Method of Identifying the State of Computer System under the Condition of Fuzzy Source Data

Svitlana Gavrylenko *gavrilenko08@gmail.com*
Computer Engineering and Programming Department
National Technical University "Kharkiv Polytechnic Institute"
Kharkiv, Ukraine

Viktor Chelak *victor.chelak@gmail.com*
Computer Engineering and Programming Department
National Technical University "Kharkiv Polytechnic Institute"
Kharkiv, Ukraine

Michael Kazarinov *kazarinov@gmail.com*
Computer Science Department
Northeastern Illinois University
Chicago, IL, USA

Abstract

The purpose of this work is developing a method for identifying the abnormal state of a computer system based on the Bayes' Fuzzy classifier. It allowed us to create a Fuzzy expert identification system with an unlimited number of controlled indicators that belong to a finite interval. Estimation of informativeness of such indicators does not depend on the type of indicator's functions and on the rule of their usage in the calculated formula. Introduced criterion allowed to estimate indices of the functioning of computer systems presented indistinctly. The quality of classification was evaluated based on ROC analysis. It was found that the method based on Bayes' Fuzzy expert system is qualitative, and its classification speed is almost independent of quantity indicators. Comparative evaluation of Bayes' Fuzzy classifier with Fuzzy clustering classifier and Fuzzy discriminant classifier are performed. In order to regulate the level of false-positive and false-negative classification, recommendations have been developed to manage the level of sensitivity and specificity of a Fuzzy expert system based on the Bayes classifier.

Keywords: State Identification, Mission Critical Computer Systems, Identification Measurements, Identification of Misuse.

1. INTRODUCTION

A characteristic feature of modern society is the active integration of computer technologies (specifically AI) into all spheres, including mission-critical systems. Failures and malfunctions of such systems can cause significant damage to the environment, economy, health, and even people's lives. The more important the area of computer system (CS) application, and the more complex the problem they intend to solve, the more critical are the requirements for key safety indicators of the proper functioning of such systems, and the efficiency and reliability of methods for identifying the system status [1]. One of the ways to solve this problem is to improve and develop new methods for identifying computer system status in general.

Cyber threat statistics show a significant increase in the annual number of attacks, which leads to significant economic, moral and reputational damages [2]. In 2019, the damages caused by malware worldwide have reached 600 billion U.S. dollars. Experts predict the number of attacks will keep increasing. This requires improvements in infrastructure, revising the information security strategy, the architecture of computer systems, and introduction of methods and means of identification of their state, especially in the context of the ever-changing global security system.

2. LITERATURE REVIEW

Analysis of computer systems has shown that one of the ways to improve the quality of their functioning is to improve and develop new methods for identification of computer system state [3,4].

Study of methods and means of identifying the state of computer systems has shown that the state identification systems typically include two classes of methods: anomaly identification methods and misuse identification methods [5].

These methods are based on technologies and procedures that solve problems of classification state. The analysis has shown that the main disadvantages of these methods are the neglect of Fuzzy data factors and poor adaptation to dynamic changes in structure of source data and external influences, which in turn lead to a decrease in reliability and efficiency of identification.

Existing computerized systems for identifying states are not always effective. When new misuses and anomalies caused by intrusions with unknown or unclear properties are discovered, these tools do not always remain effective and require a lot of resources for their adaptation [6].

Also, in a real situation, when the sample of the initial data is small, when the CS is operating in unsteady or critical mode, there is no certainty that the random initial data was distributed normally. For the same reason errors in statistical estimates of mathematical expectations and variances of controlled indicators can be unpredictably large. In such cases the use of a Fuzzy mathematics apparatus adapted to identify the state of the CS under these conditions will be more effective [7, 8].

Analysis of existing Fuzzy classification algorithms revealed several limitations on their use. For example, Fuzzy algorithms (FCM, Fuzzy C-Means), which are improved k-means methods for each element, calculate the value of the membership function for each of the clusters. These methods have difficulties when comparing Fuzzy numbers [9-11].

Methods based on regression analysis allow us to take into account the interaction between the selected indicators, but the mechanism for converting the source data to the final result is tough, the choice of the type of specific dependence is subjective (formal adjustment of the model to empirical material), and it is impossible to explain the cause-effect relationship. All together insufficient information content of the results allows for their ambiguous interpretation [12-15].

Decision tree method refers to logical classification methods. The main advantage of the method is the high performance of training and forecasting. On the other hand, the disadvantage of the method is the relatively low accuracy of forecasts [16]. To overcome this disadvantage, methods have been developed based on ensembles of classifiers [17-19]. However, such methods require a large amount of input data.

