

Sergey Subbotin - Andrii Oliinyk - Vitaly Levashenko - Elena Zaitseva*

DIAGNOSTIC RULE MINING BASED ON ARTIFICIAL IMMUNE SYSTEM FOR A CASE OF UNEVEN DISTRIBUTION OF CLASSES IN SAMPLE

The problem of development automation of classification rules synthesis on the basis of negative selection in the case of uneven distribution of classes in the sample is solved. The method for the synthesis of classification rules on the basis of negative selection in the case of uneven distribution of class instances of sample is proposed. This method uses a priori information about instances of all classes of the sample. The software implementing the proposed method is developed. Some experiments on the solution of practical problem of gas turbine air-engine blade diagnosis are conducted.

Keywords: Artificial immune system, instance, negative selection, classification, recognition error, sample.

1. Introduction

Process of building decision models for non-destructive testing, for technical or medical diagnostics, and for pattern recognition is a topical task [1-4]. The situation when most data of a training set belong to one class is typical for such a process [5 and 6]. We have to develop new models for object formalization or process descriptions. One of the perspective approaches for developing such models is based on conception of artificial immune systems [7-9]. This model can be created based on one class. The difference between numbers of instances belonging to different classes is significant in this case. Then the usage of artificial immune systems with negative selection is proposed in [10-13]. These systems involve the construction of a set of detectors (computational elements) that are capable of recognizing unknown instances [14-16]. This approach allows to detect anomalies or random variations in diagnosed objects [7 and 10], and to recognize instances of "non-self" classes (classes of objects which are not represented in the training set) [8, 12 and 15]. There are well known methods for the synthesis of artificial immune systems based on the negative selection [8-16]. These methods generate an exhaustive number of detectors (the possible solutions) and employ instances with one class only. Instances with other classes are not taken into account. Moreover, these methods have got high requirements for computing resources.

Consequently, the development of methods for the synthesis of artificial immune systems on the basis of negative selection,

which are free from these disadvantages, is a topical problem. In addition, the diagnostic models based on artificial immune systems have a low level of generalization. The detectors (rules) of the immune system are easy to understand. However, because of the low level of generalization, a detector system has a large dimension. It is difficult to understand and analyze by human, which generally leads to reduction of interoperability of the diagnostic model.

So, the purpose of this paper is to develop a method of classification rules synthesis on the basis of a set of detectors. These rules handle data of a training set with a significant difference in the number of instances which belong to different classes.

2. Problem statement

Let us assume that there is a training set $S = \langle P, T \rangle$, where P is a set of input parameters (features) of an objects and set T is a set of values of the output parameter. Set P is represented as a matrix $P = (p_{qm})_{QM}$, where p_{qm} is a value of the m -th feature of q -th instance in the set S . Variable m is a feature of the object ($m = 1, 2, \dots, M$). Variable q is a number of instance (object) in the sample ($q = 1, 2, \dots, Q$). Value M is measure of cardinality of features of set S ; Q is a cardinality of instances on set S . Set of values of the output parameter is represented as a vector $T = (t_q)_Q$, where $t_q \in T'$ is a value of the output

* ¹Sergey Subbotin, ¹Andrii Oliinyk, ²Vitaly Levashenko, ²Elena Zaitseva

¹Department of Program Tools, Faculty of Computer Sciences, Zaporizhzhya National Technical University, Ukraine

²Department of Informatics, Faculty of Management Science and Informatics, University of Zilina, Slovakia

E-mail: subbotin@zntu.edu.ua

parameter of q -th instance; T' is a set of possible values of the output parameter (usually in problems of non-destructive quality control and pattern recognition a set T' consists of two elements $T' = \{t'_0, t'_1\}$, determining class of suitability of object, such if $t_q = t'_0$ then q -th object is considered unusable, if $t_q = t'_1$ it is usable, suitable etc.).

The number of instances of the sample of one class (for example, instances of the class $t_q = t'_1$) is significantly different from the number of instances of another class, which is defined by (1):

$$0 \leq N_{t_q=t'_0} \ll N_{t_q=t'_1}, \quad (1)$$

where $N_{t_q=t'_0}$ and $N_{t_q=t'_1}$ are the numbers of instances of the sample $S = \langle P, T \rangle$, the value of the output parameter t_q of which are equal to t'_0 and t'_1 , respectively, $N_{t_q=t'_0} + N_{t_q=t'_1} = Q$.

