

Oliinyk A. O.¹, Skrupsky S. Yu.², Shkarupylo V. V.³, Blagodariov O. Yu.⁴¹PhD., Associate Professor of Department of Software Tools, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine²PhD, Associate Professor of Computer Systems and Networks Department, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine³PhD, Associate Professor of Computer Systems and Networks Department, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine⁴Postgraduate Student of Department of Software Tools, Zaporizhzhia National Technical University, Zaporizhzhia, Ukraine

PARALLEL MULTIAGENT METHOD OF BIG DATA REDUCTION FOR PATTERN RECOGNITION

Context. The problem of feature selection for big data processing based on the multi-agent approach and parallel computation has been solved. The object of research is the process of feature selection. The subject of the research are the methods of feature selection.

Objective. The purpose of the work is to create a parallel multi-agent method for reducing of big data sets.

Method. The article deals with the parallel multi-agent method for reducing of big data sets. The developed method involves splitting multiple agents into several subsets for parallel search of an informative combination of features in different areas of the search space. At the same time, it is suggested that the parallel nodes of the computer system perform the most resource-intensive operations associated with estimating the current set of agents, as well as the need to create and modify new sets of solutions based on stochastic computations. This allows to speed up the process of multi-agent search of informative combination of features, as well as to reduce the practical threshold for application of the multi-agent method with indirect communication between agents for reducing big data sets.

Results. The software which implements the proposed method and allows to select informative features based on the multi-agent approach and parallel computation has been developed.

Conclusions. The conducted experiments have confirmed the proposed software operability and allow recommending it for use in practice for solving the problems of big data processing for pattern recognition. The prospects for further research may include the modification of the developed parallel method for feature selection by using different criteria for estimation of the group information of features, as well as an experimental study of proposed method on more complex practical problems of different nature and dimensionality.

Keywords: agent, data set, feature selection, parallel computing, multi-agent approach, pattern recognition.

NOMENCLATURE

A – set of agents;

CPU – Central Processing Unit;

GPU – Graphical Processing Unit;

G_k – value of objective function of k -th agent;

$J(Xe)$ – criterion for assessing the significance of a set of features Xe ;

M – number of features in the sample of observations S ;

N_{er} – number of incorrectly recognized (classified)

observations of the sample $S = \langle P, T \rangle$ on the synthesized model;

N_{pr} – number of processes, on which task is performed;

N_χ – total number of agents in set A ;

p_{qm} – value of m -th feature (attribute) for q -th

observation ($m = 1, 2, \dots, M$, $q = 1, 2, \dots, Q$);

Q – number of observations in the given sample of observations S ;

S – sample of observations (training sample);

t_q – value of output parameter of q -th observation;

$t_{q \text{ mod}}$ – value of output parameter of q -th observation, calculated from the synthesized model;

T – set of output parameter values;

χ_k – k -th agent in set A ;

XS – set of all possible combinations of features, obtained from the initial set of characteristics P .

INTRODUCTION

The development of automated systems for pattern recognition is associated with the need of big data

processing [1–6]. Typically, the original samples of data describing the objects or processes under investigation may contain redundant and uninformative information [7–13]. The use of such information in the synthesis of recognition models leads to an increase of their complexity and redundancy, as well as a reduction in their generalizing abilities. Therefore, before synthesis of recognition models, it is relevant to perform preprocessing of data in order to exclude uninformative and redundant features from training samples [3, 5, 8].

Typically, well-known methods of feature selection use a greedy search strategy [6, 8], which often doesn't allow choosing the most informative combination of features. Stochastic methods [6, 12] are highly iterative and require substantial outlays of computing and time resources, making them difficult to use in practice. A multiagent method for feature selection that doesn't use greedy search strategy, based on modeling of agents movement in search space through stochastic computations, is offered in [14]. However, this method no longer adequate to handle big data sets due to the high iterative and consistent nature of the calculations.

It is therefore appropriate to parallelize the most computationally complex and resource-consuming operations of the multi-agent method of feature selection [14], which will reduce the practical threshold for applicability of such method when big data processing.

The purpose of the work is to create a parallel multi-agent method for big data sets reducing.

1 PROBLEM STATEMENT

Let there is a set of observations S (1):

$$S = \langle P, T \rangle. \quad (1)$$

Then the problem of informative features selection in an idealized formulation [8, 15–17] can be represented as: find a combination of features X^* from the original data set $S = \langle P, T \rangle$, at which the minimum of given criterion (2) for assessing the quality of feature set is achieved:

$$J(X^*) = \min_{Xe \in XS} J(Xe). \quad (2)$$

The error of the synthesized model, normally, is used as an criteria for assessing the significance of features set $J(Xe)$ [4–6]:

– recognition error (in problems with discrete output T) [4–6], calculated according to the formula (3):

$$E = \frac{N_{er}}{Q}; \quad (3)$$

– standard error (in the case where the output parameter T can take real values from a certain range $T \in [t_{\min}; t_{\max}]$) [6], calculated according to the formula (4):

$$E = \frac{1}{Q} \sum_{q=1}^Q (t_q - t_{q \text{ mod}})^2. \quad (4)$$

In calculating criteria values (3) and (4), the model is synthesized based on the data of the training sample S , using only the features that correspond to the combination Xe .

2 LITERATURE REVIEW

Currently, there are various methods of reducing the dimension of the feature space [3–6, 15–17]: brute-force method, depth-first search, breadth-first search, branch and bound, group method of data handling, method of sequential addition of features, method of sequential removal of features, method of alternately adding and removing of features, ranking of features, features clustering, method of random search with adaptation and evolutionary search for the selection of features [15–17].

Usually, in solving the problems of recognition, there is a system of statistically dependent features, whose set informativity isn't expressed through the informativeness of individual features. Therefore, in solving practical problems, it is necessary to evaluate a set of features, rather than each attribute separately. Thus, the use of ranking criteria on the computed individual assessment is unacceptable [6].

The application of brute-force method requires the evaluation of all possible combinations of features made up of the original data, which makes it impossible to use this approach with a large number of features in the source set, since it requires huge computational costs.

Heuristic search methods [15, 17] are not effective enough, because of sub-optimal of a greedy search strategy involving the sequential addition or deletion of one feature at each iteration, therefore of which the resulting set of characteristics contains redundant features that correlate with other features in the set.

It seems expedient to use the methods of stochastic search (in particular, the multi-agent approach) [13, 14, 18], to search such combination of informative features under

conditions of mutual dependence on each other. Since such methods are more suitable for finding new solutions by combining the best solutions that were obtained at different iterations and have the opportunity to exit from local optima.

The proposed multi-agent method [14] with an indirect communication between agents allows to select the most significant features. However, the method takes considerable time in processing of big data, because it is highly iterative and provides for sequential implementation of calculations. To address these drawbacks it is advisable to parallelize the multi-agent search of most meaningful combination of features in the processing of big data sets.

3 MATERIALS AND METHODS

In the developed parallel multi-agent method for big data reducing, it is proposed to split a set of agents $A = \{\chi_1, \chi_2, \dots, \chi_{N_\chi}\}$ into several subsets for parallel search of informative combination of features in different areas of search space. Meanwhile, to speed up the process of multi-agent search for informative combination of features in the developed method, it is suggested to fulfil the most resource-intensive operations related with evaluation of the current set of agents, including the need to create and modify of new sets of solutions based on stochastic computations on the nodes of the parallel computing system.

As noted previously, in the proposed parallel multi-agent method for big data reducing after initialization phase, a set of agents $A = \{\chi_1, \chi_2, \dots, \chi_{N_\chi}\}$ is split into several subsets

$$A^{(1)}, A^{(2)}, \dots, A^{(N_{pr})} \quad (5)$$

$$A \rightarrow \{A^{(1)}, A^{(2)}, \dots, A^{(N_{pr})}\}. \quad (5)$$

The total number of agents N_χ in the partition $A \rightarrow \{A^{(1)}, A^{(2)}, \dots, A^{(N_{pr})}\}$ remains unchanged (6):

$$|A^{(1)}| + |A^{(2)}| + \dots + |A^{(N_{pr})}| = N_\chi. \quad (6)$$

The partitioning $A \rightarrow \{A^{(1)}, A^{(2)}, \dots, A^{(N_{pr})}\}$ is done so that to ensure the separation of groups of agents $A^{(j)}$, over the search space with view to more detailed investigation of its various areas. For selected groups of compactly located agents $A^{(j)}$, it is proposed to apply cluster analysis methods [6, 9, 16].

After that, in each of the received subsets of agents $A^{(j)}$ ($j = 1, 2, \dots, N_{pr}$), it is proposed to conduct a multi-agent search to select informative features from the given data samples $S = \langle P, T \rangle$, periodically checking the stopping criteria and, if necessary, combining the subsets $A^{(j)}$.

In order to increase the efficiency of the multi-agent search for combination of informative features (reducing the search time), it is expedient to parallelize the most resource-intensive operations.

As noted above, the multi-agent method of feature selection with indirect link between agents, involves the implementation of initialization, chemotaxis modeling, reproduction, elimination and dispersion, checking the stopping criteria, and restarting agents.

In the initialization phase, the main parameters of method are defined, and the begin coordinates of agents χ_k ($k=1,2,\dots,N_\chi$) are randomly generated, after which values of the objective function $G_k = G(\chi_k)$ are calculated. This stage doesn't involve complex iterative computational procedures, so it is proposed to perform it on the main stream. Note that with a large number of agents N_χ , as well as processing complex samples $S = \langle P, T \rangle$, you can parallelize the process of assessing the initial positions of agents χ_k (calculating the values of the objective function $G_k = G(\chi_k)$, $k=1,2,\dots,N_\chi$).

The stage of chemotaxis modeling is connected with iterative implementation of tumbling, moving and sliding operators for each agent, assuming the calculation of new values of the coordinates of agents $\chi_k^{(t+1)}$ in the search space. In addition, at this stage, the evaluation of new agent's positions $G_k^{(t+1)} = G(\chi_k^{(t+1)})$, suggesting the need to data sample $S = \langle P, T \rangle$ for each of the agents. The complexity of the computational procedures of this stage necessitates its parallel implementation for the possibility of performing multi-agent search for informative combination of features in subset of agents $A^{(j)}$ at the j -th node of the parallel system.

At the stage of reproduction, new set of agents $A = \{\chi_1, \chi_2, \dots, \chi_{N_\chi}\}$ is created by performing stochastic calculations. At the same time, complex procedures are performed to select the most suitable agents, select the parent agent pair, and directly generate new agents (new coordinate points in the search space).

The stage of exclusion and dispersion also involves the need to process the entire current set of agents $A = \{\chi_1, \chi_2, \dots, \chi_{N_\chi}\}$ in order to randomly change the coordinates of some agents to exit from local extremums.

Therefore, for the possibility of implement multi-agent optimization, it is suggested that the nodes of parallel computing system perform the stages of reproduction, elimination and dispersion, as well as checking the termination criteria. These stages are associated with the need to create and modify new sets of solutions based on stochastic computations and, together with the chemotaxis simulation stage, make it possible to search for a combination of informative features in each of the subsets N_{pr} of agents $A^{(j)}$ ($j=1,2,\dots,N_{pr}$).

Based on the mentioned above, we present a model of the multiagent search process for a combination of informative features in the form of Fig. 1.

One iteration of the multi-agent search of an informative combination of features $MAS(A^{(j)})$ among a set of agents $A^{(j)}$ on the j -th node of a computer system ($j=1,2,\dots,N_{pr}$) is schematically shown in Fig. 2. In so doing, the following notation is used in the Fig. 2: $Chem(A^{(j)})$ is a chemotaxis

modeling operator over multiple agents $A^{(j)}$; $Cross(A^{(j)})$ is a reproduction operator; $Disp(A^{(j)})$ is an exclusion and scattering operator; $Crit(A^{(j)})$ is an operator of checking the criteria for the completion of multi-agent search $MAS(A^{(j)})$.

As can be seen from Fig. 1, computationally complex operations that are decided to parallelize, are related with operation of chemotaxis simulation, reproduction, exclusion and dispersion. The choice of these operations for parallel implementation of calculations is also due to the fact that information which is easily amenable to parallelization and associated with the current set of agents (coordinates of points χ_k in the search space and the corresponding values of the objective function $G_k = G(\chi_k)$) is iteratively processed on them. In addition, on the nodes of the parallel computer system are also offered to perform a check of the stopping criteria in order to cessation of the multi-agent optimization in case an optimal result is obtained (a set of characteristics that is acceptable for the solution of the current practical task). Sequences of initialization and restart of agents are consistently performed.

It is suggested that after performing multi-agent search at N_{it} iterations on nodes of parallel computer system, the information about the current sets of agents $A^{(j)}$ ($j=1,2,\dots,N_{pr}$) and the values of their cost functions $G^{(j)}$ on the main process be transmitted. As a result of which, sets A (7) and G (8) are formed:

$$A = \bigcup_{j=1}^{N_{pr}} A^{(j)} = \{\chi_1, \chi_2, \dots, \chi_{N_\chi}\}, \quad (7)$$

$$G = \bigcup_{j=1}^{N_{pr}} G^{(j)} = \{G_1, G_2, \dots, G_{N_\chi}\}. \quad (8)$$

Then, the process of restarting agents is performed on the main process. In this case, unlike the known multi-agent method with indirect connection between agents BFO [18], a new set of agents is proposed to be formed in accordance with expressions (9) and (10):

$$A = A_{elite} \cup A_{rnd} \cup A_{cross}, \quad (9)$$

$$|A_{elite}| + |A_{rnd}| + |A_{cross}| = N_\chi, \quad (10)$$

where A_{elite} is a set of elite solutions χ_k (best agents by the value of the objective function G_k) in each of the subsets $A^{(j)}$:

$$A_{elite} = \{\chi_{elite1}, \chi_{elite2}, \dots, \chi_{eliteN_{pr}}\},$$

$$\chi_{elitej} : G(\chi_{elitej}) = \min_{\chi_k \in A^{(j)}} (G(\chi_k)), |A_{elite}| = N_{pr};$$

A_{rnd} is a set of solutions χ_k , that are randomly selected in each of the subsets $A^{(j)}$:

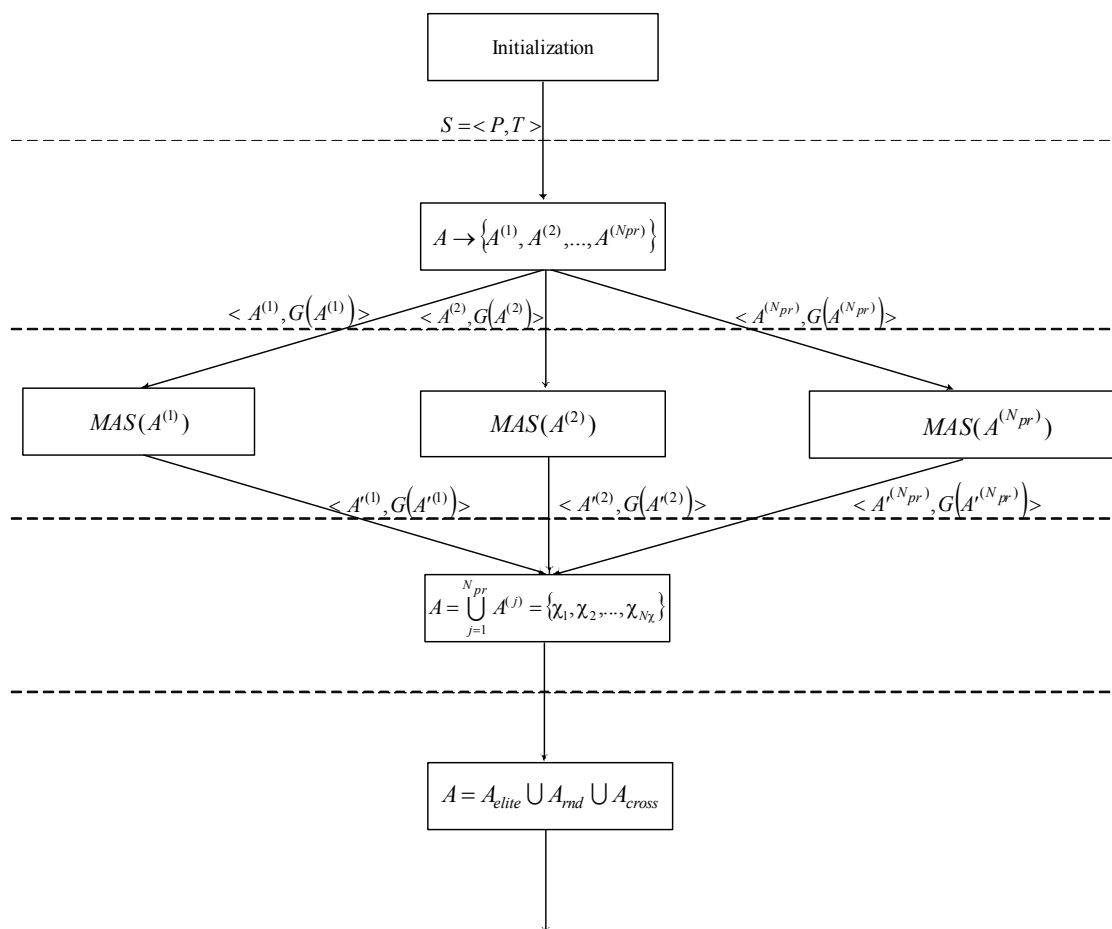


Figure 1 – The model of the multiagent search process for a feature selection

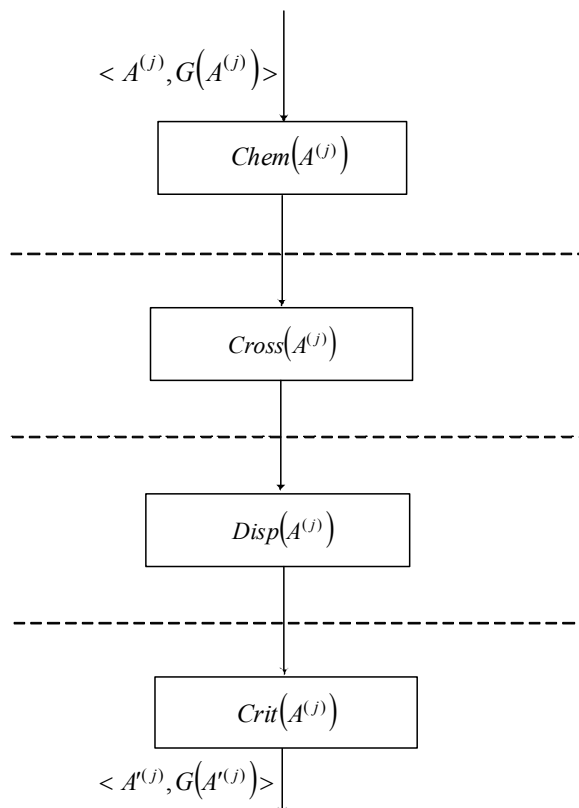


Figure 2 – One iteration of multi-agent search $MAS(A^{(j)})$ in a set of agents $A^{(j)}$

$A_{rnd} = \{\chi_{rnd1}, \chi_{rnd2}, \dots, \chi_{rnd|A_{rnd}|}\}$, $\chi_{rndj} = \text{rand}_{\chi_k \in A^{(j)}}(\chi_k)$,
 $|A_{rnd}| = \alpha_{rnd} N_\chi$, α_{rnd} is a fraction of agents in the new set $A = \{\chi_1, \chi_2, \dots, \chi_{N_\chi}\}$, randomly selected from the previous set of solutions. If $|A_{rnd}| > N_{pr}$ then random selection occurs successively among each of N_{pr} the subsets $A^{(j)}$, after which the process continues until the required number of solutions $|A_{rnd}|$ is reached, while the subset $A^{(j)}$ is randomly selected without repeating;

A_{cross} is a set of solutions χ_k , generated using a reproduction operator based on the current set of agents $A = \{\chi_1, \chi_2, \dots, \chi_{N_\chi}\}$. It is proposed to use a reproduction operator similar to that proposed in [14], using the evolutionary operators of proportional selection, arithmetic crossing and simple mutation [6, 8, 14];

$|A_{elite}|, |A_{rnd}|, |A_{cross}|$ are the total numbers of agents in sets A_{elite} , A_{rnd} and A_{cross} , accordingly.

Therefore, new set of agents $A = \{\chi_1, \chi_2, \dots, \chi_{N_\chi}\}$ in the proposed method is formed from elite agents A_{elite} , randomly selected agents from the previous set of decisions A_{rnd} , and also agents obtained by applying the reproduction operator A_{cross} . Such approach allows not only to save the current best solutions found in the search process, but also to provide a way out of local optima and the possibility of exploring different areas of search space.

The proposed parallel multiagent method of big data reduction involves splitting multiple agents into several subsets for parallel search of informative combination of features in different areas of the search space. At the nodes of parallel computing system, it is suggested to perform the most resource-intensive operations related to the evaluation of the current set of agents, as well as the need to create and modify new sets of solutions based on stochastic computations. This makes it possible to speed up the process of multi-agent search for informative combination of features, as well as to reduce the practical threshold for using a multi-agent method with indirect communication between agents to reduce big data sets.

Moreover, the proposed method fits well to SIMD (single instruction multiple data) architecture, because after the agent initialization stage the branches of algorithm perform the same actions on multiple data taking into account the stochastic component, which is data too. Hence, GPU implementation of the proposed method will probably demonstrate good computational process speedup.

4 EXPERIMENTS

The developed parallel multi-agent method of big data sets reduction has been implemented in C language using the MPI and CUDA libraries. Besides, the data exchange between the main core, which performed the initialization and agent restart stages, and other cores of cluster nodes

has been performed by multiple exchange functions of MPI library (Bcast, Gather, Scatter, Reduce). The GPU cores used the global memory of GPU, in which many agents were placed. Each core of GPU performed a separate branch of the algorithm, and since the number of cores in the GPU is counted as many hundred, as a result, a significant number of evaluations of feature combinations were performed in one cycle.

To check the effectiveness of the proposed method with the help of the developed software, a problem of selecting informative features for the recognition of vehicles was solved [19, 20]. The training sample contained information on 10,000 images obtained from highways, taken in gray. Using the recognition system [20], images of interest areas with a vehicle were identified on the images, which were manually classified by an expert person (recognition was carried out in the following classes: 0 – not recognized, 1 – motorcyclist, 2 – car, 3 – truck, 4 – bus, 5 – minivan or minibus) and were displayed into a matrix 128*128 (16384 points). The resulting images were transformed by calculating 26 characteristics that generalize graphic information about object [20]. For each class of vehicle, a different recognition model has been synthesized, which makes it possible to determine whether the recognizable means belong to this type. Thus, five training samples $S = \langle P, T \rangle$ of 10,000 instances were obtained, each of which was characterized by 26 features. The task was to reduce the number of attributes of training samples to identify the most informative combinations consisting of no more than 12 features. In the tables below, the results of selection of characteristics for the synthesis of recognition models defining the belonging of a motor vehicle to class 2 “passenger car” are presented.

To provide the experiments the following equipment has been involved:

- the cluster of Pukhov Institute for Modeling in Energy Engineering NAS of Ukraine, which involved 16 logical nodes (cores), each of which performed one process. The cluster configuration is as follows: processors Intel Xeon 5405, RAM – 4GB DDR-2 for each node, communication environment InfiniBand 20Gb/s, middleware Torque and OMPI;
- NVIDIA GTX 960 with 1024 thread’s processors.

To compare the proposed method with existing analogues, the separation of combination of features was performed using various methods: heuristic search, the method of main components analyzing, the method of group accounting of arguments, the canonical model of genetic search, multi-agent methods for selecting informative features with direct and indirect communication between agents and using the developed parallel multi-agent method of informative features selection with indirect communication between agents. Values of parameters of the multi-agent method with indirect communication between the agents were: $N_\chi = 100$; $N_{re} = 4$; $N_s = 4$; $N_c = 20$; $N_{ed} = 2$; $P_{ed} = 0,25$; $d_{attract} = 0,1$; $w_{attract} = 0,2$; $h_{repellant} = d_{attract}$; $w_{repellant} = 10$ [14, 18]. Since the optimization process of most of the listed methods is of a stochastic nature, the search for optimal solutions

during experiments was carried out 100 times, after which the average values of investigated parameters on basis of the obtained results were calculated. As criteria for assessing effectiveness of the selection of features followed were used [14, 18]:

- the number of accesses to the objective function N_{fit} necessary to achieve the result with required accuracy;
- the error of the method E , calculated by the formula (3);
- the operating time of the method T , necessary to achieve an acceptable solution;
- the number of features of k , selected by the method as informative.

As the objective function E , the probability of making erroneous decisions (3) for a neural network of direct propagation, containing 3 neurons on the first layer and one neuron on the second layer was used. The neural network was synthesized on the basis of the appropriate combination of characteristics of the sample $S = \langle P, T \rangle$, the

neuroelements had a logistic sigmoid activation function. As discriminant functions weighted sums were used.

The proposed method was performed on 1, 2, 4, 8, 12, 16 cluster nodes and the time spent on the method was recorded. In addition, speedup of the computational process and efficiency of the computer system were calculated. Moreover, communication overheads (transfers and synchronizations) were analyzed.

Similar experiments were performed using the GPU.

5 RESULTS

Table 1 presents the characteristic values of the proposed and known methods of informative features selection in the recognition of vehicles.

Figures 3 and 4 shows dependences between the proposed method execution time in seconds (T_{spent}) and the number of involved cluster nodes (fig. 3), and the number of GPU threads (fig. 4).

Table 1 – Characteristics of selection features methods for vehicles recognition

№	Method of selection of features	Values of comparison criteria			
		E	N_{fit}	T	k
1	Method of analyzing the main components (MAMC) [6, 15]	0.0412	–	524	12
2	Method of group accounting of arguments (MGAA) [5, 8, 17]	0.0358	9761	18578	11
3	Canonical method of evolutionary search (CMES) [6, 19]	0.0219	10000	23171	10
4	Method of alternately adding and removing of features (MARF) [6, 8]	0.0471	871	2199	12
5	Multiagent method with indirect connection between agents (MMICA) (Ant Colony Optimization for Feature Selection) [19]	0.0198	4872	10318	10
6	Multiagent method with direct connection between agents (MMDCA) (Bee Colony Optimization for Feature Selection) [19]	0.0191	4719	10192	11
7	Parallel multiagent method of big data sets reduction (PMMBDSR), single core implementation $N_{pr} = 1$	0.0172	5418	11461	10
	PMMBDSR, implementation in $N_{pr} = 12$		65017	1868	
	PMMBDSR, implementation in 60 threads of GPU		48216	7317	

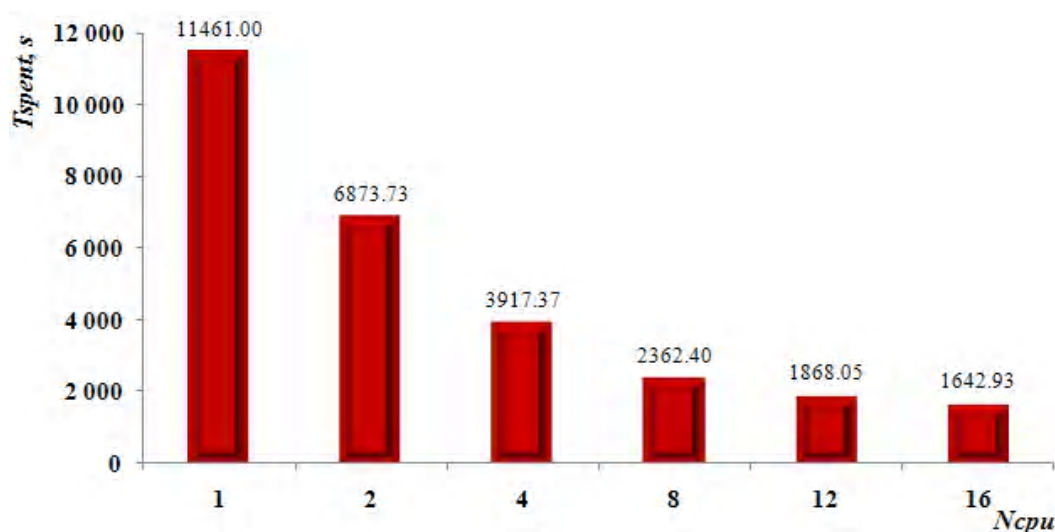


Figure 3 – Dependence between the proposed method execution time and the number of involved cluster nodes

In Figures 5 and 6 the speedup graphs of the computational process $Speedup_{pr}$ on the cluster and on the GPU, respectively, were shown.

Graph of the cluster efficiency $Efficiency_{pr}$ when performing the proposed method is shown in Figure 7.

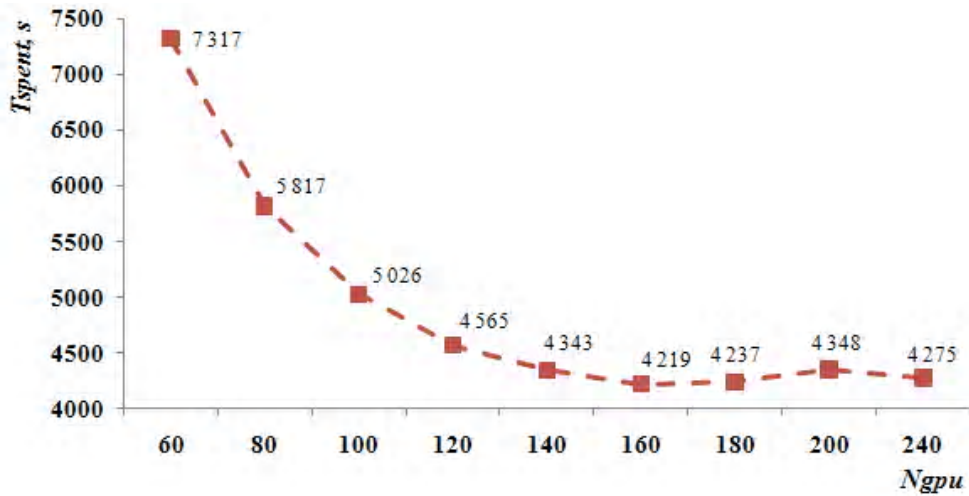


Figure 4 – Dependence between the proposed method execution time and the number of involved GPU threads

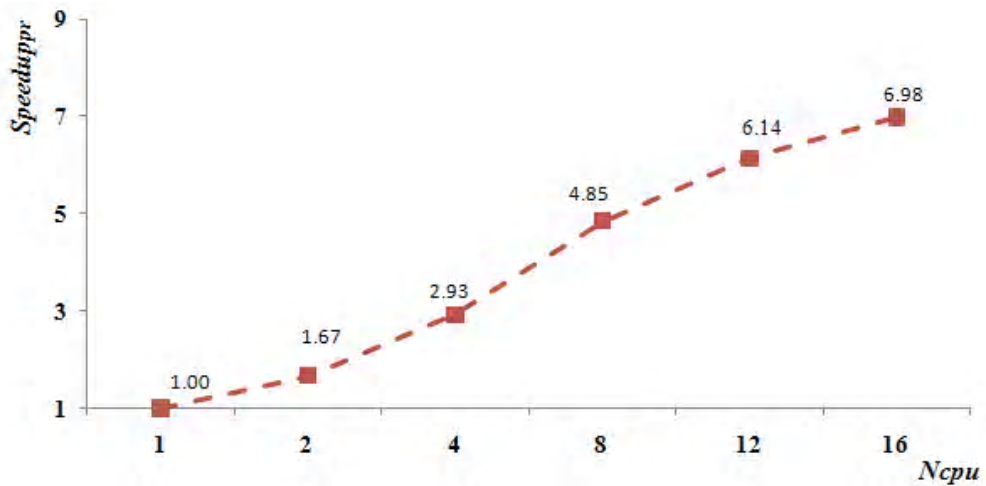


Figure 5 – Dependence between the computational speedup and the number of involved cluster nodes

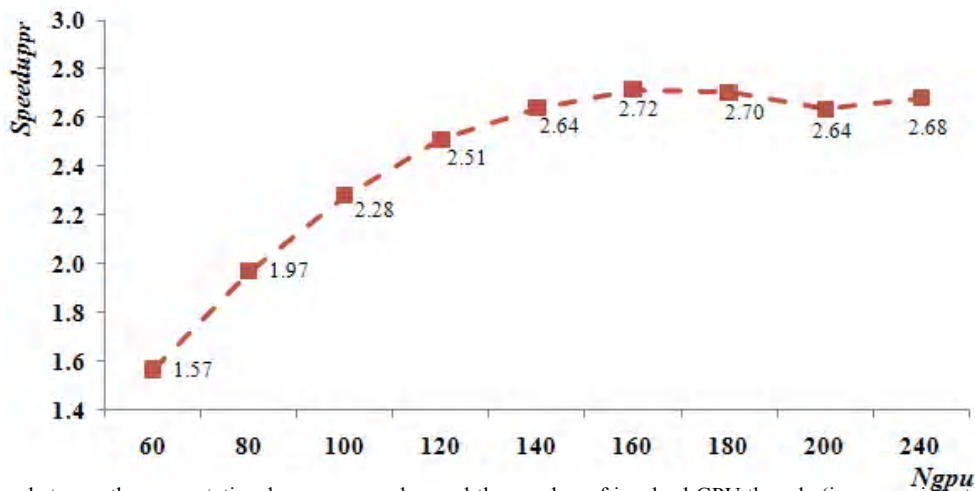


Figure 6 – Dependence between the computational process speedup and the number of involved GPU threads (in comparison to one cluster node)

As a result of the analysis of reasons of computer system efficiency decrease the corresponding graphs were formed. They show percentage of communication overheads *Overhead* (transfers and synchronizations) in the computational process from the number of involved cluster nodes (Figure 8) and GPU threads (Figure 9).

