UDC 004.85.056.5 **doi: 10.32620/reks.2022.3.09**

Anatoliy DOVBYSH[1], Volodymyr LIUBCHAK[1], Igor SHELEHOV[1,2], Julius SIMONOVSKIY[1], Alona TENYTSKA[1]

[1] ***Sumy State University, Sumy, Ukraine***
[2] ***National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine***

INFORMATION-EXTREME MACHINE LEARNING OF A CYBER ATTACK DETECTION SYSTEM

This study increases the functional efficiency of machine learning of a cyber attack detection system. An information-extreme machine learning method for a cyber attack detection system with the optimization of control tolerances for recognition features that reflect the traffic properties of the info-communication system has been developed. The method is developed within the framework of the functional approach to modeling of cognitive processes of natural intelligence at the formation and acceptance of classification decisions. This approach, in contrast to known methods of data mining, including neuron-like structures, allows giving the recognition system adaptability to arbitrary initial conditions of the learning matrix and flexibility in retraining the system by expanding the recognition class alphabet. The method idea is to maximize the information capacity of the attack detection system in the machine learning process. A modified Kullback information measure is used as a criterion for optimizing machine learning parameters. According to the proposed categorical functional model, algorithmic software for attack detection system in the mode of machine learning with the depth of the second level has been developed and implemented. However, the depth level is determined by the number of machine learning parameters, which were optimized. The geometric parameters of the recognition hyperspherical container classes and the control tolerances on the recognition features were considered as optimization parameters, which played the role of input data quantization levels in the transformation of the input Euclidean learning matrix of the type "object-property" into a working binary learning matrix given in the Hamming space. Admissible transformations of the working training matrix of the offered method allow adapting the input mathematical description of the attack detection system to the maximum full probability of the correct classification of decision acceptance. Based on the results of information-extreme machine learning within the geometric approach, decisive rules are constructed as practically invariant to the multidimensionality of the recognition feature space. The computer simulation results of information-extreme machine learning of the attack detection system to recognize four host traffic of different profiles confirm the developed method's efficiency.

Keywords*: information-extreme machine learning; information optimization criterion; machine learning parameter; cyber attack; attack detection system; traffic.*

Introduction

The computer networks significant development and the digitalization of all sectors of the socio-economic sphere of society have led to an increase in the number of cyber attacks on information systems. Looking at the recent years statistics, it can be concluded that despite the existence of a large number of mechanisms for the information protection, cyber security crimes are on the rise. Therefore, detecting various types of network attacks or unauthorized actions and protection against them is an urgent task.

In recent years, much attention has been paid to such promising research areas in this field as blockchain technology, quantum cryptography and the creation of scientific and methodological foundations of information synthesis of cyber attack detection system, which is computer-integrated into information and communication technology (ICT). The main purpose of the attack detection system (ADS) is to detect intrusions or unauthorized access attempts. In this case, the functional effectiveness of ADS significantly depends on the role of the information synthesis methods. The use of machine learning ideas and methods and pattern recognition is an unalterable way to increase the functional efficiency of ADS.

The article considers the method of deep machine learning ADS, developed within the so-called information-extreme intelligent data analysis technology, which is based on maximizing the information capacity of the system in the machine learning process.

Problem analysis

Despite the existence of various mechanisms for protecting information, statistics in recent years show a

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

sharp increase in the number of crimes, which are related to breaches of confidentiality, integrity and availability of information. This explains the considerable attention paid in recent years to, for example, quantum cryptography [1], blockchain-protected technology [2], which minimizes the time period of network insecurity, and so on. Special hopes are placed on the development of a new direction in the field of information protection, related to the creation of ADS, computer-integrated in ICS [3-5]. The purpose of such systems is to analyze both input and output information and information coming from different ICS hosts in order to detect both attempts and actual intrusions. In this case, ADS is considered a mandatory subsystem of a comprehensive cyber security management system, computer-integrated into the ICS.

Among the basic methods used to detect cyber attacks, there are two main groups:

1) signature analysis methods [6, 7];
2) detecting anomalies methods [8, 9].

Identification of the attacks by methods of signature analysis detecting anomalies is to compare incoming, outgoing and system traffic with known patterns (signatures) of attacks stored in the knowledge base. The advantage of signature analysis methods is high efficiency in detecting known attacks and a small number of "false alarms", errors of the first kind. The disadvantages of such methods are:

−the need to constantly replenish the knowledge base with signatures of new attacks, because otherwise there is a potential danger of "skipping the attack", which is characterized by the second kind error;

−high computing costs.

The essence of the methods of detecting anomalies is that ADS has a certain set of knowledge about the normal functional state of ICS. She identifies any deviations from this state as abnormal behavior of the system. The advantage of anomaly detection methods is the ability to recognize new types of attacks. The methods of detecting anomalies have the following disadvantages:

−require long-term machine learning;

−are characterized by low efficiency and high computing costs;

−often lead to errors of the first kind, i. e ."false unjustified anxieties".

According to the results of comparative analysis of the advantages and disadvantages of the above methods of detecting attacks, a promising way to increase the functional efficiency of ADS is the use of data mining based on machine learning ideas and methods of and pattern recognition [10, 11]. As a result, the prospect of creating combined ADS, in which the rules based on the results of machine learning play the role of signatures and are able to automatically fill in the mode of factor cluster analysis [12].

Analysis of modern approaches and trends in the creation of ADS showed that to improve the accuracy and reliability of their work using known methods of Data Mining technology [13, 14], including artificial neural networks [15, 16], is associated with overcoming a number of scientific and methodological character complications:

−arbitrary initial conditions of the evaluation process;

−intersection in the space of recognition classes features, which characterize the corresponding cyber attacks:

−set dimensions of the features and recognition classes dictionary;

−the impact on ICS of uncontrolled perturbations not related to cyber threats and intrusions.

In [17], to reduce the impact of the input data multidimensionality, the use of extractors based on artificial neural networks is considered, but this approach is inevitably associated with the possibility of information loss.