Then, on the basis of a training set $S = \langle P, T \rangle$ it is necessary to generate a set $RB = \{rule_1, rule_2, \dots, rule_{NR}\}$ of productions $P_r \rightarrow T_r$ that allows to provide an acceptable level of recognition error E . This error E is defined as the ratio of number of incorrectly recognized instances N_{er} to the total number of instances Q (2):

$$E = \frac{N_{er}}{Q}. \quad (2)$$

3. The method of classification rule mining based on negative selection

As noted above, the known methods of negative selection [8-16] have got such disadvantages as the generation of the exhaustive number of detectors, the usage of information of one class instances only, low interoperability of synthesized set of solutions of detectors etc. In addition, most methods based on the principles of negative selection as detectors used hypersphere with a fixed radius. This radius determines the area of feature space covered by the detector. The choice of the radius of the hypersphere-detector is a very complex task. It can be explained by the fact that for large values of the radius recognition accuracy is reduced and for low values the number of generated detectors increases. It lowers the generalization properties of the synthesized model in a form of set of detectors of artificial immune network.

These disadvantages necessitate the extension of essential requirements to the computer resources. It decreases the speed of solutions search and in some cases does not allow to find an acceptable solution. To eliminate these drawbacks it is advisable to use the method of classification rules synthesis on the basis of negative selection in the case of uneven distribution of instances of the sample classes. In this method is used:

- known information about instances of both classes $T' = \{t'_0, t'_1\}$ in generating the set of detectors $AB = \{Ab_1, Ab_2, \dots, Ab_{N_{ab}}\}$. It is forming a set of detectors with high approximation and generalizing properties;
- information about individual significance V_m of features p_m . It is eliminating irrelevant and redundant features of the sample $S = \langle P, T \rangle$;
- a hypercube of maximum possible volume as a form of detector. It is a contrast to known methods of negative selection, in which a hypersphere is used as a form of detector. This hypercube allows to eliminate the necessity of solving a resource intensive problem of search of optimal radius of hyperspheres of detectors.

Evaluation of the significance of features p_m with respect to the output parameter T is the initial stage of the proposed method. It allows to identify and to exclude irrelevant features from further consideration, thereby reducing the search space and time of the method.

As noted above, in this paper we consider the problem in which the initial sample $S = \langle P, T \rangle$ is characterized by a discrete number of classes $T' = \{t'_0, t'_1\}$. Therefore, to estimate the significance V_m of features p_m it is advisable to apply different criteria. These criteria allow to carry out an assessment of significance of features with respect to a discrete output parameter T [2, 4 and 17-22]. We propose to use entropy as an essential criterion [4 and 17]. The entropy reflects the degree of uncertainty of the state of the object. This criterion is calculated as:

$$V_m = - \sum_{n=1}^{N_{int}(p_m)} \left(\rho(p_{mn}) \sum_{t_1=1}^{N_{int}(T)} \rho(p_{mn}, t_1) \log_2 \rho(p_{mn}, t_1) \right), \quad (3)$$

where $\rho(p_{mn}) = \frac{N(p_{mn})}{Q}$ is a probability of that the feature value p_m of instances in the sample S gets to the m -th range of values; $N(p_{mn})$ is a number of instances in the sample S , the values of the m -th feature belong to the n -th interval of its range; $N_{int}(p_m)$ is a number of intervals into which the range of the m -th feature is divided; $N_{int}(T)$ is a number of possible values (intervals into which the range of values is divided) of the output parameter T ; $\rho(p_{mn}, t_1) = \frac{N(p_{mn}, t_1)}{N(p_{mn})}$ is a conditional probability that the value of the output parameter T is equal to t_1 (will be in l -th interval t_l) provided that the m -th feature p_m gets to the n -th interval p_{mn} ; $N(p_{mn}, t_1)$ is a number of instances of the sample S , whose value of output parameter T is equal to t_l (belong

to l -th interval of the range of change t_l with the proviso that the value of m -th feature belongs to n -th interval p_m .