The analysis of the efficiency of Fuzzy classifiers [20] showed a significant dependence of the classification time on the number of input parameters. This required research related to the development of a Fuzzy classifier, the identification of which is almost independent of the number of parameters.

3. DEVELOPMENT OF A FUZZY EXPERT SYSTEM OF IDENTIFICATION BASED ON THE BAYESIAN CALCULATION TECHNOLOGY

In this paper a Fuzzy expert identification system based on Bayesian technology of calculating a set of posterior probabilities of states is proposed. The computational complexity of this system practically independent of the number of input parameters.

Development of a Fuzzy expert identification system is particularly sensitive to selection of informative controlled indicators represented by Fuzzy membership functions.

The second part of Bayes' theorem is well-known in probability theory Kullback-Leibler divergence, which estimates the information content of a specific parameter by determining the distance between distributions. At the same time, the use of the Kullback measure for evaluating information content has a number of limitations, which required the development of a new criterion.

A fairly simple criterion based on an assessment of the areas of the plots that arise when two Fuzzy functions intersect was proposed recently [8]. This criterion was used in the case when membership functions are described by functions (L-R) of the type as follows

Let us introduce the criterion of informativity of Fuzzy parameters based on the evaluation of intersection area of membership functions. Let the many of possible states to be described by the set of states $H=(H_1, H_2, ..., H_m)$. For the Fuzzy parameter x we introduce the set of conditional membership functions:

$$\mu = \left(\mu(x/H_1), \mu(x/H_2), ..., \mu(x/H_m)\right),$$

where $\mu(x/H_i)$ – membership function of a Fuzzy value x for the state $H_i, i=1,2,...,m$.

Now for a pair of states $H_{i,}H_k$ we define functions:

$$\begin{aligned} \mu_{ik}^{(C)}(x) &= \min\{\mu(x/H_i), \mu(x/H_k)\} \\ \mu_{ik}^{(D)}(x) &= \max\{\mu(x/H_i), \mu(x/H_k)\}. \end{aligned} \quad (1)$$

The relations (1) are illustrated in Fig. 1 shaded areas.

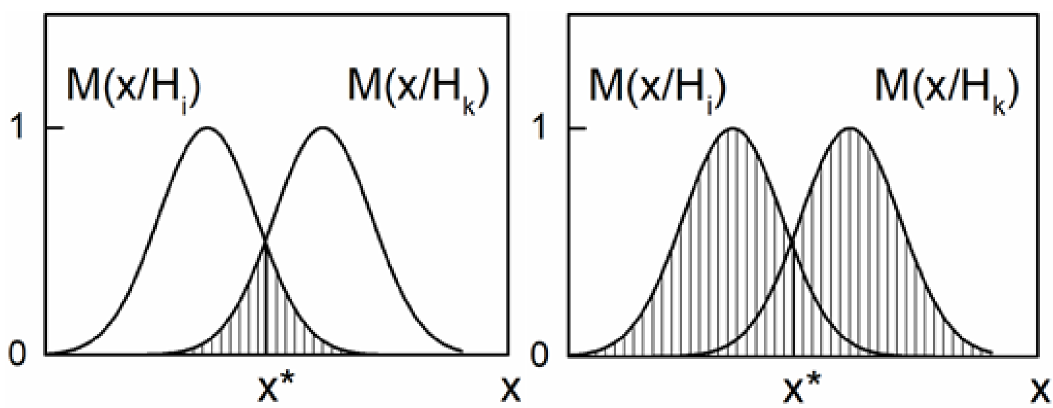

FIGURE 1: Informative criterion based on area determination.

The areas under the curves $\mu_{ik}^{(C)}(x)$ and $\mu_{ik}^{(D)}(x)$ are:

$$S\left(\mu_{ik}^{(C)}(x)\right) = \int_{-\infty}^{x^*} \mu(x/H_k)dx + \int_{x^*}^{\infty} \mu(x/H_i)dx,$$

$$S\left(\mu_{ik}^{(D)}(x)\right)=\int_{-\infty}^{x^*}\mu(x/H_i)dx+\int_{x^*}^{\infty}\mu(x/H_k)dx. \quad (2)$$

Let us determine the rule of calculation x*. For the parameter x we define the membership functions (L-R)-type corresponding to the states Hi and Hk:

$$\mu_i(x)=\begin{cases}L\left(\dfrac{a_i-x}{\alpha_i}\right), & x\le a_i,\\ R\left(\dfrac{x-a_i}{\beta_i}\right), & x>a_i;\end{cases} \qquad \mu_k(x)=\begin{cases}L\left(\dfrac{a_k-x}{\alpha_k}\right), & x\le a_k,\\ R\left(\dfrac{x-a_k}{\beta_k}\right), & x>a_k.\end{cases}$$