The modeling of cyber attacks is the importance for the machine learning ADS, which makes it possible to form an alphabet of recognition classes and retrain ADS. An example of modeling a distributed denial-of-service (DDoS) attack using a network simulator is given in the work [18].

One of the promising directions of creating intelligent ADS is applying ideas and methods of information-extreme intellectual technology (IEIT), which based on maximizing the information capacity of the system in the machine learning process [19 - 21]. The main idea of IEIT methods, as in artificial neural networks (ANN), is adapt the input mathematical description in the machine learning process to the maximum possible probability of making the correct classification decisions. But in contrast to neuro-like structures, information-extreme machine learning methods are developed within the framework of a functional approach to modeling cognitive processes inherent in man in the formation and adoption of classification decisions. This approach allows for machine learning not in interactive mode, as is the case in ANN, but in automatic mode. In addition, IEIT methods are characterized by flexibility in retraining the system due to the expansion of the recognition classes alphabet and high efficiency, which is especially important in the cyber attacks detection.

The purpose of the article is to increase the functional efficiency of ADS by developing the information-extreme machine learning method, invariant to the multidimensionality of the recognition features space.

Formalization of the cyber attack detection system information synthesis problem

Consider the formalized formulation of the problem of information synthesis, which is able to study ADS within the IEIT. Suppose that the alphabet $\{X_m^{\circ} | m=\overline{1,M}\}$ recognition classes characterizing possible ICS, traffic profiles and an object-property training matrix $\| y_{m,i}^{(j)} \|, i=\overline{1,N}, j=\overline{1,n}$, where N, n are the recognition features number and structured feature recognition vector vectors, respectively. The row of the matrix $\{y_{m,i}^{(j)} | i=\overline{1,N}\}$ determines the j-th features vector and column $\{y_{m,i}^{(j)} | j=\overline{1,n}\}$ – training random sampling of the i-th feature values. It is known that the IEI technology concept is to transform the input training matrix Y into a training binary matrix X, which adapts to the maximum possible probability of making correct classification decisions by the method of permissible transformations in machine learning the process. Therefore, for the Hamming binary space we set of $\{g_m\}$ machine learning parameters, which affect the functional efficiency of ADS. The vector of ADS machine learning parameters to recognize, for example, the characteristics the vectors of class X_m° will be represented as a structure

$$g_m = <x_m, d_m, \delta_K>, \quad (1)$$

where x_m is the average structured vector of the recognition class features X_m°;

d_m – radius of therecognition class hyperspherical container X_m°, which is restored in the radial basis of the recognition features space;

δ – parameter of the control tolerances for recognition features field, which is equal to half of the symmetric field control tolerances

Required:

1) to determine the optimal values of machine learning parameters (1), which provide the maximum alphabetically averaged information criterion classes recognition

$$\overline{E}^{*} = \frac{1}{M} \sum_{m=1}^{M} \max_{G_E \cap \{k\}} E_m^{(k)}, \quad (2)$$

where $E_m^{(k)}$ is the information criterion for optimizing the machine learning parameters ADS to recognize the implementation of class X_m°, the value of which is calculated in the k-th machine learning step;

G_E – admissible area for determining the function of the optimization information criterion (2), which will be called the working area;

{k} – ordered set of machine learning steps (recognition classes containers recovery steps in the radial basis in the radial basis of discrete feature space);

2) for a priori classified fuzzy partition $\mathfrak{R}^{|M|}$ to build by permissible transformations in the subparaceptual Hamming binary space recognition features optimal (here and further in the work in the information sense) clear partition of recognition classes $\mathfrak{R}^{|M|}$, based on which to form infallible training matrix;

3) machine learning functional efficiency to decide whether the implementation of the recognized image belongs to one of the classes of the alphabet $\{X_m^{\circ}\}$.

Thus, the task of information synthesis capable of learning ADS is to optimize in the process of information-extreme machine learning parameters of vector (1) by information criterion (2) and making in the examination mode classification decision according to the decision-making rules.

Categorical functional model

The categorical functional model of ADS information-extreme machine learning will be presented in the form of a directional graph of mapping by operators of the corresponding sets used in the learning process. The input mathematical description of the able-bodied ADS will be presented in the form of a structure

$$I = <G, T, Z, \Omega, Y, X; f_1, f_2, f_3>,$$

where G is the space of factors that affect the normal functioning of the ICS;

T − set of traffic processing time moments;

Ω − recognition features space obtained as a result of traffic analysis;

Z – space of cyber attacks possible types;

Y – input training matrix;

X – training binary training matrix, which in the process of machine learning is adapted to the maximum full probability of making the right classification decisions;

f_1 – traffic analysis operator;

f_2 – operator formation of the input training matrix Y;

f_3 – operator for converting the input training matrix Y operator for converting the input training matrix X defined in the Hamming space

Figure 1 shows a categorical functional model of information-extreme machine learning ADS with optimization of control tolerances for recognition features.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

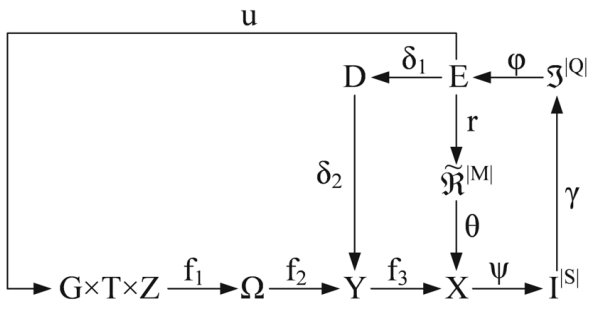

Fig. 1. Categorical functional model of machine learning

In Fig. 1 term set E, which consists of the information criterion values (2) calculated at each step of machine learning, is common to all contours of optimization of the parameters of the vector (1). Operator $r: E \to \tilde{\Re}^{|M|}$ in the machine learning process restores in the radial basis of the binary feature space containers of recognition classes, which form a partition $\tilde{\Re}^{|M|}$. Operator θ displays the partition $\tilde{\Re}^{|M|}$ on the fuzzy distribution of a priori classified binary vectors of the recognition class features. Next, operator $\psi: X \to I^{|S|}$, where $I^{|S|}$ is the set of hypotheses, tests the basic statistical hypothesis $\gamma_1: x_m^{(j)} \in X_m^{\circ}$. Operator γ determines the set of accuracy characteristics $\Im^{|Q|}$, where $Q = S^2$, and operator φ calculates the set of values E of the information optimization criterion, which is functional from the accuracy characteristics. The categorical model contains the contour of control tolerance optimization operators for recognition features, which is closed by a term set D of allowable values of the control tolerance system. In this case, the operator δ_1 at each step of machine learning changes the control field, and the operator δ_2 evaluates the dependence of the recognition features of a given control field on the tolerances. Operator u regulates the machine learning process.