Features p_m with values of the individual significance below minimum ($V_m < V_{\min}$), are considered uninformative and are excluded from the sample $S = \langle P, T \rangle$.

We propose to estimate the relationship of features as the significance of one of them in relation to another. It allows to identify groups of interdependent features. Only one highly informative feature is saved in each of these groups. Other features can be excluded from further consideration. Because these features are redundant, they complicate the synthesis of diagnostic models and reduce their interoperability. The evaluation of the significance V_{md} is provided by the use of the entropy in (3). We believe that one of the features p_d is output parameter T (interval of feature values, which is considered as an output parameter p_d). This interval is split into $N_{\text{int}}(T)$ discrete intervals). After that the analogous features are excluded from the sample S (if the value of the mutual significance V_{md} is more than the maximum permissible $V_{md} > V_{\max}$).

Next, a set of detectors (structures that can determine whether the estimated instance belongs to a particular class) is built. The use of the principles of negative selection detectors for the class $T' = t'_1$ can detect unknown instances. Some of these instances do not belong to the relevant class t'_1 [9, 11 and 13].

So, we form a set of detectors, which takes two output parameter values t'_1 (class "self") and t'_0 (class "non-self"). It is necessary for making samples S_0 and S_1 from $S = \langle P, T \rangle$. We are using instances belonging to classes t'_1 and t'_0 . $S_1 = \langle P, T = t'_1 \rangle$ (sample of "self" instances) and $S_1 = \langle P, T = t'_0 \rangle$ (sample of "non-self" instances) is true in this case.

Thereafter, the first candidate in detectors $Ab_1 = \langle Ab_{1\min}, Ab_{1\max} \rangle \in AB_1$ is created. In this case $Ab_{1\min} = \{Ab_{11\min}, Ab_{12\min}, \dots, Ab_{1M\min}\}$ and $Ab_{1\max} = \{Ab_{11\max}, Ab_{12\max}, \dots, Ab_{1M\max}\}$ are the sets of minimum and maximum values of the m -th features in the candidate detector Ab_1 . We would like to note that $Ab_{1m\min} = \min_{q=1,2,\dots,Q_1} (p_{qm})$, $Ab_{1m\max} = \max_{q=1,2,\dots,Q_1} (p_{qm})$, $m = 1, 2, \dots, M$. Q_1 is the number of instances in the sample S_1 . This first candidate in detectors Ab_1 is presented in the form of a hypercube. Set AB_1 of detectors AB_k is formed based on the set of "self" instances S_1 and it allows to detect "non-self" instances, i.e. those instances that do not belong to the class t'_1 .

We detect correspondence of every n -th instance s_q in the sample $S_1 = \langle P, T = t_1 \rangle$ to candidate detector Ab_k based on (4):

$$eq(Ab_k, s_q) = \begin{cases} 1, & \left(\sum_{m=1}^M \{1 | (Ab_{km\min} < p_{qm}) \wedge (Ab_{km\max} < p_{qm})\} \right) = M; \\ 0, & \left(\sum_{m=1}^M \{1 | (Ab_{km\min} < p_{qm}) \wedge (Ab_{km\max} < p_{qm})\} \right) \neq M; \end{cases} \quad (4)$$

where $\sum_{m=1}^M \{1 | (Ab_{km\min} < p_{qm}) \wedge (Ab_{km\max} < p_{qm})\}$ determines the number of corresponding feature values p_m of q -th instance to candidate Ab_k . If $1, \left(\sum_{m=1}^M \{1 | (Ab_{km\min} < p_{qm}) \wedge (Ab_{km\max} < p_{qm})\} \right) = M$, then it is considered that an instance $s_q = \langle p_{qm}, t_q \rangle$ corresponds to the candidate detector Ab_k . In other words, this instance is located within the space of a hypercube with coordinates $Ab_{1\min} = \{Ab_{11\min}, Ab_{12\min}, \dots, Ab_{1M\min}\}$ and $Ab_{1\max} = \{Ab_{11\max}, Ab_{12\max}, \dots, Ab_{1M\max}\}$.