Graph of execution time *Tspent* of the proposed method from the number of agents, processed on each cluster core (Agents Per cluster Core) *ApC* is shown in the Figure. 10. The *ApC* criterion was calculated by the formula (9):

$$ApC = \frac{N_{\chi}}{N_{pr}} \quad (9)$$

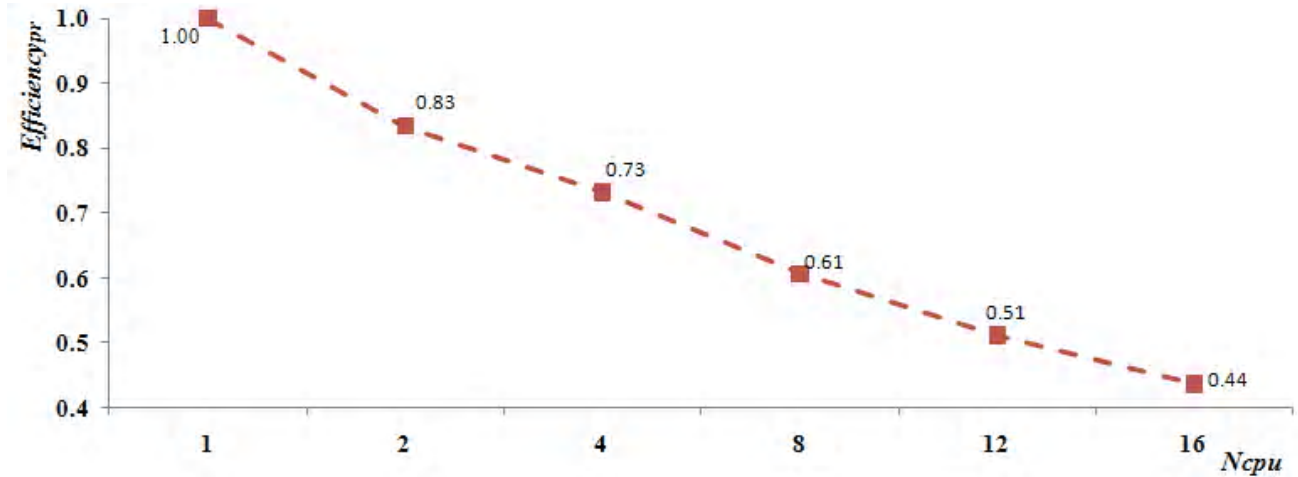


Figure 7 – Graph of the cluster efficiency when performing the proposed method

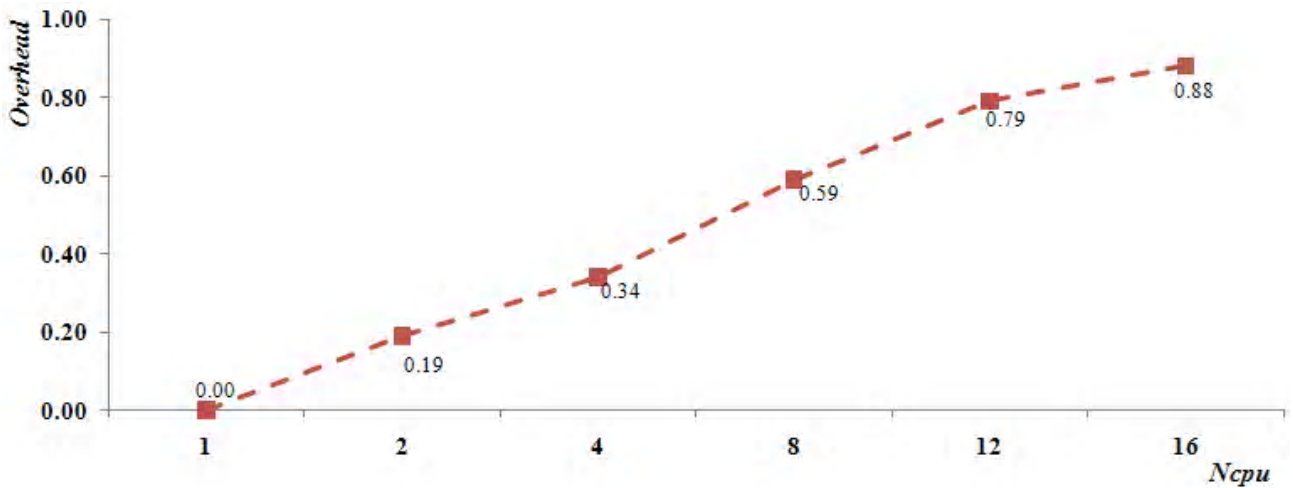


Figure 8 – Graph of communication overhead from the number of cluster nodes involved

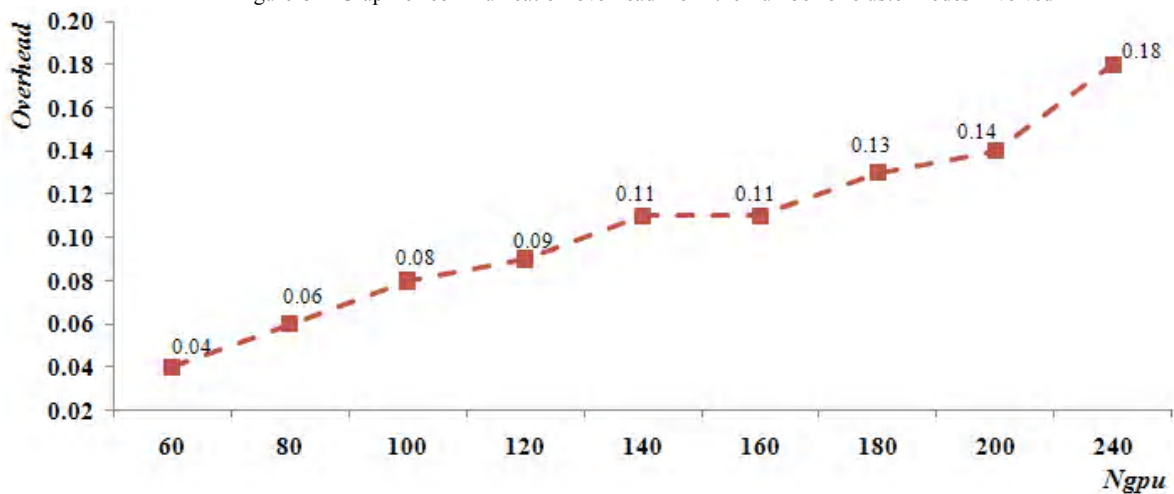


Figure 9 – Graph of communication overhead from the number of GPU threads involved

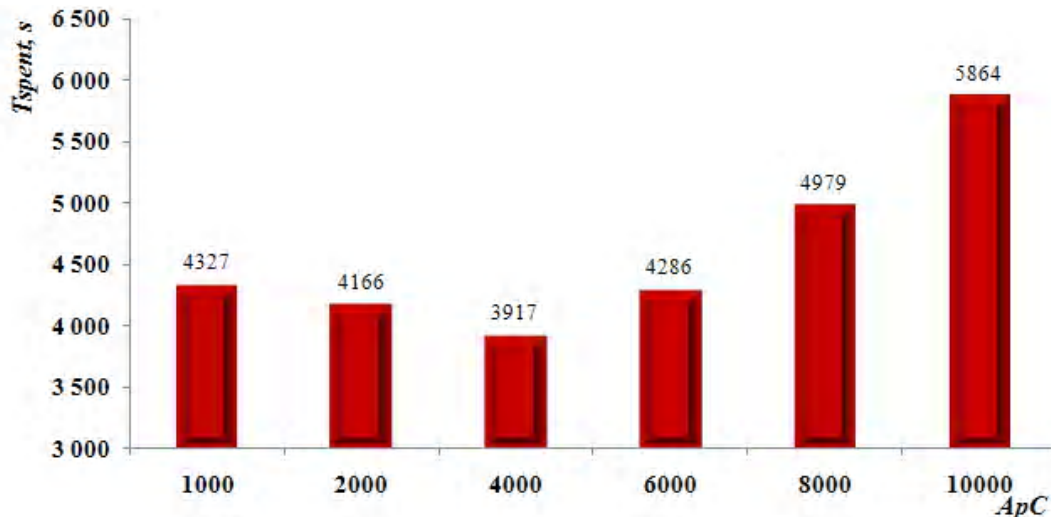


Figure 10 – Execution time graph of the proposed method from the number of agents per cluster core

6 DISCUSSION

As shown in Table 1 combinations of features which obtained using stochastic methods (CMES, MMICA, MMDCA, PMMBDSR) were characterized by more acceptable values of the objective function (for example, $E = 0.0172$ for PMMBDSR method) compared with the MAMC, MGAA and MARF methods (error value is varied in the range of $E = 0.0358$ for MGAA to $E = 0.0471$ for MARF). Such results are explained by a more effective analysis of search space using stochastic methods, including developed PMMBDSR method.

The operating time T of the methods varies from 524 s (for MAMC) to 23171 s (for CMES), which is related to number of calculations of the values of the objective function before the search stops. Such large operating time values of the selecting features methods are associated with the need to construct a recognition model based on each of the evaluated combinations of features. Obviously, sequential implementations of stochastic methods investigate a larger number of points in the search space, because of the stochastic nature, which causes large time consumption for their operation.

The results of Table 1 confirmed advisability of parallelizing the proposed method PMMBDSR: with number of parallel processes $N_{pr} = 12$, operation time of the method is 1868 s, that is less than the time of the heuristic method MARF ($T = 2199$ s)

The use of stochastic (evolutionary and multi-agent) methods allowed to select combinations that consist of fewer features ($k = 10$ for CMES, MMICA and PMMBDSR) compared with the use of other methods ($k = 11$ and 12 for other methods). This also indicates a more effective investigation of the features space using stochastic methods.

Consequently, the obtained values of the effectiveness estimation criteria of methods for selection informative features (E , N_{fit} , T , k) indicate the advisability of applying the proposed method for solving practical recognition problems.

Analysis of Figures 3–6 shows that the proposed method is well parallelized on a cluster and on a graphics processor. For instance, on 16 nodes of the cluster speedup of the computational process was 6.98, that allowed reducing the execution time of the method in computer system from 11461 s to 1643 s. On GPU, speedup of computational process compared to one core of the cluster was from 1.57 to 2.68, depending on the number of involved GPU threads.

Figure 7 shows a significant efficiency decrease of the computer system that performs the proposed method with an increase of the number of involved nodes. This is associated with an increase of overhead portion (synchronizations and transfers) with each new involved node of the system. Figure 8 confirms this fact. For example, when using 4 cluster nodes, the overhead portion in the general computing process is 0.34 and when using 16 cluster nodes is 0.88. As a result, growth in the number of involved nodes by 4 times (from 4 to 16) increases the speedup not linearly by 4 times, but only by about 2.38 times (Figure 5). On the GPU, the overheads increase with the number of involved threads not as significant as on the cluster (Figure 9). However, the capability of GPU in implementing of the proposed method is limited by the frequencies of the stream processors and by the bandwidth of data buses. As a result, the GPU managed to achieve the execution time of the proposed method, comparable to the four nodes of cluster (Figures 3, 4), which is an acceptable result.

The analysis of the runtime graph of the proposed method from the number of agents to the cluster core (Figure 10) shows that when the system's load is less than 4 000 agents per cluster core, the system is not fully loaded. For 4000 agents per core, the computer system is used efficiently, it finds a solution earlier than with a lesser load.

With a rise in the number of agents per core of more than 4 000, the time for finding solutions rises due to an increase in the overhead percentage. A concrete number of agents (in this case 4000), by which the system is used effectively depends on the capability of specific equipment. However, the form of the graph (as in Figure 10) should be preserved on other clusters.

CONCLUSIONS

The actual task of automation of large data sets reduction process based on the multi-agent approach has been solved

Scientific novelty lies in the fact that the parallel multiagent method of big data sets reduction has been proposed. The developed method involves splitting multiple agents into several subsets for parallel search of an informative combination of features in different areas of the search space. Wherein on parallel nodes of the computing system, it is suggested to perform the most resource-intensive operations related to the estimation of the current set of agents, as well as the need to create and modify new sets of solutions based on stochastic computations. This allows to speedup the process of multi-agent search for an informative combination of features, and also decreases the practical threshold for applying the multi-agent method with indirect communication between agents to reduce big amounts of data.

The practical value of the work lies in the fact that the program implementation of the proposed method for the CPU cluster and for the GPU has been developed. It allows to perform feature selection in a parallel computer system for significantly less time compared to other feature selection methods, implemented, as a rule, sequentially.

ACKNOWLEDGMENTS

The work was performed as part of research work “Methods and means of computational intelligence and parallel computing for processing large amounts of data in diagnostic systems” (number of state registration 0116U007419) of software tools department of Zaporizhzhia National Technical University.

REFERENCES

1. Salfner F. A survey of online failure prediction methods / F. Salfner, M. Lenk, M. Malek // ACM computing surveys. –2010. – Vol. 42, Issue 3. – P. 1–42. DOI: 10.1145/ 1670679.1670680.
2. Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms / J. C. Bezdek. – N.Y. : Plenum Press, 1981. – 272 p. DOI: 10.1007/978-1-4757-0450-1.
3. Bow S. Pattern recognition and image preprocessing / S. Bow. – New York: Marcel Dekker Inc., 2002. – 698 p. DOI: 10.1201/9780203903896.
4. Shin Y. C. Intelligent systems : modeling, optimization, and control / C. Y. Shin, C. Xu. – Boca Raton: CRC Press, 2009. – 456 p. DOI: 10.1201/9781420051773.
5. Bishop C. M. Pattern recognition and machine learning / C. M. Bishop. – New York : Springer, 2006. – 738 p.

6. Encyclopedia of machine learning / [eds. C. Sammut, G.I. Webb]. – New York : Springer, 2011. – 1031 p. DOI: 10.1007/978-0-387-30164-8.
7. Abonyi J. Cluster analysis for data mining and system identification / J. Abonyi, B. Feil. – Basel: Birkhäuser, 2007. – 303 p.
8. Jensen R. Computational intelligence and feature selection: rough and fuzzy approaches / R. Jensen, Q. Shen. – Hoboken: John Wiley & Sons, 2008. – 339 p. DOI: 10.1002/9780470377888.
9. Lee J. A. Nonlinear dimensionality reduction / J. A. Lee, M. Verleysen. – New York : Springer, 2007. – 308 p. DOI: 10.1007/978-0-387-39351-3.
10. Bodyanskiy Ye. A Multidimensional Cascade Neuro-Fuzzy System with Neuron Pool Optimization in Each Cascade / Ye. Bodyanskiy, O. Tyschenko, D. Kopaliani // Int. Journal of Information Technology and Computer Science (IJITCS). – 2014. – Vol. 6, No. 8. – P. 11–17. DOI: 10.5815/ijitcs.2014.08.02
11. Oliinyk A. Production rules extraction based on negative selection / A. Oliinyk // Radio Electronics, Computer Science, Control. – 2016. – № 1. – P. 40–49. DOI: 10.15588/1607-3274-2016-1-5.
12. Oliinyk A. The decision tree construction based on a stochastic search for the neuro-fuzzy network synthesis / A. Oliinyk, S. A. Subbotin // Optical Memory and Neural Networks (Information Optics). – 2015. – Vol. 24, № 1. – P. 18–27. DOI: 10.3103/S1060992X15010038.
13. Oliinyk A. Association Rules Extraction for Pattern Recognition / A. Oliinyk, S. A. Subbotin // Pattern Recognition and Image Analysis. – 2016. – Vol. 26, № 2. – P. 419–426.
14. Oliinyk A. O. Agent technologies for feature selection / A. O. Oliinyk, O. O. Oliinyk and S. A. Subbotin // Cybernetics and Systems Analysis. – 2012. – Vol. 48, Issue 2. – P. 257–267. DOI: 10.1007/s10559-012-9405-z.
15. Jolliffe I. T. Principal Component Analysis / I. T. Jolliffe. – Berlin : Springer-Verlag. – 2002. – 489 p.
16. McLachlan G. Discriminant Analysis and Statistical Pattern Recognition / G. McLachlan. – New Jersey : John Wiley & Sons. – 2004. – 526 p.
17. Guyon I. An introduction to variable and feature selection / I. Guyon, A. Elisseeff // Journal of machine learning research. – 2003. – № 3. – P. 1157–1182.
18. Kim D. H. Bacterial Foraging Based Neural Network Fuzzy Learning / D. H. Kim, C. H. Cho // Proceedings of the 2nd Indian International Conference on Artificial Intelligence (IICAI-2005). – Pune : IICAI, 2005. – P. 2030–2036.
19. Субботін С. О. Неітеративні, еволюційні та мультиагентні методи синтезу нечіткологічних і неймережних моделей: монографія / С. О. Субботін, А. О. Олійник, О. О. Олійник ; під заг. ред. С. О. Субботіна. – Запоріжжя : ЗНТУ, 2009. – 375 с.
20. Subbotin S. A. Synthesis of neuro-fuzzy models for the allocation and detection of objects on a complex background on the two-dimensional image / S. A. Subbotin // Computer modeling and intelligent systems : proceedings of the conference. – Zaporizhzhya: ZNTU, 2007. – P. 68–91.

Article was submitted 29.05.2017.

Олійник А. О.¹, Скруський С. Ю.², Шкарупило В. В.³, Благодарьов О. Ю.⁴

¹Канд.техн.наук, доцент кафедри програмних засобів, Запорізький національний технічний університет, Запоріжжя, Україна

²Канд. техн. наук, доцент кафедри комп'ютерних систем та мереж, Запорізький національний технічний університет, Запоріжжя, Україна

³Канд. техн.наук, доцент кафедри комп'ютерних систем та мереж, Запорізький національний технічний університет, Запоріжжя, Україна

⁴Аспірант кафедри програмних засобів, Запорізький національний технічний університет, Запоріжжя, Україна

ПАРАЛЕЛЬНИЙ МУЛЬТИАГЕНТНИЙ МЕТОД РЕДУКЦІЇ ВЕЛИКИХ МАСИВІВ ДАНИХ ДЛЯ РОЗПІЗНАВАННЯ ОБРАЗІВ

Актуальність. Вирішено задачу відбору інформативних ознак при обробці великих масивів даних на основі мультиагентного підходу та паралельних обчислень. Об'єкт дослідження – процес відбору інформативних ознак. Предмет дослідження – методи відбору інформативних ознак.

Мета роботи полягає в створенні паралельного мультиагентного методу редукції великих масивів даних.

Метод. Запропоновано паралельний мультиагентний метод редукції великих масивів даних. Розроблений метод передбачає розбиття множини агентів на декілька підмножин для паралельного пошуку інформативної комбінації ознак в різних областях простору пошуку. При цьому на паралельних вузлах обчислювальної системи запропоновано виконувати найбільш ресурсомісткі операції, пов'язані з оцінюванням поточної множини агентів, а також з необхідністю створення і модифікації нових множин рішень на основі стохастичних обчислень. Це дозволяє прискорити процес мультиагентного пошуку інформативної комбінації ознак, а також знизити практичний поріг застосування мультиагентного методу з непрямым зв'язком між агентами для редукції великих масивів даних.

Результати. Розроблено програмне забезпечення, яке реалізує запропонований метод і дозволяє виконувати відбір інформативних ознак на основі мультиагентного підходу і паралельних обчислень.

Висновки. Проведені експерименти підтвердили працездатність запропонованого математичного забезпечення та дозволяють рекомендувати його для використання на практиці при обробці великих масивів даних для розпізнавання образів. Перспективи подальших досліджень можуть полягати в модифікації розробленого методу шляхом використання різних критеріїв оцінювання групової інформативності ознак, а також експериментальному дослідженні запропонованого методу на більшому комплексі практичних завдань різної природи і розмірності.

Ключові слова: агент, вибірка даних, відбір ознак, паралельні обчислення, мультиагентний підхід, розпізнавання образів.

Олейник А. А.¹, Скрупский С. Ю.², Шкарупило В. В.³, Благодарев А. Ю.⁴

¹Канд. техн. наук, доцент кафедри програмних средств, Запорожский национальный технический университет, Запорожье, Украина

²Канд. техн. наук, доцент кафедры компьютерных систем и сетей, Запорожский национальный технический университет, Запорожье, Украина

³Канд. техн. наук, доцент кафедры компьютерных систем и сетей, Запорожский национальный технический университет, Запорожье, Украина

⁴Аспирант кафедры програмних средств, Запорожский национальный технический университет, Запорожье, Украина

ПАРАЛЛЕЛЬНЫЙ МУЛЬТИАГЕНТНЫЙ МЕТОД РЕДУКЦИИ БОЛЬШИХ МАССИВОВ ДАННЫХ ДЛЯ РАСПОЗНАВАНИЯ ОБРАЗОВ

Актуальность. Решена задача отбора информативных признаков при обработке больших массивов данных на основе мультиагентного подхода и параллельных вычислений. Объект исследования – процесс отбора информативных признаков. Предмет исследования – методы отбора информативных признаков.

Цель работы заключается в создании параллельного мультиагентного метода редукции больших массивов данных.

Метод. Предложен параллельный мультиагентный метод редукции больших массивов данных. Разработанный метод предполагает разбиение множества агентов на несколько подмножеств для параллельного поиска информативной комбинации признаков в различных областях пространства поиска. При этом на параллельных узлах вычислительной системы предложено выполнять наиболее ресурсоемкие операции, связанные с оценением текущего множества агентов, а также с необходимостью создания и модификации новых множеств решений на основе стохастических вычислений. Это позволяет ускорить процесс мультиагентного поиска информативной комбинации признаков, а также снизить практический порог применения мультиагентного метода с непрямой связью между агентами для редукции больших массивов данных

Результаты. Разработано программное обеспечение, которое реализует предложенный метод и позволяет выполнять отбор информативных признаков на основе мультиагентного подхода и параллельных вычислений.

Выводы. Проведенные эксперименты подтвердили работоспособность предложенного математического обеспечения и позволяют рекомендовать его для использования на практике при обработке больших массивов данных для распознавания образов. Перспективы дальнейших исследований могут заключаться в модификации разработанного метода путем использования различных критериев оценивания групповой информативности признаков, а также экспериментальном исследовании предложенного метода на большем комплексе практических задач разной природы и размерности.

Ключевые слова: агент, выборка данных, отбор признаков, параллельные вычисления, мультиагентный подход, распознавание образов.

REFERENCES

- Salfner F., Lenk M., Malek M. A survey of online failure prediction methods, *ACM computing surveys*, 2010, Vol. 42, Issue 3, pp. 1–42. DOI: 10.1145/1670679.1670680.
- Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms. N.Y. Plenum Press, 1981, 272 p. DOI: 10.1007/978-1-4757-0450-1.
- Bow S. Pattern recognition and image preprocessing. New York, Marcel Dekker Inc., 2002, 698 p. DOI: 10.1201/9780203903896.
- Shin Y. C., Xu C. Intelligent systems : modeling, optimization, and control. Boca Raton: CRC Press, 2009, 456 p. DOI: 10.1201/9781420051773.
- Bishop C. M. Pattern recognition and machine learning, New York, Springer, 2006, 738 p.
- Sammur C., Webb G. I. eds. Encyclopedia of machine learning. New York, Springer, 2011, 1031 p. DOI: 10.1007/978-0-387-30164-8.
- Abonyi J., Feil B. Cluster analysis for data mining and system identification. Basel, Birkhäuser, 2007, 303 p.
- Jensen R., Shen Q. Computational intelligence and feature selection: rough and fuzzy approaches. Hoboken, John Wiley & Sons, 2008, 339 p. DOI: 10.1002/9780470377888.
- Lee J. A., Verleysen M. Nonlinear dimensionality reduction. New York, Springer, 2007, 308 p. DOI: 10.1007/978-0-387-39351-3.
- Bodyanskiy Ye., Tyshchenko O., Kopaliani D. A Multidimensional Cascade Neuro-Fuzzy System with Neuron Pool Optimization in Each Cascade, *Int. Journal of Information Technology and Computer Science (IJITCS)*, 2014, Vol. 6, No. 8, pp. 11–17. DOI: 10.5815/ijitcs.2014.08.02
- Oliinyk A. Production rules extraction based on negative selection, *Radio Electronics, Computer Science, Control*, 2016, No. 1, pp. 40–49. DOI: 10.15588/1607-3274-2016-1-5.
- Oliinyk A., Subbotin S. A. The decision tree construction based on a stochastic search for the neuro-fuzzy network synthesis, *Optical Memory and Neural Networks (Information Optics)*, 2015, Vol. 24, No. 1, pp. 18–27. DOI: 10.3103/S1060992X15010038.
- Oliinyk A., Subbotin S. A. Association Rules Extraction for Pattern Recognition, *Pattern Recognition and Image Analysis*, 2016, Vol. 26, No. 2, pp. 419–426.
- Oliinyk A. O., Oliinyk O. O. and Subbotin S. A. Agent technologies for feature selection, *Cybernetics and Systems Analysis*, 2012, Vol. 48, Issue 2, pp. 257–267. DOI: 10.1007/s10559-012-9405-z.
- Jolliffe I. T. Principal Component Analysis. Berlin, Springer-Verlag, 2002, 489 p.
- McLachlan G. Discriminant Analysis and Statistical Pattern Recognition. New Jersey, John Wiley & Sons, 2004, 526 p.
- Guyon I., Elisseeff A. An introduction to variable and feature selection, *Journal of machine learning research*, 2003, No. 3, pp. 1157–1182.
- Kim D. H., Cho C. H. Bacterial Foraging Based Neural Network Fuzzy Learning, *Proceedings of the 2nd Indian International Conference on Artificial Intelligence (IICAI-2005)*. Pune, IICAI, 2005, pp. 2030–2036.
- Subbotin S., Oliinyk A., Oliinyk O. Noniterative, evolutionary and multi-agent methods of fuzzy and neural network models synthesis : monograph. Zaporizhzhya, ZNTU, 2009, 375 p. (In Ukrainian).
- Subbotin S. A. Synthesis of neuro-fuzzy models for the allocation and detection of objects on a complex background on the two-dimensional image, *Computer modeling and intelligent systems : proceedings of the conference*. Zaporizhzhya, ZNTU, 2007, pp. 68–91.

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PROGRESSIVE INFORMATION TECHNOLOGIES

УДК 528.29

Боровик О. В.¹, Рачок Р. В.², Дармороз М. М.³

¹Д-р техн. наук, професор, начальник кафедри інженерного забезпечення та технічних засобів охорони кордону,
Національна академія Державної прикордонної служби України
імені Богдана Хмельницького, Хмельницький, Україна

²Канд. техн. наук, доцент, докторант, Національна академія Державної прикордонної служби України імені Богдана
Хмельницького, Хмельницький, Україна

³Ад'юнкт, Національна академія Державної прикордонної служби України
імені Богдана Хмельницького, Хмельницький, Україна

ОЦІНКА ЕФЕКТИВНОСТІ ФУНКЦІОНУВАННЯ СИСТЕМИ ОПТИКО-ЕЛЕКТРОННОГО СПОСТЕРЕЖЕННЯ

Актуальність. В охороні державного кордону України використовується нова система оптико-електронного спостереження (СОЕС). Проте її потенційні можливості не реалізовані повною мірою. Однією з причин цього є відсутність науково-методичного апарату оцінки ефективності її функціонування при врахуванні різноманітних факторів природного і техногенного характеру. Останнє робить актуальним проведення відповідного дослідження.

Мета. Метою роботи є розробка методики оцінки ефективності функціонування СОЕС.

Метод. У роботі запропоновані показники ефективності функціонування окремих технічних засобів спостереження та їх комплексного використання з однієї вежі системи оптико-електронного спостереження. Також у роботі запропоновано новий показник ефективності функціонування системи оптико-електронного спостереження. У показнику враховано особливості ведення спостереження з комплексним використанням різних технічних засобів охорони кордону в різних умовах природного і техногенного характеру та з урахуванням рельєфу місцевості. Обґрунтування доцільності їх застосування передбачало використання ймовірно-статистичних методів.

Результати. На основі інформації про цифрові характеристики рельєфу досліджуваної ділянки здійснено визначення недоступних для спостереження ділянок місцевості, так званих «мертвих зон». Запропоновано методику обчислення показника ефективності функціонування системи оптико-електронного спостереження. Наведений опис алгоритмічної і програмної реалізації цієї методики. З використанням розробленого програмного забезпечення проведений аналіз особливостей функціонування існуючої системи оптико-електронного спостереження, визначені основні проблемні аспекти та напрями підвищення ефективності функціонування СОЕС.

Висновки. Сучасна побудова СОЕС не є оптимальною. Можливими напрямками підвищення ефективності її функціонування можуть бути оптимізація розстановки веж системи на місцевості, коригування висоти розміщення на вежах технічних засобів спостереження та вдалий підбір самих технічних засобів спостереження.

Ключові слова: система оптико-електронного спостереження, показник ефективності системи, математичні методи, алгоритм.

НОМЕНКЛАТУРА

E_0 – необхідний рівень ефективності СОЕС;
 E_ϕ – фактичний рівень ефективності СОЕС;
 m_R – математичне сподівання відстані до цілі;
 σ_R – середнє квадратичне відхилення відстані до цілі;
 p_0 – граничний рівень ефективності прикриття (ймовірність виявлення цілі);
 $P_1(R)$ – ймовірність виявлення цілі за допомогою РЛС;
 $P_2(R)$ – ймовірність виявлення цілі за допомогою тепловізійної камери;
 $P_3(R)$ – ймовірність виявлення цілі за допомогою тепловізійної камери;

ДПСУ – Державна прикордонна служба України;
РЛС – радіолокаційна станція;
СОЕС – система оптико-електронного спостереження;
ТЗОК – технічні засоби охорони кордону;
ТЗС – технічні засоби спостереження.

ВСТУП

Важливим фактором сучасного забезпечення прикордонної безпеки України є ефективне використання Державною прикордонною службою України технічних засобів охорони кордону. Але аналіз забезпечення підрозділів охорони кордону ТЗОК показує, що значна

частина цих засобів експлуатується понад 10–15 років і в будь-який час може вийти з ладу. Тому надзвичайно важливим є як оновлення існуючого парку ТЗОК, так і впровадження та ефективне використання кращого сучасного обладнання. Останнім часом на озброєння ДПСУ поступили нові ТЗОК в складі СОЕС. Однак виникає питання їх ефективного використання. Об'єктом дослідження є функціонування СОЕС, предметом дослідження є науково-методичний апарат оцінки ефективності функціонування СОЕС. Метою роботи є розробка методики оцінки ефективності функціонування СОЕС.

1 ПОСТАНОВКА ЗАДАЧІ

Завдяки технічній допомозі міністерства оборони США у рамках проекту міжнародної технічної допомоги ДПСУ було отримано високоточне обладнання системи оптико-електронного спостереження (СОЕС). Основою цієї системи є інтегровані вежі, на яких встановлені: радіолокаційні станції EL/M-2129 ELTA, електронно-оптичні камери, обладнання передачі даних. Однак ефективність використання цього сучасного обладнання значною мірою залежить від багатьох факторів: місць розташування веж спостереження, їх висоти, рельєфу місцевості, погодних умов, виду цілей. Попередній аналіз результатів оперативно-службової діяльності з застосуванням СОЕС свідчить про недостатню реалізованість потенціалу цієї системи. Для вдосконалення функціонування існуючої СОЕС та оптимізації побудови нових таких систем необхідно визначити показник ефективності роботи цієї системи та оцінити вплив на нього основних факторів.

2 ЛІТЕРАТУРНИЙ ОГЛЯД

Дослідженню ефективності використання засобів спостереження з урахуванням різноманітних факторів приділяли увагу багато науковців. Зокрема, в [1–3] розглядалися питання реконструкції рельєфу місцевості в задачах спостереження. Ефективність ведення спостереження прикордонниками з урахуванням рельєфу місцевості та можлива тактика дій контрабандистів досліджувалися в [4–8]. Проте в [1–8] не враховувалися особливості використання різних технічних засобів спостереження (ТЗС). Окремі аспекти функціонування таких засобів аналізувалися у роботах [9–12]. У цих працях увага в основному приділялася визначенню підходів до оцінки ефективності ведення спостереження з використанням окремих ТЗС.

Однак використання СОЕС, в складі якої комплексно задіяні різні ТЗС, гостро ставить питання оцінки ефективності її функціонування з урахуванням різних факторів, зокрема, рельєфу місцевості. Все це обумовлює актуальність саме такого дослідження СОЕС.

Зважаючи на це, метою даної роботи є визначення показника ефективності ведення спостереження з використанням СОЕС, формування підходів до його знаходження та аналіз впливу різних факторів на цю ефективність.

3 МАТЕРІАЛИ І МЕТОДИ

Виявлення цілей СОЕС будемо розглядати як випадковий процес, а дальність дії засобів спостереження – як випадкову величину, закономірності зміни якої вивчаються за допомогою методів теорії ймовірностей і математичної статистики. На вежах СОЕС застосовуються наступні засоби спостереження: радіолокаційна станція ELTA

ELM-2129, тепловізійна камера Axys EOSS 180S-DUTV, телевізійна камера. Отримаємо для цих ТЗС вирази визначення ймовірності виявлення цілі в залежності від дальності, типу цілі, умов, в яких здійснюється спостереження, стану засобу спостереження.

Якщо розглядати виявлення цілі як випадковий процес, який відбувається в достатньо однорідних типових умовах, то розподіл дальності виявлення описується нормальним законом. Тоді дальність дії засобів спостереження буде цілком описана з ймовірнісної точки зору, якщо буде встановлене математичне сподівання m_R (обумовлене впливом контрольованих факторів) і середнє квадратичне відхилення σ_R . Для отримання цих величин по кожному ТЗС в різних умовах ведення спостереження можливо використати апарат математичної статистики. Слід відмітити, що одні і ті ж погодні, техногенні умови чи період доби по-різному впливають на ймовірність виявлення для даних ТЗС. Будемо вважати їх вплив протягом певного проміжку часу незмінним. Тоді інтегральний закон розподілу, що характеризує ймовірність виявлення цілі до її підходу на відстань R , буде мати вигляд

$$F(R) = \frac{1}{\sigma_R \sqrt{2\pi}} \int_{R_{\min}}^R e^{-\frac{(R-m_R)^2}{2\sigma_R^2}} dR,$$

де $R_{\min} = m_R - 3\sigma_R$.

Ймовірність виявлення цілі на дальності R буде знаходитись за формулою

$$P(R) = 1 - \frac{1}{\sigma_R \sqrt{2\pi}} \int_{R_{\min}}^R e^{-\frac{(R-m_R)^2}{2\sigma_R^2}} dR. \quad (1)$$

У випадку, якщо умови, в яких функціонує СОЕС, є такими, що характеризуються суттєвою невизначеністю, то розподіл дальності виявлення підпорядковується закону Релея і

$$P(R) = 1 - \int_0^R \frac{R}{\sigma_R^2} e^{-\frac{R^2}{2\sigma_R^2}} dR. \quad (2)$$

Розглянемо комплексне використання всіх ТЗС, які встановлені на вежах СОЕС.

Позначимо ймовірності виявлення окремими ТЗС в певний момент часу, які обчислюються за (1) або (2), наступним чином:

– $P_1(R)$ – ймовірність виявлення цілі за допомогою РЛС;

– $P_2(R)$ – ймовірність виявлення цілі за допомогою тепловізійної камери;

– $P_3(R)$ – ймовірність виявлення цілі за допомогою телевізійної камери.

Тоді ймовірність виявлення цілі за допомогою засобів, встановлених на k -тій вежі СОЕС, може бути знайдена так:

$$P_k(R) = 1 - \bar{P}_1(R) \cdot \bar{P}_2(R) \cdot \bar{P}_3(R). \quad (3)$$

Цей вираз визначає технологію оцінки ефективності однієї вежі СОЕС, як сукупності ТЗОК, що реалізують функції виявлення цілей.

Однак СОЕС представляє собою систему веж. Розглянемо ефективність функціонування цієї системи з точки зору виявлення цілей в межах смуги прикриття, яку потрібно забезпечити. Для цього визначимо підхід щодо визначення ефективності спостереження за кожним елементом ділянки місцевості, який буде загальним для всіх засобів спостереження (рис. 1).

У лівій частині рис. 1 показана смуга прикриття, яку необхідно реалізувати. В межах цієї смуги має забезпечуватись граничний рівень ефективності прикриття (ймовірність виявлення цілі) p_0 . Наступний елемент рис. 1 – «мертві» для спостереження зони, які утворюються внаслідок перешкоджаючого впливу рельєфу місцевості. Зони ефективності на рис.1 утворюються на основі обчислення для кожної вежі СОЕС показника (3).

У правій частині рис. 1 демонструється підхід до отримання результуючої ефективності з урахуванням «мертвих» для спостереження зон. Для цих ділянок приймається $P(R)=0$.

Остаточна ефективність спостереження за певною ділянкою місцевості усіма ТЗС визначається наступним чином:

$$P(R) = 1 - \prod_k (1 - P_k^i(R)). \quad (4)$$

Слід відмітити, що P_k^i у (4) обчислюється за (3) у випадку, якщо досліджувана ділянка не належить множині «мертвих» зон. В протилежному випадку $P_k^i=0$.