The categorical functional model (Fig. 1), which reflects the construction mechanism of classification solutions by natural intelligence, can be considered as a generalized structural scheme of the information-extreme machine learning algorithm ADS. Thus, machine learning consists in search of the maximum global value of the information criterion (2) in working (admissible) area of its function definition. For two-alternative solutions, the following restrictions are imposed on the work area: the first and second reliability must be greater, respectively, errors of the first and second kind

Machine learning attack detection system

According to the categorical model (Fig. 1), the information-extreme machine learning algorithm ADS with optimization of control tolerances for recognition features corresponding to the second level of machine learning depth will be presented as a two-cycle iterative procedure for finding the global maximum information optimization criterion (2) definition of its function:

$$\{\delta^*_{K,i} \mid i = \overline{1,N}\} = \arg \max_{G_\delta}\{\max_{G_E \cap \{k\}} \bar{E}^{(k)}\}, \qquad (3)$$

where $\bar{E}^{(k)}$ is the average value of the information criterion, calculated at k-th machine learning step;

G_δ – the range of permissible values of control tolerances for signs of recognition.

The internal cycle of procedure (3) implements the so-called basic algorithm of information-extreme machine learning. The main functions of the basic algorithm are the calculation at each step of machine learning of the optimization information criterion (2) and the search for its global maximum, which determines the optimal radii of recognition classes hyperspherical containers.

The implementation of the ADS machine learning algorithm according to procedure (3) was carried out with the parallel optimization of control tolerances for recognition features, in which all tolerances for recognition features change simultaneously by a given value.

The input information for the machine learning algorithm is an array of training matrix $\{y^{(j)}_{m,i}\}$ and a system of normalized tolerances fields $\{\delta_{H,i}\}$ for recognition features, which sets the range of values of the corresponding control tolerances.

Consider the main stages of information-extreme machine learning:

1) calculation for the training matrix of the basic recognition class X_1°, for which the control tolerances are determined, the average vector of features $\{y_{1,i} | i=\overline{1,N}\}$;

2) formation of an array $\{x^{(j)}_{1,i}\}$ binary vectors of the recognition class X_1° features by the rule

$$x^{(j)}_{1,i} = \begin{cases} 1, \text{ if } y_{1,i} - \delta \le y^{(j)}_{1,i} \le y_{1,i} + \delta, \\ 0, \text{ if else;} \end{cases}$$

3) forming an array of averaged binary vectors-implementations $\{x_{m,i}|m=\overline{1,M},\ i=\overline{1,N}\}$, elements of which are calculated according to the rule

$$x_{m,i}=\begin{cases}1, \text{ if } \dfrac{1}{n}\sum\limits_{j=1}^{n}x_{m,i}^{(j)}>\rho_m,\\ 0, \text{ if else;}\end{cases}$$

where ρ_m is the selection level of binary vector coordinates x_m.

4) the set division of averages feature vectors according to the rule of "nearest neighbors" $\Re_m^{|2|}=\langle x_m, x_1\rangle$, where x_1 is the average feature vector of the neighboring class X_1° is carried out according to the following scheme:

1) the structuring of the vectors set $\{x_m\}$, is formed, starting from the vector x_1 of the base class X_1°, which characterizes the normal state of the information system functioning;

2) the matrix construction with dimension M×M of code distances between averaged feature vectors of all recognition classes;

3) determination of the minimum element for each row of the code distances matrix;

4) the structured set formation of pairwise partitioning $\{\Re_m^{|2|}\,|\,m=\overline{1,M}\}$, the elements of which are the nearest neighbors for the respective classes;

5) the code distance optimization d_m according to the iterative procedure of searching for the global maximum of the information criterion for optimizing the parameters of machine learning in the working area of determining its function:

$$d_m^*=\arg\max_{G_E\cap\{k\}} E_m^{(k)}(d), \qquad (4)$$

when the restriction on the radius value d_m of the recognition class container X_m° in the form

$$d_1<d(x_1\oplus x_2)-1;$$

6) procedure (3) is implemented and the optimal lower $A_{H,i}^*$ and upper $A_{B,i}^*$ control tolerances for recognition signs are determined according to the rules

$$A_{H,i}^*=y_{1,i}-\delta^*;\quad A_{B,i}^*=y_{1,i}+\delta^*;$$

7) STOP.

Thus, for hyperspherical containers of recognition classes, the information-extreme machine learning optimal parameters are the average implementation vectors $\{x_m^*\}$ for a given alphabet $\{X_m^\circ\}$, the recognition classes containers radii $\{d_m^*\}$ and the system of control tolerances $\{A_{H,i}^*\}$ and $\{A_{B,i}^*\}$ on recognition features.

As a criterion for optimizing of machine learning parameters was considered a modified information measure Kullback, which in equally probable two alternative hypotheses have the form

$$E_m^{(k)}=\frac{n-\left(K_{1,m}^{(k)}+K_{2,m}^{(k)}\right)}{n}\log_2\frac{2n+10^{-p}-K_{1,m}^{(k)}-K_{2,m}^{(k)}}{K_{1,m}^{(k)}+K_{2,m}^{(k)}+10^{-p}}, \qquad (5)$$

where $K_{1,m}^{(k)}$ is the number of events that indicate the non-belonging of "their" vectors of the recognition class features X_m°;

$K_{2,m}^{(k)}$ – he number of events that indicate the belonging of "foreign" vectors of the recognition class features X_m°;

10^{-p} – a small enough number that is entered to avoid division by zero;

p – the number that is recommended in practice to choose from the interval $1<p\le 3$.