If there is at least one instance $s_q = \langle p_{qm}, t_q = t'_1 \rangle \in S_1$ with $eq(Ab_k, s_q) = 1$, then the candidate Ab_k is activated by comparison with the instance S_1 . This candidate cannot be a detector. Therefore, when the condition (5) is true, then the stage of additional training of the candidate Ab_k is performed.

$$\exists s_q \in S : eq(Ab_k, s_q) = 1 \quad (5)$$

The purpose of this stage is the conversion of the candidate detector Ab_k . This conversion needs to eliminate instances S_1 , which activate the candidate Ab_k . One of the features $Ab_{km} = \langle Ab_{km\min}, Ab_{km\max} \rangle$ is selected for this purpose. Candidate detector Ab_k should coincide with an instance s_q according to this feature. Next, one of the boundary values of the m -th feature of candidate Ab_k is transformed. New boundary values equal: $Ab_{km\min} = p_{qm} + \eta_n (Ab_{km\max} - Ab_{km\min})$ if value $rnd > 0.5$ or $Ab_{km\max} = p_{qm} - \eta_n (Ab_{km\max} - Ab_{km\min})$ if value $rnd \leq 0.5$). The value rnd is calculated as a randomly generated number between [0;1): $rnd = rand[0;1)$. The volume of the hypercube Ab_k decreases based on this transformation. It is explained that an instance s_q is located outside the space described by the candidate for detector Ab_k . Ratio η_n is defined by user as a parameter of the method. We define this ratio in the range of $\eta_n \in (0; 1]$. This ratio influences the distance between instances of the sample S_1 and hypercube detector Ab_k directly. The higher value of the ratio corresponds to the greater distance.

Candidate detector Ab_k is re-checked with every instance in the sample S_1 by the condition (5) after conversion of the boundary values of one of the features. When the condition (5) is true then we have to re-transform the boundary values of one feature of the candidate Ab_k . This process will be repeated until the condition (5) is fulfilled.

We search examples s_q in set $S_1 = \langle P, T = t_1 \rangle$ which activate candidate Ab_k . If these examples are absent then the step of fitness assessment of the candidate Ab_k for generalization is started. The set of detectors $AB_1 = \{Ab_1, Ab_2, \dots, Ab_{N_{Ab}}\}$ is created by using the principles of negative selection. This set allows determining the identity of instances s_q to the same class with high accuracy [10-16]. So, we use expressions (6) and (7) as

evaluation criteria of ability of the detector Ab_k for generalization of the data:

$$G_1(Ab_k) = \frac{1}{M} \sum_{m=1}^M \frac{Ab_{km\max} - Ab_{km\min}}{p_{m\max} - p_{m\min}}, \quad (6)$$

$$G_2(Ab_k) = \frac{\prod_{m=1}^M (Ab_{km\max} - Ab_{km\min})}{\prod_{m=1}^M (p_{m\max} - p_{m\min})}, \quad (7)$$

where $p_{m\min} = \min_{q=1,2,\dots,Q_1} (p_{qm})$ and $p_{m\max} = \max_{q=1,2,\dots,Q_1} (p_{qm})$ are minimum and maximum values of the m -th feature in the sample S_1 .

Criteria (6) and (7) represent the part of search space which is covered by the detector. Criterion $G_1(Ab_k)$ shows the average fraction of coverage by the detector space in each of M dimensions of the feature space. Criterion $G_2(Ab_k)$ shows the volume part of the covered space. The values of these criteria influence the size of the search space which covers this detector. The greater values correspond to a greater part of the search space. Consequently, if the criterion for evaluating the quality of generalization $G_1(Ab_k)$ is above the threshold G_{\min} , then the candidate Ab_k has a high generalization capability. This candidate can be added to the set of detectors $AB_1 = AB_1 \cup \{Ab_k\}$.

The creation of new candidates Ab_k is carried out until the termination criteria are satisfied. These criteria can be used as follows: recognition accuracy $E(S)$, the achievement of the maximum number of detectors ($N_{Ab} = |AB| > N_{Ab\max}$), the excess of the limit of search time ($t > t_{\max}$), etc.

The set of detectors $AB_1 = \{Ab_1, Ab_2, \dots, Ab_{N_{Ab}}\}$ was generated by using negative selection. This set describes the region of the search space \bar{S}_1 . The space \bar{S}_1 is complementary to the region of space in which there is a set of "self" instances S_1 . The set $AB_1 = \{Ab_1, Ab_2, \dots, Ab_{N_{Ab}}\}$ is characterized by high approximation and generalization capability.