Let us find the point of intersection of membership functions $\mu(x/H_i)$ and $\mu(x/H_k)$, by solving equations:

$$R\left(\frac{x-a_i}{\beta_i}\right)=L\left(\frac{a_k-x}{\alpha_k}\right), \quad a_i<a_k.$$

Let us defined the membership function:

$$\mu_i(x)=\begin{cases}\exp\left\{-\left(\dfrac{\mu_i-x}{\alpha_i}\right)^2\right\}, & x\le\mu_i,\\ \exp\left\{-\left(\dfrac{x-\mu_i}{\beta_i}\right)\right\}, & x>\mu_i;\end{cases} \qquad \mu_i(x)=\begin{cases}\exp\left\{-\left(\dfrac{\mu_k-x}{\alpha_k}\right)^2\right\}, & x\le\mu_k,\\ \exp\left\{-\left(\dfrac{x-\mu_k}{\beta_k}\right)\right\}, & x>\mu_k.\end{cases}$$

Then the equation to calculate x* approach:

$$\exp\left\{-\left(\frac{x-\mu_i}{\beta_i}\right)^2\right\}=\exp\left\{-\left(\frac{\mu_k-x}{\alpha_k}\right)^2\right\},$$

where:

$$\frac{x-\mu_i}{\beta_i}=\frac{\mu_k-x}{\alpha_k},$$

$$\alpha_k x-\alpha_k\mu_i=\beta_i\mu_k-\beta_i x,$$

$$x(\alpha_k+\beta_i)=\beta_i\mu_k+\alpha_k\mu_i,$$

$$x^*=\frac{\beta_i\mu_k+\alpha_k\mu_i}{\alpha_k+\beta_i}.$$

Then the measure of informativeness of parameter x for different states Hi and Hk is determined by the relation:

$$\xi_{ik} = 1 - \frac{S\left(\mu_{ik}^{(C)}(x)\right)}{S\left(\mu_{ik}^{(D)}(x)\right)} \in [0,1]. \quad (3)$$

The information criterion (3) obtained is zero when the function of the indicator membership for different states of the CS coincides and approaches the unit as they are removed.

The following measure of informativeness of indicators was used in the development of an expert system with a non-productive logical inference mechanism based on a modified Bayes classifier. Using a number matrix

$\mu\left(x_i^{(0)} / H_k\right), i = 1,2,...,p,\ k = 1,2,...,m$ set found y_{H_k} :

$$y_{H_1} = \sum_{i=1}^{p} d_i \mu\left(x_i^{(0)} / H_1\right), y_{H_2} = \sum_{i=1}^{p} d_i \mu\left(x_i^{(0)} / H_2\right), ..., y_{H_m} = \sum_{i=1}^{p} d_i \mu\left(x_i^{(0)} / H_m\right), \quad (4)$$

where d_i – a weighting factor that determines the importance of the i indicator obtained from peer review, $i = 1,2,..., p$. Further found:

$$p_k = \frac{y_{H_k}}{\sum_{k=1}^{m} y_{H_k}}, \quad k = 1,2,...,m, \quad (5)$$

with $p_k \in [0,1]$ and $\sum_{k=1}^{m} p_k = 1$. The obtained values can be interpreted as the probabilities of the computer system being in the appropriate state. In addition, the weights that are expertly evaluated result in an unpredictable error in the final result. In this regard, this method has been modified so that the content of the calculations performed is as clear as possible.

For each element of the matrix $\mu\left(x_i^{(0)} / H_k\right)$ was applied a transformation and defined the functions:

$$f\left(x_i / H_k\right) = \frac{\mu\left(x_i / H_k\right)}{\int_{-\infty}^{\infty} \mu\left(x_i / H_k\right) dx^{(0)}}, \quad (6)$$
$$i = 1,2,..., p,\ k = 1,2,...,m.$$

.
These functions are not negative, the integral of them on the set of possible values is equal to 1, so they have all the properties of the densities of some random variables. So $f\left(x_i / H_k\right)$ can be interpreted as an analogue of the conditional density distribution of the value of the indicator, provided that the CS is in a state H_k. Then:

$$L_{H_k}(x) = \prod_{i=1}^{p} f\left(x_i / H_k\right), \quad k = 1,2,...,m \quad (7)$$

is an analog of the likelihood function. This function is the conditional density of the distribution of the values that are observed when the CS is in the state Hk.

