Corporate System Users Identification by the Keyboard Handwriting based on Neural Networks

Domanetska Iryna, Khaddad Anton, Krasovska Hanna, Yeremenko Bohdan

Abstract: ***The paper is devoted to practical information security aspects in the corporate system. Particular attention is paid for the ensuring access control problem for corporate confidential information. Solving this problem involves identifying a person trying to gain unauthorized access to the corporate system and identification of a authorized user committing illegal actions. The paper shows that a person's keyboard handwriting is determined by behavioural characteristic, which is very difficult to imitate. This means that the person identification based on the keyboard handwriting analysis is the most reliable. It is necessary to consider the possibility of accidental uncontrolled change user's keyboard handwriting settings when it works. Such changes can be caused by changes in physiological or emotional state. Using neural networks enables to perform analysis of handwriting keyboard considering these changes at the real time. Thus, the artificial neural networks using makes it possible considerably improves the security of corporate confidential information protection from authorized user committing illegal actions. Existing methods and algorithms for user identifying by his keyboard handwriting are focused on using a multilayer perceptron. However, FAM architecture and capabilities analysis of adaptive resonance theory has identified advantages such as the ability to form associative pairs and map the rules of fuzzy inferences to clusters of neural networks in this category. Considering these characteristics was decisive in choosing the neural network intelligent system for users identification by his keyboard handwriting style. As a result of the research proposed generalized architecture Corporate System Users Identification by the Keyboard Handwriting, based on the use of Cascade ARTMAP. The scheme of rules formation in Cascade ARTMAP and the knowledge base formation scheme of Users Identification Intelligent System by Keyboard Handwriting are proposed.***

Keywords: ***access control, information security, keyboard handwriting, neural network, unauthorized access.***

Revised Manuscript Received on November 06, 2019.
***** erembm@ukr.net

Domanetska Iryna, department of Intelligent Technologies, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, irinadomanetskaya@gmail.com

Khaddad Anton, department of Cybersecurity and Computer Engineering Kiev National University of Construction and Architecture Kyiv, Ukraine, symplicitystudio@gmail.com

Krasovska Hanna department of Intelligent Technologies, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, annavkrasovska@gmail.com

Yeremenko Bohdan*, department of information technology design and applied mathematics, Kyiv National University of Construction and Architecture, Kyiv, Ukraine, erembm@ukr.net

I. INTRODUCTION

Modern business is exposed to many threats associated with moral and financial losses. The security problem of data stored on the corporate network, equipment and resources is quite acute, especially by the intensification of hacking attacks using social engineering methods.

The Positive Technologies 2018 data about the main types of computer attacks in the corporate sphere shows more than 75% global companies had corporate information leaks due to unauthorized access to their information resources [1].

By the Escal Institute of Advanced Technologies investigation in 2018 the number of unauthorized access incidents increased by 10.3%. Moreover, the target for the attackers was corporate equipment in most cases of offenses (77%): organization employees' laptops and smartphones [2].

According to the Symantec Corporation's Internet Security Threat report 2018, the number of incidents caused by insiders increased by 2% and makes up 10% of the total number of information theft cases [3].

There are many different authentication mechanisms for employees, but, by statistics, the task of identifying system users in real time is relevant and timely.

II. ANALYSIS OF RECENT RESEARCH AND PUBLICATIONS

A. Abbreviations and Acronyms

- ART – Adaptive Resonance Theory.
- ARTMAP – Cascade Adaptive Resonance Theory Mapping.
- CU – Control Unit.
- IAM – Internal Associative Memory.
- IC – Internal Control.
- UIISKH – Users Identification Intelligent System by the Keyboard Handwriting.
- FAM – Fuzzy ARTMAP.
- KH – Keyboard Handwriting.
- MP – Multilayer perceptron.
- NN –Neural Networks.
- RU – Reset Unit.
- UA – Unauthorized Access.

B. Some common concepts

Concept 1. Authentication – a procedure to verify the authenticity of the access entity.

Concept 2. Authenticator is a parameter provided to the system for verification [4].