According to the optimal geometric parameters of the recognition classes containers obtained in the machine learning process, decisive rules are built, which will be presented in the form of production

$$(\forall X_m^\circ\in\tilde{\Re}^{|M|})\Big(\text{if}\quad[(\mu_m>0)\&(\mu_m=\max_{\{m\}}\{\mu_m\})]$$
$$\text{then } x^{(j)}\in X_m^\circ, \text{ else } x^{(j)}\notin X_m^\circ\Big), \qquad (6)$$

where $x^{(j)}$ is a recognizable vector;

μ_m – membership function of vector $x^{(j)}$ of the recognition class container X_m°.

In expression (6) the membership function for a hyperspherical container of recognition class X_m° is determined by the formula

$$\mu_m=1-\frac{d(x_m^*\oplus x^{(j)})}{d_m^*}, \qquad (7)$$

where $d(x_m^*\oplus x^{(j)})$ is the code distance between the vector x_m^* and the recognizable vector $x^{(j)}$.

Since the decision rules (6) are built within the geometric approach, they are practically invariant to the multidimensionality of the recognition features space and

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

are characterized by high efficiency, which is an important indicator of the functional efficiency of ADS.

Operation of the recognition system in the exam mode

The functional efficiency of information-extreme machine learning ADS is evaluated in the exam mode, the algorithm of which is similar to the system algorithm in the monitoring mode. The categorical functional model of ADS functioning in the exam mode is shown in Figure 2.

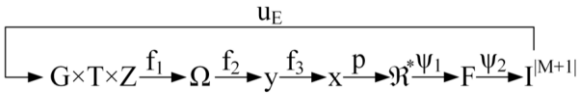

Fig. 2. Categorical model of ADS functioning in the exam mode

In the categorical model (Fig. 2), operator f_2 forms an examination vector of recognition features, similar in structure to the vectors of the training matrix. Operator f_3 generates a binary vector x according to the optimal control tolerances obtained at the machine learning stage. Operator p displays this vector for the optimal division of $\mathfrak{R}^*$ recognition classes, built at the machine learning stage. Operator ψ_1 for each feature vector sequentially calculates the value of the decisive rule (6) and forms a term set F, and operator ψ_2 on the maximum value of the decisive rule determines the affiliation of the vector x to one of the alphabet the classes $\{X_m^\circ\}$. The set of possible hypotheses $I^{|M+1|}$ contains additional hypothesis γ_{M+1}, which is accepted in case of the system failure to classify the recognizable features vector. The assignment of the u_E operator is to regulate the examination process.

Consider the main stages of the exam algorithm:

1) the recognition class counter is initialized: m := 0;

2) m := m+1;

3) the value of the membership function (7) is calculated;

4) if $m \le M$ then paragraph 2, otherwise – paragraph 5;

5) search for the maximum value of function (7):

$$\bar{\mu}_m^* = \max_{\{m\}} \bar{\mu}_m;$$

6) determination of the recognition class by the membership function, which has the maximum value;

7) if for all recognition classes the maximum values of function (7) are negative, the examination vector of features is not classified;

8) STOP.

Thus, the exam algorithm is characterized by low computational complexity, which determines its high efficiency due to the use of constructive rules built within the geometric approach. This factor is important in the operation of ADS in the monitoring mode.

Example of machine learning algorithm implementation

The above information-extreme algorithm of machine learning ADS was implemented on the example of recognizing four profiles of the physical system, system traffic, which is taken from the open data repository "Machine Learning Repository" [21]. Each traffic contains 115 characteristics, the groups description of which provides the above repository:

– H – statistics that summarize the last traffic from the host of this packet (IP);

– HH – statistics that summarize recent traffic from the host of this packet to the host destination of the packet;

– HpHp – statistics that summarize the last traffic from the host and port of this packet to the host and destination port of the packet;

– HH_jit – statistics that summarize the jitter of traffic coming from the host of this packet to the host destination of the packet;

– weight – is considered as the number of packets that are reflected in recent history (flow weight);

– radius – square root of the sum of two streams variances;

– magnitude – the square root of the sum of the two streams average values;

– cov – approximate covariance between two streams;

– pcc – correlation coefficient between two streams;

– L – time intervals.

During the implementation of the information-extreme machine learning algorithm ADS, four recognition classes were used: class X_1° – normal traffic and three infected traffic – respectively classes X_2°, X_3° and X_4°. Recognition class X_1°, was chosen as the basic one against which the control tolerances.

Figure 3 shows the graph of the dependence of the alphabetically averaged recognition classes of the normalized information criterion (4) on the parameter of the control tolerances field on traffic recognition features. In the graph, a light area indicates the work area for determining the optimization criterion, in which the first reliability is greater than the error of the first kind, and the second reliability is greater than the error of the second kind.

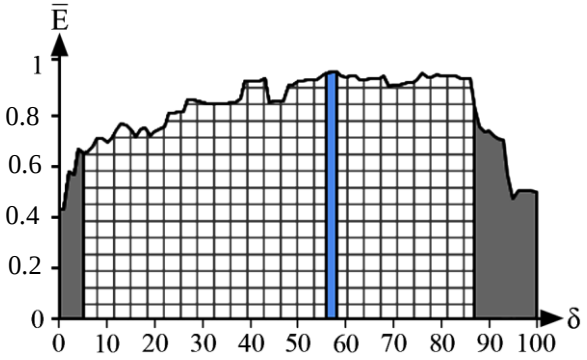

Fig. 3. Dependence graph of the information optimization criterion from the control tolerance field parameter

The analysis of the graph (Fig. 3) shows that the maximum value of the average criterion for optimizing the parameters of machine learning is equal to $\overline{E}^{*} = 0.97$ with the optimal value of the control tolerance parameter (as a deviation percentage from the nominal (average) value of recognition features).