A set of these conditional probability distributions is used to calculate Bayesian posterior probabilities of a system state, using the known sets of a priori system state probabilities $p(H_k),\ k=1,2,...,m$ for a set of controlled parameters $x_i^{(0)}$:

$$P\left(H_k / x^{(0)}\right)=\frac{L_{H_k}\left(x^{(0)}\right)\cdot p\left(H_k\right)}{\sum_{k=1}^{m} L_{H_k}\left(x^{(0)}\right)\cdot p\left(H_k\right)}. \tag{8}$$

If the a priori probabilities of the states are not known, which most closely corresponds to the real situation, then it is natural to consider them equal:

$$p\left(A_k\right)=\frac{1}{m}, k=1,2,...,m.$$

Then relation (8) is simplified and has the following form:

$$P_{H_k}=\frac{L_{H_k}\left(x^{(0)}\right)}{\sum_{k=1}^{m} L_{H_k}\left(x^{(0)}\right)}, k=1,2,...,m. \tag{9}$$

Expression (9) more accurately identifies the state of the CS compared to (5), since, according to (6), it takes into account the membership functions to the testing parameters for each state of the CS.

4. EXPERIMENTAL RESEARCH AND ASSESSMENT OF THE EFFICIENCY OF THE FUZZY BAYES CLASSIFIER

The initial data of the system are the performance characteristics of the computer system (CPU usage, memory, amount of traffic, number of read/write operations on the disk, intrusion signatures; statistics of system event analysis (number of operations with the registry, file system, number processes, etc.).

The analysis of identification efficiency was performed using of 2,3,4 indicators of CS functioning. As can be seen from Fig. 2, identification time is almost independent of the number of indicators.

As an evaluation of the quality of classification in this work used ROC analysis, which is one of the most popular quality functionals in the problems of binary classification and allows to evaluate the quality of diagnostic and prognostic methods [21]. An advantage of the ROC curve is its invariance with respect to the error rate of false-positive and false-negative identification.

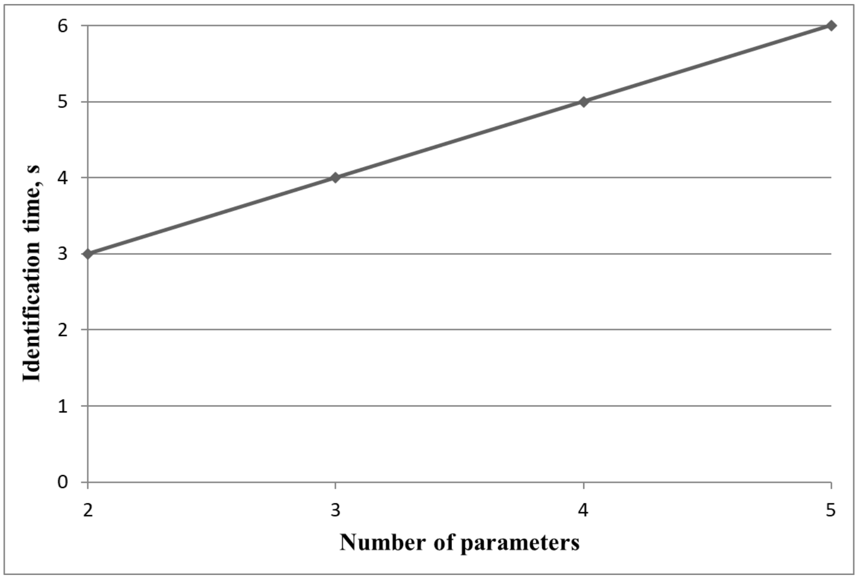

FIGURE 2: Dependence of the identification time on number of parameters.

The graph of the ROC curve for the classification of the CS state by a Fuzzy expert system based on the Bayes classifier is shown in Fig. 3.

Fig. 3 shows that the area under the ROC curve is 0.9891. This indicates that the Fuzzy expert system for identifying the state of the CS based on the Bayes classifier is excellent and is characterized by a small number of errors of false-negative and false-positive classification. The cut-off point T = 0.45 (maximum difference (TPR-FPR)) is the optimal decision threshold above which the system will be in an abnormal state.

In Fig. 4 shows a graph of the proportion of correctly classified identification objects from the cut-off threshold. As can be seen from the figure, at T = 0.45, the value of the proportion of correctly classified states is the maximum ACC = 0.95 or the total number of errors of false-positive and false-negative identification α + ß = 1 - ACC = 0.05, which is a great result for a Fuzzy expert system based on the Bayes classifier.

