency part of the modem performs PPM of the LFM signal on at a reduced frequency. Each command corresponds to the given value of the change in the time position of the LFM pulse. The duration of command transmission does not exceed one second. During this time, several LFM pulses are received, which can be used to increase the reliability of identification of the command number. However, without the development of a suitable demodulator, it is almost impossible to gain unauthorized access to the radio control channel.

Therefore, the conducted physical simulation using the ESP-32 microcontroller proved the correct operation of the PPM modulator, the generator LFM signal and the DAC of the transmitter. In the receiving part the testing were covered the ADC and the optimal processing block of the LFM signal. The practical implementation of the modem on the ESP 32 microcontroller showed that the DAC and an ACD have a rather low conversion rate. The frequency deviation of the LFM pulse does not exceed 10 kHz. Therefore, even when forming the LFM signal with an average frequency of 10 kHz, it is quite difficult to ensure the required number of transformations during the duration of the signal. This leads to slight distortions of the LFM signal at the output of the DAC of the microcontroller, which can be observed on oscillograms. However, these distortions did not significantly affect the operation of the optimal processing block of the LFM signal.

CONCLUSIONS

The scientific problem The scientific problem by the need to carry out a practical check of the signal formation and processing blocks of the wireless remote control modem by means of physical modeling using microcontrollers. The scientific novelty of obtained results is that using phase-pulse modulation of the linear-frequency modulated signal for the transmission of control commands is firstly proposed. This method of transmitting the information component, unlike the existing methods, does not require changing the parameters of the linear frequency modulated signal. The use in the receiver of optimal processing of a linear-frequency-modulated signal of sufficiently big base will allow of the wireless transmission of the commands for controlling a moving object in conditions of intentional interference.

The practical significance. Used the possibility of a PPM of the LFM signal for the wireless transmission of the commands for controlling a moving object has been proven.

Conducted a practical test of the wireless remote control modem using an ESP32 microcontroller. The results of the physical modeling of the transmitting and receiving parts of the radio modem are presented.

The direction of further research consists in the development of the high-frequency part of the modem and checking the possibility of using of the wireless transmission of the commands for controlling a moving object in conditions of intentional interference.

ACKNOWLEDGEMENTS

We thank the management of Zhytomyr Polytechnic State University for the opportunity to conduct scientific research. University has a patent number 118728 from 02/25/2019 “Device for receiving broadband signals with linear frequency modulation” on which the proposed method of the wireless transmission of the control commands is based.

REFERENCES

1. Romanenko I., Shyshatskyi A. Analysis of modern condition of military radiocommunication system, *Advanced Information Systems*, 2017, 1(1), pp. 28–33. DOI: 10.20998/2522-9052.2017.1.05
2. Yaacoub J. P., Noura H., Salman O., Chehab A. Security analysis of drones systems: Attacks, limitations, and recommendations, *Internet of Things*, 2020, 11. DOI: 10.1016/j.iot.2020.100218
3. Iman A., Bernard K. An overview of Ultra-Wideband Technology and its Applications, 2022. [Electronic Publication] URL: https://www.lumenci.com/research-articles/ultrawideband.
4. Wang J. J. H. Stealth Communication Via Smart Ultra-Wide-Band Signal in 5G, Radar, Electronic Warfare, etc., *2020 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting*. Montreal, QC, Canada.-2020, pp. 1825–1826. DOI: 10.1109/IEEECONF35879.2020.9330108.
5. Kuvshynov O., Shyshatskyi A., Zhuk O. et al. Development of a method of increasing the interference immunity of frequency-hopping spread spectrum radio communication devices, *Eastern-European Journal of Enterprise Technologie*, 2019, 2/9 (98), pp. 74–84. DOI: 10.15587/1729-4061.2019.160328
6. Parekha C.D., Patel J.M. IEEE 802.11ac WLAN Simulation in MATLAB, *International Journal of Emerging Technologies and Innovative Research*, 2018, 5(10), pp. 96–100. URL: https://www.jetir.org/papers/JETIR1810621.pdf
7. LoRa and LoRaWAN: A Technical Overview, 2020. URL: https://lora-developers.semtech.com/uploads/documents/files/LoRa_and_LoRaWAN-A_Tech_Overview-Downloadable.pdf.
8. Hanif M. and Nguyen H. H. Slope-Shift Keying LoRa-Based Modulation, *IEEE Internet of Things Journal*, 2021, 8(1), pp. 211–221. DOI: 10.1109/JIOT.2020.3004318.
9. Augustin A., Yi J., Clausen T., Townsley W. M. A Study of LoRa: Long Range & Low Power Networks for the Internet of Things, *Sensors*, 2016, 16(9):1466. DOI: 10.3390/s16091466.
10. Andreiev O. V., Dubyna O. F., Nikitchuk T. M., Tsyporenko V. V. Using Walsh Functions for Increase the Stealth Communication in a Digital Radio Channel”, *Visnyk NTUU KPI Seria – Radiotekhnika Radioaparatobuduvan-*