To construct the decision rules (5) at optimal parameter δ^{*} the optimal radii of hyperspherical containers of recognition classes were determined by procedure (4). Figure 4 shows the dependence graphs of the information criterion (5) of machine learning parameters optimization on the hyperspherical containers radii of recognition classes, obtained in the process of information-extreme machine learning ADS.

Analysis of Figure 4 shows that the optimal radii of containers of recognition classes are equal to: $d_1^{*}=38$ (hereinafter in code units) for class X_1°; $d_2^{*}=10$ for class X_2°; $d_3^{*}=19$ for class X_3° and $d_4^{*}=11$ for class X_4°.

Since in Figures 4, b and 4, c the maximum values of the optimization criterion had plateau-type areas, of the radii optimal values of the recognition classes containers were determined according to the work [12] by the minimum value of the coefficient characterizing the intersection degree of two nearest neighboring recognition classes:

$$\eta_{\delta} = \frac{d_m}{d(x_m \oplus x_c)} \to \min_{\{d\}},$$

where $d(x_m \oplus x_c)$ – is the Hamming code distance between the implementation of the x_m recognition class X_m° and the implementation x_c of the nearest neighboring recognition class X_c°.

At the maximum value of the normalized criterion, ($\overline{E}^{*}=0.96$, as shown in Fig. 3, the first and second averaging values were $\overline{D}_1^{*}=0.98$ and $\overline{D}_2^{*}=0.96$. respectively. With these exact characteristics making the right classification decisions is equal to $P_t^{*}=0.97$.

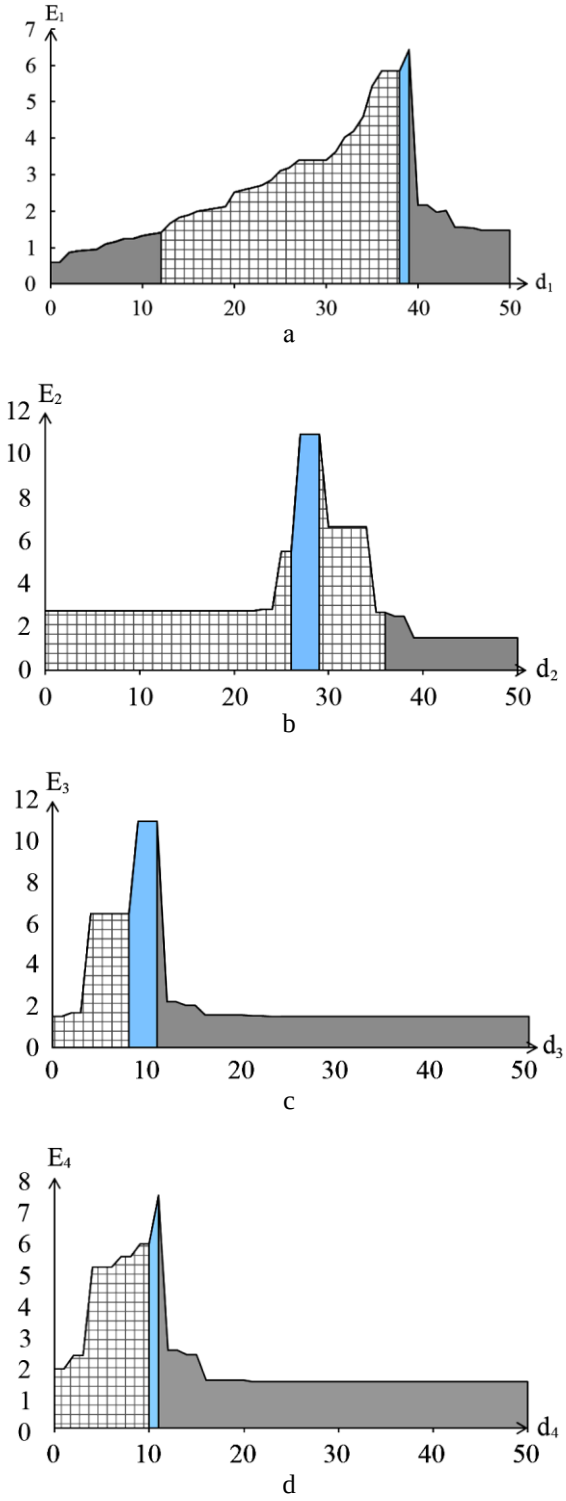

Fig. 4. Dependence graphs of the optimization criterion on the radii of the recognition classes containers:
a – class X_1°; b – class X_2°; c – class X_3°; d – class X_4°

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

In the examination mode, the following values of membership functions (7) were obtained when recognizing the recognition class implementation X_1°: $\mu_1 = 0.10$ for class X_1°; $\mu_2 = -0.78$ for class X_2°; $\mu_3 = -3.18$ for class X_3°; and $\mu_4 = -1.9$ for class X_4°. As a result, the ADS decided that the examination implementation belonged to class X_1°, which was true.

Thus, the results of machine learning have constructed a sufficiently reliable rule.

Machine learning of a typical CNN with backpropagation of the error on similar training matrix and alphabet of recognition classes was implemented in order to evaluate the advantages of the proposed method. As a result, high reliability, the same as in our case ($P_t^* = 0.98$), was achieved only when recognizing traffic of the class X_2^0 that characterizes a DDoS attack. The analysis showed that the traffic of this attack has the least amount of overlap with the traffic of other recognition classes, which ensures its good separation from the normal traffic. The total probabilities of correct traffic recognition of other recognized classes ranged from 0.62 to 0.67, which is significantly lower than the indicators given in the article. In addition, the result of such a negative application of CNN can be an insufficient volume of the training matrix, since machine learning of neuron-like structures requires more traffic. Thus, it was experimentally confirmed that, the information-extreme machine learning method, which proposed in the article and developed as a functional approach to modelling cognitive decision-making processes, is characterized by higher reliability and speed of recognition with a significantly smaller volume of the training matrix in comparison with neuro-like structures.