The graph of the dependence of the overall OVR error on the cut-off threshold (Fig. 5) also confirmed the high degree of classification of the Fuzzy expert system based on the Bayes classifier. As can be seen from the figure, at T = 0.45 the value of the total error rate is minimal, OVR = 0.05.

Simulation modeling allowed us to develop appropriate guidelines for managing the level of sensitivity and specificity of the classifier in order to regulate the level of false-positive and false-negative identification.

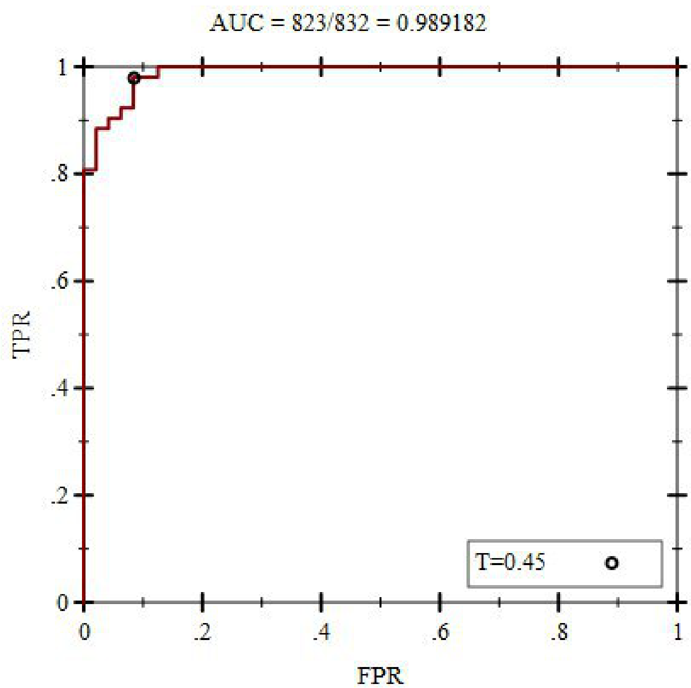

FIGURE 3: ROC curve for classification of CS state by Fuzzy expert system based on Bayes classifier.

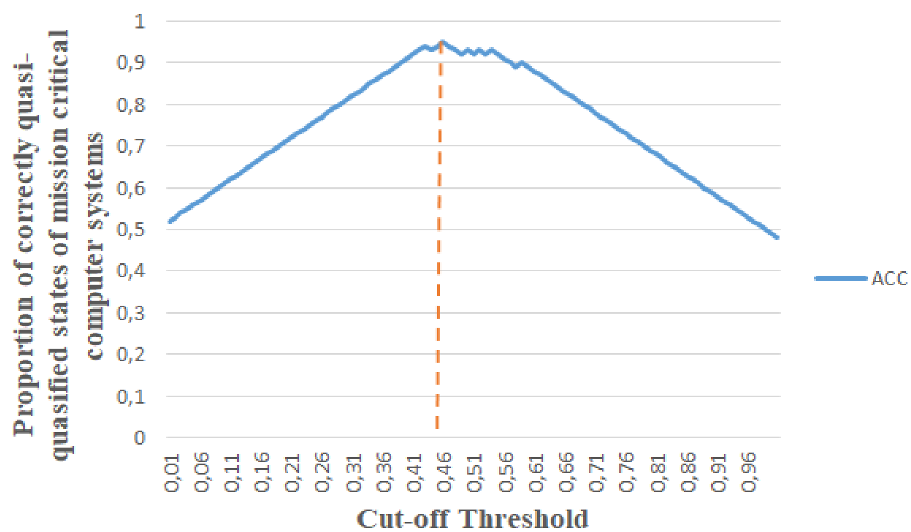

FIGURE 4: Dependence of the part of correctly classified states on the cut-off value.