DOI 10.15588/1607-3274-2023-3-20

nia, 2021, 85, pp. 27–32. DOI: 10.20535/RADAP.2021.85.27-32.
11. Recommendation ITU-R SM.329-12. Unwanted emissions in the spurious domain. SM Series Spectrum management. Electronic Publication. Geneva, 2014. URL: https://www.itu.int/dms_pubrec/itu-r/rec/sm/R-REC-SM.329-12-201209-I!!PDF-E.pdf
12. ESP32 datasheet, URL:http://espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf.
13. ESP32-CAM camera development board, URL: https://docs.ai-thinker.com/en/esp32-cam.
14. Lobanchykova1 N. M., Pilkevych I. A., Korchenko O. Analysis and protection of IoT systems: Edgecomputing and decentralized decision-making, *Journal of Edge Computing*, 2022, 1, pp. 55–67. DOI: 10.55056/jec.573
15. Mobasshir Mahbub Design and Implementation of Multi-purpose Radio Controller Unit Using nRF24L01 Wireless Transceiver Module and Arduino as MCU, *International Journal of Digital Information and Wireless Communications*, 2019, 8(1), pp. 61–72. DOI: 10.17781/P002598
16. Pasic R., Kuzmanov I., Atanasovski K. ESP-NOW communication protocol with ESP32, *Journal of Universal Excellence*, 2021, 1, pp. 53–60. DOI: 10.37886/ip.2021.019.
17. Guan Chengyu, Zhou Zemin, Zeng Xinwu Optimal Waveform Design Using Frequency-Modulated Pulse Trains for Active Sonar, *Sensors*, 2019, 19(19):4262. DOI: 10.3390/s19194262

Received 23.05.2023.
Accepted 15.08.2023.

УДК 621.396.2

ВИКОРИСТАННЯ МІКРОКОНТРОЛЕРА ESP32 ДЛЯ ФІЗИЧНОГО МОДЕЛЮВАННЯ БЕЗДРОТОВОГО МОДЕМА ДИСТАНЦІЙНОГО КЕРУВАННЯ

Вакалюк Т. А. – д-р пед. наук, професор, професор кафедри інженерії програмного забезпечення, Державний університет «Житомирська політехніка», Житомир, Україна.

Андреєв О. В. – канд. техн. наук, доцент, доцент кафедри комп'ютерних технологій у медицині та телекомунікаціях, Державний університет «Житомирська політехніка», Житомир, Україна.

Нікітчук Т. М.– канд. техн. наук, доцент, декан факультету інформаційно-комп'ютерних технологій, Державний університет «Житомирська політехніка», Житомир, Україна.

Осадчий В. В. – д-р пед. наук, професор, декан факультету економіки та управління, Київський університет імені Бориса Грінченка, Київ, Україна.

Дубина О. Ф. – канд. техн. наук, доцент, доцент кафедри комп'ютерних технологій у медицині та телекомунікаціях, Державний університет «Житомирська політехніка», Житомир, Україна.

АНОТАЦІЯ

Актуальність обумовлена необхідністю практичної реалізації теоретичних положень запропонованого способу бездротової передачі команд керування рухомим об'єктом в умовах навмисного втручання.

Метою роботи є практична перевірка можливості використання фазово-імпульсної модуляції лінійно-частотно модульованого сигналу для передачі команд керування шляхом фізичного моделювання з використанням мікроконтролерів.

Метод. Аналітичні розрахунки зміни в часі напруги на вході та виході пристрою оптимальної обробки лінійно-частотно модульованого сигналу проводились виключно з використанням обчислювальних можливостей мікроконтролера. Графіки змін у часі відповідних параметрів побудовані за допомогою програми Excel з використанням даних, що виводились у монітор послідовного порту програмного середовища Arduino IDE. Для контролю роботи бездротового модему передачі команд управління рухомому об'єкту використовувався цифровий осцилограф з USB-хостом.

Результати. Аналітичними розрахунками та фізичним моделюванням, з використанням сучасного мікроконтролера, доведено працездатність низькочастотної частини бездротового модему дистанційного керування з використанням фазово-імпульсної модуляції лінійно-частотно модульованого сигналу.

Висновки. Розглянуто можливість використання фазово-імпульсної модуляції лінійно-частотно модульованого сигналу для передачі команд керування. Цей спосіб передачі інформаційної складової, на відміну від існуючих способів, не потребує зміни параметрів лінійно-частотно модульованого сигналу. Використання в приймачі оптимальної обробки лінійно-частотно модульованого сигналу достатньо великої бази дозволить здійснювати бездротове передавати команд керування рухомим об'єктом в умовах навмисного втручання. Використання сучасних мікроконтролерів дозволило провести практичну перевірку працездатності низькочастотної частини бездротового модему дистанційного керування з фазово-імпульсною модуляцією лінійно-частотно модульованого сигналу шляхом фізичного моделювання.

КЛЮЧОВІ СЛОВА: бездротовий зв'язок, дистанційне керування, лінійно-частотна модуляція, фазово-імпульсна модуляція, моделювання.