There are three types of authenticators:

- Unique knowledge (pin code, password).
- Unique item (key, smart card, token).
- Unique characteristic of the subject (static such as: fingerprints, retina shots and behavioural such as voice, pattern of action, keyboard handwriting).

Concept 3. Keyboard handwriting – behavioural characteristic, which is described by the parameters [5]:

- Input speed – the number of characters entered divided by typing time.
- Input dynamics – the parameter characterized by the time between keystrokes and the time they are held.
- Input Errors Frequency – the parameter, which is measured by the number of errors made by the user on the computer when typing the certain amount of text.
- Using keys – using varied function keys are pressed to enter capital letters.

Concept 4. Identification – the process of establishing a subject according to certain parameters.

Concept 5.Social Engineering – a method of obtaining the necessary access to information, based on the people psychology characteristics in order to gain access to confidential information [6]..

Concept 6. Unauthorized Access – access to information in violation of the official authority of the employee, access to information closed to public access by persons who do not have permission to access this information [7].

Also, in some cases, unauthorized access is called gaining access to information by a person who has the right to access this information in an amount exceeding that necessary for the performance of official duties. Upon receipt of UA to classified information, as a rule, insufficient security of authorization means is used. In addition to the classical methods for obtaining UA (theft of passwords and smart cards), cybercriminals use social engineering techniques and methods, which are described in detail at the works [6 – 8]. One of the offenses with social engineering methods using is access to employees' unlocked workplaces when the employees are absence. In such situations, the methods of users identifying by KH will be the most effective.

C. KH analysis methods

Applying the KH analysis methods has been considered by a lot of works. In [5] described the recognizing KH methods, models and algorithms in key systems. The topic was not left without attention. Then there were other works about user authentication using the keystrokes analysis on the keyboard [9].

Source [10] includes several researches of the user keystrokes dynamics. The researches were carried out on computer crimes investigations.

These methods have several disadvantages [5, 9, 10]:

- Does not contain a user identification and authentication mechanism description in real time.
- Need application training.
- Has a strong dependence on the ergonomics of the keyboard.
- Has a significant dependence on the psychophysical state of the operator.

In addition, other important user features are missed in the work: the manipulator use, the directories using features and the system operation as a whole.

KH is used to collect and store information about users features of keyboard control sequences entering.

Any use of the keyboard makes it possible to form an associative image to identify the user by KH.

Thus, it is reasonable collection of such data:

- Record the frequency of use for keyboard control sequences entering.
- Fixate delays between clicks.
- Determine how long to press certain keys.
- Measure the of keystrokes strength.
- Track the special keys use pattern and their combinations.

To collect the mouse using and similar manipulators information is necessary to fixate indicators such as [5]:

- Style of cursor manipulation;
- Cursor holding time;
- Cursor movement speed;
- Frequency of mouse clicks.

An analysis of the appropriate accuracy collected data allows us to correctly determine such fuzzy factors as the psychological and emotional state of the user, such as [11]:

- Temperament;
- Degree of assuredness;
- Stress level;
- Tiredness level;
- Degree of concentration.

The data is fixated and saved to special database. Subsequently, these data are used to determine the legality of the user's actions and provide him access to the system.

In this case, it is necessary to consider stochastic uncontrolled changes in user's behaviour during work. However, algorithms based on classical identification methods are unable to identify users whose behaviour changes over time. Papers [7, 8] do not contain a description of the user identification mechanism in real time. At the same time, the main NNs based models advantages are capability to UA independently qualify in fuzzy conditions. NNs have gained such an advantage due to their capability to retrain in real time. Therefore, the object of research in this work is the modelling of the means for user identification by keyboard handwriting based on Fuzzy ARTMAP and Cascade ARTMAP neural net.

III. BASIC MATERIAL

The main task of the NN is users identification.

In this study, the task posed to the NN is formulated as follows:

- "NN according to the real input must either select one of the reference source patterns contained in its memory or conclude that the real data does not match any of the reference source patterns".

In other words, the NN must determine whether the user belongs to one of three classes.