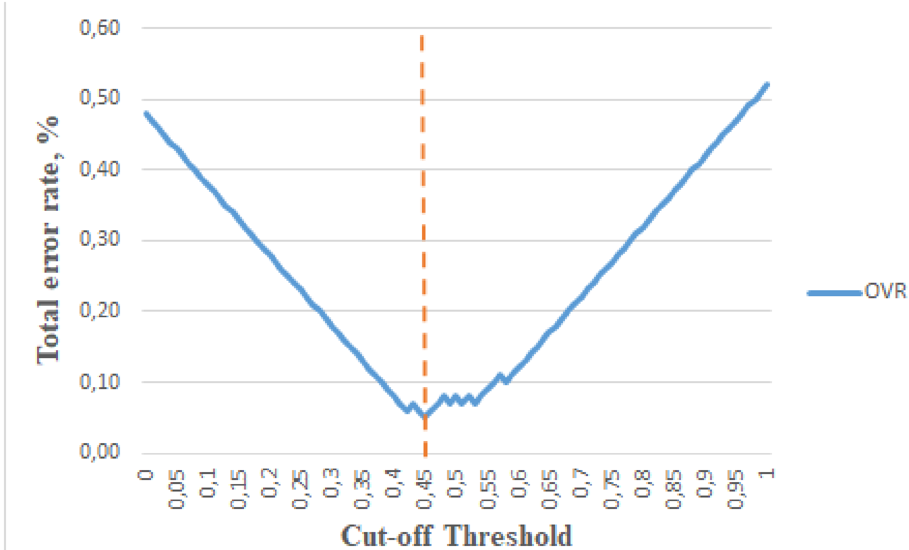

FIGURE 5: The dependence of the total error rate on the cut-off value.

Comparative evaluation of Fuzzy Bayes classifier with Fuzzy clustering classifier and Fuzzy discriminant classifier [8] is shown in Tab. 1. Thus, we can conclude that the Fuzzy Bayes classifier is more qualitative and can be used as a classifier of the method for identifying the state of a computer system.

Type of classifier	AUC	ACC
Fuzzy discriminant classifier	0,83	0,75
Fuzzy clustering classifier	0,75	0,72
Fuzzy Bayes classifier	0,99	0,95

TABLE 1: Comparative evaluation of Fuzzy Bayes classifier.

In Fig. 6 shows the dependence of false-positive and false-negative identification on the cut-off threshold, and highlights the cut-off area for which the classification system operates qualitatively. Changing the cut-off threshold within this area is also affected by the specificity and sensitivity of the classification system, which in turn leads to a change in the level of false-positive and false-negative identification.

In Fig. 7 shows a graph of the dependence of false-positive and false-negative classification on the cut-off threshold. For the fixed cut-off = 0.48 the probability of false-positive classification will be equal to 0.083, and the probability of false-negative identification will be equal to 0.077. Increasing the cut-off threshold (cut-off = 0.53) will reduce the probability of a false-positive classification, which will be equal to 0.026, and the probability of a false-negative classification will increase to 0.13.

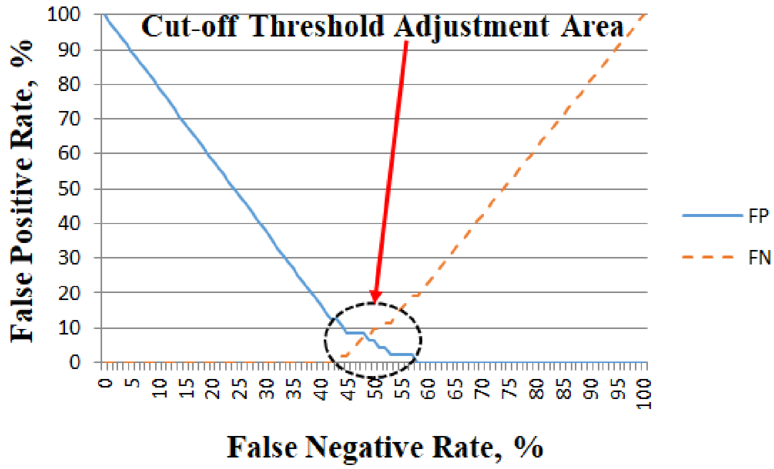

FIGURE 6: The area of control of the level of sensitivity and specificity in order to regulate the level of false-negative and false-positive identification.

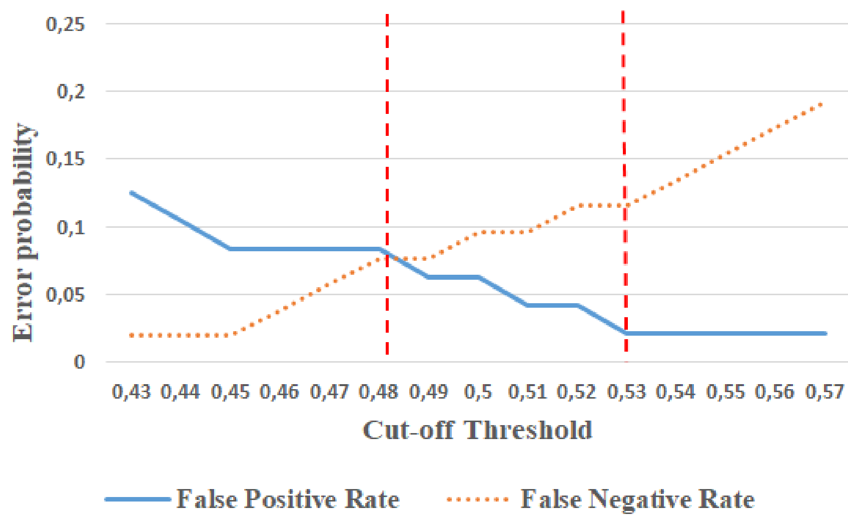

FIGURE 7: Graph of dependence of false-positive and false-negative classification on cut-off threshold.