Необхідний рівень ефективності СОЕС приймемо

$$E_0 = S_0 \cdot p_0. \quad (5)$$

Слід відмітити, що такий рівень забезпечуватиметься у випадку, коли відсутні «мертві» зони і по всій площі смуги перекриття забезпечується висока ймовірність виявлення цілей p_0 (звичайно $p_0=0,95$). На жаль, наявність

недоступних для спостереження зон та зниження ймовірності виявлення цілі зі збільшенням відстані суттєво знижує фактичний рівень ефективності СОЕС

$$E_{\Phi} = \sum_i S_i \cdot p_i. \quad (6)$$

Слід відмітити, що p_i у (6) обчислюється на основі (4) відповідно до координат i -тої ділянки. Коли відомий фактичний рівень ефективності СОЕС (6) та необхідний рівень ефективності СОЕС (5) можливо остаточно оцінити ефективність прикриття необхідної смуги перекриття наступним чином:

$$W = \frac{E_{\Phi}}{E_0} \cdot 100\%. \quad (7)$$

Слід відмітити, що процедура обчислення показника (7) є достатньо складною. Це обумовлено необхідністю визначення «мертвих», недоступних для спостереження зон, які залежать від місць розташування веж СОЕС, висоти вежі та рельєфу місцевості. Тому необхідна алгоритмічна і програмна реалізація визначення (7).

Блок-схема знаходження (7), яка представляє відповідну методику, наведена на рис. 2.

4 ЕКСПЕРИМЕНТИ

З метою програмної реалізації блок-схеми (рис. 2) було розроблене програмне забезпечення, в якому алгоритмічно реалізовано всі етапи розрахунку (7). Головне вікно програми представлено на рис. 3.

У верхній частині вікна знаходяться кнопки, з використанням яких реалізується завантаження у растрових форматах карти місцевості, висот та смуги перекриття. Після завантаження цих даних і визначення місць розташування веж СОЕС стає активною кнопка «Зони видимості». Обчислення цих зон проводиться на основі інформації про рельєф місцевості, координати веж СОЕС та висоти веж. Висоти місцевості отримані з проекту SRTM з просторовою роздільною здатністю 90 м.

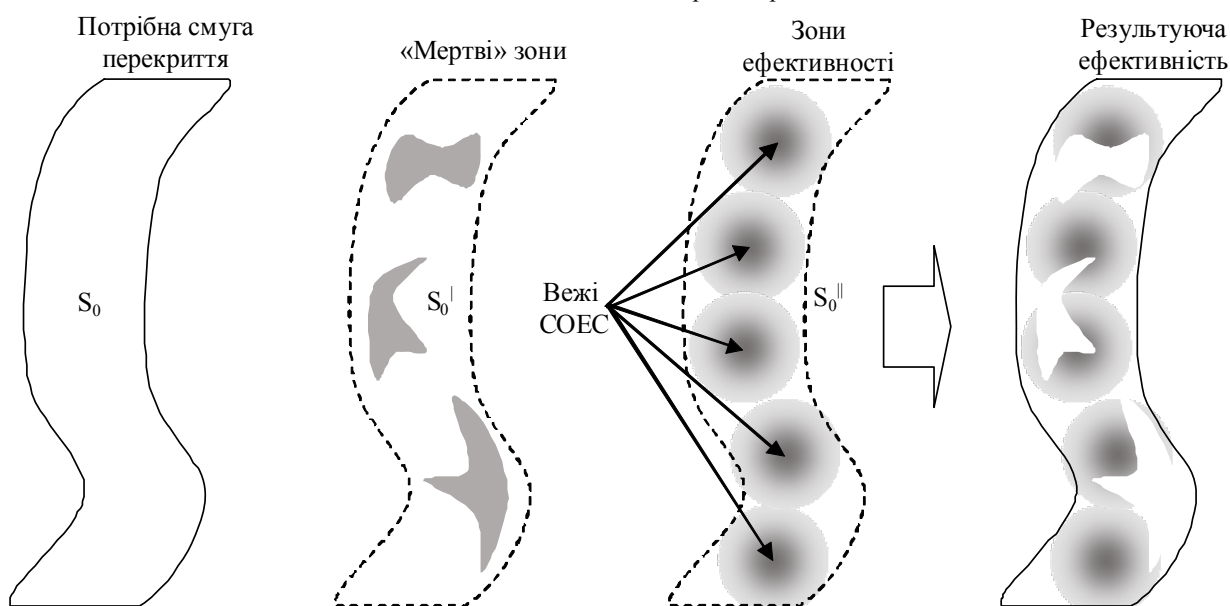


Рисунок 1 – Схема визначення ефективності виявлення цілі з використанням СОЕС

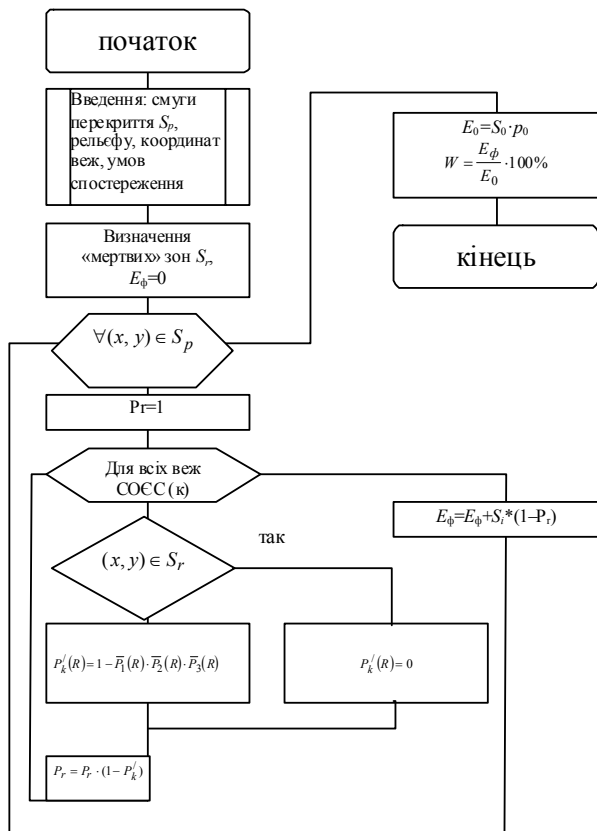


Рисунок 2 – Блок-схема обчислення ефективності прикриття необхідної смуги перекриття

Для визначення видимості ділянки місцевості, яка відповідає кожному пікселю, використовується авторський алгоритм, що ґрунтується на швидкому алгоритмі комп'ютерної графіки побудови відрізків на растрі – алгоритмі Брезенхема. Для кожної точки на місцевості від вежі спостереження до вказаної ділянки здійснюється обчислення висоти проходження променя (використовується рівняння прямої) і порівнюється з висотою рельєфу в цій точці. Якщо висота рельєфу є більшою – умова видимості порушується, перебір точок місцевості припиняється і досліджувана ділянка відноситься до «мертвої» зони. Порівняння програмної реалізації запропонованого алгоритму з програмною реалізацією інших аналогічних алгоритмів (r.viewshed в GRASS GIS) визначення зон видимості на одній обчислювальній платформі показала в декілька разів більшу швидкість.

Після визначення зон видимості стає доступною кнопка «Ефективність», яка запускає процедуру обчислення (7). При визначенні ефективності враховуються умови проведення спостереження (час доби, погодні умови), вид цілі і необхідний граничний рівень ефективності прикриття. Введення цих даних передбачене в головному вікні програми. Для кожного з варіантів умов проведення спостереження у програмі задані параметри m_R і σ_R , необхідні для обчислення (3).

З використанням картографічного сервісу Google Mars було визначено розміщення веж спостереження СОЕС. Результат обчислення показника (7) показав низьку ефективність існуючого прикриття необхідної смуги перекриття. У випадку малої цілі ця ефективність складає 47,57 що менше 50%. Однією з причин цього є нера-

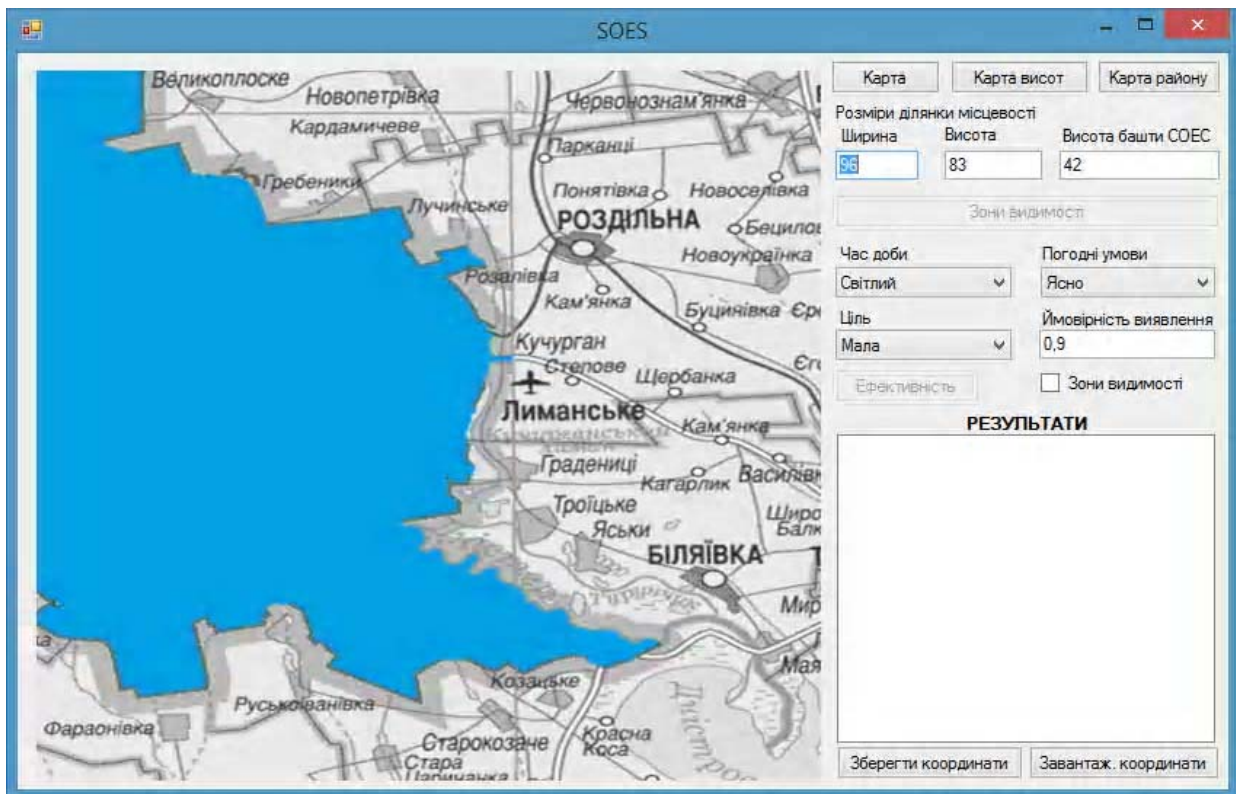


Рисунок 3 – Програма для визначення ефективності прикриття необхідної смуги перекриття з використанням СОЕС

ціональне розташування веж спостереження, яке не повною мірою враховує особливості рельєфу місцевості і форму необхідної смуги перекриття. В окремих випадках вежі встановлені на краю смуги. Звичайно, для малих цілей ефективність зі збільшенням відстані достатньо швидко спадає. Тому при невеликому розташуванні веж ефективне спостереження здійснюється лише по малій частці потрібної смуги перекриття. Одна з веж встановлена в місці, де смуга перекриття є вузькою. Тому значна частка потенційної ефективної площі спостереження втрачається, оскільки вона знаходиться поза межами смуги перекриття.

Після вибору іншого можливого більш раціонального розташування веж спостереження були отримані значно кращі результати. Цей спосіб розміщення веж отриманий шляхом незначних переміщень окремих з них. Для одного з варіантів умов спостереження вікно програми з результатами представлено на рис. 4. Для зручної інтерпретації результатів у програмі кольором відмічаються «мертві зони» та ефективність ведення спостереження.

5 РЕЗУЛЬТАТИ

Результати проведених розрахунків для різних варіантів умов спостереження (табл. 1) свідчать про можливість вдосконалення існуючої СОЕС за рахунок більш ефективного розташування веж спостереження.

З даних, наведених у табл. 1, випливає, що найбільше зростання W (практично у два рази) за рахунок кращого розташування веж досягається по малих цілях. У випадку середніх цілей значення W є кращими приблизно на 50%. Для великих цілей таке зростання є більш скромним і становить лише декілька відсотків.

Таблиця 1 – Оцінювання ефективності СОЕС

Погодні умови	Ціль	W початкова, %	W можлива, %
ясно	мала	47,57	75,37
	середня	59,11	85,07
	велика	97,47	99,16
дощ	мала	38,67	64,58
	середня	47,57	75,37
	велика	94,87	99,16

6 ОБГОВОРЕННЯ

Це пояснюється тим, що ефективна дальність спостереження ТЗС по великих цілях забезпечує значне перекриття зон для сусідніх веж. Тому остаточна ефективність обмежується лише негативним впливом рельєфу, що приводить до появи «мертвих» зон. Однак навіть у цьому випадку, як випливає з табл. 1, можливе більш ефективне розташування веж, при якому зменшується площа цих зон.

ВИСНОВКИ

Таким чином, за результатами проведеного дослідження в роботі запропоновані показники для оцінки ефективності ведення спостереження з використанням СОЕС та наведений можливий підхід до їх знаходження (методика обчислення). На основі використання програмної реалізації запропонованої методики, проведений аналіз впливу різних факторів на ефективність виявлення цілей. Проведені дослідження показали нераціональність існуючої побудови СОЕС та визначили можливі шляхи підвищення ефективності системи за рахунок покращення розстановки веж спостереження.

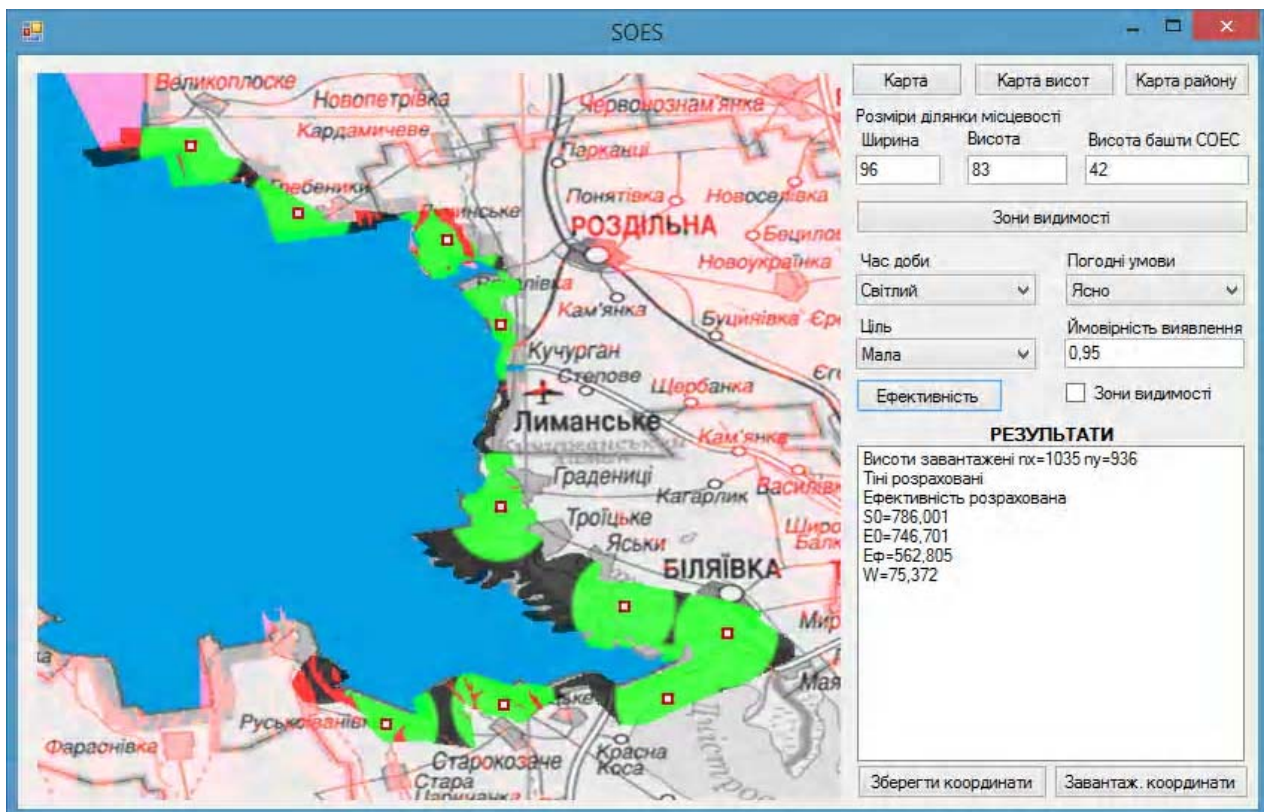


Рисунок 4 – Розрахунок ефективності прикриття необхідної смуги перекриття для можливого розташування веж СОЕС

ПОДЯКИ

У статті визначений науково-методичний апарат оцінки ефективності функціонування системи оптико-електронного спостереження. Роботу виконано в рамках спільних наукових досліджень кафедри інженерного забезпечення та технічних засобів охорони кордону і кафедри загальнонаукових та інженерних дисциплін Національної академії Державної прикордонної служби України. Дослідження здійснювались у рамках науково-дослідних робіт за темами «Опис рельєфу місцевості в задачах статистичної радіофізики з застосуванням геоінформаційних систем» та «Формалізація задач геоінформаційної обробки даних в системі охорони кордону».

СПИСОК ЛІТЕРАТУРИ

1. Tamal K. Curve and Surface Reconstruction : Algorithms with Mathematical Analysis / K. Tamal. – New York : Cambridge University Press, 2006. – 230 p.
2. Owen M. The compression of digital terrain elevation data (DTED) using JPEG 2000 : technical report : DSTO-TR-1548 / M. Owen, M. Grigg. / Command and Control Division Information Sciences Laboratory. – Australia, 2004. – 21 p.
3. Parallel ODETLAP for terrain compression and reconstruction / [J. Stookey, Z. Xie, B. Cutler, W. Randolph, D. Tracy, M. Andrade et al.] // ACM SIGSPATIAL international conference on Advances in geographic information systems : 16th international conference, Irvine, 5–7 November 2008 : proceedings. ACM New York, 2008, – P. 135–144.
4. Smugglers and border guards – The GeoStar Project at RPI / [W. Franklin, M. Inanc, Z. Xie, D. Tracy, B. Cutler, M. Andrade, F. Luk et al.]. // ACM International Symposium on Advances in Geographic Information Systems : 15th international symposium,

- Seattle, 7–9 November 2007 : proceedings. ACM New York, 2007. – P. 228–236.
5. Multiple observer siting and path planning on a compressed terrain / [D. Tracy, W. Franklin, B. Cutler et al.] // Advanced Signal Processing Algorithms, Architectures, and Implementations : 17th international conference, San Diego, 26–27 August 2007: proceedings. Bellingham, 2007. – P. 6697–16.
 6. Tradeoffs when multiple observer siting on large terrain cells / [W. Franklin, C. Vogt, A. Riedl, W. Kainz, G. Elmes et al.] // Progress in Spatial Data Handling : 12th international symposium, University of Vienna, 12–14 July 2006: proceedings. Springer Berlin Heidelberg, 2006. – P. 845–861.
 7. Slope accuracy and path planning on compressed terrain / [W. Franklin, D. Tracy, M. Andrade et al.] // Spatial Data Handling : 13 th international symposium, Montpellier, 23–25 July 2008: proceedings. Springer Berlin Heidelberg, 2008. – P. 335–349.
 8. Tracy D. Path planning and slope representation on compressed terrain : thesis ... doctor of philosophy / Tracy Daniel. – Troy, Rensselaer Polytechnic Institute, 2009. – 106 p.
 9. Горбунов В. А. Эффективность обнаружения целей / В. А. Горбунов. – Москва : Воениздат, 1979. – 160 с.
 10. Кривошеев В. А. Эффективность охраны государственной границы СССР и основные направления ее повышения / В. А. Кривошеев. – Москва : Воениздат, 1988. – 256 с.
 11. Царьов Ю. О. Засоби візуального спостереження охорони державного кордону / Ю. О. Царьов, Д. А. Купрієнко. – Хмельницький : Вид-во НАДПСУ, 2010. – 164 с.
 12. Дем'янюк С. М. Покращення характеристик радіолокаційних засобів охорони кордону при застосуванні спектрально-фазового методу обробки сигналів : дис. ... канд. техн. наук : 20.02.14 / Дем'янюк Сергій Миколайович. – Хмельницький, 2004. – 235 с.

Стаття надійшла до редакції 23.12.2016.
Після доробки 12.01.2017.

Боровик О. В.¹, Рачок Р. В.², Дармороз М. М.³

¹Д-р техн. наук, професор, начальник кафедри інженерного забезпечення і технічних засобів охорони границь, Національна академія Государственной пограничной службы Украины имени Богдана Хмельницкого, Хмельницький, Україна

²Канд. техн. наук, доцент, докторант, Національна академія Государственной пограничной службы Украины имени Богдана Хмельницкого, Хмельницький, Україна

³Ад'юнк, Національна академія Государственной пограничной службы Украины имени Богдана Хмельницкого, Хмельницький, Україна

ОЦЕНКА ЭФФЕКТИВНОСТИ ФУНКЦИОНИРОВАНИЯ СИСТЕМЫ ОПТИКО-ЭЛЕКТРОННОГО НАБЛЮДЕНИЯ

Актуальность. В охране государственной границы Украины используется новая система оптико-электронного наблюдения (СОЭН). Однако ее потенциальные возможности не реализованы в полной мере. Одной из причин этого является отсутствие научно-методического аппарата оценки эффективности ее функционирования с учетом различных факторов природного и техногенного характера. Последнее делает актуальным проведение соответствующего исследования.

Цель. Целью работы является разработка методики оценки эффективности функционирования СОЭН.

Метод. В работе предложены показатели эффективности функционирования отдельных технических средств наблюдения и их комплексного использования с одной башни системы оптико-электронного наблюдения. Также в работе предложен новый показатель эффективности функционирования системы оптико-электронного наблюдения. В показателе учтены особенности ведения наблюдения с комплексным использованием различных технических средств охраны границы в различных условиях природного и техногенного характера и с учетом рельефа местности. Обоснование целесообразности их применения предусматривало использование вероятностно-статистических методов.

Результаты. На основе информации о цифровых характеристиках рельефа исследуемого участка осуществлено определение недоступных для наблюдения участков местности, так называемых «мертвых зон». Предложена методика вычисления показателя эффективности функционирования системы оптико-электронного наблюдения. Приведено описание алгоритмической и программной реализации этой методики. С использованием разработанного программного обеспечения проведен анализ особенностей функционирования существующей системы оптико-электронного наблюдения, определены основные проблемные аспекты и направления повышения эффективности функционирования СОЭН.

Выводы. Текущее построение СОЭН не является оптимальным. Возможными направлениями повышения эффективности ее функционирования могут быть оптимизация расстановки башен системы на местности, корректировка высоты размещения на башнях технических средств наблюдения и удачный подбор самих технических средств наблюдения.

Ключевые слова: система оптико-электронного наблюдения, показатель эффективности системы, математические методы, алгоритм.

Borovik O. V.¹, Rachok R. V.², Darmoroz M. M.³

¹Dr.Sc., Professor, Department of Engineering and Technical Means of Border Protection, National Academy of State Border Guard Service of Ukraine Named After Bohdan Khmelnytsky, Khmelnytsky, Ukraine

²PhD, Associate Professor, Doctorate, National Academy of State Border Guard Service of Ukraine Named After Bohdan Khmelnytsky, Khmelnytsky, Ukraine

³Adjunct, National Academy of State Border Guard Service of Ukraine Named After Bohdan Khmelnytsky, Khmelnytsky, Ukraine

EVALUATING THE EFFECTIVENESS OF THE OPTICAL-ELECTRONIC SURVEILLANCE SYSTEM

Context. In the protection of the state border of Ukraine used the new system of optical-electronic observation (SOEO). However, its potential is not fully realized. One of the reasons for this is the lack of scientific and methodological apparatus of evaluation of the effectiveness of its functioning in various factors of natural and technogenic character. The latter makes it relevant to conduct appropriate research.

Objective. The aim of this work is to develop a methodology for assessing the efficiency of SOEO.

Method. The proposed indicators of efficiency of functioning of individual surveillance technology and their integrated use with one tower system opto-electronic surveillance. The work also proposed a new indicator of efficiency of functioning of system of optical-electronic observation. The indicator takes into account features of conducting observations with the comprehensive use of various technical means of border protection in various conditions of natural and technogenic character, and given the terrain. Substantiation of expediency of their application involved the use of probabilistic-statistical methods.

Results. On the basis of information about digital characteristics of the topography of the investigated area carried out determining unobservable areas, the so-called "dead zones". The proposed method of calculating the index of efficiency of functioning of system of optical-electronic observation. The description of algorithmic and software implementation of this technique. Using the developed software the analysis of the functioning of the current system optical-electronic surveillance, identified key problematic aspects and directions of improving the efficiency of SOEO.

Conclusions. Modern building of SOEO is not optimal. Possible directions of increase of efficiency of its functioning can be optimization of the placement of the towers system on the ground, adjusting the height placement on the towers of surveillance technology and a good selection of the technical means of observation.

Keywords: system of optical-electronic observation, system performance indicator, mathematical methods, algorithm.

REFERENCES

1. Tamal K. Curve and Surface Reconstruction : Algorithms with Mathematical Analysis. New York, Cambridge University Press, 2006, 230 p.
2. Owen M., Grigg M. The compression of digital terrain elevation data (DTED) using JPEG 2000 : technical report : DSTO-TR-1548, *Command and Control Division Information Sciences Laboratory*. Australia, 2004, 21 p.
3. Stookey J., Xie Z., Cutler B., Randolph W., Tracy D., Andrade M. et al. Parallel ODETLAP for terrain compression and reconstruction, ACM SIGSPATIAL international conference on Advances in geographic information systems : 16th international conference, Irvine, 5–7 November 2008 : proceedings. ACM New York, 2008, pp. 135–144.
4. Franklin W., Inanc M., Xie Z., Tracy D., Cutler B., Andrade M., Luk F. et al. Smugglers and border guards – The GeoStar Project at RPI, *ACM International Symposium on Advances in Geographic Information Systems : 15th international symposium, Seattle, 7–9 November 2007 : proceedings*. ACM New York, 2007, pp. 228–236.
5. Tracy D., Franklin W., Cutler B., Andrade M., Luk F., Inanc M., Xie Z. et al. Multiple observer siting and path planning on a compressed terrain, *Advanced Signal Processing Algorithms, Architectures, and Implementations : 17th international conference*. San Diego, 26–27 August 2007: proceedings. Bellingham, 2007, pp. 6697–16.
6. Franklin W., Vogt C., Riedl A., Kainz W., Elmes G. et al. Tradeoffs when multiple observer siting on large terrain cells, *Progress in Spatial Data Handling : 12th international symposium, University of Vienna, 12–14 July 2006: proceedings*. Springer Berlin Heidelberg, 2006, pp. 845–861.
7. Franklin W., Tracy D., Andrade M., Muckell J., Inanc M., Xie Z., Cutler B. et al. Slope accuracy and path planning on compressed terrain, *Spatial Data Handling : 13th international symposium, Montpellier, 23–25 July 2008: proceedings*. Springer Berlin Heidelberg, 2008, pp. 335–349.
8. Tracy D. Path planning and slope representation on compressed terrain : thesis ... doctor of philosophy. Troy, Rensselaer Polytechnic Institute, 2009, 106 p.
9. Gorbunov V. A. *Jeffektivnost' obnaruzhenija celej*. Moscow, Voenizdat, 1979, 160 p.
10. Krivosheev V. A. *Jeffektivnost' ohrany gosudarstvennoj granicy SSSR i osnovnye napravlenija ee povyshenija*. Moscow, Voenizdat, 1988, 256 p.
11. Car'ov Ju. O., Kuprienko D. A. *Zasobi vizual'nogo sposterezhenija ohoroni derzhavnogo kordonu*. Hmel'nic'kij, Vid-vo NADPSU, 2010, 164 p.
12. Dem'janjuk S. M. *Pokrashhenija karakteristik radiolokacijnih zasobiv ohoroni kordonu pri zastosuvanni spektral'no-fazovogo metodu obrobki signaliv : dis. ... kand. tehn. nauk : 20.02.14*. Hmel'nic'kij, 2004, 235 p.

INTEGRATED METHOD OF EXTRACTION, FORMALIZATION AND AGGREGATION OF COMPETITIVE AGENTS EXPERT EVALUATIONS IN A GROUP

Context. The problem of extraction, formalization and aggregation of expert evaluations performed during selection of the best solution from possible alternatives set was considered. The problem actuality is defined by different application areas, additional analysis necessity and group evaluation under uncertainty.

Objective. The research objective was to raise quality of the decisions made by group of experts due to increase of individual expert evaluation process efficiency and improvement of evaluation aggregation process.

Method. The integrated method which consists of individual and group evaluation was proposed for the problem solution.

The modified method of extraction and formalization of individual expert evaluations which is based on the analytic hierarchy process modification and includes absolute and relative evaluation phases was proposed. It evaluates expert competence based on confidence coefficient for expert judgments.

The modified method of agent group evaluation based on summation of individual expert evaluations of every alternative was proposed. It makes total evaluation based on relative quantitative importance of the agents and confidence coefficients for judgments of every group participant. It gives preference to judgments of qualified experts to increase solution quality.

Results. The experimental investigation of the proposed methods confirmed availability of the developed mathematical support. The proposed methods are able to detect and control nontrivial issues in tasks which are solved.

Conclusions. Scientific novelty of the paper consists in the proposed integrated method of extraction, formalization and aggregation of expert evaluations in group which enables to define integrated assessment of competitive agents as well as to evaluate confidence coefficient for expert judgments directly during individual evaluation process and to use it in the following group decision-making phase.

Practical significance of the paper results consists in the developed information technology which made it possible to put the method into practice for solving of the tender support tasks and experiment results.

Keywords: expert evaluation, group evaluations, individual evaluations, expert competence, competitive agents.

NOMENCLATURE

A – pairwise comparison matrix which elements are criteria or agents;

Ag – set of competitive agents;

a_{rp} – result of comparison of the r object with the p object according to general criterion;

B – pairwise comparison matrix of competitive agents according to criteria;

b_{ij} – evaluation of the i -th agent according to the j -th criterion;

C – criterion set;

D – set of competitive agents evaluations;

d_{best} – agent which is considered most preferable;

d_i – part of the i -th agent in the total estimate;

f_1 – function which extracts element with the largest key value;

f_2 – functional dependence;

g^w_g – global priority vector for the g -th expert;

$g^w_g^{best}$ – agent which is considered most preferable by the g -th expert;

g^w_{gi} – element of global priority vector which corresponds to the evaluation of the i -th agent by the g -th expert;

KA – number of compared elements (criteria or agents);

KM – number of evaluation criteria;

KN – number of agents;

KS – number of experts;

PCM – pairwise comparison matrix;

Q – set of “control” questions;

q_k – comparative evaluation of random agent pair from the k matrix;

S – set of experts;

s_g – the g -th expert from the set S ;

Sum_i – total evaluation of the i -th competitive agent;

v_g – confidence coefficient for judgments of the expert s_g ;

$v_g(k)$ – confidence coefficient for judgment of the g -th expert on the k -th question;

W – local priority vector;

w_r – the r -th element of local priority vector W .

INTRODUCTION

The problems of selection of one solution from the set of possible variants are known in different activity areas. These problems are often characterized by the fact that the choice has significant impact on problem-solving environment state. At the same time most commonly it isn't possible to select the best solution from the set of alternatives in such problems without additional analysis of all possible variants as well as factors which impact on solution quality.

As factors which are evaluation criteria can be not only quantitative but also qualitative, and evaluation is executed under uncertainty, automatic decision-making isn't possible.

Hence expert knowledge are used for selection of the best variant from the set of alternatives.

To raise objectivity of evaluation procedure it should be done by more than one expert. However one part of experts in group may turn out to be more competent than others in the problem which is solved, so judgments of this part of expert group should be preferred during aggregation of group evaluation.

So the development of integrated method which enables to extract and aggregate expert evaluations taking into account competence of every expert is actual.

Object of the research is process of formalization and aggregation of expert evaluations. Subject of the research is methods of evaluation performed during selection of solutions from the set of alternatives (which are considered as competitive agents).

The research objective is to raise quality of the decisions made by group of experts due to increase of individual expert evaluation process efficiency and improvement of evaluation aggregation process.

The following tasks were defined to achieve the research objective:

- improvement of the method of extraction and formalization of expert judgments;
- development of the method of expert competence evaluation during individual evaluation process;
- improvement of the method of group evaluation based on individual expert evaluations taking into account expert competence;
- practical investigation of the proposed methods efficiency using the developed information technology.

1 PROBLEM STATEMENT

Suppose we have the set Ag of competitive agents, the set S of experts who can estimate these agents and the set C of the evaluation criteria.

Then the problem of competitive agents selection is to define estimates $D = \{d_i | i = 1, \dots, KN\}$ of each agent from the set Ag using transformation $i = f_1(Ag_j), j = 1..KN$, which executes descending ordering according to the values of d_i .

Values of elements from the set D are defined by the dependence $D = f_2(C, S)$. Method of determination of the dependence should be proposed in the paper for competitive agents selection problem solving.

It should be noted that the given problem includes also competitive agent selection problem which consists in selection of competitive agent with the biggest part in total evaluation $d_{best} = \max_i d_i$.

2 REVIEW OF THE LITERATURE

When decision is made by expert group, collaboration using open dynamic discussion has substantial disadvantages which can impact on decision, making it subjective [1–4]. Therefore in some circumstances it is more efficient to question experts individually, permitting them to use only personal experience making judgments [5–6]. At the same time expert should be proposed to place all possible solutions in descending order of its correspondence to the task as well as to estimate solutions quantitatively determining this correspondence level or relative priority of

each alternative comparing it with another one. To obtain such estimations there is a lot of traditional methods: the Schulze method [7], the method of preference [8], the analytic hierarchy process [9], fuzzy methods and the COPRAS method [10]. But the described methods have substantial disadvantages which are critical for the problem which is solved. So the Schulze method and the method of preference make evaluations only by one criterion and limit scale of correspondence to the task by natural numbers. The analytic hierarchy process is sufficiently laborious and doesn't enable to take into account correspondence level of considered variants to the problem objective. Fuzzy methods may be unefficient for practical tasks, because these methods can make inaccuracy for similar alternative evaluations during transformation of triangular numbers to real ones. The COPRAS method solves tasks of multicriterion alternatives evaluation based on its correspondence to the task objective, but this method doesn't enable to detect its priorities comparing with each other. The methods [11–19] enable to make evaluations according to several criteria based on association rules synthesis [11], decision trees [12], radial basis neural networks [13], production rules [14], neuro-fuzzy networks [15], cascade systems [16] and sequential [11, 17] or parallel [18, 19] computing. However these methods are applicable only when training data samples, which contain collected information about objects and processes under consideration, are available, that makes difficult and in some circumstances eliminates possibility of its applicability for solving of tasks which need expert evaluation.

As analysis result it should be noted that regarding the problem which is considered decomposition approach of the analytic hierarchy process enables to develop the most appropriate problem model [20], since experts can use quantitative and qualitative criteria for evaluation of agents using qualitative criteria based on clear scale.

Thus disadvantages of existing expert evaluation methods cause necessity of improvement of methods which conduct individual evaluation of competitive agents by expert group and further evaluations aggregation taking into account competency of every expert.