The first class includes users with authorized access. The second class includes employees who are denied access to this information. The third class includes external violators. It consists of unauthorized users who are not employees of the company.

The data examples of both first and second groups are necessary for properly work of artificial neural network. In addition, the network must be able to identify a user who is not a user of the corporate system. Research of different classifiers in [12, 13] showed that most often models as MP is adapted to solve such problems.

A. Corporate System Users Identification by the keyboard handwriting based on MP

In the paper [14], it was proposed to use MP to solve the user identification problem by KH. The main task of MP in these researches was to separate the data into two classes. One of them contains authorized corporate system users. Second class contains unauthorized users.

To adapt MP to the solution of this problem a three-layer perceptron was used [14]:

- The input layer consisted of two neurons;
- The output layer NN consisted of one neuron;
- The number of neurons in hidden layer was determined experimentally;
- The Gauss function was used as a function of input variables;
- A sigmoid function with a curvature coefficient equal to one was used as an activation function;
- As the method of fuzzy inference, the method of the centre of gravity of Mamdani was chosen.

The practical experiment was conducted for the model proposed in [14]. During the experiment, to computer users identification data about the input speed and assuredness were fed to the network input. The «input speed» parameter values range belonged from 0 to 1000 actions per minute and was described by fuzzy characteristics, which took the values of «low», «medium» and «high». The «confidence of input» parameter values ranged from 0 to 1 and were described by fuzzy characteristics, which took the values of «uncertain», «average» and «assuredness».

The output layer neuron state was characterized by such fuzzy values as:

- «True» – an authorized user;
- «False» – an unauthorized user;
- «Evolution» – an authorized user, but networks need retraining.

This paper recent research analysis has shown that existing KH identification tools don't consider such parameters as:

- Cursor manipulation style;
- Using directories features;
- System Operation as a whole.

In addition, apart from separating the corporate system users into authorized and unauthorized, NN must identify external violators. This category users are distinguished by the fact that in a special database there is no their KH data.

The latter requirement was the main reason for finding another neural network capable to solve the classifying data problem in cases where the number of classes is not determined in advance. Research of NN's functionality with different architectures in [15, 16] showed that most often models as Self-Organizing Map and ART are adapted to solve such problems.

Self-Organizing Map is designed to solve cluster analysis problems without a teacher. This neural network is used to project multidimensional space into smaller space. However, the implementation of the network requires the determination the output clusters number, but this number is unknown in advance when corporate network users identifying.

The advantages of ART are provided by its possibility to perceive new abnormal objects. Another advantage of ART is that it can be train both «with the teacher» and «without the teacher». Standard ART is an NN that is train without a teacher. Unlike most neural networks ART does not provide for a strict separation of the lifecycle at the learning and operation stages. These NNs are trained throughout including the operation phase.

Theoretical researches of the ART various types for the different objects classifying showed that development the system for users identification based on FAM and Cascade ARTMAP is expediency.

FAMs and Cascade ARTMAP allow to implement such important adaptive qualities of classifiers as [11, 13]:

- Stability;
- Plasticity;
- Possibility to adjust the apriority knowledge base formation throughout the life cycle;
- Ability to detect new patterns that are unlike any of the standards stored in IAM.

In this way, developed system will be able to detect users who are unknown for NN. That's why FAM neural networks categories were selected for further researches.

B. Model of FAM

The FAM's architecture is the result of combining a self-learning ART network with a transform field.

The FAM architecture model is shown in Fig. 1.