To increase the machine learning functional efficiency, it is necessary to increase its depth by optimizing the ADS additional parameters, including the parameters of formation of input information description of the system. Further development of the synthesized ADS is to expand its functionality by increasing the power of the recognition classes alphabet. In this case, there is a need to move from the above linear algorithms of information-extreme machine learning to hierarchical. At the same time, it is important to provide ADS flexibility properties in machine retraining, which will be the subject of signature methods, the role of which is in further research.

Conclusions

1. Developed composite method of information synthesis unites the advantages of basic approaches to detecting cyber attacks:

− the signatures methods, the role of which is played by decisive rules, built on the geometric parameters of recognition classes container obtained as a result of information-extreme machine learning, which allows to ensure high efficiency of cyber attack detection system in monitoring mode;

− anomaly detection methods that are make it possible to detect new cyber attacks. This property in the proposed method is provided by the fact that it is developed as part of a functional approach to modeling the cognitive processes of natural intelligence in the formation and adoption of classification decisions.

2. The developed method provides:

− practical invariance of decisive rules to the multidimensionality of the traffic recognition features space. This property of ADS is due to the construction of decisive rules within the geometric approach. It is known that modern computer systems are capable of processing vectors that contain 285 recognition features;

− flexibility to retrain ADS through the expansion of the recognition classes alphabet;

− in contrast to neuro-like structures, the implementation of the machine learning proposed method is carried out in automatic mode and requires an order of magnitude less training matrix.

3. The proposed method implementation on the example of the four-host traffic of different profiles recognition confirmed a fairly high reliability of classification solutions. This paves the way for further improving the functional efficiency of machine learning by optimizing additional parameters of the system.

The research was partially carried out within the project *"Fulfillment of tasks of the perspective plan of development of a scientific direction "Technical sciences" Sumy State University"* funded by the Ministry of Education and Science of Ukraine (State reg. no. 0121U112684).

Contribution of the authors: review and analysis of references, formulation of conclusions – **Anatoliy Dovbysh**; formulation (groundation) of the purpose and tasks of research – **Volodymyr Liubchak**; development of methods – **Igor Shelehov**; development of mathematical models and analysis of research results – **Julius Simonovskiy**; selection and application of software and hardware tools for modeling and presentation of results – **Alona Tenytska**.

All authors have read and approved the published version of the manuscript.

References (GOST 7.1:2006)

1. Hibrid quantum random number generator for cryptographic algorithms [Text] / M. Iavich, T. Kuchukhidze, G. Lashvili, S. Gnatyuk //

Radioelectronic and computer systems, – 2021. – No. 4. – P. 103-118. DOI:10.32620/reks.2021.4.09.

2. Bhardwaj, A. Security Incidents & Response Against Cyber Attacks [Text] / A. Bhardwaj, V. Sapra. – Springer, 2021. – 250 p.

3. Intrusion Detection Systems Explained: 13 Best IDS Software Tools Reviewed. [Electronic resource]. – Access mode: https://www.comparitech.com/net-admin/network-intrusion-detection-tools/ – 21.05.2022.

4. Top 10 BEST Intrusion Detection Systems (IDS) [2021 Rankings] [Electronic resource]. – Access mode: https://www.softwaretestinghelp.com/intrusion-detection-systems/ – 21.05.2022.

5. Best FREE Intrusion Detection Software in 2021. [Electronic resource]. – Access mode: https:// addictivetips.com/net-admin/intrusion-detection-tools/ – 21.05.2022.

6. Toliupa, S. Signature and statistical analyzers in the cyber attack detection system [Text] / S. Toliupa, V. Nakonechnyi, O. Uspenskyi // Information Technology and Security. – 2019. – Vol. 7, Iss. 1(12). – P. 69–79.

7. Snehi, J. Diverse Methods for Signature based Intrusion Detection Schemes Adopted [Text] / J. Snehi // International Journal of Recent Technology and Engineering. – 2020. – Vol. 9, Iss. 2. – P. 44-49.

8. Ananin, E. Anomalies and intrusions detection methods [Electronic resource] / E. Ananin, I. Kozhevnikova, A. Lysenko, A. Nikishova // Problems of Science. – 2016. – No. 34 (76). – P. 48-50.

9. Manasi, G. Taxonomy of Anomaly Based Intrusion Detection System: A Review [Electronic resource] / G. Manasi // International Journal of Scientific and Research Publications. – 2012. – Vol. 2, Iss. 12. – Access mode: http://www.ijsrp.org/research-paper-1212.php?rp=P12460. – 21.05.2022.

10. Dua, S. Data Mining and Machine Learning in Cybersecurity [Text] / S. Dua, X. Du. – 1st Edition. – Auerbach Publications, 2011. – 256 p.

11. Honglin, H. A Network Traffic Classification Method Using Support Vector Machine with Feature Weighted-degree [Text] / H. Honglin // Journal of Digital Information Management. – 2017. – Vol. 15(2). – P. 76-83.

12. Functional diagnostic system for multichannel mine lifting machine working in factor cluster analysis mode [Text] / V. I. Zimovets, N. I. Kalashnykova, D. E. Olada, S. V. Shamatrin // Journal of Engineering Sciences. – 2020. – Vol. 7, Iss. 1. – P. E20–E27. DOI: 10.21272/jes.2020.7(1).e4.

13. Xu, G. Applied Data Mining [Text] / G. Xu, Y. Zong, Z. Yang. – CRC Press, 2013. – 284 p.

14. Bai, J. A Deep Neural Network Based on Classification of Traffic Volume for Short-Term Forecasting [Text] / J. Bai, Y. Chen // Mathematical Problems in Engineering. – 2019. – Article ID 6318094. DOI: 10.1155/2019/6318094.