3 MATERIALS AND METHODS

Currently only one approach is usually used for extraction and formalization of expert judgments at a time. Possible approaches include the following ones: expert can estimate agent as consistent entity based on its correspondence to the given problem or expert can compare all agents with each other creating relative rating. Both these approaches give only one-track evaluation. For integrated evaluation of alternatives as competitive agents within the scope of the current problem it was proposed to make evaluation using two approaches [21–22]:

- absolute evaluation as determination of competitive agents correspondence to the customization demands for each criterion;
- relative evaluation based on comparison of agents with each other according to each criterion.

At the beginning of the proposed method of extraction and formalization of individual expert evaluations level of

correspondence of competitive agents to the problem statement according to each criterion is evaluated (absolute evaluation). This phase is made by absolute percentage evaluation (from 0 to 100%) of each agent according to each criterion comparing to the problem statement requirements. After evaluation obtained values has to be normalized using division by 100. The result of this stage is the matrix B of the size $KN \times KM$ which consists of evaluations $b_{ij} \in [0,1]$ for $i=1, \dots, KN, j=1, \dots, KM$:

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1KM} \\ b_{21} & b_{22} & \dots & b_{2KM} \\ \dots & \dots & \dots & \dots \\ b_{KN1} & b_{KN2} & \dots & b_{KNKM} \end{pmatrix}.$$

Then the modified analytic hierarchy process is used to evaluate agents comparatively according to each criterion (relative evaluation).

Relative evaluation process is generally represented on activity diagram (fig. 1).

The method of paired comparisons [9] is used to compare criteria according to the given task and to compare agents according to criteria. According to this method pairwise comparison matrices (PCM) are obtained from the results of comparison of every row element with every column element.

So the matrix A of size $KA \times KA$ is obtained from elements a_{rp} which are presented by real numbers:

$$A = \begin{pmatrix} 1 & a_{12} & \dots & a_{1KA} \\ a_{21} & 1 & \dots & a_{2KA} \\ \dots & \dots & \dots & \dots \\ a_{KA1} & a_{KA2} & \dots & 1 \end{pmatrix}.$$

Every expert that takes part in the evaluation process has to compare all elements with each other presenting evaluations as numeric values in accordance with fundamental ratio scale [23] and has to complete corresponding PCM.

One of the main disadvantages of the method of paired comparisons is its laboriousness. It is necessary to analyze fundamental properties of PCM [9] to reduce the number of comparison operations which have to be executed by expert.

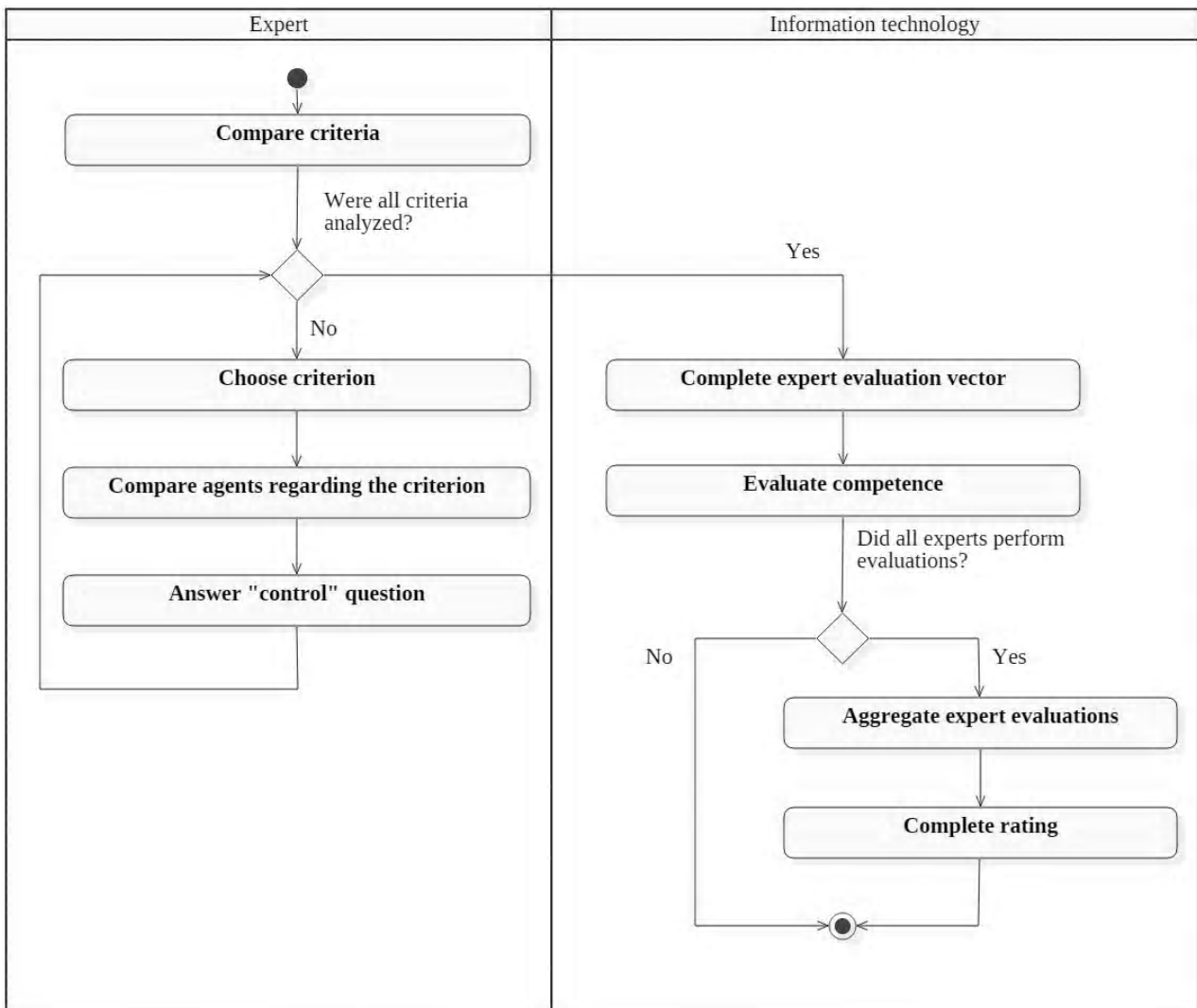


Figure 1 – Activity diagram for relative agent evaluation (UML 2.5 notation)

The main property of PCM is the following: for each $r, p=1, \dots, KA$ elements $a_{rp} > 0$. Besides that under $r=p$ $a_{rp}=1$, because these elements present comparisons of each element with oneself and have to correspond to equal significance.

The third property is inverse symmetry of the matrix about the main diagonal. For each $r, p=1, \dots, KA$ $a_{pr}=1/a_{rp}$, so expert has to complete only evaluations which are situated above the main diagonal.

Thus the PCM A will take the following form:

$$A = \begin{pmatrix} 1 & a_{12} & \dots & a_{1KA} \\ 1/a_{12} & 1 & \dots & a_{2KA} \\ \dots & \dots & \dots & \dots \\ 1/a_{1KA} & 1/a_{1KA} & \dots & 1 \end{pmatrix}.$$

The phase of PCM filling has to be done for every hierarchy level.

The results of this phase are PCMs where criteria of the same hierarchy level are compared with each other, criteria of lower hierarchy level are compared with each other according to criteria of higher hierarchy level, if there are ones, and agents are compared with each other according to criteria.

Then local priority vectors are obtained from the collection of PCMs. Local priority vector $W = (w_1, w_2, \dots, w_{KA})^T$ has to be determined for every obtained PCM. The main property of the obtained local priority vectors is that sum of vector components equals 1,

$$\text{that is } \sum_{r=1}^{KA} w_r = 1.$$

To obtain global priority vector $gW_g = \{gW_{g1}, gW_{g2}, \dots, gW_{gKN}\}$ for each g -th expert, local priorities of each evaluated agent are multiplied by priority of the corresponding criterion and results are added to each other.

Then the corresponding values of the global priority gW_{gi} of every agent are compared with each other. Agent which is considered most preferable by the g -th expert will have the biggest value of the global priority:

$$gW_g^{best} = \max_i gW_{gi}, \quad i = 1, \dots, KN.$$

The tasks which are solved in the context of the considered problem are seldom completely identical, because every task has its own features and differences: for example technical or time constraints. Application of the analytic hierarchy process demands familiarity with these features, constraints and its influence on the final result from experts. However it isn't always possible because of the wide specialization of experts who take part in questioning in practice.

In this case it is necessary to evaluate additionally expert competence and confidence in decisions made by him [24]. Expert competence level should be then used for aggregation of group evaluation [25].

Confidence coefficient v_g corresponds to quantification of the level of confidence in competence of expert $s_g \in S$ ($g=1, \dots, KS$) in the presented problem.

It is proposed to add set of «control» questions Q to pairwise comparison sequence to evaluate confidence coefficients for expert judgments. Each element of the set Q is defined as $q_k = a_{rp}$ ($r \in \{2, KN\}$, $p \in \{1, KN-r\}$) which is comparative evaluation of random pair of agents from the bottom part of the k -th PCM where agents are compared according to criteria. So the set Q is defined in the following form:

$$Q = \{q_k \mid k = 1, \dots, KM\}.$$

«Control» question is a question which was restated and repeated to expert. Such questions enable to determine confidence of expert answering main questions in agents evaluation process.

Since all elements of the matrix A are inverse symmetrical that is for each $r, p=1, \dots, KA$ $a_{pr}=1/a_{rp}$, it is possible to verify if expert judgments which were gotten comparing random pair of agents are equal to expert judgments which were gotten during primary comparisons. This verification is based on evaluation of PCM elements which are situated under the main diagonal as inverse proportion to the values of elements which are situated in the matrix part which was filled by expert.

For each k -th PCM random pair of agents p and r is selected from the elements of the matrix which are situated under the main diagonal and is proposed to expert for comparison.

If values a_{rp} and a_{pr} are bigger or lower than 1 at the same time, then confidence coefficient for the given question should be specified as $v_g(k) = 0$ (the g -th expert showed his incompetence answering the k -th question). Otherwise confidence coefficient for the given question should be calculated using the following expression:

$$v_g(k) = \begin{cases} 1 - \frac{|a_{rp} - 1/a_{pr}|}{\max(a_{rp}, 1/a_{pr})}, & (a_{rp} \geq 1) \cap (a_{pr} \leq 1), \\ 1 - \frac{|a_{pr} - 1/a_{rp}|}{\max(a_{pr}, 1/a_{rp})}, & (a_{rp} < 1) \cap (a_{pr} > 1) \end{cases}$$

Overall confidence coefficient v_g for judgments of the expert s_g is calculated using the following formula:

$$v_g = \frac{\sum_{k=1}^{KM} v_g(k)}{KM}.$$

After questioning is finished, final value of v_g should be evaluated. If $v_g=1$ expert should be considered as competent and his judgments can be trusted. If $v_g=0$ expert should be considered as incompetent relatively to the

solved task. In this case evaluations of this expert should not be taken into account in aggregation of group evaluation.

After all experts of the group made own evaluations, global priority vector g^w_g and confidence coefficient for expert judgments v_g are defined for each expert from number KS , the method of group agent evaluation has to be executed. This method calculates total evaluation Sum_i for each i -th agent based on summation of evaluations g^w_{gi} which were given to the agent by the g -th expert. At the same time evaluations has to be multiplied by confidence coefficients for expert judgments:

$$Sum_i = \sum_{g=1}^{KS} v_g \cdot g^w_{gi}.$$

Thus the bigger value confidence coefficient v_g for expert judgments has, more impact g -th expert will have on overall evaluation.

After summation is done, overall rating of competitive agents D is defined. Each individual evaluation is multiplied by confidence coefficient for expert judgments and part d_i of each i -th agent in the total evaluation is calculated using the following formula:

$$d_i = \frac{Sum_i}{\sum_{i=1}^{KN} Sum_i}, \sum_{i=1}^{KN} d_i = 1.$$

Agents are ordered by decrease of d_i values, in that way defining function f_1 which can be implemented as binary heap. The result of this operation is agent rating, where agent which is considered most preferable by expert group heads the list. Based on the obtained rating responsible person makes final decision on the best agent. In most cases it is agent which is the first in the list and it becomes solution of the competitive agents selection problem.

4 EXPERIMENTS

The proposed methods were used for the development of the competitive agents selection information technology.

The tasks of competitive agents selection were solved for experimental investigation of practical availability of the proposed methods using the developed information technology.

The investigation was made based on data collected from tender tasks, participant offers and tender results accumulated by design companies LLC “SK Inzhenerne systemy” and “VST” (Ukraine) and evaluations of experts who were selected and invited based on the tender tasks.

The following competitive agents (companies) were considered in the task of tender for design, supply and mounting of metalware and barriers:

- AO “Metallist SMK”, Kharkiv;
- Atlas Word (Germany), branch in Kyiv;
- Borga Hale, Gdynia, Poland.

The following competitive agents were considered in the task of supply of materials and equipment for mounting of shopping center buildings:

- Master Profi, Dnipro;
- Remko, Kyiv;
- Atlas Word, Kyiv;
- Askon, Kamyanske;
- IBT, Kyiv.

Design companies traditionally used pricing appraisal or oral discussion method for evaluation during tender process. However since tender results can be determined in long time horizon [26] it is necessary from one to six months after tender is ended for its efficiency evaluation. During experimental investigation the results of traditional approach were compared with the results obtained by the proposed integrated method.

5 RESULTS

Solving the task of tender for design, supply and mounting of metalware and barriers the derived proposals were formalized as competitive agents and were proposed to the group of three experts for evaluation. Expert evaluations were used for agents rating aggregation.

The rating of competitive agents obtained by the traditional approach (without taking into account confidence coefficient for expert judgments) can be summarized in the following list (fig. 2a):

- Borga Hale – 35,60%;

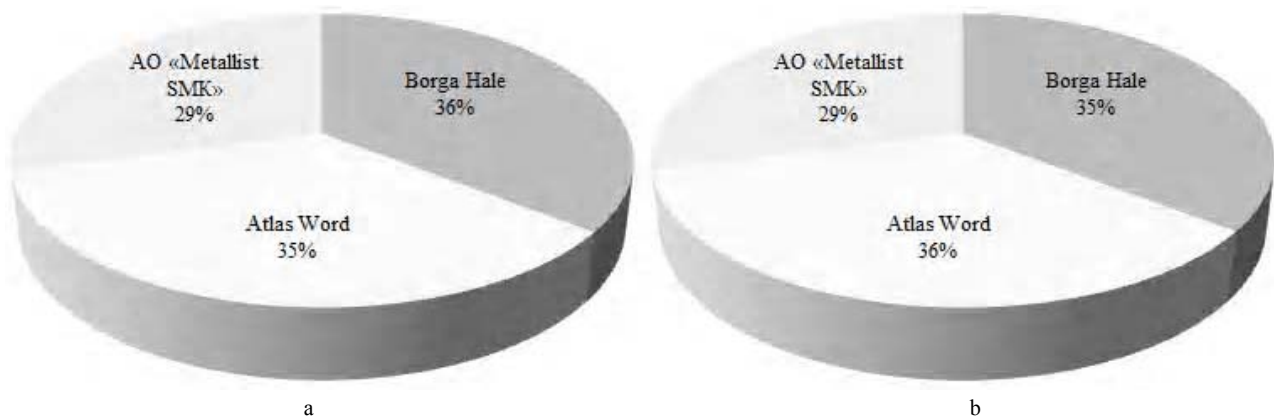


Figure 2 – The results of competitive agents evaluation in the task of tender for design, supply and mounting of metalware and barriers:
 a – without taking into account confidence coefficient for expert judgments;
 b – taking into account confidence coefficient for expert judgments

- Atlas Word – 35,35%;
- AO “Metallist SMK” – 29,05%.

The rating of competitive agents obtained using the modified methods (taking into account confidence coefficient for expert judgments) can be summarized in the following list (fig. 2b):

- Atlas Word – 35,60%;
- Borga Hale – 35,50%;
- AO “Metallist SMK” – 28,9%.

Solving the task of supply of materials and equipment for mounting of shopping center buildings the derived proposals were formalized as competitive agents and were proposed to the group of four experts for evaluation.

Before the beginning of evaluation process experts were proposed to examine tender task using the developed information technology. Then traditional approach based on oral discussion was used for evaluation. Based on oral discussion and simple voting methods expert commission elected the tender offer of “Master Profi” because of its minimum cost and delivery time.

Then agents evaluation was repeated by the same expert commission using competitive agents information technology based on the proposed methods (fig. 3).

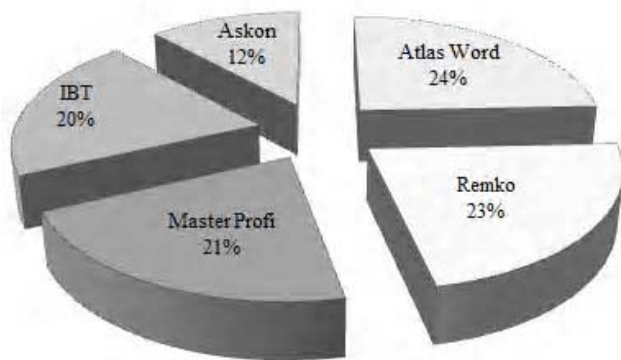


Figure 3 – The results of competitive agents evaluation in the task of supply of materials and equipment for mounting of shopping center buildings taking into account confidence coefficient for expert judgments

Besides that individual expert preferences and confidence coefficients for expert judgments were defined (fig. 4).

The rating of competitive agents obtained using the integrated method and taking into account confidence coefficient for expert judgments can be summarized in the following list:

- Atlas Word – 24.35%;
- Remko – 22.40%;
- Master Profi – 21.31%;
- IBT – 20.13%;
- Askon – 11.81%.

6 DISCUSSION

Results of competitive agents evaluation in the task of tender for design, supply and mounting of metalware and barriers showed that taking into account confidence coefficient for expert judgments changed positions of participants in rating. It should be noted that the bigger the dispersion of the confidence coefficients for expert judgments is, more impact experts with bigger values of confidence coefficient will have on tender results.

As the result of the proposed integrated method application to the tender for supply of materials and equipment for mounting of shopping center buildings the tender offer of the company «Master Profi» which was chosen during oral discussion occupied only the third place in the rating. It is caused by its failures which were missed during basic analysis: delayed starting date of execution phase and small production volume regarding competitors.

Thus investigation results confirm availability of the developed competitive agents selection information technology in whole and the proposed integrated method particularly for practical tasks solving since the proposed problem solution helps to detect nontrivial problems and failures of the considered competitive agents.

Analysis of the obtained results showed that the proposed integrated method of extraction, formalization and aggregation of expert evaluations enabled to raise accuracy

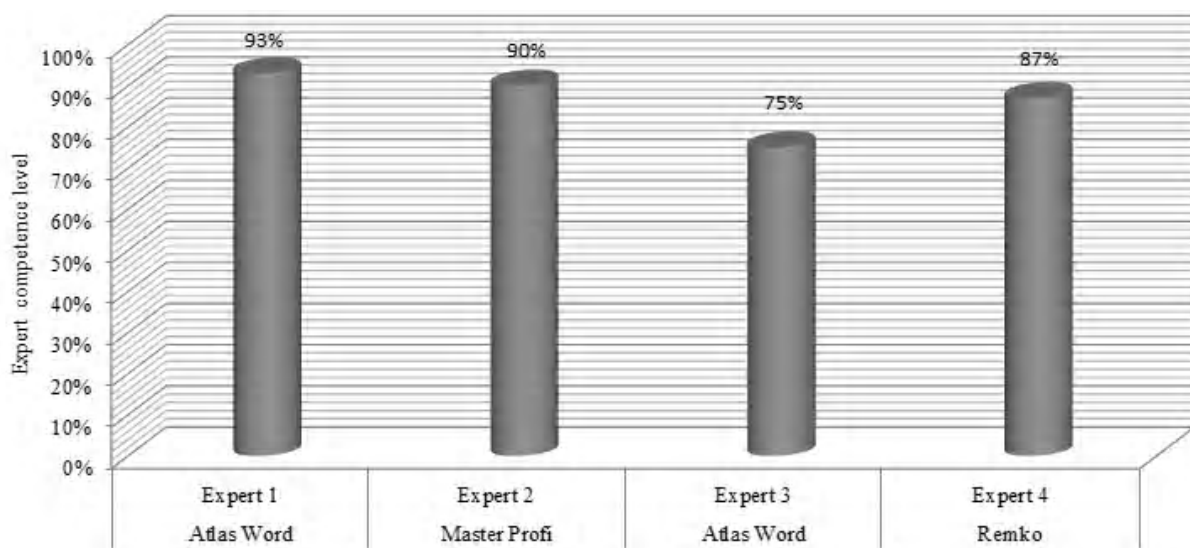


Figure 4 – Expert competence level

of group evaluations due to formalization of individual evaluations derivation process as well as implementation of expert competence evaluation mechanism. In turn it ensures group decision making process efficiency and enables to raise quality of the final solution.

The proposed integrated method can be used for practical expert evaluation tasks solving in many areas of industry, economics and medicine: tender processes in all areas of state and business activity, complicated clinical decision support, expert evaluation of techware quality in industry.

In what follows it is planned to investigate availability of the developed method for solving of practical tasks in other activity areas and the dependence of the method efficiency on application domain.

CONCLUSIONS

In this paper actual problem of competitive agents selection was solved based on group expert evaluations.

Scientific novelty of the paper consists in the proposed integrated method of extraction, formalization and aggregation of expert evaluations in group which enables to define integrated assessment of competitive agents as well as to evaluate confidence coefficient for expert judgments directly during individual evaluation process and to use it in the following group decision-making phase. The proposed method gives preference to judgments of competent experts in the overall evaluation, at the same time increasing quality of the decision.

Practical significance of the paper results consists in the developed information technology and experiment results. Information technology implemented the proposed integrated method that made it possible to put the method into practice for solving of the tender support tasks. Experiments confirmed availability of the developed mathematical support and made it possible to recommend it for practical application.

Application of the proposed mathematical support and the developed information technology for other practical tasks directly concerned with competitive agents evaluation should be noted as further research directions.

ACKNOWLEDGEMENTS

The research was made in the context of the state budget research and development project “Methods and tools of computational intelligence and parallel computing for big data processing in diagnostic systems” (state registration number 0116U007419) of Zaporizhzhya National Technical University with the support of the international project “Centers of Excellence for young REsearchers” of the Tempus programme of the European Union (registration number 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES).

REFERENCES

- Карнышев А. Д. Психология деятельности и управления : учеб. пособие / А. Д. Карнышев, Е. А. Иванова. – Иркутск : ИГЭА, 2001. – 186 с.
- Мулен Э. Кооперативное принятие решений : аксиомы и модели / Э. Мулен. – М. : Мир, 1991. – 464 с.
- Майерс Д. Социальная психология / Д. Майерс. – СПб. : Питер, 2013. – 800 с.
- Ambrusa A. How individual preferences are aggregated in groups: An experimental study / A. Ambrusa, B. Greinerb, P. A. Pathakc // *Journal of Public Economics*. – 2015. – Volume 129. – P. 1–13.
- Budziński R. Model of competence of experts in the computer decision support system / R. Budziński, J. Becker // *Quantitative Methods In Economics*. – 2013. – Volume XIV, Issue 1. – Pp. 53-64.
- Patel H. T. A Study on the Effectiveness of Group Activity and Group Discussion Method in English / H. T. Patel // *International Journal of Research in Humanities and Social Sciences*. – 2014. – Volume 2, Issue 1. – P. 13–15.
- Parkes D. C. A Complexity-of-Strategic-Behavior Comparison between Schulze’s Rule and Ranked Pairs / D. C. Parkes, L. Xia // *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. – 2012. – P. 1429–1435.
- Roszkowska E. Rank ordering criteria weighting methods – a comparative overview / E. Roszkowska // *Optimum. Studia Ekonomiczne*. – 2013. – № 5 (65). – P. 14–33.
- Saaty T. L. Models, Methods, Concepts & Applications of the Analytic Hierarchy Process / T. L. Saaty, L. G. Vargas. – Springer US, 2012. – 346 p.
- An Integrated Approach of Fuzzy Linguistic Preference Based AHP and Fuzzy COPRAS for Machine Tool Evaluation / [H. T. Nguyen, S. Z. M. Dawal, Y. Nukman at al.] // *PLOS One*. – 2015. – № 10 (9). – 24 p.
- Oliinyk A. Training Sample Reduction Based on Association Rules for Neuro-Fuzzy Networks Synthesis / A. Oliinyk, T. Zaiko, S. Subbotin // *Optical Memory and Neural Networks (Information Optics)*. – 2014. – Vol. 23, № 2. – P. 89–95.
- Oliinyk A. The decision tree construction based on a stochastic search for the neuro-fuzzy network synthesis / A. Oliinyk, S. A. Subbotin // *Optical Memory and Neural Networks (Information Optics)*. – 2015. – Vol. 24, № 1. – P. 18–27.
- Subbotin S. A. The method of diagnostic model synthesis based on radial basis neural networks with the support of generalization properties / S. A. Subbotin // *Radio Electronics. Computer Science. Control*. – 2016. – № 2. – P. 64–69.
- Oliinyk A. Production rules extraction based on negative selection / A. Oliinyk // *Radio Electronics, Computer Science, Control*. – 2016. – № 1. – P. 40–49.
- Oliinyk A.O. Experimental Investigation with Analyzing the Training Method Complexity of Neuro-Fuzzy Networks Based on Parallel Random Search / A. O. Oliinyk, S. Yu. Skrupsky, S. A. Subbotin // *Automatic Control and Computer Sciences*. – 2015. – Vol. 49, Issue 1. – P. 11–20.
- Bodyanskiy Ye. V. An evolving cascade system based on neurofuzzy nodes / Ye. V. Bodyanskiy, O. K. Tyshchenko, O. O. Boiko // *Radio Electronics. Computer Science. Control*. – 2016. – № 2. – P. 40–45.
- Oliinyk A. Association Rules Extraction for Pattern Recognition / A. Oliinyk, S.A. Subbotin // *Pattern Recognition and Image Analysis*. – 2016. – Vol. 26, № 2. – P. 419–426.
- Oliinyk A. O. Using Parallel Random Search to Train Fuzzy Neural Networks / A. O. Oliinyk, S. Yu. Skrupsky, S. A. Subbotin // *Automatic Control and Computer Sciences*. – 2014. – Vol. 48, Issue 6. – P. 313–323.
- Subbotin S. Individual prediction of the hypertensive patient condition based on computational intelligence / S. Subbotin, A. Oliinyk, S. Skrupsky // *Information and Digital Technologies: International Conference IDT’2015, Zilina, 7–9 July 2015: proceedings of the conference*. – Zilina: Institute of Electrical and Electronics Engineers, 2015. – P. 336–344.
- Pangeran M. H. Conceptual model of analytical network process for prioritizing risk in a PPP infrastructure project / M. H. Pangeran, K. S. Pribad // *Proceedings of the First Makassar International Conference on Civil Engineering, March 9–10, 2010*. – P. 1217–1227.

21. Jao C. S. Efficient Decision Support Systems – Practice and Challenges in Multidisciplinary Domains / Chiang S. Jao. – InTech, 2011. – 478 p.
22. Rush C. Expert Judgement in Cost Estimating: Modelling the Reasoning Process / Christopher Rush, Rajkumar Roy. – Concurrent Engineering. – 2001. – № 9. – P. 271–284.
23. Андрейчиков А. В. Анализ, синтез, планирование решений в экономике / А. В. Андрейчиков, О. Н. Андрейчикова. – М. : Финансы и статистика, 2002. – 368 с.
24. Shrotriya S. Prediction of the Winner by Using a Weighted Approach of Preferential Balloting Systems on the Basis of Their Satisfied Criteria and Artifice Behavioral Complexity / S. Shrotriya, A. Pandey // 3rd International Conference on Information Security and Artificial Intelligence (ISAI 2012), Singapore. – 2012. – P. 142–146.
25. Wang B. Determining decision makers' weights in group ranking: a granular computing method / B. Wang, J. Liang, Y. Qian // International Journal of Machine Learning and Cybernetics. – 2015. – Volume 6, Issue 3. – P. 511–521.
26. Tao L. Decision Support for Contractor Selection: Incorporating consolidated Past Performance Information / Lijuan Tao. – University of Hong Kong. – 2010. – 184 p.
- Article was submitted 14.12.2016.
After revision 29.12.2016.

Колпакова Т. О.¹, Олійник А. О.², Льовкін В. М.³

¹Ст. викладач кафедри програмних засобів Запорізького національного технічного університету, Запоріжжя, Україна

²Канд. техн. наук, доцент, доцент кафедри програмних засобів Запорізького національного технічного університету, Запоріжжя, Україна

³Канд. техн. наук, доцент кафедри програмних засобів Запорізького національного технічного університету, Запоріжжя, Україна

КОМПЛЕКСНИЙ МЕТОД ВИДОБУВАННЯ, ФОРМАЛІЗАЦІЇ ТА АГРЕГАЦІЇ ЕКСПЕРТНИХ ОЦІНОК КОНКУРУЮЧИХ АГЕНТІВ У ГРУПІ

Актуальність. Розглянуто проблему видобування, формалізації та агрегації експертних оцінок під час пошуку найкращого рішення з множини можливих. Актуальність даної проблеми визначається різноманітністю галузей застосування, необхідністю додаткового аналізу та груповим оцінюванням за умов невизначеності.

Мета. Мета даного дослідження полягала в підвищенні якості рішень, що приймаються групою експертів, за рахунок підвищення ефективності процесу отримання індивідуальних експертних оцінок та вдосконалення процесу їх подальшої агрегації.

Метод. Запропоновано комплексний метод, що складається з індивідуального та групового оцінювання, для розв'язання даної проблеми.

Удосконалено метод видобування та формалізації індивідуальних експертних оцінок. Запропонований метод ґрунтується на модифікації методу аналізу ієрархій, містить етапи абсолютного та відносного оцінювання, дозволяє визначити компетентність експерта шляхом обчислення коефіцієнта довіри до його суджень.

Удосконалено метод визначення групових оцінок агентів на основі підсумовування індивідуальних експертних оцінок за кожним рішенням. Даний метод враховує під час обчислення сумарної оцінки кількісну перевагу агентів один над одним та коефіцієнти довіри до суджень кожного учасника групи. Це дозволяє вважати судження компетентних експертів більш значущими з метою підвищення якості отриманого рішення.

Результати. Проведене експериментальне дослідження підтверджує працездатність розробленого математичного забезпечення. Розроблені методи продемонстрували здатність визначати і усувати неочевидні проблеми в задачах, що розв'язуються.

Висновки. Наукова новизна роботи полягає в тому, що запропоновано комплексний метод видобування, формалізації та агрегації експертних оцінок у групі, який дозволяє визначати комплексну оцінку конкуруючих агентів, а також обчислювати кількісне значення коефіцієнту довіри до суджень експерта безпосередньо в процесі отримання індивідуальної оцінки і в подальшому використовувати її для прийняття групових рішень.

Практична значимість результатів роботи полягає в тому, що розроблено інформаційну технологію, яка реалізує запропонований комплексний метод, що дозволило застосувати на практиці даний метод для розв'язання задачі проведення тендерів, та проведено експерименти.

Ключові слова: експертне оцінювання, групові оцінки, індивідуальні оцінки, компетентність експертів, конкуруючі агенти.

Колпакова Т. А.¹, Олейник А. А.², Левкин В. Н.³

¹Ст. преподаватель кафедры программных средств Запорожского национального технического университета, Запорожье, Украина

²Канд. техн. наук, доцент, доцент кафедры программных средств Запорожского национального технического университета, Запорожье, Украина

³Канд. техн. наук, доцент кафедры программных средств Запорожского национального технического университета, Запорожье, Украина

КОМПЛЕКСНИЙ МЕТОД ИЗВЛЕЧЕНИЯ, ФОРМАЛИЗАЦИИ И АГРЕГАЦИИ ЭКСПЕРТНЫХ ОЦЕНОК КОНКУРИРУЮЩИХ АГЕНТОВ В ГРУППЕ

Актуальность. Рассмотрена проблема извлечения, формализации и агрегации экспертных оценок при поиске наилучшего решения из множества возможных. Актуальность данной проблемы определяется разнообразием областей применения, необходимостью дополнительного анализа и групповым оцениванием в условиях неопределенности.

Цель. Цель данного исследования заключалась в повышении качества решений, принимаемых группой экспертов, за счет повышения эффективности процесса получения индивидуальных экспертных оценок и усовершенствования процесса их последующей агрегации.

Метод. Предложено решение рассматриваемой проблемы в виде комплексного метода, состоящего из индивидуального и группового оценивания.

Усовершенствован метод извлечения и формализации индивидуальных экспертных оценок. Предложенный метод основан на модификации метода анализа иерархий, включает в себя этапы абсолютного и относительного оценивания, позволяет определить компетентность эксперта путем вычисления коэффициента доверия к его мнению.

Усовершенствован метод определения групповых оценок агентов на основе суммирования индивидуальных экспертных оценок по каждому решению. Этот метод учитывает при вычислении суммарной оценки количественное преимущество агентов друг над другом и коэффициенты доверия к мнению каждого участника группы. Это позволяет считать мнение компетентных экспертов более значимым с целью повышения качества получаемого решения.

Результаты. Проведенное экспериментальное исследование подтвердило работоспособность разработанного математического обеспечения. Разработанные методы продемонстрировали возможность выявлять и устранять неочевидные проблемы в решаемых задачах.

Выводы. Научная новизна работы заключается в том, что предложен комплексный метод извлечения, формализации и агрегации экспертных оценок в группе, который позволяет определить комплексную оценку конкурирующих агентов, а также вычислить количественное значение коэффициента доверия к мнению эксперта непосредственно в процессе получения индивидуальной оценки и в дальнейшем использовать его для принятия групповых решений.

Практическая значимость результатов работы заключается в том, что разработана информационная технология, реализующая предложенный комплексный метод, что позволило применить на практике данный метод для решения задач проведения тендеров, и проведены эксперименты.

Ключевые слова: экспертное оценивание, групповые оценки, индивидуальные оценки, компетентность экспертов, конкурирующие агенты.