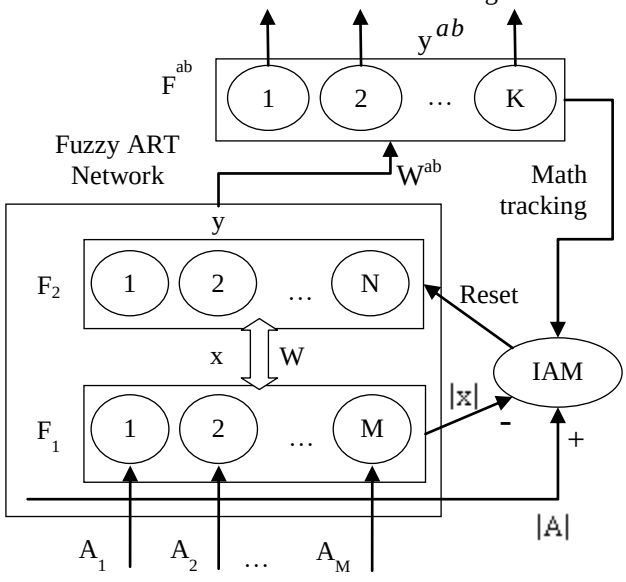

Fig. 1 FAM Architecture Model

The ART Network consists of M fully connected input neurons (F_1) and N competitive neurons (F_2).

Each neuron of the F_2 layer represents a recognition category that corresponds with a vector-prototype.

Layer F_2 connected through trained associative connections with K neurons of the output layer displays the transform field (F^{ab}).

Weight matrix (W) associates layer F_1 with c layer F_2 by direct bonds:

The matrix W^{ab} connects the F_2 layer with output layer.

The associative pattern is defined by pairs of vectors $\{a^{(p)}, b^{(p)}\}$, (p=1, 2, ...) formed during the system training and stored in the IAM (Fig. 2).

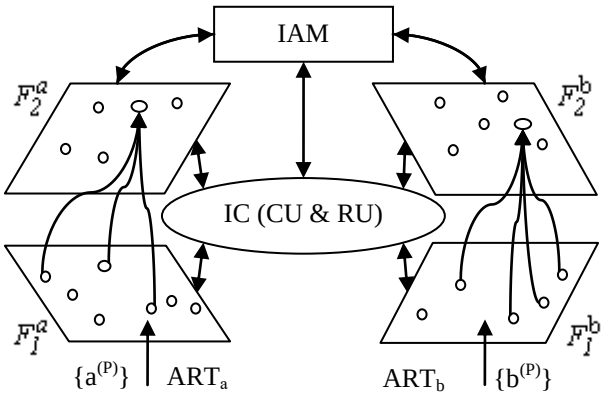

Fig. 2. Vectors associative pairs formation scheme

IAM works in the following way:

- Input vector a is processed by the weight matrix W in the ARTa, resulting in the output vector b=F(aW);
- Output vector b is processed by the transposed weight matrix W^T in the ART_b, resulting in the new output vector a=F(bW^T);

In adaptation mode FAM uses a training method with the teacher over the normalized vectors of the training sample set according to the output vector.

The General system functioning scheme is follow:

- Input vector is fed to the neurons of the first layer.
- Weighed outputs of the first layer are fed to the inputs of the second layer.
- Weighed outputs of the second layer return to the first layer.
- Current outputs of the first layer are compared with the input vector, and the resonance state is checked.
- If resonance is not established, the second layer is modified, and the procedure repeats from the second step.

The resonance system state is determined by comparing the outputs ratio of the first layer neurons with the vigilance parameter. Usually this control function is performed by internal control module [10].

Internal control module includes a CU and a RU [10].

To IC, each first layer neuron has an input signal from the control unit in the case of an active input and inactive second layer at the initial state of the system.

Further:

- If there is a resonance in the system, the pattern is successfully classified and user access authorization is set in real time.
- If the resonance is not observed, it may mean that the system has detected a new pattern which is not similar to any reference pattern stored in the associative memory of the system. In this case, the neural network classifies the user as an external intruder.
- In a case of fake winner neuron detecting the incorrect pattern recommendations will be choosen. This event should trigger a RU blocking the current winner neuron. In this case NN, I will continue to search for the standard that best fits the user's KH.

Cascade FARTMAP, as a generalization of FAM [13], allows reflecting the fuzzy inference rules as a knowledge base part to the NN topology (fig. 3).