15. Abbasi, M. Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey [Text] / M. Abbasi, A. Shahraki, A. Taherkordi // Computer Communications. – 2021. – Vol. 170. – P. 19-41

16. Enhanced Network Intrusion Detection System [Text] / Ketan Kotecha, Raghav Verma et al. // Sensors. – 2021. – Vol. 21, Iss. 23. – Article ID 7835. DOI: 10.3390/s21237835.

17. Moskalenko, V. V. Extreme algorithm of the system for recognition of objects on the terrain with optimization parameter feature extraction [Text] / V. V. Moskalenko, A. G. Korobov // Radio Electronics, Computer Science, Control. – 2017. – No. 2. – P. 38-45.

18. Using graphic network simulator for ddos attacks simulation [Text] / A. Balyk, M. Karpinski, A. Naglik, G. Shangytbayeva, I. Romanets // International Journal of Computing. – 2017. – Vol. 16, Iss. 4. – P. 219-225. DOI: 10.47839/ijc.16.4.910.

19. Dovbysh, A. S. Information-Extreme Method for Classification of Observations with Categorical Attributes [Text] / A. S. Dovbysh, V. V. Moskalenko, A. S. Rizhova // Cybernetics and Systems Analysis. – 2016. – Vol. 52, Iss. 2. – P. 45-52. DOI: 10.1007/ s10559-016-9818-1.

20. Information-Extreme Machine Learning of On-Board Vehicle Recognition System [Text] / A. S. Dovbysh, M. M. Budnyk, V. Yu. Piatachenko, M. I. Myronenko // Cybernetics and Systems Analysis. – 2020. – Vol. 56, Iss. 4. – P. 534-543. DOI: 10.1007/ s10559-020-00269-y.

21. Dovbysh, A. S. Information-extreme learning algorithm for a system of recognition of morphological images in diagnosing oncological pathologies [Text] / A. S. Dovbysh, M. S. Rudenko // Cybernetics and Systems Analysis. – 2014. – Vol. 50, Iss. 1. – P. 157-163. DOI: 10.1007/s10559-014-9603-y.

22. Machine Learning Repository. [Electronic resource]. – Access mode: https://archive.ics.uci.edu/ml/ datasets/detection_of_IoT_botnet_attacks_N_BaIoT. – 21.05.2022.

References (BSI)

1. Iavich, M., Kuchukhidze, T., Lashvili, G., Gnatyuk, S. Hibrid quantum random number generator for cryptographic algorithms. *Radioelectronic and computer systems*, 2021, no. 4, pp. 103-118. DOI: 10.32620/reks.2021.4.09.

2. Bhardwaj, A., Sapra, V. *Security Incidents & Response Against Cyber Attacks*. Springer, 2021. 250 p.

3. *Intrusion Detection Systems Explained: 13 Best IDS Software Tools Reviewed*. Available at: https://www.comparitech.com/net-admin/network-intrusion-detection-tools/ (accessed 21.05.2022).

4. *Top 10 BEST Intrusion Detection Systems (IDS)* [2021 Rankings]. Available at: https://www.software testinghelp.com/intrusion-detection-systems/ (accessed 21.05.2022).

5. *Best FREE Intrusion Detection Software in 2021*. Available at: https://www.addictivetips.com/net-admin/intrusion-detection-tools/ (accessed 21.05.2022).

6. Toliupa, S., Nakonechnyi, V., Uspenskyi, O. Signature and statistical analyzers in the cyber attack

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

detection system. *Information Technology and Security*, 2019, vol. 7, iss. 1(12), pp. 69-79.

7. Snehi, J. Diverse Methods for Signature based Intrusion Detection Schemes Adopted. *International Journal of Recent Technology and Engineering*, 2020, vol. 9, iss. 2, pp. 44-49.

8. Ananin, E., Kozhevnikova, I., Lysenko, A., Nikishova, A. Anomalies and intrusions detection methods. *Problems of Sciense*, 2016. no. 34 (76), pp. 48-50.

9. Manasi, G. Taxonomy of Anomaly Based Intrusion Detection System: A Review. *International Journal of Scientific and Research Publications*, 2012. vol. 2, iss. 12. Available at: http://www.ijsrp.org/research-paper-1212.php?rp=P12460. (accessed 21.05.2022).

10. Dua, S., Du, X. *Data Mining and Machine Learning in Cybersecurity*. 1st Edition. Auerbach Publications, 2011. 256 p.

11. Honglin, H. A Network Traffic Classification Method Using Support Vector Machine with Feature Weighted-degree. *Journal of Digital Information Management*, 2017, vol. 15(2), pp. 76-83.

12. Zimovets, V. I., Kalashnykova, N. I., Olada, D. E., Shamatrin, S. V. Functional diagnostic system for multichannel mine lifting machine working in factor cluster analysis mode. *Journal of Engineering Sciences*, 2020, vol. 7, no. 1, pp. E20–E27. DOI: 10.21272/jes.2020.7(1).e4.

13. Xu, G., Zong, Y., Yang, Z. *Applied Data Mining*. CRC Press, 2013. 284 p.

14. Bai, J., Chen, Y. A Deep Neural Network Based on Classification of Traffic Volume for Short-Term Forecasting. *Mathematical Problems in Engineering*, 2019, article id 6318094. DOI: 10.1155/2019/6318094.

15. Abbasi, M., Shahraki, A., Taherkordi, A. Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey. *Computer Communications*, 2021, vol. 170, pp. 19-41.

16. Kotecha, K., Verma, R. et al. Enhanced Network Intrusion Detection System. *Sensors*, 2021, vol. 21, iss. 23, article id 7835. DOI: 10.3390/s21237835.

17. Moskalenko, V. V., Korobov, A. G. Extreme algorithm of the system for recognition of objects on the terrain with optimization parameter feature extraction. *Radio Electronics, Computer Science, Control*, 2017, no 2, pp. 38-45.

18. Balyk, A., Karpinski, M., Naglik, A., Shangytbayeva, G., Romanets, I. Using graphic network simulator for ddos attacks simulation. *International Journal of Computing*, 2017, vol. 16, iss. 4, pp. 219-225. DOI: 10.47839/ijc.16.4.910.