Thus, in order to minimize the errors of false-negative classification, a model with high sensitivity should be used. In order to minimize the errors of false-positive classification, a model with high specificity must be used. This will allow the most accurate recognition of the positive or negative consequences and give the least number of false-positive or false-negative errors in the identification of the state of the CS.

5. DISCUSSION

When solving problems related to the diagnosis and protection of computer information resources, a number of limitations of the use of existing computerized systems for identifying the state of the compressor were identified. So, the presence of an insufficient volume of marked output parameters of the functioning of the CS leads to difficulties in developing an assessment criterion corresponding to the selected indicators. In addition, with the emergence of new abuses and anomalies generated by intrusions with unidentified or vaguely defined properties, these tools do not always remain effective and require long time resources for their appropriate adaptation, which leads to a decrease in the classification quality indicators.

That is why, research has made it possible to propose a method of identifying a computer system status in conditions of Fuzzy initial data. Method is based on Bayesian technology of calculating a set of posterior probabilities of states.

The quality of classification was evaluated based on ROC analysis. It was found that the method based on Bayes' Fuzzy expert system is qualitative (AUC = 0.9891; ACC = 0.95; OVR = 0.05), and its classification speed is almost independent of quantity indicators. In order to regulate the level of false-positive and false-negative classification, recommendations have been developed to manage the level of sensitivity and specificity of a Fuzzy expert system based on the Bayes classifier.

6. CONCLUSIONS AND RECOMMENDATION

When solving problems related to the diagnosis and protection of computer information resources, a number of limitations of the use of existing computerized systems for identifying the state of the compressor were identified. So, the presence of an insufficient volume of marked output parameters of the functioning of the CS leads to difficulties in developing an assessment criterion corresponding to the selected indicators. In addition, with the emergence of new abuses and anomalies generated by intrusions with unidentified or vaguely defined properties, these tools do not always remain effective and require long time resources for their appropriate adaptation, which leads to a decrease in the classification quality indicators.

That is why, research has made it possible to create a Fuzzy expert identification system with an unlimited number of controlled indicators that belong to a finite interval. Estimation of informativeness of such indicators does not depend on the type of membership functions and on the rule of their usage in the calculated formula. Introduced criterion allowed to estimate indices of functioning of computer systems represented by Fuzzy values.

The quality of classification was evaluated based on ROC analysis. It is found that the method based on Bayes' Fuzzy expert system is qualitative (AUC = 0.9891; ACC = 0.95; OVR = 0.05), and its identification speed almost independent of number of the parameters.

Comparative evaluation of Bayes' Fuzzy classifier with Fuzzy clustering classifier and Fuzzy discriminant classifier are performed. The following results were obtained: the method based on Bayes' Fuzzy expert system is more qualitative and can be used as a classifier of the method for identifying the state of a computer system.

In order to regulate the level of false-positive and false-negative classification, recommendations have been developed to manage the level of sensitivity and specificity of a Fuzzy expert system based on the Bayes classifier.

Further researches of the problem of state identification can be carried out in the following directions:
1) development of methods of Fuzzy optimization;
2) development of Fuzzy ensemble classifiers.

7. ACKNOWLEDGEMENTS

The work was supported by an international project Ерасмус + "Digital competence framework for Ukrainian teachers and other citizens", dComFra (598236-EPP-1-2018-1-IT-EPPKA2-CBHE-SP).

8. REFERENCES

[1] V.S. Kharchenko, M.A. Yastrebeneczkij and V.V. Sklyar. Novy`e informaczionny`e tekhnologii i bezopasnost` informaczionno-upravlyayushhikh sistem AE`S. “Yadernaya i radiaczionnaya bezopasnost”, Ukraine, Vol. 2, pp. 19-28, 2003.

[2] Trend Micro predstavila statistiku po kiberugrozam v mire za 2018 god. Internet: https://itc.ua/news/trend-micro-predstavila-statistiku-po-kiberugrozam-v-mire-za-2018-god, Feb. 2019 [May. 29, 2020].