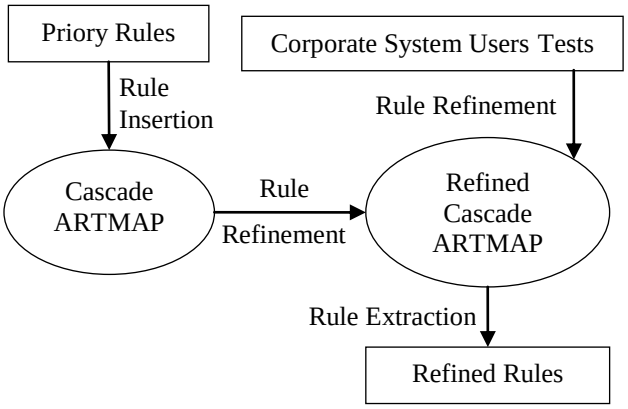

Fig. 3. Rules formation Scheme in Cascade ARTMAP

The structure of such a neural network is compatible with the representation of knowledge in expert systems. This means that reflecting the experience of information security experts can be:

- formalized as fuzzy implications
  if «$x_1, x_2, \ldots, x_M$» then «$y_1, y_2, \ldots, y_n$»;
- converted to recognition category;
- mapped to clusters Cascade ARTMAP.

Initialization of Cascade ARTMAP by a priori rules forms the initial structure of the network even before the beginning of training at the Rule Insertion stage. This can significantly speed up the NN learning process.

In the course of the Cascade ARTMAP adaptation, Map Field adjustments are made in the test cases. It is possible to modify the rules by experts in the UIISKH knowledge base, which is being developed.

C. Modelling of the UIISKH

The basis of fuzzy knowledge base UIISKH is Map Field of the artificial neural network Cascade ARTMAP

The diagram of the process of forming the initial Map Field of the Cascade ARTMAP network in the composition of UIISKH is shown in Fig. 4.

It is assumed that there is a users 'patterns bank in which each user's computer handwriting samples are saved in the form of a standard, which is formed when applying for a job.

Thus, an a priori rule base is formed (Fig. 3) to identify employees by their keyboard handwriting.

One neuron is created, when the advent of a new user in Map Field F_2^a (Fig. 2) of the Cascade ARTMAP network. Thus, priori knowledge base is modified.

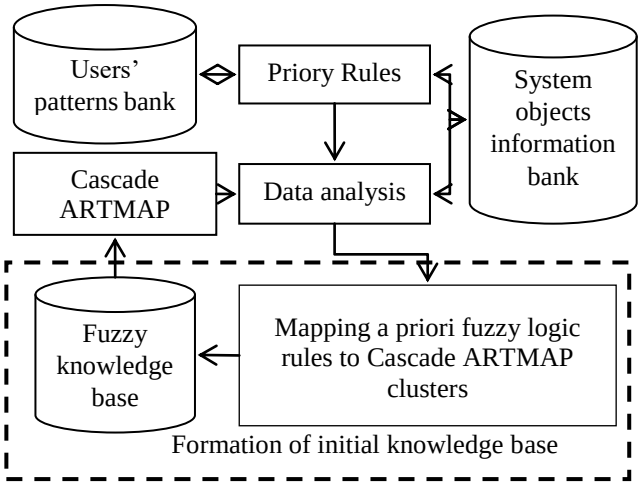

Fig. 4 Cascade ARTMAP training scheme

Then, for a while the Corporate User System is tested. The result of the system operation in testing mode is a patterns matrix (Tab. 1).

Table- I: Output patterns matrix

Pattern No	Source binary variable No.					
	1	2	…	k	…	K
1	y_{11}	y_{12}	…	y_{1k}	…	y_{1K}
2	y_{21}	y_{22}	…	y_{2k}	…	y_{2K}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nk}	…	y_{nK}
…	…	…	…	…	…	…
N	y_{N1}	y_{N2}	…	y_{Nk}	…	y_{NK}

Information is accumulated in "System objects information bank" and processed by the subsystem "Data analysis" for the rule refinement.

«Data analysis» module is intended for analogous objects information semantic analysis. To store statistical data a system objects information bank is assigned.