REFERENCES

- Karnyishev A. D., Ivanova E. A. *Psihologiya deyatelnosti i upravleniya: ucheb. posobie*. Irkutsk, IGEA, 2001, 186 p.
- Mulen E. *Kooperativnoe prinyatie resheniy: aksiomy i modeli*. Moscow, Mir, 1991, 464 p.
- Mayers D. *Sotsialnaya psihologiya*. Piter, 2013, 800 p.
- Ambrusa A., Greinerb B., Pathakc P. A. How individual preferences are aggregated in groups: An experimental study. *Journal of Public Economics*, 2015, Volume 129, pp. 1–13.
- Budziński R., Becker J. Model of competence of experts in the computer decision support system. *Quantitative Methods In Economics*, 2013, Volume XIV, Issue 1, pp. 53–64.
- Patel H. T. A Study on the Effectiveness of Group Activity and Group Discussion Method in English. *International Journal of Research in Humanities and Social Sciences*, 2014, Volume 2, Issue 1, pp. 13–15.
- Parkes D. C., Xia L. A Complexity-of-Strategic-Behavior Comparison between Schulze's Rule and Ranked Pairs. *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012, pp. 1429–1435.
- Roszkowska E. Rank ordering criteria weighting methods – a comparative overview. *Optimum. Studia Ekonomiczne*, 2013, No. 5 (65), pp. 14–33.
- Saaty T. L., Vargas L. G. *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. Springer US, 2012, 346 p.
- Nguyen H. T., Dawal S. Z. M., Nukman Y., Aoyama H., Case K. An Integrated Approach of Fuzzy Linguistic Preference Based AHP and Fuzzy COPRAS for Machine Tool Evaluation. *PLOS One*, 2015, No. 10 (9), 24 p.
- Oliinyk A., Zaiko T., Subbotin S. Training Sample Reduction Based on Association Rules for Neuro-Fuzzy Networks Synthesis. *Optical Memory and Neural Networks (Information Optics)*, 2014, Vol. 23, No. 2, pp. 89–95.
- Oliinyk A., Subbotin S. A. The decision tree construction based on a stochastic search for the neuro-fuzzy network synthesis. *Optical Memory and Neural Networks (Information Optics)*, 2015, Vol. 24, No. 1, pp. 18–27.
- Subbotin S. A. The method of diagnostic model synthesis based on radial basis neural networks with the support of generalization properties. *Radio Electronics. Computer Science. Control*, 2016, No. 2, pp. 64–69.
- Oliinyk A. Production rules extraction based on negative selection. *Radio Electronics. Computer Science. Control*, 2016, No. 1, pp. 40–49.
- Oliinyk A. O., Skrupsky S. Yu., Subbotin S. A. Experimental Investigation with Analyzing the Training Method Complexity of Neuro-Fuzzy Networks Based on Parallel Random Search. *Automatic Control and Computer Sciences*, 2015, Vol. 49, Issue 1, pp. 11–20.
- Bodyanskiy Ye. V., Tyshchenko O. K., Boiko O. O. An evolving cascade system based on neurofuzzy nodes. *Radio Electronics. Computer Science. Control*, 2016, No. 2, pp. 40–45.
- Oliinyk A. A., Subbotin S. A. Association Rules Extraction for Pattern Recognition. *Pattern Recognition and Image Analysis*, 2016, Vol. 26, No. 2, pp. 419–426.
- Oliinyk A. O., Skrupsky S. Yu., Subbotin S. A. Using Parallel Random Search to Train Fuzzy Neural Networks. *Automatic Control and Computer Sciences*, 2014, Vol. 48, Issue 6, pp. 313–323.
- Subbotin S., Oliinyk A., Skrupsky S. Individual prediction of the hypertensive patient condition based on computational intelligence. *Proceedings of the International Conference on Information and Digital Technologies*, 2015, pp. 336–344.
- Pangeran M. H., Pribad K. S. Conceptual model of analytical network process for prioritizing risk in a PPP infrastructure project. *Proceedings of the First Makassar International Conference on Civil Engineering*, 2010, pp. 1217–1227.
- Jao C. S. Efficient Decision Support Systems – Practice and Challenges in Multidisciplinary Domains. *InTech*, 2011, 478 p.
- Rush C., Roy R. Expert Judgement in Cost Estimating: Modelling the Reasoning Process. *Concurrent Engineering*, 2001, No. 9, pp. 271–284.
- Andreychikov A. V., Andreychikova O. N. *Analiz, sintez, planirovanie resheniy v ekonomike*. M., Finansyi i statistika, 2012, 368 p.
- Shrotriya S., Pandey A. Prediction of the Winner by Using a Weighted Approach of Preferential Balloting Systems on the Basis of Their Satisfied Criteria and Artifice Behavioral Complexity. *3rd International Conference on Information Security and Artificial Intelligence (ISAI 2012)*, Singapore, 2012, pp. 142–146.
- Wang B., Qian Y. Determining decision makers' weights in group ranking: a granular computing method. *International Journal of Machine Learning and Cybernetics*, 2015, Volume 6, Issue 3, pp. 511–521.
- Tao L. *Decision Support for Contractor Selection: Incorporating consolidated Past Performance Information*. University of Hong Kong, 2010, 184 p.

DEVELOPMENT OF A SUPPORT SYSTEM FOR MANAGING THE CYBER SECURITY

Context. In this paper the urgent problem of development of software of decision making support systems in information security is solved. Approach is based on a choice of rational options of response to events taking into account operational state-of-health data of a subject to protection.

Objective. Goal of the research is developing a cyber-threats counterwork model using decision support system, choosing rational variants of reactions on the occurrences in cybersecurity, and taking into account current operational data.

Method. The information object cyber security operational management system and the formation of the protection methods rational sets model which is based on a morphological approach is developed. It is proposed to find an optimal variant of the information security perimeter sets using an object function that maximizes the correlation of a consolidated figure of "information security" to consolidated figure "costs".

Results. A model for the operational management of cyber security-critical computer systems was developed. This model allows us to generate different variants of protection sets that are compliant with a computer system taking into account morphological matrices for each security perimeter prepared with the intelligent decision support system. It is proved that the use of the developed decision support systems can significantly reduce the costs planned for the complex means of cyber defense, as well as reduce the time to inform decision-makers on how to counter the identified information security incidents.

Conclusions. Scientific novelty of research consists that the model of operational management of cyber security of an information objects and formation of a rational complex of security features based on morphological approach is for the first time offered. The practical value of the developed methods and instruments is that they allow: to reduce time of development of systems of cyber security, to increase efficiency of planning of rational modular composition of security features due to creation of information and software environment in case of design; to increase validity of the made decisions on operational and to organizational technical control by protection.

Keywords: information safety, information security management, decision support system, morphological approach.

NOMENCLATURE

ACS – automated control systems;
AEIS – audit of CS events;
AVP – antivirus protection;
B – backup;
CIIO – critically important information object;
CONTS – controlling system;
CA – cyber-attacks;
CS – cybersecurity;
DA – data array;
DIC – data integrity control;
DSS – decision support system;
ES – expert systems;
ICS – information-communicative systems;
IDSS – intelligent decision of information security OM support system;
IO – information objects;
IPM – information protection means;
IPR – information protection;
IS – informational security;
ISMC – information security method complex rational structure;
ISMS – information security management system;
ISS – information security systems;
MACS – monitoring and analysis of CS;
MIP – means of information protection;
NLAC – network-level access control;
OM – operational management;
OPIO (V) – the outer perimeter of information object;
PCOI (II) – perimeter of control of information object;
PIS (I) – the perimeter of the information system;
PNE (IV) – the perimeter of the network equipment;

PSIO – physical security of IO;
SDCA – subsystem of detection of CA;
UAC – User Access Control;
UAP (III) – User Access Perimeter;
 A – linguistic variable "number of unusual events in network against the spreading of CA";
 AL – the number of alternative variants
 As_i – DA criticality in i -IO crosspoint;
 At_i – IS breach level in i -IO crosspoint;
 B – linguistic variable «number of unusual events in the host»;
 C – linguistic variable «number of unusual events in IO perimeter»;
 CE – communication equipment in an information channel;
 C_{ICR} – coefficient that allows to represent a result in a range [0; 1];
 DA_j – damage assessment;
 DPT_{lm} – security tool for the realization of functional subsystem of l ;
 EST – external source of threat;
 $ICA_{l(m)}$ – internal CA against IA with k -level of criticality;
 IMI_{cis_i} – IS incident importance in i -IO crosspoint;
 IND – network or IO perimeter indicator;
 IST_l^{k-1} – insider source of threat;
 l, m – numbers of crosspoints;
 L – number of functional subsystems for the perimeter of the IS;

LS_i – security measures level in i -IO crosspoint;
 MA_{m_l} – indicator of “expenses” for functional subsystem l ;
 $MA_{K_m}^{ST_{li}}$ – indicator value “expenses” for security tools ST_{lm} ;
 $MA_{K_{LS}}^{ST_{li}}$ – indicator value “security”;
 MA_l – data for the choice of rational variants;
 MA_{LS_l} – indicator of “information security”;
 n – quantity of crosspoints in IO structure;
 NN_m^k – IO crosspoint, on which information with the highest level of criticality (k);
 O – access object;
 OF – object function of choice;
 P_{CA} – CA probability;
 $P(z_l)$ – (l) environment status probability;
 PP – protocols and packets;
 PUR – purpose of decision making;
 RCA – remote CA against IO;
 RE_j – result;
 RO_i – reaction variant;
 $RO^{rat}(P_{CA})$ – rational variant of reaction;
 RSV – range of synthetic variants of setup;
 RUL – model of decision making regarding the choice of the optimal variant of CS tools
 SFS – range of functional sub-systems for perimeter IS;
 SS_{ne}, SS_h – security services against the method of an CA spreading (networked and host);
 ST_{lm} – security tool for the realization of functional subsystem of l ;
 TL_i – trust level of a device, which reports about IS breaches in i -IO crosspoint;
 WCA – possible spreading ways of CA against IO crosspoints;
 X – IS events indicators numeric evaluation;
 z – environment status uncertainty characteristic;
 $\tau_a(p_i)$ – crossing;
 $\zeta_a(wca_i)$ – crossing that determines an indicators set which reacts against the CA.

INTRODUCTION

It is impossible to imagine modern attitudes and perspectives of further ICS development in different fields of human activity without the increased attention of questions regarding IS and CS particularly because of the increasing number of CA and the destructive influence on IO. The rapid increase of incidents in the field of IS has shown that existing ISS, which are built on the basis of known threats and emerging attacks, are not always effective in cases of new CA which are created against the widespread enterprise information system, ACS in electronics, industry, transport, the banking system etc.

Goal of the research – developing a cyber-threats counterwork model using DSS, choosing rational variants of reactions on the occurrences in CS, and taking into account current operational IO data.

1 PROBLEM STATEMENT

Suppose that in the process of organizational and technical cyber security management of the CIIO, the protection methods rational sets model planning stage (information protection means) is considered as a process of sequential removal of uncertainty of ISS structure and composition. Thus, the planning of rational compatible software and hardware sets IPM is a consideration of alternatives $AL: PL = SFS \rightarrow CS_{al}$. Then the decision selection by the IDSS is regarded as forming a subset of the best options set $CS' \subseteq CS$.

In the study, the problem of comparing sets of IPM options is examined using morphological matrix sets in terms of “information security” in the perimeter ISS CIIO and “costs” for l functional subsystem ISS, which operate in conditions of uncertainty, inconsistency and lack of knowledge about the state of the object which is protected.

2 REVIEW OF THE LITERATURE

The increasing number of IS and CS threats has given rise to the surge of research in the field of development of uncovering and preventing CA systems [1–4], and also DSS [5, 6] and ES [7–9] in this field. Publication analysis [10, 11], allows us to uncover the increasing popularity of ISS risk assessment automated methods [12] and program sets of IS and CS risk management [13]. It was mentioned in the works [14, 15] that ISMS, in which intelligent technologies of cyber-threat identification and reacting to occurrences of IS breaches are realized, are products of private companies, and that a customer in general doesn't have any information about methods and models of leading effects forming in systems [16]. It is shown in the works [17–20] that it is appropriate to equip existing DSS and ES in field of IS (excluding tasks of cyber-security management) with functional models that allow us to increase efficiency of enumeration and investigation of illegitimate interferences to the work of ICS crimes.

In such a way, according to the disputes in publications [5, 6, 8, 10, 16, 17], dedicated to the potential of using integrated DSS or ES in ISMS, the task of developing methods, models for using them in practice in intelligent support of ISS rational structure planning and the task of assessment and prediction of IS and CS risks became relevant.

3 MATERIALS AND METHODS

There is one main problem creating the CONTS – development of the threat model [7, 15, 21], which is connected to the specification of a management object interaction – ISS IO with the environment. IDSS, which develops a threat model building method, is based on a qualified scheme of goal-oriented destructive influences on IS and CS IO [22–24]. A generalized architecture of ISMS and CS is offered according to the results of the control strategy in conditions of uncertainty analysis [4, 15, 24].

Level of safety is used in the capacity of an operated variable. The LS value depends on the maximum level of information urgency which is being updated according to recent changes in ICS. Models [4, 15] consist of five perimeters for decentralized architecture of IO, fig. 1.

Mechanisms of IPR control are created in the circuit with organizational-technical control governing changing business applications, DA processing plans, infrastructure, and all the corresponding requests to the information safety level. The circuit contains: IDSS in regards to choosing a security strategy and a system of safety level assessment. Managing influence in the circuit is realized by the staff of the IS department. The task of ISMC rational structure choice for IO is made according to the following criteria [4, 9, 7, 15]: minimum probability of achieving goals by an attacker; minimum of IO losses should the attacker's goals be achieved; maximum probability of successful ISMC counteraction to the actions of an attacker; minimum "cost-risk" integrated index value [4, 9].

Quantity assessment of IO safety can be found the following way:

$$LS_{CIS} = \prod_{i=1}^n (1 - C_{ICR} \cdot At_i \cdot As_i \cdot TL_i \cdot DPT_{Im_i}). \quad (1)$$

Quantity of insider and external CA against IO are given in the form of tuples:

$$RCA = \langle EST, CE, SS_{ne}, SS_h, PP, O(NN) \rangle, \quad (2)$$

$$ICA_{l(m)} = \langle IST_l^{k-1}, CE, SS_{ne}, SS_h, PP, O^k(NN_m^k) \rangle. \quad (3)$$

It is proven in works [4, 10, 12, 20] that the only effective way to identify an attack is in the analysis of a combination of unusual events. That is why in IDSS, an attack spreading WCA possible ways, quantity is compared to a quantity of

indicators IND . The probability of the fact that suspicious action is a attack is assessed with the indicators quantity which reacted against the attack spreading method. Crossing $\tau_a(p_i)$ determines an indicators set. We get the following expression:

$$\zeta_a \subseteq WCA \times IND = \{ (wca_i, ind_j) : wca_i \in WCA \wedge ind_j \in IND \} \quad (4)$$

In conditions when the status of the information environment is unknown, the threat counteraction model is enabled in IDSS, which has an opportunity to choose a controlling influence that better corresponds to the management object status. A process of choosing an optimal safety events reaction variant are given in a form of a tuple:

$$\langle RO_i, RE_j, DA(RE_j), P_{CA}, P(z_l), OF, RO^{rat}(P_{CA}) \rangle. \quad (5)$$

Safety events [4, 6, 9] reaction variants probability analysis $\{RO_i\}$ has shown that the number of control influences for each situation is limited $i \in [1,3]$. An alternative advantages evaluation with a damage assessment model is used in IDSS – $\{RE_j\}, j \in [1,4]$ taking into account that the IS events reaction variants choice is made in conditions of a potential CA: no harm, losses for a certain user, losses for a group of users, loss for all ICS from attack realization.

Define a function with which we choose an optimal reaction variant:

$$OF(RO_i, z) = \sum_{l=1}^s DA_j(RE_j(RO_i, z_l)) \cdot \prod_{i=1}^I p_{ij}(RE_j(RO_i), P_{CA}). \quad (6)$$

The probability p_{ij} of getting every j -result choosing every i -reaction variant is determined the following way:

$$p_{ij} = p_{ij}(RE_j(RO_i), P_{CA}), \forall i: \sum_j p_{ij} = 1. \quad (7)$$

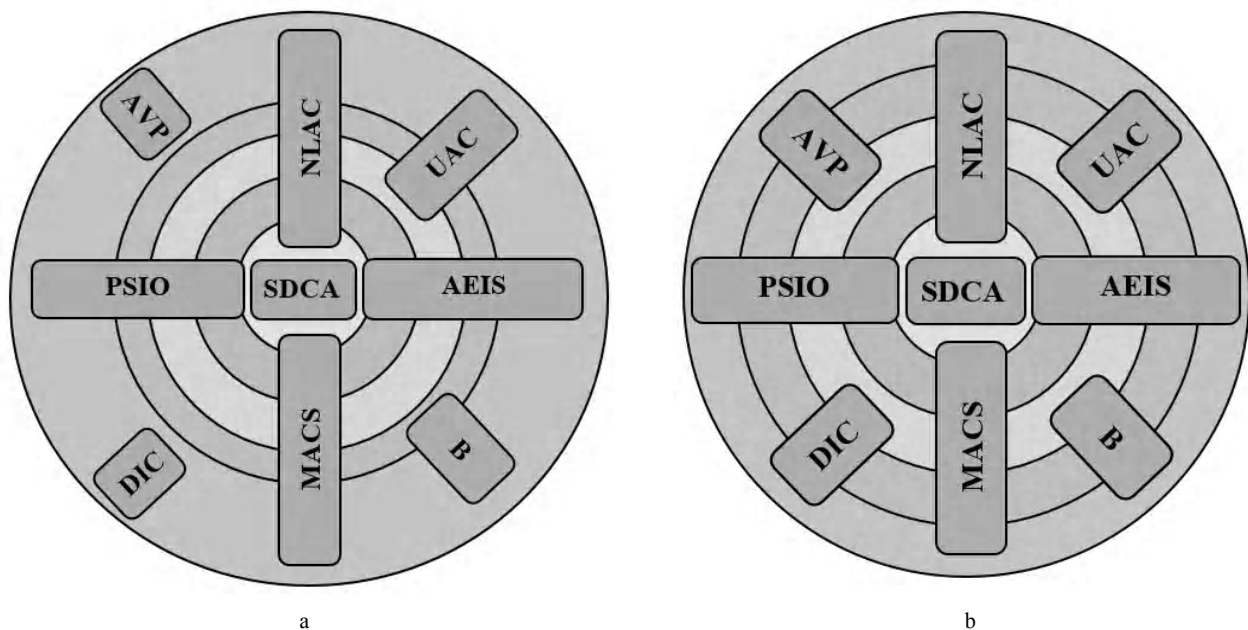


Figure 1 – IS subsystems:
a – Centralized IO variant; b – Decentralized IO variant

Control influence rational variant $RO^{rat}(P_{CA})$ is determined this way:

$$RO^{rat}(P_{CA}) = RO\left(\arg \min_i (OF(RO_i, z))\right). \quad (8)$$

An IS OM intelligent support subsystem contains: a fuzzy inference mechanism for CA probability numeric evaluation; organized structure information about knowledge database events; threat recognition and counteraction models [4, 5, 9]; algorithm for making a decision regarding choosing an optimal safety events reaction variant [8].

During the organizational-technical management process, the stage of planning of storage for information security tools, and the process of gradual removal of uncertainty about the structure and the storage of information security tools in the information security system is being considered. The process of planning PL rational sets MIP is described with the formula:

$$PL = SFS \rightarrow RSV_{al}. \quad (9)$$

With the help of the system for intelligent support, the process of choosing optimal variant of MIP setup for perimeters of CS is considered as the formation of a sub-range for the best variants of setup $RSV' \subseteq RSV$. The range of the setup variants is described as

$$RSV = \{RSV_1, \dots, RSV_{AL}\}. \quad (10)$$

For the choice of the optimal variant of CS tools the objective function OF is used:

$$RSV_{al} = OF(RSV). \quad (11)$$

The population of data, which make it possible to compare variants of setups, includes two sub-ranges:

$$MA_{LS_l} \subset MA_l \quad \& \quad MA_{in_l} \subset MA_l. \quad (12)$$

Usage a morphological approach, the model of decision making regarding the choice of the optimal variant of CS tools, is presented in the form of the sequence:

$$RUL : \langle PUR, SFS, RUL_s, RSV, MA_l, OF, RSV_r(RSV') \rangle. \quad (13)$$

The starting data for the synthesis of variants of CS tools sets: $SFS = \{SFS_1, \dots, SFS_L\}$. Accepted:

$$RSV = SFS_1 \times \dots \times SFS_L \quad \& \quad SFS_l = \{ST_{l1}, \dots, ST_{lK_l}\}. \quad (14)$$

The choice of the rational variants of information security tools setups is realized on the basis of experts' knowledge in the field of CS. The process of formation of a rational structure of information security tools is divided into five stages: 1) The variants of MIP setup are put under development. The range of possible variants of the solution is set with the help of the morphological matrix. For the considered perimeters of information security, the morphological matrices of CS tools are developed; 2) The intermediary matrices are filled in, which indicate compatibility with the firmware. For each couple of CS tools

for different functional subsystems their compatibility is defined. The result is filled in the table. If MIP are consistent, then the function of compatibility is $s(ST_{lm}, ST_{pr}) = 1$, in other case $s(ST_{lm}, ST_{pr}) = 0$; 3) The range of decisions concerning the choice of variants of setup MIP is generated. The range of tool setups is reduced to the sub-range which is known to be compatible with each other. The range $RSV = \{RSV_1, \dots, RSV_R\}$, which consists of all possible variants of formulation of setup MIP for the considered perimeter, is the Descartes's production of setups of alternatives (morphological matrix ranges).

The element of the range is presented as follows:

$$RSV_r = \left\{ (ST_{li}, ST_{2j}, \dots, ST_{Ln}) : ST_{lm} \in SFS_l, \forall l = \overline{1, L} \right\}. \quad (15)$$

The generation of the range of decisions concerning the choice of the variants, which consists of compatibility with each other MIP , is done in the following way. The iterated synthesis of variants of setup is made, which consists of compatibility MIP : on the first step the variants of CS tools for the first sub-system are consequently enumerated, after the choice of the alternative ST_{li} the transition to the second stage is made. On the second stage the consequent enumeration of variants of CS tools for the second subsystem is made, but the choice is made only for those alternatives ST_{2j} , for which the compatibility function is $s(ST_{li}, ST_{2j}) = 1$ etc. For the choice of alternatives for l subsystem, the choice is made only from those alternatives ST_{lm} , for which the compatibility function is equal to one: $s(ST_{l-1,m}, ST_{lm}) = 1, \dots, s(ST_{li}, ST_{lm}) = 1$.

Hence, the choice of MIP from each set of morphological matrix (one from each range) for the formation of a variant of setup is made only from compatibility with each other's firmware.

The further reduction of the CS range in the system of intelligent support for decision making is made in the form of a full enumeration with the given objective function. As the objective function for the choice of the setup variant, $CS_r = \{ST_{li}, ST_{2j}, \dots, ST_{lm}, \dots, ST_{Ln}\}$ the function is used

$$OF = \max_r \frac{MA_{LS}^{ST_{li}} + \dots + MA_{LS}^{ST_{Ln}}}{MA_{in}^{ST_{li}} + \dots + MA_{in}^{ST_{Ln}}}. \quad (16)$$

The criteria of the quality of information "security" indicator is divided into two groups: the indicators of the effectiveness of the operative methods of the security and the indicators of the functional fitness.

4 EXPERIMENTS

The software package ("Decision Support System of Management protection of information – DMSSCIS") was developed for check of working capacity and practical applicability of the offered model of operational management of cyber security [4, 15]. In the course of the experimental

check of SP reaction options (RO_i) decision making support systems on different classes of CA for the current parameters of probability of implementation of the attack of P_a were researched. Also sets of instruments of information protection $CS' \subseteq CS$ for the purpose of a choice of rational option were researched. Restrictions on the cost of a set are accepted and minimum probabilities of successful implementation attacking all are more whole than CA for the selected set. DMSSCIS was also used in the modernization of existing information security systems in data centers of transport companies in Dnipro (2014) and several industrial enterprises in Kyiv.

5 RESULTS

On the software “DMSSCIS”, that particular selection method implemented an efficient option for responding to security events. The results are shown in Table 1.

During the research the possibility was taken into account of an attack that implements remote intrusion through the perimeter, the availability of internal and external users, and abusers that have high privileges and violate the safety of information. After the formation of efficient information security in enterprises which took part in the study, with the help of intelligent decision support “DMSSCIS” the predicted value was $P_a = 1,78-1,91\%$ risk that there was an average value of 5,9–6,2 times less risk to information security systems compared to before.

The amount of expenditure by the organization on information security for critical nodes of information objects order from 5200–5500 \$. The likelihood of the offender achieving all their goals is 10^{-2} . The increase in appropriations for the organization of information security above a certain level (above 13,000 \$.) is inappropriate because it does not lead to a significant increase in the efficiency of information security.

During the research it was shown that the implementation of the intelligent decision support “DMSSCIS” allows an increased level of automation and centralized monitoring of CS facility and reduces the time to inform those responsible for information security incidents by 6,9–7,2 times.

6 DISCUSSION

The approach of building a comprehensive information security system for the information object makes it possible to reduce the cost of data protection by 32–35% compared to alternative methods [6–9, 12, 13]. A certain lack of intelligent systems of decision support of “DMSSCIS”, required the involvement in the initial study of several independent experts to build membership functions of production and assembly rules. At the current stage of research for this instrument, the fuzzy logic Fuzzy Toolbox (Matlab) was employed, which calculated “security information” MIP parameters for everyone involved in perimeter protection.

Overall, based on the studies, we can ascertain the effectiveness of the proposed models and software for information security management (information systems and automated control system) in examined enterprises.

CONCLUSIONS

In this paper the urgent problem of development of software of decision making support systems in CS of OI.

Scientific novelty of research consists that the model of operational management of CS of an OI and formation of a rational complex of security features based on morphological approach is for the first time offered. Unlike existing solutions, the model prepared on the basis of intelligent decision support, a morphological matrix for each facility’s perimeters of information protection, and can generate a set of options for remedies which take into account the compatibility of software and hardware. The choice of the optimal option set for that perimeter protection of information, implements an objective function that maximizes the ratio of the sum “security information” to the total rate “cost”. It provides a range of remedies for a given class of certified security, and satisfies the requirements for eligible costs for implementation of CS.

The practical value of the developed methods and instruments is that they allow: to reduce time of development of systems of CS, to increase efficiency of planning of rational modular composition of security features due to

Table 1 – The results of testing the software system “DMSSCIS”

Class of CA	Options for responding to the current settings information of environment of information object RO_i		
U2R	$A=2, B=3, P_a = 0,54$		$A=1, B=1, P_a = 0,242$
	The end of session attack source node	Sending a warning message to the user	
R2L	$A=1, B=3, P_a = 0,43$		$A=1, B=1, P_a = 0,192$
	The end of session attack source node	Sending a warning message to the user	
DOS/DDOS	$A=2, B=3, P_a = 0,62$		$A=1, C=2, P_a = 0,4$
	The end of session attack source node	Sending a warning message to the user	
The external attack (Wi Fi)	$A=3, C=3, P_a = 0,678$	$A=1, C=2, P_a = 0,4$	$A=1, C=1, P_a = 0,3$
	Blocking access point	DOS-attack on stations	The lack of response
A remote attack via lines on the perimeter	$A=3, B=4, C=2, P_a = 0,82$		$A=1, B=1, C=1, P_a = 0,224$
	Blocking access to the server in the network	Reconfiguration of security services to block IP	$A=1, P_a = 0,076$
Cost of a rational set of means of information protection $CS' \subseteq CS$ for OI			
Experts		“DMSSCIS”	
1000–11000 \$		7000–8000 \$	

creation of information and software environment in case of design; to increase validity of the made decisions on operational and to organizational technical control by protection. By using the developed system of intelligent decision support, networks of enterprises using DMSSCIS reduced the projected cost of the planned system of protection to 35 %. Further development of this work may be improving the interaction of traditional mechanisms of information security IO, which, in particular, are working on primary information system modules and intelligent decision support “DMSSCIS”.

ACKNOWLEDGEMENTS

The work is supported by the state budget scientific research project of Dnipropetrovsk National University of Railway Transport named after academician V. Lazaryan “Automation of processes of complex technical systems in conditions of uncertainty” (state registration number 0104U005401).

REFERENCES

1. Distributed Intrusion Detection System in a Multi-Layer Network Architecture of Smart Grids / [Y. Zhang, L. Wang, W. Sun, R. C. Green II et al] // *IEEE Transactions on Smart Grid*. – 2011. – Vol. 2, No. 4. – P. 796–808. DOI:10.1109/TSG.2011.2159818
2. Al-Jarrah O. Network Intrusion Detection System using attack behavior classification / O. Al-Jarrah, A. Arafat // *5th International Conference Information and Communication Systems (ICICS)*. – 2014. – P. 1–6. DOI: 10.1109/IACS.2014.6841978
3. Louvieris P. Effects-based feature identification for network intrusion detection / P. Louvieris, N. Clewley, X. Liu // *Neurocomputing*. – 2013. – Vol. 121, Iss. 9. – P. 265–273. DOI:10.1016/j.neucom.2013.04.038
4. Lakhno V. Creation of the adaptive cyber threat detection system on the basis of fuzzy feature clustering / V. Lakhno // *Eastern-European Journal of Enterprise Technologies*. – 2016. – Vol. 2, No 9(80). – P. 18–25. DOI: 10.15587/1729-4061.2016.66015
5. Cybersecurity Games and Investments: A Decision Support Approach / [E. Panaousis, A. Fielder, P. Malacaria, C. Hankin et al] // *Chapter Decision and Game Theory for Security*. – 2014. – Vol. 8840. – P. 266–286. DOI: 10.1007/978-3-319-12601-2_15
6. Cavusoglu H. Decision-theoretic and game-theoretic approaches to IT security investment / H. Cavusoglu, R. Srinivasan, T. Y. Wei // *Journal of Management Information Systems*. – 2008. – Vol. 25(2). – P. 281–304.
7. Li-Yun Chang. Applying fuzzy expert system to information security risk Assessment – A case study on an attendance system / Li-Yun Chang, Zne-Jung Lee // *2013 International Conference on Fuzzy Theory and Its Applications*. – P. 346–351. DOI: 10.1109/iFuzzy.2013.6825462
8. Atymtayeva L. Building a Knowledge Base for Expert System in Information Security / L. Atymtayeva, K. Kozhakhmet, G. Bortsova // *Chapter Soft Computing in Artificial Intelligence*. – 2014. – Vol. 270. – P. 57–76. DOI:10.1007/978-3-319-05515-2_7
9. Kanatov M. Expert systems for information security management and audit / M. Kanatov, L. Atymtayeva, B. Yagaliyeva // *Implementation phase issues, Soft Computing and Intelligent Systems, Joint 7th International Conference on and Advanced Intelligent Systems (ISIS), 15th International Symposium on 3–6 Dec. 2014*. – P. 896–900. DOI:10.1109/SCIS-ISIS.2014.7044702
10. Yanga Y.-P. Ou. A VIKOR technique based on DEMATEL and ANP for information security risk control assessment / Y.-P. O. Yanga, H. Shieha, G. Tzeng // *Information Sciences*. – 2013. – Vol. 232. – P. 482–500. <http://dx.doi.org/10.1016/j.ins.2011.09.012>
11. Bulgurcu B. Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness / B. Bulgurcu, H. Cavusoglu, I. Benbasat // *MIS Quarterly*. – 2010. – No. 34(3). – P. 523–548.
12. Fuzzy logic based anomaly detection for embedded network security cyber sensor / [O. Linda, M. Manic, T. Vollmer, J. Wright] // *Computational Intelligence in Cyber Security (CICS), IEEE Symposium on 11–15 April 2011*. – P. 202–209. DOI: 10.1109/CICYBS.2011.5949392
13. Demetz L. To Invest or Not to Invest? Assessing the Economic Viability of a Policy and Security Configuration Management Tool / L. Demetz, D. Bachlechner // *The Economics of Information Security and Privacy*, Springer, 2013. – P. 25–47. DOI:10.1007/978-3-642-39498-0_2
14. Oglaza A. Authorization Policies: Using Decision Support System for Context-Aware Protection of User’s Private Data, Trust, Security and Privacy in Computing and Communications (TrustCom) / A. Oglaza, R. Laborde, P. Zarate // *2013 12th IEEE International Conference on 16–18 July 2013, Melbourne, VIC Australia*. – P. 1639–164. DOI: 10.1109/TrustCom.2013.202.
15. Lakhno Design of adaptive system of detection of cyber-attacks, based on the model of logical procedures and the coverage matrices of features / [V. Lakhno, S. Kazmirchuk, Y. Kovalenko, L. Myrutenko et al] // *Eastern-European Journal of Enterprise Technologies*. – 2016. – No. 3/9 (81). – P. 30–38. DOI: 10.15587/1729-4061.2016.71769
16. Gamal M. M. A Security Analysis Framework Powered by an Expert System / M. M. Gamal, B. Hasan, A. F. Hegazy // *International Journal of Computer Science and Security*. – 2011. – Vol. 4, No. 6. – P. 505–527.
17. Goztepe K. Designing Fuzzy Rule Based Expert System for Cyber Security / K. Goztepe // *International Journal of Information Security Science*. – 2012. – Vol. 1, No. 1. – P. 13–19.
18. Gutzwiller R. S. A task analysis toward characterizing cyber-cognitive situation awareness (CCSA) in cyber defense analysts / R. S. Gutzwiller, S. M. Hunt, D. S. Lange // *2016 IEEE International Multi-Disciplinary Conference on 21–25 March 2016*, DOI: 10.1109/COGSIMA.2016.7497780.
19. Decision support for Cybersecurity risk planning / [L. P. Reesa, J. K. Deanea, T. R. Rakesa, W. H. Bakerb] // *Decision Support Systems*. – 2011. – Vol. 51, Iss. 3. – P. 493–505. DOI.org/10.1016/j.dss.2011.02.013
20. Paliwal S. Denial-of-Service, Probing & Remote to User (R2L) Attack Detection using Genetic Algorithm / S. Paliwal, R. Gupta // *International Journal of Computer Applications*. – 2012. – Vol. 60, No.19. – P. 57–62.
21. Ben-Asher N. Effects of cyber security knowledge on attack detection / N. Ben-Asher, C. Gonzalez // *Computers in Human Behavior*. – 2015. – Vol. 48. – P. 51–61. DOI: 10.1016/j.chb.2015.01.039
22. Buryachok V. L. Algoritm ocinyuvannya stupenya zaxishhenosti special'nix informacijno-telekomunikacijnix sistem // *Zaxist informacii*. – 2011. – No. 3. – P. 19–27.
23. Valenzuela J. Real-Time Intrusion Detection in Power System Operations / J. Valenzuela, J. Wang, N. Bissinger // *IEEE Transactions on Power Systems*, 2013. – Vol. 28, No. 2. – P. 1052–1062. DOI:10.1109/TPWRS.2012.2224144
24. Потій О. В. Дослідження методів оцінки ризиків безпеки інформації та розробка пропозицій з їх вдосконалення на основі системного підходу / О. В. Потій, А. В. Леншин [Текст] // *Збірник наукових праць Харківського університету Повітряних Сил*. – 2010. – Вип. 2(24). – С. 85–91.

Article was submitted 16.12.2016.

After revision 28.12.2016.

Ляхно В. А.

Д-р техн. наук, доцент, зав. кафедри організації комплексного захисту інформації, Європейський університет, Київ, Україна
РОЗРОБКА СИСТЕМИ ПІДТРИМКИ РІШЕНЬ З УПРАВЛІННЯ КІБЕРЗАХИСТОМ

Актуальність. В роботі вирішена актуальна задача розвитку математичного забезпечення систем підтримки прийняття рішень з кібербезпеки на основі вибору оптимізованих варіантів реагування на інциденти. При цьому враховуються експлуатаційні параметри об'єкта захисту.

Мета. Розробка моделі протидії кіберзагрозам на основі застосування системи підтримки рішень по вибору оптимізованих варіантів реагування на інциденти кібербезпеки з урахуванням поточної інформації, яка стосується об'єктів захисту.

Метод. Розроблено модель оперативного менеджменту кібербезпекою критично важливих комп'ютерних систем і синтезу раціональних комплексів засобів захисту. Модель базується на морфологічній парадигмі. Запропоновано здійснювати вибір оптимізованих варіантів комплексів для периметрів кіберзахисту за допомогою цільової функції, яка максимізує відношення узагальненого індексу «захищеність інформації» до підсумкового показника «витрати».