19. Dovbysh, A. S., Moskalenko, V. V., Rizhova, A. S. Information-Extreme Method for Classification of Observations with Categorical Attributes. *Cybernetics and Systems Analysis*, 2016, vol. 52, iss. 2, pp. 45-52. DOI: 10.1007/s10559-016-9818-1.

20 Dovbysh, A. S., Budnyk, M. M., Piatachenko, V. Yu., Myronenko, M. I. Information-Extreme Machine Learning of On-Board Vehicle Recognition System. *Cybernetics and Systems Analysis*, 2020, vol. 56, iss. 4, pp. 534-543. DOI: 10.1007/s10559-020-00269-y.

21 Dovbysh, A. S., Rudenko, M. S. Information-extreme learning algorithm for a system of recognition of morphological images in diagnosing oncological pathologies. *Cybernetics and Systems Analysis*, 2014, vol. 50, iss. 1, pp. 157-163. DOI 10.1007/s10559-014-9603-y.

22 *Machine Learning Repository*. Available at: https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT (accessed 21.05.2022).

Надійшла до редакції 25.05.2022, розглянута на редколегії 25.08.2022

ІНФОРМАЦІЙНО-ЕКСТРЕМАЛЬНЕ МАШИННЕ НАВЧАННЯ СИСТЕМИ ВИЯВЛЕННЯ КІБЕРАТАК

А. С. Довбиш, В. О. Любчак, І. В. Шелехов, Ю. В. Симоновський, А. О. Теницька

Метою дослідження є підвищення функціональної ефективності машинного навчання системи виявлення кібератак. Розроблено метод інформаційно-екстремального машинного навчання системи виявлення кібератак з оптимізацією контрольних допусків на ознаки розпізнавання, які відбивали властивості трафіка інфокомунікаційної системи. Метод розроблено в рамках функціонального підходу до моделювання когнітивних процесів природнього інтелекту при формуванні та прийнятті класифікаційних рішень. Такий підхід на відміну від відомих методів інтелектуального аналізу даних, включаючи нейроподібні структури, дозволяє надати системі розпізнавання властивості адаптивності до довільних початкових умов формування навчальної матриці та гнучкості при перенавчанні системи через розширення алфавіту класів розпізнавання. Ідея методу полягає в максимізації інформаційної спроможності системи виявлення атак в процесі машинного навчання. Як критерій оптимізації параметрів машинного навчання використовується модифікована інформаційна міра Кульбака. Згідно із запропонованою категорійною функціональною моделлю розроблено і програмно реалізовано алгоритмічне забезпечення системи виявлення атак в режимі машинного навчання з глибиною другого рівня. При цьому рівень глибини визначався кількістю параметрів машинного навчання, що оптимізувалися. Як параметри оптимізації розглядалися геометричні параметри гіперсферичних контейнерів класів розпізнавання і контрольні допуски на ознаки розпізнавання. які відігравали роль рівнів квантування вхідних даних при перетворенні вхідної евклідової навчальної матриці типу «об’єкт-властивість»

в робочу бінарну навчальну матрицю, задану в просторі Хеммінга. Шляхом допустимих перетворень робочої навчальної матриці запропонований метод дозволяє адаптувати вхідний математичний опис системи виявлення атак до максимальної повної ймовірності прийняття правильних класифікаційних рішень. За результатами інформаційно-екстремального машинного навчання в рамках геометричного підходу побудовано вирішальні правила, практично інваріантні до багато вимірності простору ознак розпізнавання. Результати комп'ютерного моделювання інформаційно-екстремального машинного навчання системи виявлення атак для розпізнавання чотирьох хостових трафіків різного профілю підтверджують працездатність розробленого методу.

Ключові слова: інформаційно-екстремальне машинне навчання; інформаційний критерій оптимізації; параметр машинного навчання; кібератака; система виявлення атак; трафік.

Довбиш, Анатолій Степанович – д-р техн. наук, проф., зав. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна.

Любчак Володимир Олександрович – канд. фіз.-мат. наук, доц., зав. каф. кібербезпеки, Сумський державний університет, Суми, Україна.

Шелехов Ігор Володимирович – канд. техн. наук, доц., доц. каф. комп'ютерних наук, Сумський державний університет, Суми, Україна; докторант каф. комп'ютерних систем, мереж і кібербезпеки, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Симоновський Юлій Віталійович – асистент кафедри комп'ютерних наук, Сумський державний університет, Суми, Україна

Теницька Альона Олексіївна – здобувач кафедри комп'ютерних наук, Сумський державний університет, Суми, Україна.

Anatoliy Dovbysh – Doctor of Technical Science, Head of the Computer Science Department, Sumy State University, Sumy, Ukraine,
e-mail: a.dovbysh@cs.sumdu.edu.ua, ORCID: 0000-0003-1829-3318, Scopus Author ID: 36052468600,
ResearcherID: AAH-1630-2021.

Volodymyr Liubchak – Ph.D., Chair of the Department of Cybersecurity, Sumy State University, Sumy, Ukraine,
e-mail: v.liubchak@dcs.sumdu.edu.ua, ORCID: 0000-0002-7335-6716, Scopus Author ID: 55654127800.

Igor Shelehov – PhD, Associate Professor at the Computer Science Department, Sumy State University, Sumy; DrS-student at the Department of Computer Systems, Networks and Cybersecurity, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,
e-mail: i.shelehov@cs.sumdu.edu.ua, i.shelehov@csn.khai.edu, ORCID: 0000-0003-4304-7768,
Scopus Author ID: 55537177800, ResearcherID: AAD-4757-2019.

Julius Simonovskiy – Assistant at the Computer Science Department, Sumy State University, Sumy, Ukraine, e-mail: julius.simonovskii@gmail.com, ORCID: 0000-0002-1228-3103.

Alona Tenytska – PhD Candidate, Computer Science Department, Sumy State University, Sumy, Ukraine, e-mail: tenickajaalena@gmail.com, ORCID: 0000-0002-2526-8842, ResearcherID: ACR-8377-2022.