At this research stage, a reference patterns bank is being developed; the patterns are maps of authorized users for the getting the access based on the keyboard handwriting analysis [2, 11]. In this case, the module «Knowledge formalization» operates as a «white box» which reliability is guaranteed by a human expert who provides an assessment or forms a rule [4, 6].

At this stage of UIISKH's development, this work is done by experts, and they need to find out how the user's psychological and emotional state influences changes in his keyboard handwriting. For the convenience of experts, test results can be visualized as maps (Fig. 6). [17].

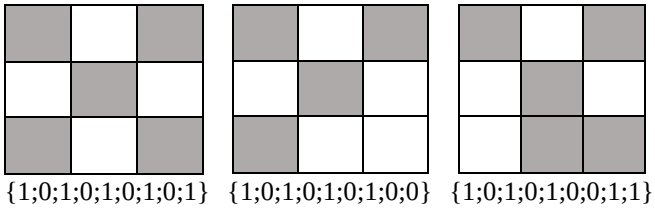

Fig. 6. Examples of test results of corporate system users

Expert conclusions with explanations are formalized in the form of fuzzy rules that seed the fuzzy knowledge base of the system. Thus, the ISIUKH fuzzy knowledge base consists of integrated expert knowledge [12].

The scheme of the development of fuzzy inference system is shown in Fig. 4.

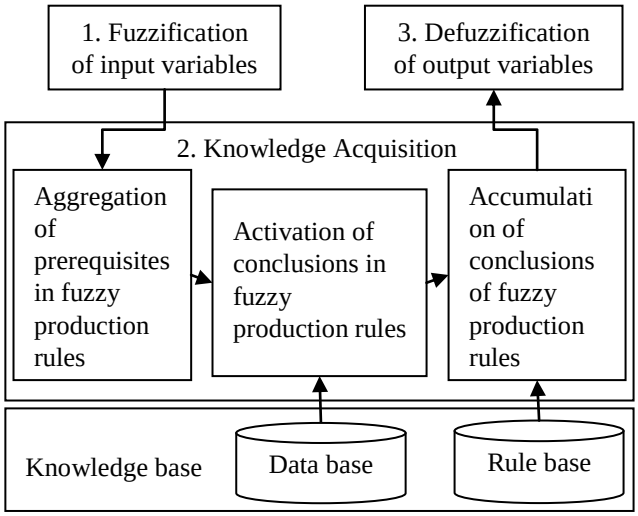

Fig.4 Scheme of Rule Refinement

IV. CONCLUSION

The paper proposes to build an intelligent system for users identifying by their keyboard handwriting based on Cascade FARTMAP.

A comparative analysis of MP, SOM and ART capabilities revealed expediency of implementation in UIISKH the Cascade ARTMAP. Cascade ARTMAP can implement a knowledge base optimization procedure by removing nonessential identification factors]. And the network ability to correct fuzzy inferences at the operation provides stage gives an opportunity to consider the user's psychological and emotional state fuzzy factors.

REFERENCES

1. Hack at any cost: how much APT can cost https://www.ptsecurity.com/ru-ru/research/pt-esc-threat-intelligence/ch
2. Internet Security Threat Report: Volume 21 by Symantec Corporation https://www.symantec.com/content/dam/symantec/docs/reports/istr21-2016-en.pdf
3. W.-K. Incident Response Capabilities in 2016: The 2016 SANS https://www.sans.org/reading-room/whitepapers/incident/incidentresponse-capabilities-2016-2016-incident-response-survey-37047.
4. V. Ivanov, E. Lubova, D. Cherkasov, "Authentication And Authorization", Problems of modern science and education 2017.
5. Keyboard handwriting as a means of authentication, I. Aguryanov , https://www.securitylab.ru/blog/personal/aguryanov/29985.php
6. K. Bossanova, Social Engineering Or How To Hack The Human System, Scientists Of Labor On The Union At The Student - Plovdiv. Series A: Public Sciences, Art And Culture.
7. V.E. Snityuk "The problem of choosing the optimal alternative in the conditions of compositional uncertainty", Cherkasy: ChITI Bulletin, 2000, №2, pp. 140-145.
8. S. Osowski "Sieci neuronowe do przetwarzania informacji", Warszawa, 2000, p 342. (in Polish).