Результати. Розроблено модель оперативного менеджменту кібербезпекою критично важливих комп'ютерних систем. Модель дозволяє з урахуванням сукупності морфологічних матриць, підготовлених в ході роботи системи підтримки прийняття рішень для кожного з розглянутих в роботі периметрів, згенерувати варіативні набори комплексів захисту, в яких врахована їх апаратно-програмна сумісність. Розроблено програмний комплекс для інтелектуальної підтримки прийняття рішень в задачах управління кібербезпекою об'єкта інформатизації. Доведено, що використання розробленої системи підтримки рішень дозволяє істотно зменшити плановані витрати на комплекси засобів кіберзахисту, а також скоротити час інформування відповідальних осіб про способи протидії виявленим інцидентам з інформаційної безпеки.

Висновки. Наукова новизна досліджень полягає в тому, що вперше запропонована модель оперативного менеджменту кібербезпекою критично важливих комп'ютерних систем і синтезу оптимізованих комплексів засобів захисту, що базується на морфологічній парадигмі. Практична цінність розроблених методів і засобів полягає в тому, що вони забезпечують: економію часу на етапі побудови комплексних систем кібербезпеки об'єктів інформатизації; підвищення ефективності вибору оптимізованих модульних систем кіберзахисту; аргументованість прийнятих рішень в процесі стратегічного, оперативного та організаційного менеджменту захистом об'єктів інформатизації.

Ключові слова: інформаційна безпека, управління захистом інформації, система підтримки рішення, раціональний набір засобів захисту.

Ляхно В. А.

Д-р техн. наук, доцент, зав. кафедри організації комплексної захисту інформації, Європейський університет, Київ, Україна
РАЗРАБОТКА СИСТЕМЫ ПОДДЕРЖКИ РЕШЕНИЙ ПО УПРАВЛЕНИЮ КИБЕРЗАЩИТОЙ

Актуальность. В работе решена актуальная задача развития математического обеспечения систем поддержки принятия решений по кибербезопасности на основе выбора оптимизированных вариантов реагирования на инциденты с учетом эксплуатационных параметров объекта защиты.

Цель. Разработка модели противодействия киберугрозам на основе применения системы поддержки решений по выбору оптимизированных вариантов реагирования на инциденты кибербезопасности с учетом текущей информации, касающейся объектов защиты.

Метод. Разработана модель оперативного менеджмента кибербезопасностью критически важных компьютерных систем и синтеза рациональных комплексов средств защиты. Модель базируется на морфологической парадигме. Предложено осуществлять выбор оптимизированных вариантов комплексов для периметров киберзащиты с помощью целевой функции, которая максимизирует отношение обобщенного индекса «защищенность информации» к итоговому показателю «затраты».

Результаты. Разработана модель оперативного менеджмента кибербезопасностью критически важных компьютерных систем. Модель позволяет с учетом совокупности морфологических матриц, подготавливаемых в ходе работы системы поддержки принятия решений для каждого из рассматриваемых в работе периметров, сгенерировать вариативные наборы комплексов защиты, в которых учтена их аппаратно-програмная совместимость. Разработан программный комплекс для интеллектуальной поддержки принятия решений в задачах управления кибербезопасностью объекта информатизации. Доказано, что использование разработанной системы поддержки решений позволяет существенно уменьшить планируемые расходы на комплексы средств киберзащиты, а также сократить время информирования ответственных лиц о способах противодействия выявленным инцидентам информационной безопасности.

Выводы. Научная новизна исследования состоит в том, что впервые предложена модель оперативного менеджмента кибербезопасностью критически важных компьютерных систем и синтеза оптимизированных комплексов средств защиты, базирующаяся на морфологической парадигме. Практическая ценность разработанных методов и средств заключается в том, что они обеспечивают: экономию времени на этапе построения комплексных систем кибербезопасности объектов информатизации; повышение эффективности выбора оптимизированных модульных систем киберзащиты; аргументированность принимаемых решений в процессе стратегического, оперативного и организационного менеджмента защитой объектов информатизации.

Ключевые слова: информационная безопасность, управление защитой информации, система поддержки решения, рациональный набор средств защиты.

REFERENCES

1. Zhang Y., Wang L., Sun W., Green R. C., Alam M. Distributed Intrusion Detection System in a Multi-Layer Network Architecture of Smart Grids, *IEEE Transactions on Smart Grid*, 2011, Vol. 2, No. 4, pp. 796–808. DOI:10.1109/TSG.2011.2159818
2. Al-Jarrah O., Arafat A. Network Intrusion Detection System using attack behavior classification, *Information and Communication Systems (ICICS), 2014 5th International Conference*, 2014, pp. 1–6. DOI: 10.1109/IACS.2014.6841978
3. Louvieris P., Clewley N., Liu X. Effects-based feature identification for network intrusion detection, *Neurocomputing*, 2013, Vol. 121, Iss. 9, P. 265–273. DOI:10.1016/j.neucom.2013.04.038
4. Lakhno V. Creation of the adaptive cyber threat detection system on the basis of fuzzy feature clustering, *Eastern-European Journal of Enterprise Technologies*, 2016, Vol. 2, No. 9(80): Information and controlling system, pp. 18–25. DOI: 10.15587/1729-4061.2016.66015
5. Panaousis E., Fielder A., Malacaria P., Hankin C., Smeraldi F. Cybersecurity Games and Investments: A Decision Support Approach, Chapter Decision and Game Theory for Security of

- the series Lecture Notes in Computer Science, 2014, Vol. 8840, pp. 266–286. DOI: 10.1007/978-3-319-12601-2_15
6. Cavusoglu H., Srinivasan R., Wei T. Y. Decision-theoretic and game-theoretic approaches to IT security investment, *Journal of Management Information Systems*, 2008, Vol. 25(2), pp. 281–304.
 7. Li-Yun, Chang, Zne-Jung Lee Applying fuzzy expert system to information security risk Assessment – A case study on an attendance system, *2013 International Conference on Fuzzy Theory and Its Applications*, 2013, pp. 346–351. DOI: 10.1109/iFuzzy.2013.6825462
 8. Atymtayeva L., Kozhakhmet K., Bortsova G. Building a Knowledge Base for Expert System in Information Security, *Chapter Soft Computing in Artificial Intelligence*, 2014, Vol. 270, pp. 57–76. DOI:10.1007/978-3-319-05515-2_7
 9. Kanatov M., Atymtayeva L., Yagaliyeva, B. Expert systems for information security management and audit. Implementation phase issues, *Soft Computing and Intelligent Systems (SCIS), Joint 7th International Conference on and Advanced Intelligent Systems (ISIS), 15th International Symposium on 3–6 Dec. 2014*, pp. 896–900. DOI:10.1109/SCIS-ISIS.2014.7044702
 10. Yu-Ping Ou Yanga, How-Ming Shieha, Gwo-Hshiung Tzeng A VIKOR technique based on DEMATEL and ANP for information security risk control assessment, *Information Sciences*, 2013, Vol. 232, pp. 482–500. <http://dx.doi.org/10.1016/j.ins.2011.09.012>
 11. Bulgurcu B., Cavusoglu H. and Benbasat I. Information security policy compliance: An empirical study of rationality-based beliefs and information security awareness, *MIS Quarterly*, 2010, No. 34(3), pp. 523–548.
 12. Linda O., Manic M., Vollmer T., Wright J. Fuzzy logic based anomaly detection for embedded network security cyber sensor, *Computational Intelligence in Cyber Security (CICS), IEEE Symposium on 11–15 April 2011*, 2011, pp. 202–209. DOI: 10.1109/CICYBS.2011.5949392
 13. Demetz L., Bachlechner D. To Invest or Not to Invest? Assessing the Economic Viability of a Policy and Security Configuration Management Tool, *The Economics of Information Security and Privacy*, Springer, Heidelberg, 2013, pp. 25–47. DOI:10.1007/978-3-642-39498-0_2
 14. Oglaza A., Laborde R., Zarate P. Authorization Policies: Using Decision Support System for Context-Aware Protection of User's Private Data, Trust, Security and Privacy in Computing and Communications (TrustCom), *12th IEEE International Conference on 16–18 July 2013*, 2013, pp. 1639–164. DOI: 10.1109/TrustCom.2013.202.
 15. Lakhno V., Kazmirchuk S., Kovalenko Y., Myrutenko L., Zhmurko T. Design of adaptive system of detection of cyber-attacks, based on the model of logical procedures and the coverage matrices of features, *Eastern-European Journal of Enterprise Technologies*, 2016, No. 3/9 (81), pp. 30–38. DOI: 10.15587/1729-4061.2016.71769
 16. Gamal, M. M., Hasan, B., Hegazy, A.F. A Security Analysis Framework Powered by an Expert System, *International Journal of Computer Science and Security*, 2011, Vol. 4, No. 6, pp. 505–527.
 17. Goztepe, K. Designing Fuzzy Rule Based Expert System for Cyber Security, *International Journal of Information Security Science*, 2012, Vol. 1, No. 1, pp. 13–19.
 18. Robert S. Gutzwiller, Sarah M. Hunt, Douglas S. Lange A task analysis toward characterizing cyber-cognitive situation awareness (CCSA) in cyber defense analysts, *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), IEEE International Multi-Disciplinary Conference on 21–25 March 2016*, 2016. DOI: 10.1109/COGSIMA.2016.7497780.
 19. Loren Paul Reesa, Jason K. Deanea, Terry R. Rakesa, Wade H. Bakerb Decision support for Cybersecurity risk planning, *Decision Support Systems*, 2011, Vol. 51, Iss. 3, pp. 493–505. DOI.org/10.1016/j.dss.2011.02.013
 20. Paliwal, S., Gupta, R. Denial-of-Service, Probing & Remote to User (R2L) Attack Detection using Genetic Algorithm, *International Journal of Computer Applications*, 2012, Vol. 60, No. 19, pp. 57–62.
 21. Ben-Asher N., Gonzalez C. Effects of cyber security knowledge on attack detection, *Computers in Human Behavior*, 2015, Vol. 48, pp. 51–61. DOI: 10.1016/j.chb.2015.01.039
 22. Burachok, V. Algorithm for evaluating the degree of protection of special information and telecommunication systems, *Information Security*, 2011, No. 3, pp. 19–27.
 23. Valenzuela J., Wang J., Bissinger N. Real-Time Intrusion Detection in Power System Operations, *IEEE Transactions on Power Systems*, 2013, Vol. 28, No. 2, pp. 1052–1062. DOI:10.1109/TPWRS.2012.2224144
 24. Potij O. V., Ljenshyn A. V. Doslidzhennja metodiv ocinky ryzykiv bezpeci informacii' ta rozrobka propozycij z i'h vdoskonalennja na osnovi systemnogo pidhodu, *Zbirnyk naukovyh prac' Harkivs'kogo universytetu Povitrjanyh Syl*, 2010, Vyp. 2(24), pp. 85–91.

УДК 004.056.5:004.738.5(045)

Молодецька-Гринчук К. В.

Канд. техн. наук, доцент, доцент кафедри комп'ютерних технологій і моделювання систем Житомирського національного агроєкологічного університету, Житомир, Україна

МЕТОД ВИЯВЛЕННЯ ОЗНАК ІНФОРМАЦІЙНИХ ВПЛИВІВ У СОЦІАЛЬНИХ ІНТЕРНЕТ-СЕРВІСАХ ЗА ЗМІСТОВНИМИ ОЗНАКАМИ

Актуальність. Сьогодні соціальні інтернет-сервіси перетворилися на ефективний інструмент комунікації між учасниками віртуальних спільнот – акторами. З огляду на високу швидкість поширення контенту, транскордонність процесів взаємодії акторів, наявність засобів для їх організації у групи, соціальні-інтернет сервіси можуть застосовуватися як дієвий засіб проведення інформаційних операцій проти людини, суспільства, держави. В процесі комунікації акторів у віртуальних спільнотах найчастіше використовується текстовий контент. Розроблення ефективних алгоритмів функціонування системи забезпечення інформаційної безпеки держави для вирішення проблеми моніторингу текстового контенту соціальних інтернет-сервісів є актуальним теоретико-прикладним завданням.

Мета. Метою досліджень є обґрунтування й розроблення методу виявлення інформаційних впливів у соціальних інтернет-сервісах за змістовними ознаками для підвищення ефективності системи забезпечення інформаційної безпеки держави.

Метод. Розроблений метод ґрунтується на сучасних підходах до інтелектуального аналізу текстового контенту – латентно-семантичному індексуванню і семантичному аналізі на базі онтологій.

Результати. В результаті латентно-семантичного індексування відбирається релевантний заданій тематиці текстовий контент на основі його змісту і без врахування щільності ключових слів. На етапі семантичного аналізу реалізовано процедури подальшого виявлення небезпечних семантичних конструкцій у відібраному текстовому контенті. У випадку детектування суперечливих фрагментів текстового контенту і онтології та після їх аналізу експертами виконується наповнення онтологічних баз знань шаблонами загроз інформаційній безпеці у соціальних інтернет-сервісах. Таким чином, комбінація методів латентно-семантичного індексування і семантичного аналізу забезпечує взаємну компенсацію їх недоліків та виявлення прихованих залежностей між мовними одиницями.

Висновки. Завдяки розробленому методу виявлення ознак інформаційних впливів у соціальних інтернет-сервісах автоматизуються процедури ідентифікації у віртуальних спільнотах інформаційних операцій, направлених проти інформаційної безпеки людини, суспільства, держави. Застосування запропонованого методу виявлення ознак інформаційних впливів дозволяє підвищити ефективність і швидкість системи забезпечення інформаційної безпеки держави в соціальних інтернет-сервісах як складової національної безпеки України.

Ключові слова: соціальний інтернет-сервіс, інформаційна безпека, загроза, текстовий контент, латентно-семантичне індексування, семантичний аналіз, онтологія.

НОМЕНКЛАТУРА

i – кількість ключових слів;

j – кількість публікацій акторів;

k – кількість сингулярних значень матриці;

$W = \langle w_i \rangle$ – ключові слова;

M – частотна матриця;

U і V^T – ортогональні матриці;

S – діагональна матриця;

P_n – скінченна множина концептів;

R_n – скінченна множина відношень між концептами;

$P_s(p_n) \in P_n$ – підмножина множини концептів P_n , суміжних до деякого концепту p_n ;

$P_{in}(r_n) \in P_n$ – підмножина множини концептів P_n , інцидентних до відношення r_n ;

$R_{in}(p_n) \in R_n$ – підмножина множини відношень R_n , інцидентних до деякого концепту p_n ;

$R_z(p_n) \in R_n$ – підмножина множини відношень R_n , яка вказує на небезпеку для концепту p_n ;

$r_i(p_i)$ – деяке відношення з текстового контенту СІС, що аналізується;

$p_i \in P_n$ – концепт онтології досліджуваної віртуальної спільноти;

$r_i \in R_z(p_n)$ – множина відношень, які вказують на небезпеку для деякого концепту p_n ;

D – колекції публікацій;

LSI – латентно-семантичне індексування;

Ont – онтологія;

СІС – соціальний інтернет-сервіс.

ВСТУП

На сучасному етапі значну роль в процесі комунікації суспільства відіграють соціальні інтернет-сервіси (СІС), які забезпечують учасників віртуальних спільнот – акторів, новітніми засобами взаємодії [1]. Завдяки комунікаційним перевагам [2] СІС перетворилися на потужний інструмент взаємодії громадянського суспільства і держави, формування суспільної думки з багатьох актуальних питань. Однак, внаслідок широкої популярності СІС стали і дієвим засобом проведення інформаційних операцій проти людини, суспільства, держави. Встановлено, що ознаки інформаційних впливів у СІС доцільно об'єднати в групи за такими характеристиками – організаційні, змістовні та маніпулятивні. Виявлення згаданих ознак, як показано в [1, 3], у першу чергу пов'язане з моніторингом їх контенту.

В свою чергу, контент, який генерується акторами віртуальних спільнот, може бути представлений тексто-

вими, аудіо, відео, графічними та іншими типами даних. Дослідження співвідношення видів контенту в СІС показують, що найбільшу частку інформаційного середовища займає текстовий тип даних [4]. В першу чергу, це пов'язано з високою швидкістю його споживання акторами віртуальних спільнот завдяки універсальності, відсутності залежності якості відображення від типу кінцевого пристрою користувача. Популярність текстового контенту визначається закладеною в його зміст ідеєю, яка приймаючи різні форми покликана донести до актора потрібні картини суспільних чи політичних подій.

Однак, як показано в [1], зміст текстового контенту може містити деструктивний інформаційний вплив у явному або прихованому вигляді. Висока складність процедур аналізу змісту текстового контенту й особливо автоматичного аналізу, призводить до ускладнення процесу виявлення початку та самого факту інформаційного впливу в СІС. Тому розроблення ефективних алгоритмів функціонування системи забезпечення інформаційної безпеки держави для вирішення проблеми моніторингу текстового контенту СІС є актуальним теоретико-прикладним завданням.

Мета статті полягає в обґрунтуванні й розробленні методу виявлення інформаційних впливів в СІС за змістовними ознаками, який дозволить підвищити ефективність функціонування системи забезпечення інформаційної безпеки держави.

Для досягнення поставленої в статті мети необхідно розв'язати такі частинні завдання:

- проаналізувати сучасні підходи до виявлення загроз інформаційній безпеці держави інформаційно-психологічного характеру в розрізі аналізу текстового контенту СІС;
- встановити напрямки підвищення ефективності функціонування системи забезпечення інформаційної безпеки держави для завчасного виявлення ознак інформаційних впливів у СІС;
- розробити метод ідентифікації інформаційних впливів на основі виявлення їх ознак в змісті текстового контенту, який генерується або поширюється акторами в віртуальних спільнотах СІС;
- перевірити достовірність запропонованого методу на реальних прикладах.

1 ПОСТАНОВКА ЗАДАЧІ

Поставимо задачу розробити метод виявлення інформаційних впливів на акторів віртуальних спільнот СІС, який забезпечить задану достовірність функціонування, виявлення відомих та нових загроз в явному або прихованому вигляді, реалізованість за умови обмеженого забезпечення ресурсами. Вибір текстового контенту для досягнення мети дослідження проводиться відповідно до вимог критичності і значущості його тематики для інформаційної безпеки держави. Для розв'язку задачі необхідно узагальнити і систематизувати змістовні ознаки для виявлення інформаційного впливу на акторів. Використати сучасні підходи до інтелектуального аналізу текстового контенту, а для побудови онтологій обираються предметні області, пов'язані із забезпеченням інформаційної безпеки держави. Розроблений метод повинен забезпе-

чити задану швидкодію для автоматизації процедур виявлення загроз інформаційній безпеці держави і функціонування підсистеми моніторингу інформаційного середовища системи забезпечення інформаційної безпеки держави у СІС.

2 ОГЛЯД ЛІТЕРАТУРИ

Аналіз останніх досліджень і публікацій показав, що для обробки і аналізу текстового контенту використовуються методи статистичного й лінгвістичного аналізу [3, 5–9]. Перша група методів ґрунтується на аналізі змісту контенту за частотою слів, які в ньому використовуються. Спільним недоліком групи статистичних методів є неможливість врахування зв'язності текстового контенту і, як наслідок, його представлення у вигляді множини не пов'язаних між собою слів або Bag of Words [3, 5]. Для усунення цього недоліку використовують методи лінгвістичного аналізу, які включають такі рівні: графематичний для виділення абзаців, речень і окремих слів; морфологічний, який має на меті визначення морфологічних характеристик і слівформ вхідного контенту; синтаксичний аналіз для встановлення синтаксичної залежності слів у реченні; семантичний аналіз для змістовного розуміння текстового контенту.

Семантичний аналіз є складною процедурою, яка ґрунтується на використанні баз знань і тезаурусів для відображення зв'язку між окремими словами й словосполученнями [3]. Результатом семантичного аналізу є формалізоване подання текстового контенту, який досліджується, для подальшого інтелектуального аналізу. До недоліків сучасних семантичних аналізаторів можна віднести високу вартість їх підтримки для кожної окремої мови, істотну обчислювальну складність, неоднозначність результатів функціонування [3, 7].

Проблема виявлення ознак інформаційних впливів за змістовними ознаками не обмежується лінгвістичним аналізом текстового контенту СІС. Детектування загроз системою забезпечення інформаційної безпеки держави в СІС належить, зокрема, до задач інформаційного пошуку [7]. Серед класичних методів інформаційного пошуку виділяють теоретико-множинні, алгебричні та ймовірнісні. Однак, їх застосування для пошуку в СІС текстового контенту з деструктивним інформаційним впливом обмежується рядом недоліків [3]: неможливістю ранжування результатів пошуку; розрахунки пов'язані зі значними об'ємами даних; низька обчислювальна масштабованість і необхідність навчання системи. Тому перспективним напрямком є використання моделей семантичного пошуку та аналізу, які враховують зміст текстового контенту.

Встановлено, що в існуючих дослідженнях проблеми виявлення інформаційних впливів у СІС не використовується комплексний підхід, який би систематизував і узагальнив їх частинні ознаки. Протиріччя між рівнем розвитку сучасних інформаційних технологій і науковим базисом автоматизованого виявлення загроз в СІС, відсутність дієвих методик змістовного аналізу текстового контенту СІС на предмет деструктивного інформаційного впливу додатково актуалізують обраний напрямок досліджень.

2 ОГЛЯД ЛІТЕРАТУРИ

Аналіз останніх досліджень і публікацій показав, що для обробки і аналізу текстового контенту використовуються методи статистичного й лінгвістичного аналізу [3, 5–9]. Перша група методів ґрунтується на аналізі змісту контенту за частотою слів, які в ньому використовуються. Спільним недоліком групи статистичних методів є неможливість врахування зв'язності текстового контенту і, як наслідок, його представлення у вигляді множини не пов'язаних між собою слів або Bag of Words [3, 5]. Для усунення цього недоліку використовують методи лінгвістичного аналізу, які включають такі рівні: графематичний для виділення абзаців, речень і окремих слів; морфологічний, який має на меті визначення морфологічних характеристик і слів форм вхідного контенту; синтаксичний аналіз для встановлення синтаксичної залежності слів у реченні; семантичний аналіз для змістовного розуміння текстового контенту.

Семантичний аналіз є складною процедурою, яка ґрунтується на використанні баз знань і тезаурусів для відображення зв'язку між окремими словами й словосполученнями [3]. Результатом семантичного аналізу є формалізоване подання текстового контенту, який досліджується, для подальшого інтелектуального аналізу. До недоліків сучасних семантичних аналізаторів можна віднести високу вартість їх підтримки для кожної окремої мови, істотну обчислювальну складність, неоднозначність результатів функціонування [3, 7].

Проблема виявлення ознак інформаційних впливів за змістовними ознаками не обмежується лінгвістичним аналізом текстового контенту СІС. Детектування загроз системою забезпечення інформаційної безпеки держави в СІС належить, зокрема, до задач інформаційного пошуку [7]. Серед класичних методів інформаційного пошуку виділяють теоретико-множинні, алгебричні та ймовірнісні. Однак, їх застосування для пошуку в СІС текстового контенту з деструктивним інформаційним впливом обмежується рядом недоліків [3]: неможливістю ранжування результатів пошуку; розрахунки пов'язані зі значними об'ємами даних; низька обчислювальна масштабованість і необхідність навчання системи. Тому перспективним напрямком є використання моделей семантичного пошуку та аналізу, які враховують зміст текстового контенту.

Встановлено, що в існуючих дослідженнях проблеми виявлення інформаційних впливів у СІС не використовується комплексний підхід, який би систематизував і узагальнив їх частинні ознаки. Протиріччя між рівнем розвитку сучасних інформаційних технологій і науковим базисом автоматизованого виявлення загроз в СІС, відсутність дієвих методик змістовного аналізу текстового контенту СІС на предмет деструктивного інформаційного впливу додатково актуалізують обраний напрямок досліджень.

3 МАТЕРІАЛИ І МЕТОДИ

З [5, 6] відомо, що в загальному вигляді сучасні системи виявлення загроз як в текстовому контенті, так і в технічних системах, функціонують на основі таких базових методів:

– сигнатурні, ідея яких полягає в ідентифікації загроз із використанням відомих значень їх параметрів;

– виявлення аномалій на базі відхилень від еталонної моделі функціонування об'єкта досліджень.

Сигнатурні методи відрізняються високою ефективністю, а загроза, в свою чергу, описується у вигляді набору правил чи формальної моделі. Однак, у разі відсутності моделі деякої загрози в системі, її ідентифікація стає неможливою. Детектування системою виявлення загроз аномальної поведінки об'єкта дослідження пов'язане з його навчанням еталонній поведінці, відхилення від якої вказує на появу загрози. В окремих випадках опис загрози може збігатися з еталонною поведінкою об'єкта, що істотно впливає на достовірність функціонування системи виявлення загроз. Тому перспективним напрямком є розробка таких методів виявлення загроз інформаційній безпеці держави в СІС за змістовними ознаками, які забезпечать задану достовірність функціонування, виявлення відомих та нових загроз в явному або прихованому вигляді, реалізованість за умови обмеженого забезпечення ресурсами.

У результаті узагальнення відомих підходів до інформаційного пошуку і лінгвістичного аналізу контенту СІС розроблено метод виявлення інформаційних впливів на основі встановлення їх ознак у змісті текстового контенту, який зводиться до такого.

Етап 1. Пошук текстового контенту в СІС за заданим інформаційним приводом. На цьому етапі визначаються ключові слова $W = \langle w_i \rangle, i = \overline{1, n}$ для пошуку текстового контенту в СІС за критерієм актуальності, критичності та рівня обговорення у суспільстві його тематики. В даному випадку поставлена задача зводиться до вибору методу інформаційного пошуку текстового контенту, який задовольняє вимогам релевантного пошуку текстового контенту, ефективній обробці, зокрема, коротких публікацій. Тому пропонується скористатися методом латентно-семантичного індексування (LSI) (латентно-семантичного аналізу (LSA)) [7–9]. Особливостями LSI є пошук контенту в СІС на основі його змісту, а не щільності ключових слів, і пошук прихованих семантичних зв'язків між ключовими словами й безпосередньо контентом. Суть методу LSI, адаптованого під задачі дослідження, наведена нижче [8].

Крок 1.1 методу полягає у попередній підготовці досліджуваного контенту СІС шляхом видалення стоп-слів, стеммінгу або лематизації слів. Стоп-слова зустрічаються у всьому контенті й не мають змістовного навантаження, наприклад, сполучники, частки, прийменники тощо. Стеммінг полягає у виділенні основи слова виключенням закінчень і суфіксів та є обов'язковим на великих наборах публікацій контенту СІС. Для випадку невеликих наборів публікацій доцільно скористатися алгоритмом Портера [10], який не вимагає морфологічних словників, а виділення основи слова реалізоване на основі визначених правил. Лематизація – це приведення слова до словникового виду.

Крок 1.2. Призначений для виключення з досліджуваного текстового контенту СІС слів, які вживаються тільки один раз. Даний крок не є обов'язковим, однак зменшує кількість подальших обчислень і, як наслідок, підвищує швидкодію.

Крок 1.3. Формування частотної матриці M ключових слів W , які індексуються. Рядками i цієї матриці є ключові слова W семантичного ядра, за яким виконується моніторинг текстового контенту СІС, а стовпцями j -публікації акторів чи віртуальних спільнот. Елементи матриці m_{ij} представляють собою частоту вживання деякого ключового слова W_i в j -й публікації.

Крок 1.4 полягає в сингулярному розкладанні початкової матриці M на три компоненти

$$M = U \times S \times V^T, \quad (1)$$

де матриці U і V^T мають розмірність $i \times k$ та $k \times j$ відповідно; S – діагональна матриця розмірністю $k \times k$, причому k – кількість прихованих тематик контенту, а її елементи впорядковані за спаданням.

На кроці 1.5 виділяють ті рядки матриці U і стовпці V^T , які відповідають найбільшим сингулярним числам k , а їх величина – ступеню прояву ключових слів у колекції публікацій в СІС.

У результаті виконання етапу 1 вдається одержати колекцію публікацій у СІС, які відповідають семантичному ядру пошукового запиту. Такі публікації підлягають подальшому семантичному аналізу на основі онтологій.

Етап 2. Виявлення ознак інформаційних впливів у СІС на основі сигнатурного методу і методу виявлення аномалій. Суть даного етапу полягає у виявленні загроз інформаційній безпеці держави в СІС, які містяться в текстовому контенті, який генерується або поширюється віртуальними спільнотами, і мають на меті вплив на акторів, їх свободу вибору, дискредитацію органів влади тощо [11]. Контент віртуальних спільнот, проіндексований і відібраний на попередньому етапі, підлягає семантичному аналізу на базі онтологій [12–17]. В роботі [12] визначено, що онтологія є експліцитною (явною) специфікацією опису множини об'єктів, які називають концептами, і зв'язків між ними – відношеннями. Зміст другого етапу полягає в наступному.

На кроці 2.1 складаються онтологія функціонування віртуальної спільноти в СІС

$$Ont = \langle P_n, R_n \rangle. \quad (2)$$

При цьому в онтології Ont (2) для автоматизації подальшого аналізу текстового контенту віртуальних спільнот виділяються підмножини [5] $P_s(p_n) \in P_n$, $P_{in}(r_n) \in P_n$, $R_{in}(p_n) \in R_n$, $R_z(p_n) \in R_n$.

Крок 2.2 полягає у побудові семантичного опису текстового контенту, виявленого на першому етапі методу

$$Sem_i = \langle P_i, R_i \rangle. \quad (3)$$

Для забезпечення швидкодії функціонування методу доцільно застосовувати автоматичні засоби генерації онтологій. До такого класу програмних продуктів відносять Protégé, Ontology Learning Framework, LOGS, ABBYY Compeno тощо [13–16].

Крок 2.3 зводиться до виявлення ознак загрози у попередньо проіндексованому на першому етапі методу текстовому контенті СІС. У формальному вигляді правила виявлення загроз інформаційній безпеці зводяться до такого [5]:

– детектування на основі сигнатурного методу і семантичних ознак

$$\exists r_i(p_i): p_i \in P_n \wedge r_i \in R_z(p_n). \quad (3)$$

Суть правила (3) полягає в наступному: для виявлення загрози інформаційній безпеці у змісті текстового контенту віртуальної спільноти СІС необхідно виявити зв'язок між об'єктом публікації з його характеристиками у контенті та негативними ознаками для цього об'єкта внаслідок реалізації загрози.

Виявлення суперечностей на основі виявлення аномалій шляхом співставлення семантичного опису проіндексованого текстового контенту і семантичного шаблону загрози. При цьому встановлюється невідповідність фактів у контенті СІС, що проявляється як:

– протиріччя понять в змісті контенту віртуальних спільнот – вживання концепту в конкретному відношенні не передбачене онтологією

$$\exists r_i(p_i): p_i \in P_n \wedge r_i \in R_n \wedge p_i \notin P_{in}(r_n). \quad (4)$$

Зміст правила (4) полягає в такому: деякий концепт p_i текстового контенту СІС не може вживатися у конкретному відношенні $r_i(p_i)$.

– протиріччя відношень в контенті – вжите відношення між концептами не визначене онтологією

$$\forall p_n \in P_n \neg \exists r_n \in R_n : r_i = r_n. \quad (5)$$

Правило (5) трактується як: множина відношень R_n онтології функціонування віртуальної спільноти, в які можуть вступати концепти $p_n \in P_n$, які пов'язані між собою відношеннями r_i , не містить цих відношень.

Крок 2.4 призначений для розгорнутого аналізу експертами проіндексованого текстового контенту на предмет наявності загроз. Даний крок необхідний, якщо контент СІС є релевантним семантичному ядру пошукового запиту до контенту СІС з етапу 1, однак на етапі 2 не виявлено суперечливих фрагментів такого контенту і онтології. Тоді ідентифіковані експертами небезпечні семантичні конструкції додаються до шаблонів загроз і використовуються в онтології. В результаті таких дій метод LSI забезпечує ефективне виявлення залежностей між лексичними одиницями в контенті СІС для наповнення онтологічних баз знань [18].

Метод виявлення змістовних ознак інформаційних впливів у СІС з використанням латентного семантичного аналізу і семантичного аналізу на базі онтологій поданий у вигляді структурної схеми на рис. 1.

Застосування розробленого методу виявлення ознак інформаційних впливів дозволяє ідентифікувати у віртуальних спільнотах СІС інформаційні операції, направлені проти інформаційної безпеки людини, суспільства, держави. Перспективним є використання запропонованого методу для функціонування системи забезпечення інформаційної безпеки держави в СІС, зокрема автоматизації процесів моніторингу інформаційного середовища віртуальних спільнот. Окреме місце в забезпеченні функціонування методу належить експертам і співробітникам спеціальних підрозділів, перед якими стоїть завдання виявлення деструктивних інформаційних впливів у попередньо проіндексованому й відібраному контенті СІС. На основі таких

ідентифікованих впливів створюються семантичні шаблони загроз і додаються до онтологічних баз знань.

4 ЕКСПЕРИМЕНТИ

Відповідно до розробленого методу дослідимо контент SIC Facebook на предмет наявності ознак інформаційного впливу на акторів. Для цього задамо семантичне ядро

$W = \langle \text{майдан; бандити; влада; поліція; корупція; диктатура; війна; злидні; Порошенко; олігархи} \rangle$

для його пошуку в колекції публікацій $D_i, i = \overline{1,10}$ актора Миколи Гайдука в соціальній мережі Facebook. Після виконання першого етапу розробленого методу дослідимо

мо релевантні до запиту W публікації, використовуючи семантичний аналіз і онтологічні бази знань.

5 РЕЗУЛЬТАТИ

Результати першого етапу функціонування методу виявлення інформаційних впливів подано в табл. 1.

Застосуємо до матриці в табл. 1 операцію двовимірного сингулярного розкладу, внаслідок якого отримаємо три матриці (рис. 2).

З метою візуалізації результатів розрахунків зобразимо їх на графіку (рис. 3).

З рис. 3 видно, що проіндексовані документи D_2, D_3, D_5 і D_{10} , не зважаючи на наявність слів із заданого семантичного ядра W , не релевантні пошуковому запиту.