ЛІТЕРАТУРА

1. Analysis of modern condition of military radiocommunication system / [I. Romanenko, A. Shyshatskyi] // Advanced Information Systems. – 2017. – 1(1). – P. 28–33. DOI: 10.20998/2522-9052.2017.1.05
2. Security analysis of drones systems: Attacks, limitations, and recommendations / [J. P. Yaacoub, H. Noura, O. Salman, A. Chehab] // Internet of Things. – 2020. – 11. DOI: 10.1016/j.iot.2020.100218
3. Iman A. An overview of Ultra-Wideband Technology and its Applications / A. Iman, K. Bernard // 2022. [Electronic Publication] URL: https://www.lumenci.com/research-articles/ultrawideband.
4. Wang J. J. H. Stealth Communication Via Smart Ultra-Wide-Band Signal in 5G, Radar, Electronic Warfare, etc. / J. J. H. Wang // 2020 IEEE International Symposium on Antennas and Propagation and North American Radio Science Meeting, Montreal, QC, Canada. – 2020. – P. 1825–1826. DOI: 10.1109/IEEECONF35879.2020.9330108.
5. Development of a method of increasing the interference immunity of frequency-hopping spread spectrum radio communication devices / [O. Kuvshynov, A. Shyshatskyi,

DOI 10.15588/1607-3274-2023-3-20

O. Zhuk et al.] // Eastern-European Journal of Enterprise Technologie. – 2019. – 2/9 (98). – P.74–84. DOI: 10.15587/1729-4061.2019.160328
6. Parekha C. D. IEEE 802.11ac WLAN Simulation in MATLAB / [C. D. Parekha, J. M. Patel] // International Journal of Emerging Technologies and Innovative Research. – 2018. – 5(10). – P. 96–100. URL: https://www.jetir.org/papers/JETIR1810621.pdf
7. LoRa and LoRaWAN: A Technical Overview, 2020. URL: https://lora- developers.semtech.com/uploads/documents /files/LoRa_and_LoRaWAN-A_Tech_Overview-Downloadable.pdf.
8. Hanif M. Slope-Shift Keying LoRa-Based Modulation / [M. Hanif and H. H. Nguyen] // IEEE Internet of Things Journal. – 2021. – 8(1). – P. 211–221. DOI: 10.1109/JIOT.2020.3004318.
9. A Study of LoRa: Long Range & Low Power Networks for the Internet of Things / [A. Augustin, J. Yi, T. Clausen, W.M. Townsley] // Sensors.–2016.–16(9):1466. DOI: 10.3390/s16091466.
10. Using Walsh Functions for Increase the Stealth Communication in a Digital Radio Channel"/ [O. V. Andreiev, O. F. Dubyna, T. M. Nikitchuk, V. V. Tsyporenko] // Visnyk NTUU KPI Seria – Radiotekhnika Radioaparatobuduvannia. – 2021. – 85. – P. 27–32. DOI: 10.20535/RADAP.2021.85.27-32.
11. Recommendation ITU-R SM.329-12. Unwanted emissions in the spurious domain. SM Series Spectrum management. Electronic Publication, Geneva, 2014. URL: https://www.itu.int/dms_pubrec/itu-r/rec/sm/R-REC-SM.329-12-201209-I!!PDF-E.pdf
12. ESP32 datasheet, URL:http://espressif.com/sites/default/ files/documentation/esp32_datasheet_en.pdf.
13. ESP32-CAM camera development board, URL: https://docs.ai-thinker.com/en/esp32-cam.
14. Analysis and protection of IoT systems: Edgecomputing and decentralized decision-making / [N. M. Lobanchykova1, I. A. Pilkevych, O. Korchenko] // Journal of Edge Computing. – 2022. – 1. – P. 55–67. DOI: 10.55056/jec.573
15. Mobasshir Mahbub Design and Implementation of Multipurpose Radio Controller Unit Using nRF24L01 Wireless Transceiver Module and Arduino as MCU // International Journal of Digital Information and Wireless Communications. – 2019. – 8(1). – P. 61–72. DOI: 10.17781/P002598
16. Pasic R. ESP-NOW communication protocol with ESP32 / R. Pasic, I. Kuzmanov, K. Atanasovski // Journal of Universal Excellence. – 2021. – 1. – P. 53–60. DOI: 10.37886/ip.2021.019.
17. Optimal Waveform Design Using Frequency-Modulated Pulse Trains for Active Sonar / [Chengyu Guan, Zemin Zhou, Xinwu Zeng] // Sensors. – 2019. – 19(19):4262. DOI: 10.3390/s19194262s

DOI 10.15588/1607-3274-2023-3-20

Наукове видання

Радіоелектроніка,

інформатика,

управління

№ 3/2023

Науковий журнал

Оригінал-макет підготовлено у редакційно-видавничому відділі НУ «Запорізька політехніка»

Свідоцтво про державну реєстрацію
КВ № 24220-14060 ПР від 19.11.2019.

Підписано до друку 14.09.2023. Формат 60×84/8.
Папір офс. Різогр. друк. Ум. друк. арк. 24,99.
Тираж 300 прим. Зам. № 684.

69063, м. Запоріжжя, НУ «Запорізька політехніка», друкарня, вул. Жуковського, 64

Свідоцтво суб’єкта видавничої справи
ДК № 6952 від 22.10.2019.