9. G.A. Carpenter, S. Grossberg, J.H. Reynolds “ARTMAP: Supervised Real-Time Learning and Classification of Nonstationary Data by a Self-Organizing Neural Network“, Neural Networks, 4, 1991.
10. A. Arutiunian “Fuzzy Neural Network Sorter”, Technological Systems, Kiev,2009, vol.1(45), pp. 54–57.
11. B. Kosko, C. Guest “Optical bi-directional associative memories“, Sosiety for Photo-optical and Instrumentation Engineers Proceedings: Image Understanding, 1987, pp. 11-18
12. J. J. Hopfield “Learning algorithms and probability distributions in feedforward and feed-back networks“, PNAS December 1, 1987 84 (23) pp. 8429-8433. https://doi.org/10.1073/pnas.84.23.8429
13. A. H. Tan “Cascade ARTMAP: Integrating Neural Computation and Symbolic Knowledge Processing“, IEEE Trans, Neural Networks, 1997, vol. 8, n.2.
14. Terenchuk, N. Poltorachenko, Yu. Kosharna “Analysis of the ability of artificial neural networks to solve the problems of assessing the technical condition of building structures”. “Construction production”. Kiev, vol.63/1, 2017, pp.85-90
15. Terenchuk, A. Pashko, B. Yeremenko, S. Kartavykh, N. Ershova, “Modeling an intelligent system for the estimation of technical state of construction structures“, Eastern-European Journal of Enterprise Technologies, 2018, Vol3, #2(93), pp. 47-53
16. H.T. Hguen, M. Sugeno, R. Tong, R. R. Yager “Theoretical aspects of fuzzy control“, New York, John Wiley & Sons, 1995, 359 p.
17. Yu. Riabchun “Intellectualization of decision support systems for choosing the specialization of training”. Management of Development of Complex Systems. Kiev, 2019, vol. 39

AUTHORS PROFILE

Domanetska Iryna works in Department of Intelligent Technologies, Taras Shevchenko National University of Kyiv. PhD, Associate Professor. She graduated from Kyiv Civil Engineering Institute, majoring "Automated control systems"; got her PhD in Computer Science from Kyiv Civil Engineering Institute, Ukraine. She has an experience teaching computer science courses and has authored several tutorials for these courses, has more than 100 scientific and educational works. She conducts collaborative research with students, is the supervisor of undergraduate Bachelor’s and Master’s theses. She has published many scientific papers in national, international journals and conferences. Research interests: machine learning, data mining, Neuro Fuzzy technology, artificial neural networks in decision support systems.

Krasovska Hanna graduated from the Kyiv Civil Engineering Institute majoring in Computer aided design systems in construction, PhD in Technology, Associate Professor. Now works in the department of Intelligent Technologies, Taras Shevchenko National University of Kyiv, Ukraine. Lectures and provides the practical training for a number of courses in the field of programming, object-oriented analysis and design, multi-agent systems development. She supervises of scientific research for undergraduate and graduate students. Has more than 65 scientific papers in national, international journals and conferences and several tutorials for computer science students. Research interests: intelligent decision-support systems, prediction technologies and scenario analysis, adaptive intelligent systems in education, multi-agent systems and technologies.

Khaddad Anton graduated from Taras Shevchenko National University of Kyiv, majoring in Information Systems, qualification of Master of Computer Science. He works in department of Cybersecurity and Computer Engineering, Kiev National University of Construction and Architecture. Research interests: networks, web technologies, data security.

Yeremenko Bohdan graduated from Kiev National University of Construction and Architecture, majoring in Information Technology of Design, qualification of Master of Computer Science. He works in department of information technology design and applied mathematics, Kiev National University of Construction and Architecture. He has more than 50 published scientific works. PhD. Research interests: data science, artificial neural networks, machine learning, integrated technologies of complex dynamic modeling, intelligent decision-support systems.