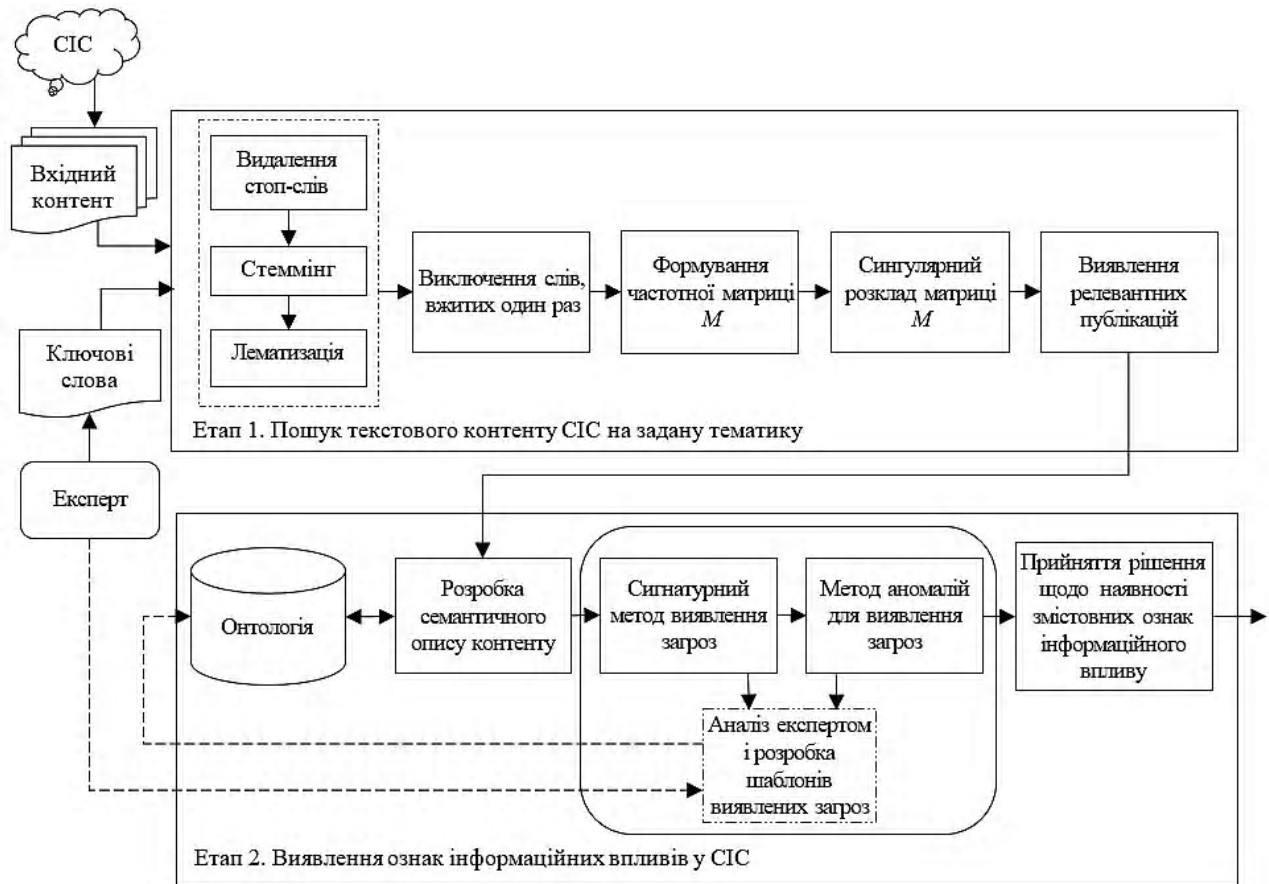


Рисунок 1 – Узагальнена структурна схема методу виявлення інформаційних впливів у SIC за змістовними ознаками

Таблиця 1 – Частотна матриця індексованих слів

Ключові слова	Публікації									
	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}
майдан	1	0	1	0	1	0	0	1	0	0
бандити	0	1	1	1	1	0	1	1	0	0
влада	1	1	0	0	1	1	1	0	0	0
поліція	1	1	1	0	1	1	0	0	1	1
корупція	0	0	1	1	0	0	1	1	0	0
диктатура	1	1	0	0	0	0	1	0	1	0
війна	0	0	1	0	1	0	1	1	0	0
злидні	0	0	0	0	0	0	0	0	0	1
Порошенко	1	0	0	1	0	0	0	1	1	1
олігархи	1	1	0	0	1	1	0	1	0	0

$U=$	-0,316	-0,140
	-0,411	-0,402
	-0,354	0,289
	-0,427	0,430
	-0,248	-0,513
	-0,257	0,258
	-0,299	-0,395
	-0,029	0,076
	-0,264	0,055
	-0,368	0,235

$S^T=$	4,957	0
	0	2,723

$V^T=$	-0,367	0,297	0,253	-0,336	0,035	-0,302	-0,440	-0,398	0,004	0,395
	-0,343	-0,375	-0,002	0,231	0,570	0,274	-0,322	0,320	0,041	0,283

Рисунок 2 – Результати розрахунків

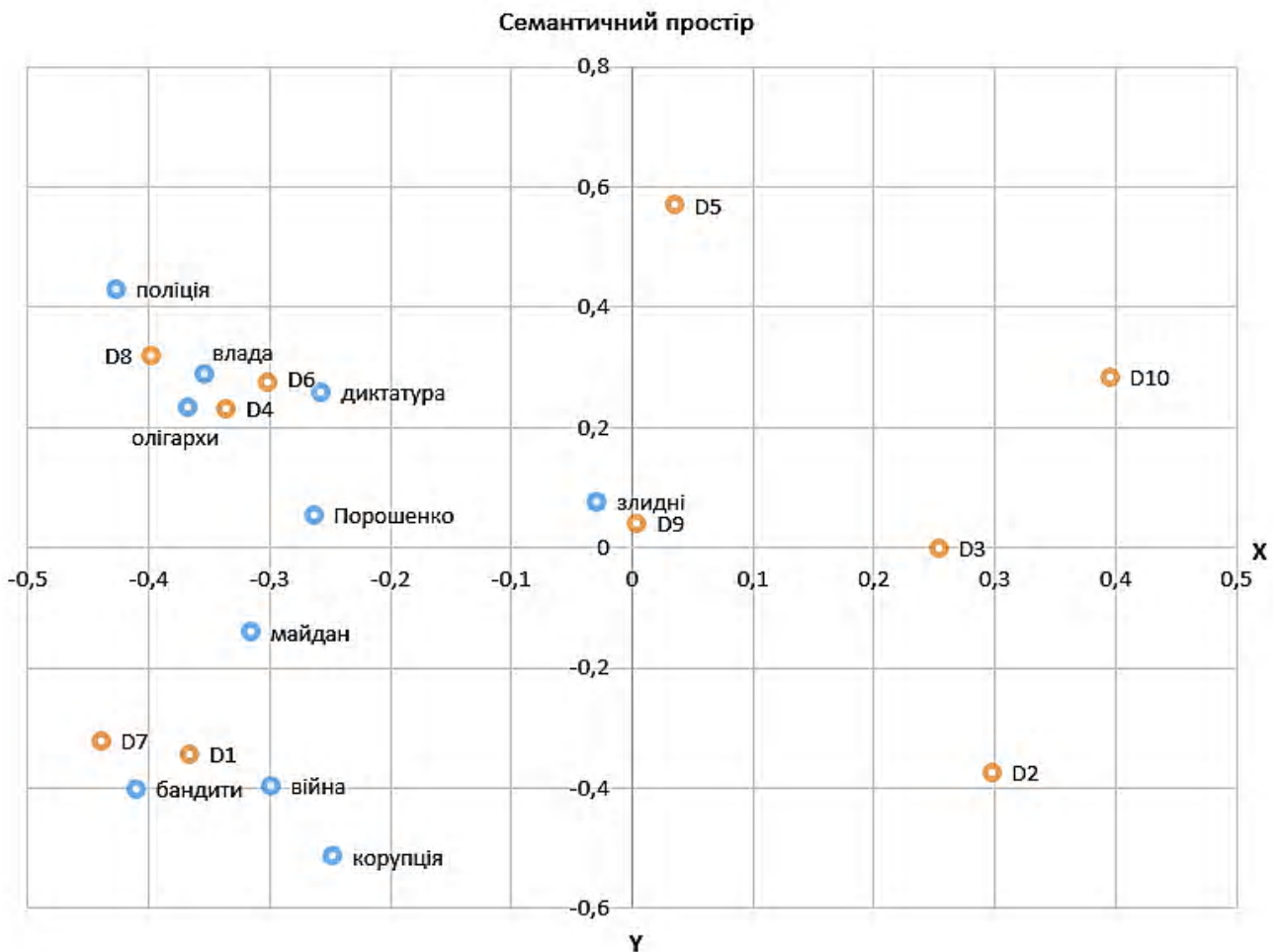


Рисунок 3 – Розподіл колекції публікацій в семантичному просторі

Документи D_4 , D_6 , D_8 утворюють неочевидний зв'язок з термінами, що пов'язані зі «встановленням диктатури олігархату Порошенка». Близькими для змісту документів D_1 і D_7 є терміни «бандіти», «війна», «корупція» та «майдан», які їх об'єднують в окрему групу. Зміст документу D_9 найтісніше пов'язаний з низьким рівнем життя населення і не містить прихованих залежностей від інших слів семантичного ядра.

Отже, подальшому семантичному аналізу підлягають релевантні до семантичного ядра W документи D_1 , D_4 ,

D_6 , D_7 , D_8 . Таким чином зменшується загальна кількість документів для подальшого дослідження модулем моніторингу контенту системи забезпечення інформаційної безпеки в СІС, підвищуються ефективність і швидкість її функціонування.

Розглянемо окремі етапи семантичного аналізу документу D_6 , зміст якого наведений на рис. 4.

Фрагмент семантичного опису цього документа наведено на рис. 5.



Рисунок 4 – Досліджуваний контент в СІС Facebook

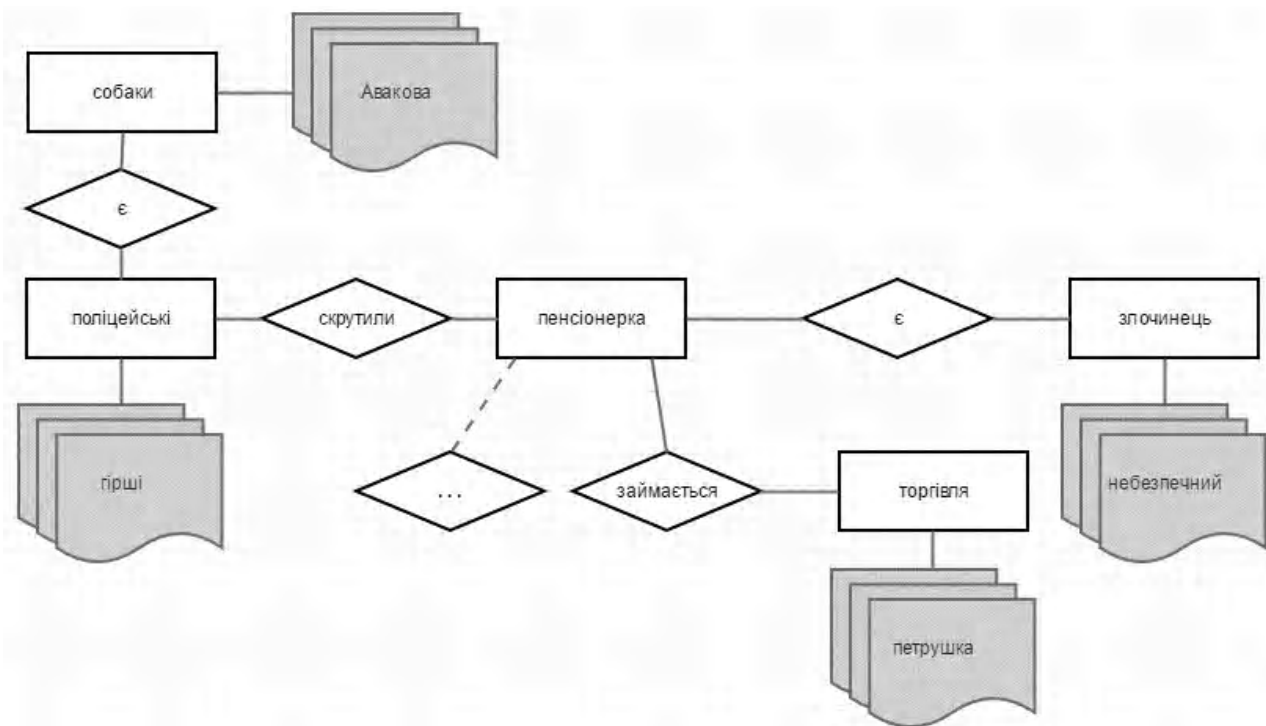


Рисунок 5 – Фрагмент семантичного опису документа D_6

Сигнатурне виявлення загроз в текстовому контенті СІС детектувало використання з концептом «пенсіонерка» відношення «скрутили», яке становить фізичну небезпеку для об'єкта. Відповідно до методу аномалій встановлено, що концепт «пенсіонерка» не може вживатися разом з відношенням «є» і концептом «злочинець» з атрибутом «небезпечний». Також в семантичному описі виявлено відношення «скрутили» між концептами «по-

ліцейські» та «пенсіонерка», що не визначено онтологією. В досліджуваному документі вжито порівняння поліцейських з «собаками Авакова», яке необхідно додати до семантичних шаблонів загроз онтологічної бази знань.

6 ОБГОВОРЕННЯ

Розглянутий метод поєднує в собі сучасні підходи до обробки текстового контенту – семантичний і латентно-семантичний аналіз. Така комбінація підходів реалізує

взаємну компенсацію їх недоліків [18]: полісемії, омонімії та інших видів лінгвістичних неоднозначностей для LSI; виявлення латентних залежностей між концептами для семантичного аналізу.

Одержані результати збіжні, наприклад, з дослідженнями аналітичного центру Тексти.org, яким встановлено взаємопов'язані віртуальні спільноти у Вконтакте і Facebook, актори яких закликали до перевороту й протестів в Україні [19]. У результаті розкрито адміністратора таких груп, який зареєстрований у СІС під іменем Микола Гайдук. Отже, дієвість використання запропонованого у статті методу виявлення ознак інформаційного впливу доведено.

ВИСНОВКИ

Метод виявлення ознак інформаційного впливу у СІС за змістовними ознаками ґрунтується на сучасних підходах інтелектуальної обробки текстового контенту. Перевагою розглянутого методу є застосування LSI для пошуку текстового контенту віртуальних спільнот, який містить деструктивний інформаційний вплив в явному або прихованому вигляді. Завдяки додатковому використанню сигнатурного методу і методу аномалій для детектування відомих та нових загроз інформаційній безпеці досягається взаємна компенсація недоліків таких підходів. Таким чином, досягається підвищення ефективності й швидкодії системи забезпечення інформаційної безпеки держави в СІС, що є сьогодні вкрай актуальним завданням для України.

ПОДЯКИ

Робота виконана на кафедрі комп'ютерних технологій і моделювання систем Житомирського національного агрокологічного університету в рамках госпдоговірної науково-дослідницької теми «Методологія побудови сучасних інформаційних технологій аналізу і відображення стану інформаційної та екологічної безпеки держави» ДР№ 0115U004181 за сприяння науково-дослідного відділу інформаційної та кібернетичної безпеки наукового центру Житомирського військового інституту ім. С. П. Корольова в особі начальника відділу – д.т.н., с.н.с. Р. В. Грищука. За результати наукових досліджень у напрямку розробки і впровадження новітньої методології побудови систем забезпечення інформаційної безпеки держави в соціальних інтернет-сервісах автора відзначено стипендією Кабінету Міністрів України для молодих учених.

СПИСОК ЛІТЕРАТУРИ

1. Грищук Р. В. Основи кібернетичної безпеки : моногр. / Р. В. Грищук, Ю. Г. Даник ; за заг. ред. проф. Даника Ю. Г. – Житомир : ЖНАЕУ, 2016. – 636 с.
2. Соціальні мережі як інструмент взаємовпливу влади та громадянського суспільства : [монографія] / [О. С. Онищенко, В. М. Горвий, В. І. Попик та ін.] ; НАН України, Нац. б-ка України ім. В. І. Вернадського. – Київ, 2014. – 260 с.
3. Лапшин С. В. Методы повышения показателей качества фильтрации DLP-систем на основе предметно-ориентированной морфологической модели естественного языка : дисс. ... канд. техн. наук : 05.13.19 / Лапшин Сергей Владимирович. – Санкт-Петербург, 2014. – 115 с.
4. Какие бывают типы контента и как с ними работать для реализации успешной стратегии контент-маркетинга / Alphaland.in Интернет Маркетинг [Электронный ресурс]. – Режим доступа : <http://alphaland.in/kakie-by-vayut-tipy-kontenta-i-kak-s-nimi/> (дата обращения: 29.10.16). – Название с экрана.
5. Чернишук С. В. Методика виявлення кібернетичних загроз у природномовних текстах / С. В. Чернишук // Проблеми створення, випробування, застосування та експлуатації складних інформаційних систем. – 2013. – Вип. 8. – С. 112–121.
6. Грищук Р. В. Теоретичні основи моделювання процесів нападу на інформацію методами теорій диференціальних ігор та диференціальних перетворень: монографія / Р. В. Грищук. – Житомир : Рута, 2010. – 280 с.
7. Manning Chr. Introduction to Information Retrieval / Chr. Manning, P. Raghavan, H. Schütze. – Cambridge University Press, 2008. – 544 p.
8. Латентно-семантический анализ / Сайт Хабрхабр [Электронный ресурс]. – Режим доступа : <https://habrahabr.ru/post/110078/> (дата обращения : 29.10.16). – Название с экрана.
9. Indexing by latent semantic analysis / [S. Deerwester, S. T. Dumais, G. W. Furnas et al], // Journal of the American Society for Information Science. – 1990. – 41(6). – P. 391–407. DOI: 10.1002/(sici)1097-4571(199009)41:6<391::aid-asi1>3.0.co;2-9.
10. Стеммер Портера для русского языка / О программировании, алгоритмах и не только [Электронный ресурс]. – Режим доступа : <http://www.algorithmist.ru/2010/12/porter-stemmer-russian.html/> (дата обращения: 29.10.16). – Название с экрана.
11. Молодецька К. В. Узагальнена класифікація загроз інформаційній безпеці держави в соціальних інтернет-сервісах / К. В. Молодецька // Защита информации. – 2016. – Вып. 23. – С. 75–87.
12. Gruber T. R. A Translation Approach to Portable Ontology Specifications / T. R. Gruber // Knowledge Acquisition. – 1993. – 5(2). – P. 199–220. DOI: 10.1006/knac.1993.1008.
13. Bouiadja A. B. A framework for evaluating and ranking ontologies / A. B. Bouiadja, S. M. Benslimane // International Journal of Metadata, Semantics and Ontologies. – 2013. – 8(2). – P. 155. DOI: 10.1504/ijmso.2013.056600.
14. Vargas-Vera M. State of the art on Ontology alignment / M. Vargas-Vera, M. Nagy // International Journal of Knowledge Society Research. – 2015. – 6(1). – P. 17–42. DOI: 10.4018/ijksr.2015010102.
15. Priya M. A survey of state of the art of Ontology construction and merging using formal concept analysis / M. Priya, C. Aswani Kumar // Indian Journal of Science and Technology. – 2015. – 8(24). DOI: 10.17485/ijst/2015/v8i24/82808.
16. Василюк Я. Р. Аналіз моделей, методів, принципів і засобів для побудови онтологій області мікроелектромеханічних систем / Я. Р. Василюк, В. М. Теслюк, А. Я. Зелінський // Наук. вісн. Нац. лісотехн. ун-ту Укр. : зб. наук. -техн. пр. – 2011. – Вип. 21(12). – С. 322–330.
17. Wang W. Probabilistic topic models for learning Terminological Ontologies / W. Wang, P. M. Barnaghi, A. Bargiela // IEEE Transactions on Knowledge and Data Engineering. – 2010. – 22(7). – P. 1028–1040. DOI: 10.1109/tkde.2009.122.
18. Марченко О. О. Порівняння методів онтологічного семантичного аналізу та алгоритмів латентного семантичного аналізу / О. О. Марченко // Вісн. Київського нац. ун-ту ім. Т. Шевченка. Серія фізико-математичні науки. – 2012. – 2. – С. 169–174.
19. Тролесфера / Тексти.org.ua [Електронний ресурс]. – Режим доступа : <http://texty.org.ua/d/fb-trolls/> (дата звернення: 29.10.16). – Назва з екрану.

Стаття надійшла до редакції 08.12.2016.

Після доробки 26.12.2016.

Молодецкая-Гринчук К. В.

Канд. техн. наук, доцент, доцент кафедры компьютерных технологий и моделирования систем Житомирского национального агроэкологического университета, Житомир, Украина

МЕТОД ВЫЯВЛЕНИЯ ПРИЗНАКОВ ИНФОРМАЦИОННОГО ВОЗДЕЙСТВИЯ В СОЦИАЛЬНЫХ ИНТЕРНЕТ-СЕРВИСАХ ПО СОДЕРЖАТЕЛЬНЫМ ПРИЗНАКАМ

Актуальность. Сегодня социальные интернет-сервисы превратились в эффективный инструмент коммуникации между участниками виртуальных сообществ – акторами. Учитывая высокую скорость распространения контента, трансграничность процессов взаимодействия акторов, наличие средств для их организации в группы, социальные-интернет сервисы могут применяться как действенное средство проведения информационных операций против человека, общества, государства. В процессах коммуникации акторов в виртуальных сообществах часто используется текстовый контент. Разработка эффективных алгоритмов функционирования системы обеспечения информационной безопасности государства для решения проблемы мониторинга текстового контента социальных интернет-сервисов является актуальной теоретико-прикладной задачей.

Цель. Целью исследований является обоснование и разработка метода выявления информационных воздействий в социальных интернет-сервисах по содержательным признакам для повышения эффективности системы обеспечения информационной безопасности государства.

Метод. Разработанный метод основывается на современных подходах к интеллектуальному анализу текстового контента – латентно-семантическом индексировании и семантическом анализе на базе онтологий.

Результаты. В результате латентно-семантического индексирования отбирается релевантный заданной тематике текстовый контент на основе его содержания и без учета плотности ключевых слов. На этапе семантического анализа реализовано процедуры последующего обнаружения опасных семантических конструкций в отобранном текстовом контенте. В случае детектирования противоречивых фрагментов текстового контента и онтологии и после их анализа экспертами выполняется наполнение онтологических баз знаний шаблонами угроз информационной безопасности в социальных интернет-сервисах. Таким образом, комбинация методов латентно-семантического индексирования и семантического анализа обеспечивает взаимную компенсацию их недостатков и выявления скрытых зависимостей между языковыми единицами.

Выводы. Благодаря разработанному методу выявления признаков информационных воздействий в социальных интернет-сервисах автоматизируются процедуры идентификации в виртуальных сообществах информационных операций, направленных против информационной безопасности человека, общества, государства. Применение предложенного метода выявления признаков информационных воздействий позволяет повысить эффективность и быстрдействие системы обеспечения информационной безопасности государства в социальных интернет-сервисах как составляющей национальной безопасности Украины.

Ключевые слова: социальный интернет-сервис, информационная безопасность, угроза, текстовый контент, латентно-семантическая индексация, семантический анализ, онтология.

Molodetska-Hrynchuk K.

Candidate of Technical Sciences, Docent, Associate Professor at the Information Technologies and System Modelling sub-Department, Zhitomir National Agroecological University, Ukraine

OUTREACHES CONTENT TRACING TECHNIQUE FOR SOCIAL NETWORKING SERVICES

Context. Today's social networking services have turned into an efficient communication tool for the members of virtual communities termed actors. The high rate of content dissemination, transboundary interactions between the actors and the drives to group the latter up, could make the services efficient cyberwarfare against the person, society and country. Text content is the most frequently used means of communication between the actors of virtual communities. The actual theoretical and applied research, aimed at designing maintenance procedures for the national cybersecurity system, is to solve the problem of text content monitoring.

Objective. The research objective is to suggest and design a method for detecting outreaches by their content in the collection of text to solve the problem of efficient monitoring the social media.

Method. The method designed is based on the modern content analysis approaches, i. e. the latent semantic indexing and the semantic analysis based on ontology.

Results. The latent semantic indexing selects the concept relevant context by its content ignoring key word density. The semantic analysis will specify the further techniques of detecting harmful semantic structures in the body of text selected. When any conceptual mismatches in the text content and ontology detected and professionally analyzed, the ontology knowledge bases get updates of the social networking service cyberthreat patterns. Thus, the combination of both produces cancelling effect on the defects of and uncovers latent relationships between linguistic units.

Conclusions. The outreaches tracing method proactively identifies cyberwarfare in virtual communities waged against individual, social and national cybersecurity. The suggested method increases efficiency and effectiveness of the national cybersecurity system on social networking services.

Keywords: social networking services; cybersecurity; threats; text content; latent semantic indexing; semantic analysis; ontology.

REFERENCES

1. Hryshchuk R. V., Danyk Yu. H. za zah. red. prof. Danyka Yu. H. Osnovy kibernetichnoi bezpeky : monohr. Zhytomyr, ZhNAEU, 2016, 636 p.
2. Onyshchenko O. S., Horovyi V. M., Popyk V. I. ta in. Sotsialni merezhi yak instrument vziaimovplyvu vlady ta hromadianskoho suspilstva : [monohrafiia] ; NAN Ukrainy, Nats. b-ka Ukrainy im. V. I. Vernadskoho. Kyiv, 2014, 260 p.
3. Lapshin S. V. Metody povysheniya pokazatelej kachestva fil'tracii DLP-sistem na osnove predmetno-orientirovannoj morfologicheskoy modeli estestvennogo yazyka : diss. ... kand. texn. nauk : 05.13.19. Sankt-Peterburg, 2014, 115 p.
4. Kakie byvayut tipy kontenta i kak s nimi rabotat' dlya realizacii uspeshnoj strategii kontent-marketinga / Alphaland. in Internet Marketing [E'lektronnyj resurs]. Rezhim dostupa: <http://>

- alphaland.in/kakie-by-vayut-tipy-kontenta-i-kak-s-nimi/ (data obrashheniya: 29.10.16). Nazvanie s e'krana.
5. Chernyshuk S. V. Metodyka vyavleniia kibernetichnykh zahroz u pryrodnomovnykh tekstakh, *Problemy stvorennia, vyprobuvannia, zastosuvannia ta ekspluatatsii skladnykh informatsiinykh system*, 2013, Vyp. 8, pp. 112–121.
 6. Hryshchuk R. V. Teoretychni osnovy modeliuvannia protsesiv napadu na informatsiiu metodamy teorii dyferentsialnykh ihor ta dyferentsialnykh peretvoren: monohrafiia. Zhytomyr, Ruta, 2010, 280 p.
 7. Manning Chr., Raghavan P., Schütze H. Introduction to Information Retrieval. Cambridge University Press, 2008, 544 p.
 8. Sajt Xabrxabr Latentno-semanticheskij analiz. [E'lektronnyj resurs]. Rezhim dostupa : <https://habrahabr.ru/post/110078/> (data obrashheniya : 29.10.16). Nazvanie s e'krana.
 9. Deerwester S., Dumais S. T., Furnas G. W., Landauer T. K., Harshman R. Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, 1990, 41(6), pp. 391–407. DOI: 10.1002/(sici)1097-4571(199009)41:6<391::aid-asi1>3.0.co;2-9.
 10. Stemmer Portera dlya russkogo yazyka / O programmirovani, algoritmax i ne tol'ko [E'lektronnyj resurs]. Rezhim dostupa: <http://www.algorithmist.ru/2010/12/porter-stemmer-russian.html/> (data obrashheniya: 29.10.16). Nazvanie s e'krana.
 11. Molodetska K. V. Uzahalnena klasyfikatsiia zahroz informatsiinii bezpetsi derzhavy v sotsialnykh internet-servisakh, *Zashhita informacii*, 2016, Vyp. 23, pp. 75–87.
 12. Gruber T. R. A Translation Approach to Portable Ontology Specifications, *Knowledge Acquisition*, 1993, 5(2), pp. 199–220. DOI: 10.1006/knac.1993.1008.
 13. Bouiadra A. B., Benslimane S. M. A framework for evaluating and ranking ontologies, *International Journal of Metadata, Semantics and Ontologies*, 2013, 8(2), pp. 155. DOI: 10.1504/ijmso.2013.056600.
 14. Vargas-Vera M., Nagy M. State of the art on Ontology alignment, *International Journal of Knowledge Society Research*, 2015, 6(1), pp. 17–42. DOI: 10.4018/ijksr.2015010102.
 15. Priya M., Aswani Kumar C. A survey of state of the art of Ontology construction and merging using formal concept analysis, *Indian Journal of Science and Technology*, 2015, 8(24). DOI: 10.17485/ijst/2015/v8i24/82808.
 16. Vasyliuk Ya. R., Tesliuk V. M., Zelinskyi A. Ya. Analiz modelei, metodiv, pryntsyviv i zasobiv dlia pobudovy ontolohii oblasti mikroelektromekhanichnykh system, *Nauk. visn. Nats. lisotekhn. un-tu Ukr. : zb. nauk.-tekhn. pr.*, 2011, Vyp. 21(12), pp. 322–330.
 17. Wang W., Barnaghi P. M., Bargiela A. Probabilistic topic models for learning Terminological Ontologies, *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(7), pp. 1028–1040. DOI: 10.1109/tkde.2009.122.
 18. Marchenko O. O. Porivniannia metodiv ontolohichnoho semantichnoho analizu ta alhorytmiv latentnoho semantichnoho analizu, *Visn. Kyivskoho nats. un-tu im. T. Shevchenka. Seriiia fizyko-matematychni nauky*, 2012, 2, pp. 169–174.
 19. Trolesfera Teksty.org.ua [Elektronnyi resurs]. Rezhym dostupu : <http://texty.org.ua/d/fb-trolls/> (data zvernennia: 29.10.16). Nazva z ekranu.

УДК 004.932

Сердюк М. Є.¹, Беркут В. Г.²

¹Канд. техн. наук, доцент, доцент кафедри комп'ютерних технологій Дніпропетровського національного університету імені Олеся Гончара, Дніпро, Україна

²Студент магістратури кафедри комп'ютерних технологій Дніпропетровського національного університету імені Олеся Гончара, Дніпро, Україна

КОМП'ЮТЕРНА СИСТЕМА ЛОКАЛІЗАЦІЇ ТА ВИДАЛЕННЯ ТІНЕЙ У ЦИФРОВИХ ЗОБРАЖЕННЯХ

Актуальність. Наявність затінених ділянок на цифрових зображеннях може спричиняти проблеми в процесах розпізнавання, відстеження, класифікації об'єктів. Тому виявлення та видалення тіней є важливим етапом попередньої обробки зображень, яка застосовується в багатьох алгоритмах комп'ютерного зору.

Мета. Метою роботи є пошук ефективних методів автоматичного виявлення та видалення тіней та розробка програмної системи, яка реалізує відповідну обробку зображень.

Методи. В роботі розглянуто декілька підходів до процесів виявлення та перетворення затінених ділянок, проаналізовано застосовність їх до тіней різних типів. Для локалізації тіней в роботі використано метод з переходом до кольорного простору *Lab* та метод з попереднім видаленням текстур. Для видалення тіней використано чотири методи: адитивний, на основі математичної моделі тіні, з переходом до *Lab*-простору, константний. Запропоновано модифікації деяких етапів з метою покращення практичних результатів.

Результати. Розроблена комп'ютерна система, в якій реалізовано автоматичне виявлення та усунення тіней з використанням зазначених методів, які обираються в залежності від характеру затінених областей. Деякі параметри методів доступні для налаштування. Система протестована на низці зображень. Проаналізовано якість та швидкість роботи реалізованих методів. Наведені результати обробки зображень різними методами.

Висновки. Реалізовані в представленій комп'ютерній системі методи локалізації та видалення тіней з запропонованими модифікаціями ефективно працюють на зображеннях різних типів. Додаткове налаштування параметрів дозволяє покращити результати в деяких складних випадках. Система може бути застосована для попередньої обробки зображень з метою запобігання помилкам в подальшій роботі алгоритмів комп'ютерного зору.

Ключові слова: виявлення тіні, видалення тіні, обробка зображень, кольорний простір *Lab*, сегментація, модель тіні.

НОМЕНКЛАТУРА

a – хроматична складова в діапазоні від зеленого до червоного кольору у колірній моделі *Lab*;

B – хроматична складова в діапазоні від синього до жовтого кольору у колірній моделі *Lab*;

H – значення порогу для кольорних складових в методі з переходом до простору *Lab*;

I – вихідне зображення з тінювими областями;

\tilde{I} – результуюче зображення без тіней;

I_i – інтенсивність *i*-го пікселя;

K_C – значення константи для кольорного каналу *C* в константному методі;

k_i – коефіцієнт тіні *i*-го пікселя;

K_L – коефіцієнт для розрахунку порогового значення в процедурі бінаризації;

L – світлота пікселя в колірній моделі *Lab*;

$\text{mean}(x)$ – середнє значення по масиву *x*;

n – розмір масиву пікселів зображення;

Q – освітлена область вихідного зображення;

R_i – коефіцієнт, що характеризує відбивні властивості поверхні в *i*-му пікселі;

r – коефіцієнт освітленості *i*-го пікселя;

S – тінюва область зображення;

$\text{SD}(x)$ – стандартне відхилення для масива *x*;

T – значення порогу в процедурі бінаризації зображення;

t_i – фактор розсіювання направленої світла в *i*-му пікселі;

δ – ширина смуги пікселів уздовж границі тіні;

θ_i – кут між напрямком направленої світла та нормаллю поверхні в *i*-му пікселі.

ВСТУП

В багатьох задачах комп'ютерного зору проблемою є зайва інформативність вхідних зображень, що є наслідком присутності тіней на цих зображеннях. Тіні утворюють помилкові об'єкти через значну різницю яскравостей тіней і навколишнього фону, а це призводить до небажаних проблем і навіть помилок в роботі алгоритмів розпізнавання, виявлення, відстеження і класифікації об'єктів. Тому виявлення та видалення тіней є важливою задачею попередньої обробки зображень. Крім того, видалення тіней з зображень представляє інтерес з естетичної точки зору та використовується при редагуванні зображень чи в обчислювальній фотографії. Отже, актуальною є розробка інформаційної технології виявлення та видалення тіней з зображень, яка б відрізнялась високим ступенем автоматизації процесу локалізації та усунення тіней.

Тіні можна визначити як частини сцени, які безпосередньо не освітлені джерелом світла через присутність об'єкта, що загороджує це джерело [1]. Можна виділити два основні класи тіней: власна тінь та відкидвана тінь. Власна тінь з'являється на частині об'єкта, яка не освітлюється прямим світлом. Відкидвана тінь – це область, яка проектується об'єктом на поверхню в напрямку прямого світла. На основі інтенсивності тіні можна розділити на м'які та жорсткі. М'які тіні зберігають текстуру поверхні фону, але мають розмиті та нечіткі границі, що зазвичай свідчить про плавний перехід від тіні до освітленої частини зображення, а це викликає труднощі при виявленні таких тіней та їх локалізації. Жорсткі тіні мають різкі границі та інтенсивність, що значно відрізняється

від оточуючої поверхні. Але такі тіні є занадто темними і містять мало інформації про текстуру поверхні, що ускладнює їх видалення зі збереженням природного вигляду текстури. Крім того, жорсткі тіні можуть бути помилково прийняті за темні об'єкти.

Видалення тіней з зображень може істотно покращити та спростити виконання ряду завдань комп'ютерного зору. З цих причин постійно розробляються та удосконалюються різні підходи розв'язання даної задачі, які направлені на знаходження способів правильного виявлення та видалення тіней при збереженні незмінними інших деталей вихідного зображення.

Метою роботи є пошук ефективних шляхів автоматичного виявлення тіні та її видалення з використанням тільки одного вихідного зображення, а також розробка програмної системи, яка реалізує ці методи для обробки зображень.

1 ПОСТАНОВКА ЗАДАЧІ

Нехай є вихідне зображення I , яке має затінені та освітлені ділянки. Це зображення можна подати у вигляді $I = S + Q$. На основі аналізу зображення I необхідно визначити область з тінювими пікселями S , обробити її та отримати результуюче зображення \tilde{I} , яке не містить тіней: $\tilde{I} = \tilde{Q}$.

Розв'язання задачі будемо проводити в два етапи. Перший етап – виявлення тіні, тобто локалізація області S . Результатом має бути бінарне зображення, на якому білі пікселі відповідають тінювими, а чорні – нетінювими пікселям. Другий етап – видалення тіні, тобто обробка пікселів області S на основі бінарної маски та отримання безтінювого зображення \tilde{I} .

2 ОГЛЯД ЛІТЕРАТУРИ

В літературі представлені різні підходи до розв'язання задачі усунення тіней в зображеннях. Так, в роботі [1] запропоновано метод для виявлення та видалення тіней з використанням переходу до кольорного простору $YCbCr$. Виявлення тіні здійснюється на основі аналізу компоненти інтенсивності, зображення сегментується на декілька областей, які мають однакову щільність тіні. Для видалення тіні інтенсивність відповідних пікселів підвищується, а потім здійснюється корекція кольору затінутих ділянок вже у кольорному просторі RGB . Перевагою методу є те, що в результуючому зображенні немає різких переходів між затіненими та незатіненими деталями, і всі деталі в затінених областях залишаються незмінними. Але метод працює добре на зображеннях з різкими границями тіні.