[3] M.A. Yastrebeneczkij, V.V. Sklyar and M.A. Yadernaya. “Novy`e informaczionny`e tekhnologii i bezopasnost` informaczionno-upravlyayushhikh sistem AE`S”, Yadernaya i radiaczionnaya bezopasnost, Vol. 2., pp. 19-28, 2003.

[4] P.V. Shkody`rev, K.I. Yagafarov and V.A. Bashtovenko, “Obzor metodov obnaruzheniya anomalij v potokakh danny`kh”, in Proc. Software Engineering and Information Management, Sankt-Peterburg, Rossiya, Vol.18, pp. 64–70, 2017.

[5] O.I. Shelukhin, A.S. Filinova and Sakalema, Zh Obnaruzhenie vtorzhenij v komp`yuterny`e seti. Moscow, Garyachaya liniya-Telekom, 220 p, 2013.

[6] A.O. Korchenko. Metodi i`dentifi`kaczi`yi anomal`nikh stani`v dlya sistem viyavlennya vtorgnen`. Dis.dok.tekh.n Speczi`al`ni`st` 05.13.21 – «Sistemi zakhistu i`nformaczi`yi», Kiev, 405 p., 2019.

[7] L.G. Raskin, O.V. Seraya. Nechetkaya matematika. Osnovy` teorii. Prilozheniya. Ukraine, Parus, 2008, 352 p

[8] S. Semenov, O. Sira, S. Gavrylenko and N. Kuchuk. “Identification of the state of an object under conditions of Fuzzy input data”, Eastern-European Journal of Enterprise Technologies, Vol. 1, no 4 (97), pp. 22-29, 2019.

[9] K. Krishna and M. Murty. “Genetic K-mean algorithm”. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol.19, pp. 513-524, 2011.

[10] S.A. Suslov. “Klasterny`j analiz: sushhnost`, preimushhestva i nedostatki”,Vestnik NGIE`I, N. Novgrod, Vol.1, pp. 51-57, 2010.

[11] Y. Hong and S. Kwong. “To combine steady-state genetic algorithm and ensemble learning for data clustering”. Pattern Recognition Letters, Vol.29(9) pp.1416-1423, 2008

[12] A. Streke and J Ghosh. “Cluster Ensembles. A Knowledge Reuse Framework for Combining Multiple Partitions”,Journal of Machine Learning Research, pp. 583-617, 2012.

[13] R. A. Fisher. “Statisticheskie metody` dlya issledovatelej”, Russia, Gosstatizdat, 1958, 267 p

[14] F. Horia. A. and Pop. “New Fuzzy Discriminant Analysis”, Communications in Mathematical and in Computer Chemistry, pp. 391-412, 2013.

[15] V. Chandola, A. Banerjee and V. Kumar. “Anomaly detection for discrete sequences: A survey”, IEEE Transactions on Knowledge and Data Engineering, Vol. 24, №. 5, pp. 823–839, 2012.

[16] S.Yu. Gavrylenko, M.S. Melnyk and V. V. Chelak “Development of a heuristic antivirus scanner based on the file's PE-structure analysis”, Informacijni texnologiyi ta komp'yuterna inzheneriya. Mizhnarodny`j naukovo-texnichny`j zhurnal, Vol. 3 (40), pp. 23-29, 2017.

[17] Aggarwal C. C. Outlier ensembles: position paper, ACMSIGKDD Explorations Newsletter, 2017, V. 14, No. 2, pp. 49-58.

[18] Gavrylenko S., Sheverdin I., Kazarinov M. The ensemble method development of classification of the computer system state based on decisions trees / S. Gavrylenko, I. Sheverdin, Kazarinov M. // Advanced Information Systems Vol.4,No.2.–2020–P5-10 doi: 10.20998/2522-9052.2020.3.01

[19] Bolshakov A.S., Gubankova E.V. Obnaruzhenie anomalij v kompyuternyh setyah s ispolzovaniem metodov mashinnogo obucheniya. Telekommunikacionnye ustrojstva i sistemy, 2020, V. 10., No. 1, pp. 37-42.

[20] Chowdhury M. Malware Analysis and Detection Using Data Mining and Machine Learning Classification / M. Chowdhury, A. Rahman, R. Islam // International Conference on Applications and Techniques in Cyber Security and Intelligence. – ATCI. – 2018. – P. 266-274.

[21] Fawcett T. “An Introduction to ROC Analysis” Pattern Recognition Letters, 27 (8), pp. 861–874, 2006.