В роботі [2] запропоновано метод, оснований на використанні градієнтних полів вихідного зображення та спеціального зображення, інваріантного до освітлення, отримання якого є окремою задачею. В результаті порівняння градієнтних полів двох зображень будується нове градієнтне поле безтінювого зображення. Результуюче зображення без тіней відновлюється за градієнтним полем з використанням рівняння Пуассона. Метод дає непогані результати для тіней з різкими границями, але потребує послідовності зображень, що містять одну сцену при різних умовах денного світла. Недоліком методу є також те, що результат виходить помітно розмитим.

Підхід, запропонований в роботі [3], базується на припущенні, що тінюві області відрізняються від освітлених на деяку константу. Значення константи обчислюється для кожного каналу RGB зображення та додається до значень пікселів тіні, і таким чином зменшується різниця між тінювою областю та сусідніми освітленими областями. Перевагами даного методу є його швидкість, простота реалізації та ефективність. Але в досить складних сценах тінь не видаляється, а лише послаблюється.

У роботі [4] запропоновано метод, в якому видалення тіней здійснюється в три етапи. Спочатку створюється 1D безтінюве зображення з постійним освітленням. Потім отримується еквівалентне 2D зображення, яке зберігає деяку інформацію про колір. В кінці генерується безтінюве кольорове 3D зображення, а границі тіні коригуються за допомогою процедури зафарбовування. Метод вимагає калібрування, крім того може помилково класифікувати деякі межі об'єктів вихідного зображення як границі тіні.

У роботі [5] запропоновано спосіб видалення тіней з криволінійних областей, що зберігає текстуру фону. Видалення ефекту тіней досягається шляхом обчислення різних масштабних коефіцієнтів для тінювих та напівтінювих регіонів.

Ще одним способом, який використовується для локалізації тіней, є інтерактивне виокремлення області тіні користувачем, наприклад, за допомогою миші. Такий спосіб, очевидно, не є автоматичним, але може застосовуватися для виявлення як жорстких, так і м'яких тіней.

3 МАТЕРІАЛИ ТА МЕТОДИ

Для реалізації етапу виявлення тінювих областей були обрані два методи. Перший з них оснований на ідеях, викладених в роботі [6]. Для виявлення тіні здійснюється перехід від кольорного простору RGB до простору Lab , у якому компонента L характеризує світлоту, що асоціюється з яскравістю пікселя, а хроматична складова пікселя описується компонентами a і b . Оскільки тінь менш освітлена, ніж нетінюва частина зображення, то можливо її виявити при розгляді каналу L , який зберігає інформацію про яскравість. Для більшості зображень, отриманих на відкритому повітрі, значення параметра b менше в тінювих областях. Отже, для виявлення тіней аналізуються значення параметрів L , a та b кожного пікселя. Для цього розраховуються середні значення за всіма каналами простору Lab : $\text{mean}(L)$, $\text{mean}(a)$, $\text{mean}(b)$, стандартне відхилення для каналу L :

$$SD(L) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (L_i - \text{mean}(L))^2}, \quad (1)$$

а також порогове значення

$$T = \text{mean}(L) - K_L \cdot SD(L). \quad (2)$$

Коефіцієнт K_L у формулі (2) підібрано емпіричним шляхом, в роботі використовується значення $K_L = 1/3$. Виявлення тіні здійснюється за наступним правилом. Якщо $\text{mean}(a) + \text{mean}(b) \leq H$, то піксель зі значенням $L < T$ позначається як тінювий, а решта – як нетінюві. Якщо $\text{mean}(a) + \text{mean}(b) > H$, то пікселі з найбільш

низькими значеннями параметрів L та b можуть бути класифіковані як ті, що належать тіні, а інші – як нетіньові. Значення порогу H може налаштуватися в залежності від розкиду значень a та b , але в загальному випадку $H=256$ дає задовільний результат. Отримана бінарна маска часто містить велику кількість неправильно класифікованих пікселів та шум. Для їх видалення можна застосувати процедуру з використанням порогових значень для областей бінарної маски тіні. Але в даній роботі пропонується підхід, що базується на використанні медіанного фільтру для видалення шуму на бінарній масці тіні. Він не потребує додаткових параметрів і дає задовільний результат.

Другий метод для виявлення тіні розроблений з використанням ідей, наведених в [7]. Метод складається з декількох кроків. На першому кроці зображення розмивається за допомогою фільтра Гауса. На другому – застосовується алгоритм середнього зсуву для видалення текстури. Алгоритм середнього зсуву групує об'єкти з близькими ознаками. В якості компонент вектора ознак обираються координати (x, y) пікселя та його RGB компоненти. Пікселі зі схожими ознаками поєднуються в один сегмент, в результаті отримуємо зображення з однорідними областями. На третьому кроці отримане зображення конвертується у відтінки сірого. На останньому кроці застосовується процес бінаризації для перетворення зображення у відтінках сірого в монохромне, де присутні тільки чорні та білі пікселі. Порогове значення T обчислюється за формулою (2), де середнє значення та стандартне відхилення (формула (1)) розраховуються на масиві інтенсивностей пікселів зображення у відтінках сірого. Після виконання описаних кроків отримуємо бінарне зображення, що визначає виявлену тінь: білі пікселі відповідають тіні, а чорні – освітленій частині. Слід відзначити, що ефективність даного методу суттєво залежить від параметрів, що використовуються в алгоритмі середнього зсуву, та від значення порогу T , що застосовується при бінаризації.

Другий етап розв'язання задачі – видалення тіні. Основна ідея видалення тіні полягає в її освітленні у відповідності до обраного підходу. Припускаємо, що тінь вже виявлено, тобто тіннова область S та освітлена область Q вже відомі. Для видалення тіні пропонується використовувати один з чотирьох методів, який обирається в залежності від типу зображення та тіні.

Перший метод – адитивний [8]. В його основі лежить ідея збалансування інтенсивності тіннових та освітлених областей. Це досягається шляхом обчислення середніх значень по червоному, зеленому та синьому каналам в RGB -зображенні для освітлених та затінених областей окремо та додаванні різниці цих значень до пікселів тіні. Отже, алгоритм цього методу складається з таких кроків:

- розрахувати середні значення для освітлених пікселів $\text{mean}(Q(C))$ та для тіннових $\text{mean}(S(C))$, де $C \in \{R, G, B\}$.
- знайти різницю відповідних середніх значень за формулою:

$$\text{del}_C = \text{mean}(Q(C)) - \text{mean}(S(C)), \text{ де } C \in \{R, G, B\}.$$

- додати отримані різниці до каналів RGB -зображення:

$$C = C + \text{del}_C, C \in \{R, G, B\}.$$

Метод простий, але добре працює лише для зображень, що містять тінь на деякій однорідній поверхні.

Другий метод оснований на моделі тіні, наведений у роботі [1]. Проста модель тіні враховує два типи світла: розсіяне та направлене. Направлене світло надходить безпосередньо від джерела світла, а розсіяне багаторазово відбивається від оточуючих поверхонь. Модель зображення з тінями можна надати формулою

$$I_i = (L_Q t_i \cos \theta_i + L_S) R_i. \quad (3)$$

L_Q та L_S описують інтенсивність освітлених та затінених пікселів. Якщо $t_i = 1$, то об'єкт знаходиться в освітленій області, якщо $t_i = 0$, то в тіні. Коефіцієнт тіні для i -го пікселя визначається як $k_i = t_i \cos \theta_i$, а його коефіцієнт освітленості – як $r = L_Q / L_S$. З урахуванням цих коефіцієнтів формула (3) набуває вигляду:

$$I_i = L_S R_i (k_i r + 1). \quad (4)$$

Якщо тінь виявлена, то кожному i -му пікселю присвоюємо значення k_i , яке дорівнює нулю, якщо піксель тінновий, або одиниці в протилежному випадку. Тоді для отримання безтіньових пікселів необхідно збільшити інтенсивність кожного тіннового пікселя. Для цього, використовуючи формулу (4), отримуємо наступний вираз для розрахунку значень пікселів у безтіньовому зображенні:

$$\tilde{I}_i = I_i (r + 1) / (k_i r + 1). \quad (5)$$

Отже, алгоритм видалення тіні за даним методом складається з таких кроків:

- обчислити середні значення для освітлених пікселів $\text{mean}(Q(C))$ та для тіннових $\text{mean}(S(C))$, де $C \in \{R, G, B\}$.

- знайти коефіцієнт $r_c = \frac{\text{mean}(Q(C))}{\text{mean}(S(C))}$, де $C \in \{R, G, B\}$.

- обчислити нові значення тіннових пікселів для кожного колірної каналу: $\tilde{C} = (r_c + 1) \cdot C$, $C \in \{R, G, B\}$.

Розглянуті два методи працюють швидко та демонструють прийнятні результати на зображеннях, що містять частково освітлені однорідні поверхні. Загальним недоліком цих методів є те, що в них враховуються лише середні значення інтенсивності і кольору для всього зображення при видаленні тіні. Тому, якщо сцена складається з великої кількості об'єктів з різною освітленістю та кольором, ці методи не дають коректного освітлення тіні та вирівнювання її кольору.

Для більш складних випадків зображень з тінями пропонується використовувати наступні два методи. Один із них застосовує ідеї роботи [9]. Слід відзначити, що в багатьох випадках на зображеннях з тінями інтенсивність затінених областей поступово зростає від тіні до світла. Тому на зображенні можуть бути наявними два види тіннових областей: частково освітлені та неосвітлені. Отже, освітлення області тіні часто приводить до неповного видалення тіні. Для повного видалення тіні необхідно враховувати не тільки інтенсивність, але і колірний тон та наси-

ченість тіньової області. Значення колірних характеристик області тіні повинні бути вирівняні до відповідних характеристик освітлених областей, що її оточують. Алгоритм, що реалізує наведені міркування, складається з таких кроків:

- вихідне *RGB*-зображення конвертується в *Lab*-простір;

- *Lab*-зображення сегментується з використанням алгоритму середнього зсуву у відповідності до кольору пікселів;

- якщо до сегменту потрапляють і освітлені, і тіньові пікселі, то такий сегмент розбивається на два окремих. В результаті тіньові сегменти будуть сусідніми до сегментів освітленої області;

- для кожного тіньового сегменту S_k серед усіх сусідніх освітлених сегментів обирається найближчий за кольором – сегмент Q_k . Нові значення каналів L , a та b сегменту S_k розраховуються за формулами:

$$\tilde{L}_{S_k} = L_{S_k} + (\text{mean}(L_{Q_k}) - \text{mean}(L_{S_k})),$$

$$\tilde{a}_{S_k} = a_{S_k} \frac{\text{mean}(a_{Q_k})}{\text{mean}(a_{S_k})}, \quad \tilde{b}_{S_k} = b_{S_k} \frac{\text{mean}(b_{Q_k})}{\text{mean}(b_{S_k})}.$$

- на останньому кроці всі границі між затіненими та освітленими регіонами розмиваються за допомогою фільтра Гауса або обробляються з використанням процедури зафарбовування.

Слід зауважити, що вирівнювання значень кольорів областей, які повністю затінені, до найближчого за кольором сусіднього освітленого сегменту може призвести до появи на зображенні «хибних» кольорів, тобто після корекції колір області буде суттєво відрізнятися від початкового, і сцена не буде виглядати природньо. Для попередження цього ефекту необхідно перевіряти різницю кольорів між сегментами S_k та Q_k і, якщо вона перевищує деякий поріг, вирівнювання значення кольору сегменту S_k не виконується. Однак на практиці не вдалось знайти універсального значення різниці кольорів, тому цей поріг налаштовується для кожного конкретного зображення.

Ще один метод видалення тіней – константний, оснований на припущенні, що тіньова область відрізняється від освітленої лише на значення деякої константи [3]. Головна ідея полягає у знаходженні значення цієї константи та освітленні тіні з її використанням. Для цього після виявлення тіні знаходять границі тіні (в роботі для цього використовується метод Кенні [10]) та визначають два масиви пікселів: масив S містить тіньові пікселі, що знаходяться вздовж границі тіні, масив Q містить нетіньові пікселі, сусідні до границі тіні. Відзначимо, що на відміну від базового методу, де в масив обираються пікселі, що безпосередньо сусідні до границі тіні, в даній роботі для покращення результатів видалення, зокрема для зображень з нечіткими границями тіні, масиви Q та S будуються з пікселів, які розташовані в межах смуги шириною δ уздовж границі тіні. Величина δ може варіюватися в залежності від характеру тіні. Крім того, якщо масиви Q та S містять різну кількість елементів, то їх розміри вирів-

нюються так, щоб їх довжини були рівні. Значення константи знаходять за формулою:

$$K_C = \min_{a_c} \|Q(C) - (S(C) + a_c)\|^2, \text{ де } C \in \{R, G, B\}.$$

Далі здійснюється видалення тіні з зображення шляхом додавання значення константи до відповідного каналу кожного пікселя тіні. Для подальшого згладжування переходу між освітленими та тіньовими областями в даній роботі пропонується застосувати медіанний фільтр або процедуру зафарбовування границі, використовуючи інформацію, представлену в інших частинах зображення. Якщо на зображенні присутні декілька тіньових областей, то описана процедура повторюється для пошуку константи для кожного регіону. Це дозволило отримати кращі результати, в порівнянні з використанням лише одного значення константи K_C для всіх тіньових регіонів. Варто зауважити, що пошук різних констант для різних тіньових областей збільшив час обробки зображення, але якість результуючого зображення при цьому суттєво вище.

4 ЕКСПЕРИМЕНТИ

Описані вище методи реалізовані в комп'ютерній системі, яка може виконувати такі функції: виявлення тіней, видалення тіней, кластеризація, обробка границь, збір статистики. Програма розроблена на мові C++ з використанням бібліотеки комп'ютерного зору OpenCV. Вікно програми складається з області керування та області виводу. Область керування містить кнопки та прапорці для вибору методів та параметрів. Всі засоби управління програмою згруповані в панелі відповідно до свого призначення. Область виводу поділяється на три частини: в першій відображається вихідне зображення, в другу виводиться результат виявлення тіні (бінарна маска тіні), в третю частину виводиться результуюче зображення. Відзначимо, що обрати можна не тільки методи виявлення та видалення тіні, а й методи обробки границь (медіанний фільтр, фільтр Гауса, операція зафарбування), а також задати деякі параметри. На рис. 1 наведено вікно програми з вихідним зображенням та результатами його обробки.

5 РЕЗУЛЬТАТИ

Представлена комп'ютерна система була протестована на низці зображень. Приклади виявлення тіней програмою наведені на рис. 2, приклади видалення тіней різними методами – на рис. 3.

Крім якості обробки зображень була проаналізована й швидкодія реалізованих методів. Результати вимірювання часу виявлення та видалення тіней наведені в табл. 1.

6 ОБГОВОРЕННЯ

Тестування програми на різних зображеннях показало, що перший метод виявлення тіні з використанням простору *Lab* добре працює на зображеннях з чіткими границями тіні. Більш складні види тіней краще визначає другий метод. Він є більш універсальним, але потребує більше часу. При видаленні тіней перші два методи дають гарні результати для зображень, які містять деяку однорідну поверхню (об'єкт), на яку частково падає тінь. *Lab*-метод успішно видаляє тіні при слабкіших обмеженнях на зображення. Найбільш універсальним показав себе константний метод, який дав задовільні результати видалення на досить складних зображеннях.

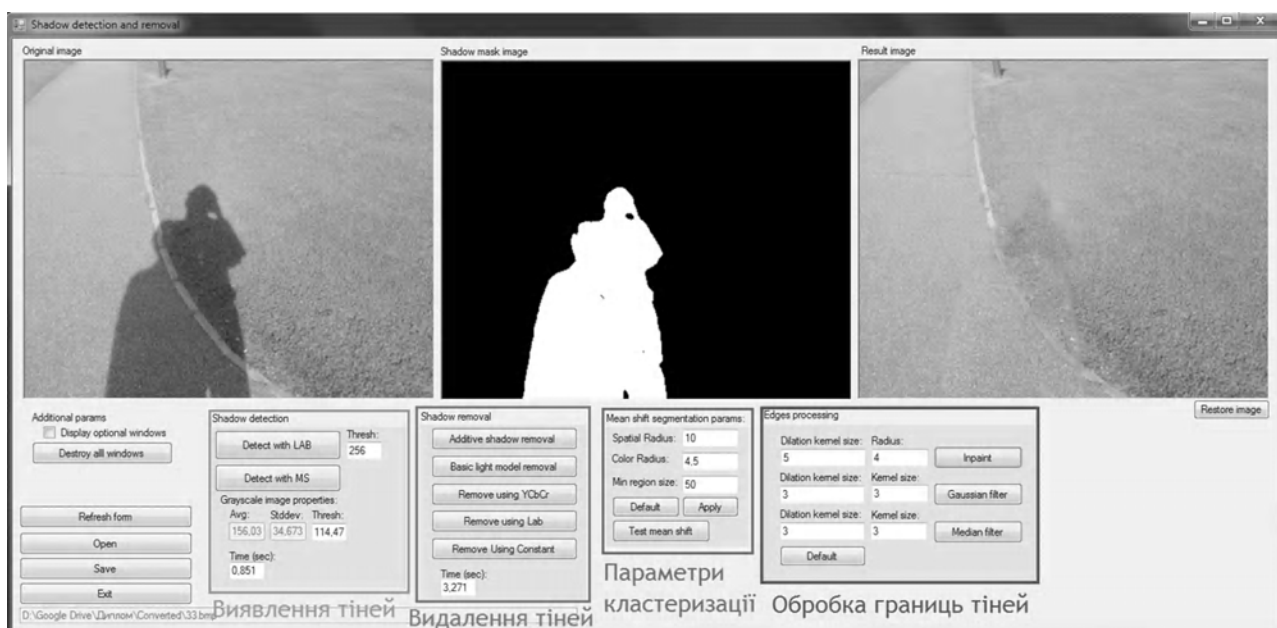


Рисунок 1 – Вікно програми

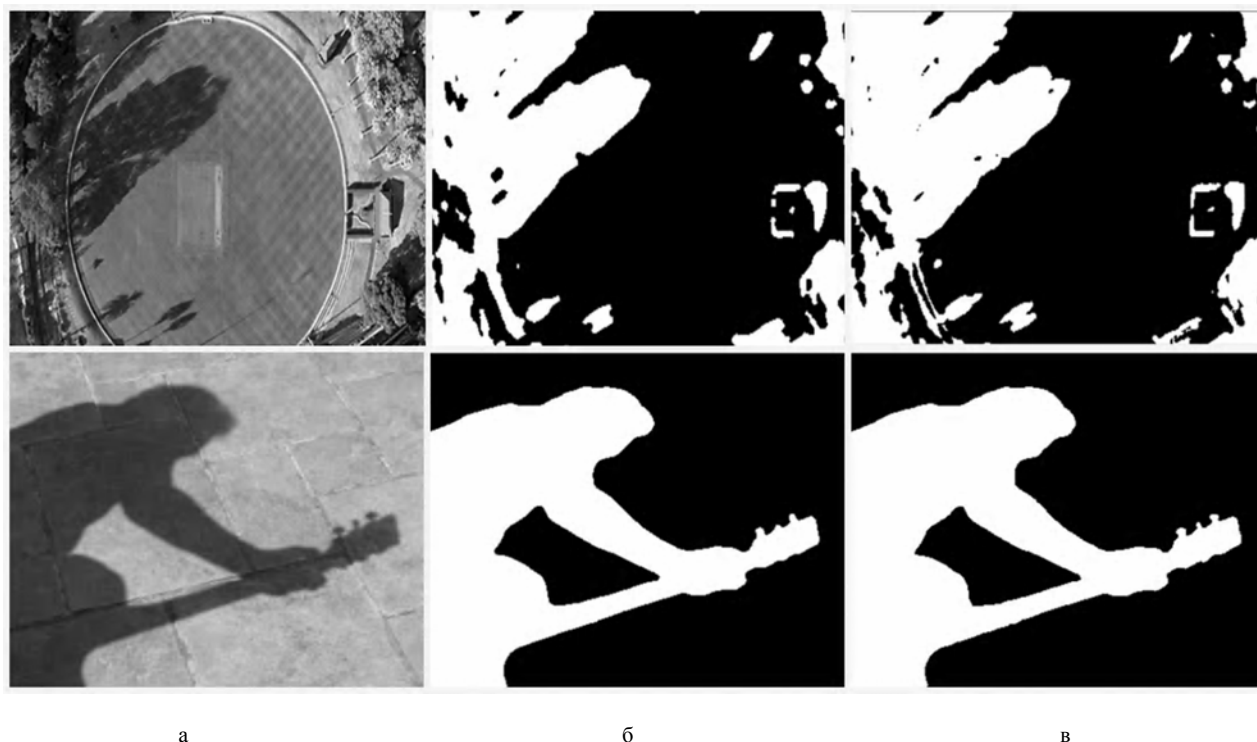


Рисунок 2 – Результати виявлення тіней:

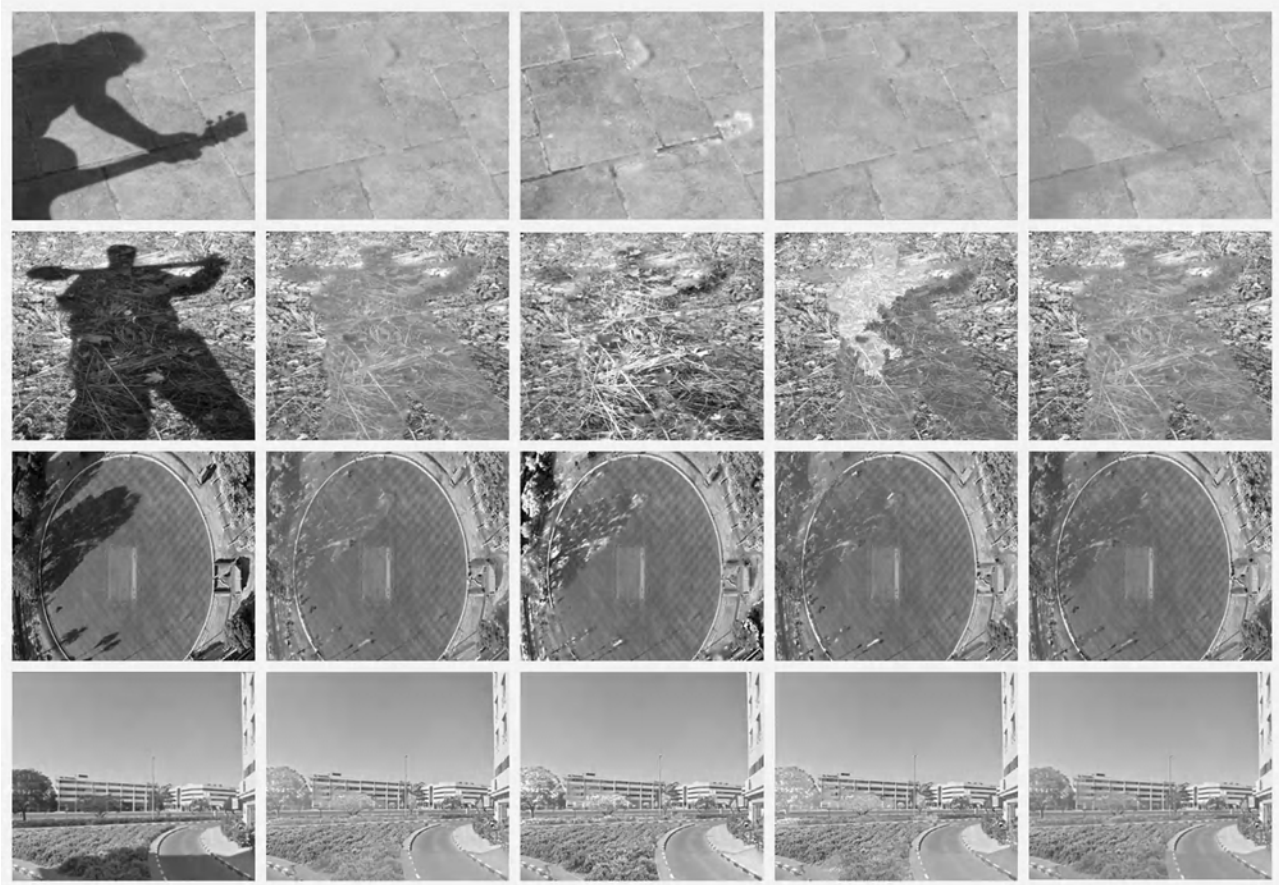
а – вихідне зображення, б – виявлення методом з переходом до *Lab*-простору, в – виявлення методом з видаленням текстури

Аналіз швидкодії методів показав, що швидше працює *Lab*-метод виявлення тіней, а для видалення тіней – адитивний та метод з використанням математичної моделі тіні, в основі яких лежать нескладні алгоритми. Найбільш затратним за ресурсами та часом виявився метод видалення тіней з переходом до *Lab*-простору, що зумовлено використанням процедури кластеризації, а також аналізом та обробкою кожного з регіонів тіні окремо.

ВИСНОВКИ

Запропоновано комп'ютерна система, яка здійснює автоматичне виявлення та видалення тіней на цифровому

зображенні. Реалізовані в системі методи локалізації та перетворення тінювих ділянок з запропонованими модифікаціями ефективно працюють на зображеннях різних типів. В залежності від типу зображення та тіні можливий вибір відповідного методу та налаштування параметрів його роботи. Практична цінність системи полягає в тому, що вона може бути застосована для попередньої обробки зображень з метою запобігання помилкам в подальшій роботі алгоритмів комп'ютерного зору. Надалі передбачається пошук шляхів для покращення результатів виявлення та видалення тіней з нечіткими границями.



а б в г д

Рисунок 3 – Приклади видалення тіней: а – вихідне зображення, б – адитивний метод, в – метод на основі моделі тіні, г – Lab-метод, д – константний метод

Таблиця 1 – Оцінка швидкодії реалізованих методів

Метод	Середній час (с)	Мінімальн. час (с)	Максим. час (с)
Виявлення тінювих областей			
Перехід до Lab-простору	0,085	0,07	0,098
Видалення текстури	2,17212	0,662	6,432
Видалення тіней			
Адитивний метод	0,013	0,011	0,015
Метод на основі моделі тіні	0,014	0,011	0,015
Перехід до Lab-простору	4,607	2,773	7,270
Константний метод	2,214	0,363	8,530

ПОДЯКИ

Роботу виконано в рамках науково-дослідної роботи кафедри комп'ютерних технологій Дніпропетровського національного університету імені Олеся Гончара «Дослідження математичних моделей фізичних процесів методами ідентифікації та рекурентного аналізу із застосуванням інформаційних технологій» (шифр 0113U004203).

СПИСОК ЛІТЕРАТУРИ

1. Deb K. Shadow Detection and Removal Based on YCbCr Color Space / K. Deb, A.H. Suny // Smart Computing Review. – 2014. – Vol. 4, № 1. – P. 23–32.
2. Finlayson G. D. Removing Shadows from Images / G. D. Finlayson, S. D. Hordley, M. S. Drew // In: Proceedings of 7 th European Conference on Computer Vision – Part IV, ECCV'02, London, UK, Springer-Verlag, 2002. – P. 823–836.
3. Fredembach C. Simple Shadow Removal / C. Fredembach, G. D. Finlayson // In: Proceedings of 18th International Conference

- on Pattern Recognition, ICPR'06 IEEE Computer Society, 2006. – P. 832–835.
4. Finlayson G. D. On the Removal of Shadows from Images / G. D. Finlayson, S. D. Hordley, C. Lu, M. S. Drew // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2006. – Vol. 28, № 1. – P. 59–68.
5. Arbel E. Texture-Preserving Shadow Removal in Color Images Containing Curved Surfaces [Electronic resource] / E. Arbel, H. Hel-Or. – Access mode: http://cs.haifa.ac.il/hagit/papers/CONF/CVPR07_ArbelHelOr-shadowRemoval.pdf.
6. Murali S. Shadow Detection and Removal from a Single Image Using LAB Color Space/ S. Murali, V. K. Govindan // Cybernetics and Information Technologies. – 2013. – Vol. 13, № 1. – P. 95–103.
7. Image Shadow Removal in Mathematica [Electronic resource]. – Access mode: <http://mathematica.stackexchange.com/questions/7414/image-shadow-removal-in-mathematica>.
8. Blajovici C. Shadow detection and removal from a single image [Electronic resource] / C. Blajovici, P. J. Kiss, Z. Bonus, L. Varga. – Access mode: <http://www.inf.u-szeged.hu/projectdirs/ssip2011/teamF/ShadowRemoval/Documentation/Shadow%20detector%20and%20removal%20from%20a%20single%20image.pdf>

9. Suny A. H. A Shadow Detection and Removal from a Single Image Using LAB Color Space/ A. H. Suny, N. H. Mithila // *IJCSI International Journal of Computer Science Issues*. – 2013. – Vol. 10, № 2. – P. 270–273.
10. Canny J. A Computational Approach to Edge Detection [Electronic resource] / J. Canny. – Access mode: https://perso.limsi.fr/vezien/PAPIERS_ACS/canny1986.pdf.

Стаття надійшла до редакції 21.12.2016.
Після доробки 28.12.2016.

Сердюк М. Е.¹, Беркут В. Г.²

¹Канд. техн. наук, доцент, доцент кафедри комп'ютерних технологій Дніпропетровського національного університету імені Олеся Гончара, Дніпр, Україна

²Студент магістратури кафедри комп'ютерних технологій Дніпропетровського національного університету імені Олеся Гончара, Дніпр, Україна

КОМПЬЮТЕРНАЯ СИСТЕМА ЛОКАЛИЗАЦИИ И УДАЛЕНИЯ ТЕНЕЙ В ЦИФРОВЫХ ИЗОБРАЖЕНИЯХ

Актуальность. Наличие затененных участков на цифровых изображениях может вызывать проблемы в процессах распознавания, отслеживания, классификации объектов. Поэтому выявление и удаление теней является важным этапом предварительной обработки изображений, которая применяется во многих алгоритмах компьютерного зрения.

Цель. Целью работы является поиск эффективных методов автоматического выявления и удаления теней и разработка программной системы, которая реализует соответствующую обработку изображений.

Методы. В работе рассмотрено несколько подходов к процессу выявления и преобразования затененных участков, проанализирована применимость их к теням разных типов. Для локализации теней в работе использован метод с переходом к цветовому пространству *Lab* и метод с предварительным удалением текстуры. Для удаления теней использовано четыре метода: аддитивный, на основе математической модели тени, с переходом к *Lab*-пространству, константный. Предложены модификации некоторых этапов с целью улучшения практических результатов.

Результаты. Разработана компьютерная система, в которой реализовано автоматическое выявление и устранение теней с использованием указанных методов, которые выбираются в зависимости от характера затененных областей. Некоторые параметры методов доступны для настройки. Система протестирована на ряде изображений. Проанализированы качество и скорость работы реализованных методов. Приведены результаты обработки изображений разными методами.

Выводы. Реализованные в представленной компьютерной системе методы локализации и удаления теней с предложенными модификациями эффективно работают на изображениях разных типов. Дополнительная настройка параметров позволяет улучшить результаты в некоторых сложных случаях. Система может быть использована для предварительной обработки изображений с целью предотвращения ошибок в дальнейшей работе алгоритмов компьютерного зрения.

Ключевые слова: выявление тени, удаление тени, обработка изображений, цветовое пространство *Lab*, сегментация, модель тени.

Serdiuk M. E.¹, Berkut V. G.²

¹PhD, Associate Professor, Associate Professor of Computer Technologies Department, Oles Honchar Dnipropetrovsk National University, Dnipro, Ukraine

²Graduate student of Computer Technologies Department, Oles Honchar Dnipropetrovsk National University, Dnipro, Ukraine

THE COMPUTER SYSTEM OF SHADOW LOCATING AND REMOVAL IN DIGITAL IMAGES

Context. Presence of shaded areas in digital images may cause problems in processes of objects recognition, tracking, classification. Thus shadows identification and removal are important pre-processing stages, which are used in many computer vision algorithms.

Objective. The work objective is to find effective methods for automatic shadow detection and removal and development of a software system that implements the appropriate image processing.

Method. In this paper we have discussed several approaches to the process of identifying and converting shaded areas, analyzed their applicability for different types of shadows. For shadow localization two methods are used: the method with the transition to the *Lab* color space and the method with preliminary removal of texture. For shadow removing four methods are used: additive method, method based on a mathematical model of shadows, method with the transition to the *Lab*-space, constant method. We have proposed some stages modifications to improve the practical results.

Results. Computer system with automatic detection and elimination of shadows, using the above methods, was developed. The choice of methods is carried out depending on the type of shaded areas. Some options of methods are available for configuration. The system was tested on series of images. We have analyzed quality and speed of the implemented methods. Image processing results by different methods are shown.

Conclusions. Implemented in the presented computer system methods of shadow locating and removing with the proposed modifications work effectively in different types of images. Additional settings allow to improve the results in some difficult cases. The system can be used for image pre-processing in order to avoid errors in further work of the computer vision algorithms.

Keywords: shadow detection, shadow removal, image processing, *Lab* color space, segmentation, shadow model.

REFERENCES

1. Deb K., Suny A. H. Shadow Detection and Removal Based on YCbCr Color Space, *Smart Computing Review*, 2014, Vol. 4, No. 1, pp. 23–32.
2. Finlayson G. D., Hordley S. D., Drew M. S. Removing Shadows from Images, In: *Proceedings of 7th European Conference on Computer Vision. Part IV, ECCV'02*. London, UK, Springer-Verlag, 2002, pp. 823–836.
3. Fredembach C., Finlayson G. D., Fredembach C. Simple Shadow Removal, In: *Proceedings of 18th International Conference on Pattern Recognition, ICPR'06. IEEE Computer Society*, 2006, pp. 832–835.
4. Finlayson G. D., Hordley S. D., Lu C., Drew M. S. On the Removal of Shadows from Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, Vol. 28, No. 1, pp. 59–68.
5. Arbel E., Hel-Or H. Texture-Preserving Shadow Removal in Color Images Containing Curved Surfaces [Electronic resource]. Access mode: http://cs.haifa.ac.il/hagit/papers/CONF/CVPR07_ArbelHelOr-shadowRemoval.pdf.
6. Murali S., Govindan V. K. Shadow Detection and Removal from a Single Image Using LAB Color Space, *Cybernetics and Information Technologies*, 2013, Vol. 13, No. 1, pp. 95–103.
7. Blajovic C., Kiss P. J., Bonus Z., Varga L. Image Shadow Removal in Mathematica [Electronic resource]. Access mode: <http://mathematica.stackexchange.com/questions/7414/image-shadow-removal-in-mathematica>.
8. Shadow detection and removal from a single image [Electronic resource]. Access mode: <http://www.inf.u-szeged.hu/projectdirs/ssip2011/teamF/ShadowRemoval/Documentation/Shadow%20detection%20and%20removal%20from%20a%20single%20image.pdf>
9. Suny A. H., Mithila N. H. A Shadow Detection and Removal from a Single Image Using LAB Color Space, *IJCSI International Journal of Computer Science Issues*, 2013, Vol. 10, No. 2, pp. 270–273.
10. Canny J. A Computational Approach to Edge Detection [Electronic resource]. Access mode: https://perso.limsi.fr/vezien/PAPIERS_ACS/canny1986.pdf.