Similarly, a set of detectors AB_0 for the set S_0 can be created. However, there are tasks with the uneven distribution of class instances in set $S = \langle P, T \rangle$. Problems with the creation of detectors, which adequately reflect the space of instances S_0 , may occur in these tasks. In particular, the detectors Ab_k in a form of hypercube with too large volume can be generated. These detectors cannot be able to summarize the data adequately. This is due to the insufficient number of instances in the sample $S_0(Q_0 \ll Q_1)$.

Therefore, we propose to calculate information on the size of the detectors which built on the basis of sample S_1 . This information will also be used for generation of detectors for instances S_0 .

These detectors will show information about the presence of instances of the sample S_0 in the hypercube. Information about their absence is not shown in this case. It is an essential difference of the proposed method from the approach based on classic negative selection. These detectors will be fully consistent with

the detectors built previously for the sample S_1 based on negative selection. These detectors will contain information about areas of the search space in which items S_1 are not arranged.

Detectors $Ab_k^{(0)}$ of sample S_0 are generated so that their centers correspond to the coordinates of instances $s_k = \langle p_{km}, t_k = t'_0 \rangle \in S_0$ of sample S_0 . As the metric for estimation of the coordinates of detectors $Ab_k^{(0)}$ we used average detector length that can be considered as normalized Manhattan distance. The size of their hypercube edges correspond to the same sizes of detectors created on the basis of the sample S_1 .

Consequently, the coordinates of the detector $Ab_{km}^{(0)} = \langle Ab_{km\min}^{(0)}, Ab_{km\max}^{(0)} \rangle$ are determined using (8) and (9):

$$Ab_{km\min}^{(0)} = p_{km} - \frac{1}{2} \Delta Ab_m, \quad (8)$$

$$Ab_{km\max}^{(0)} = p_{km} + \frac{1}{2} \Delta Ab_m, \quad (9)$$

where ΔAb_m is an average length of the edges of detectors $AB_1 = \{Ab_1, Ab_2, \dots, Ab_{N_{Ab}}\}$ which were created on the basis of the set S_1 . Value ΔAb_m was calculated on the basis of the information about detectors $AB_1 = \{Ab_1, Ab_2, \dots, Ab_{N_{Ab}}\}$ by (10):

$$\Delta Ab = \frac{1}{N_{AB} M} \left(\sum_{k=1}^{N_{Ab}} \sum_{m=1}^M (Ab_{km\max} - Ab_{km\min}) \right). \quad (10)$$

Then, the comparison of the generated detectors $Ab_k^{(0)}$ to the instances of sample S_1 is performed by (4). If the condition (5) is true then the detectors $Ab_k^{(0)}$ are converted similarly to the above stage of additional training. Then the value of one of criteria $G(Ab_k^{(0)})$ is calculated. This criteria estimates the ability of detector to generalize the data. If its value is above the threshold, the detector $Ab_k^{(0)}$ is added to the set $AB_0 = AB_0 \cup \{Ab_k^{(0)}\}$.

Thus the set of detectors AB_0 is generated. Both set AB_0 and set AB_1 describe the area of the search space \bar{S}_1 . This space is complementary to the region of arranging instances S_1 . Therefore, a recognizing model can be represented as a set of detectors $AB = AB_0 \cup AB_1$. These detectors permit to recognize membership of unknown instances $s'_q = \langle p'_{qm}, t'_q \rangle \notin S$ to the class of "non-self", i.e. to refer them to the class $t'_0 : t'_q = t'_0$.

Lets us improve the level of interoperability of the synthesized recognition model which was presented as a set of detectors $AB = \{Ab_1, Ab_2, \dots, Ab_{N_{Ab}}\}$. A set of classification rules $PR_r : P_r \rightarrow T_r$ based on a set AB is formed for this goal. Note that the left part P_r of the implication is a set of conditions (11):

$$\text{If } (p_1 \in [Ab_{k1\min}; Ab_{k1\max}]) \wedge (p_2 \in [Ab_{k2\min}; Ab_{k2\max}]) \wedge \dots \wedge (p_M \in [Ab_{kM\min}; Ab_{kM\max}]) \quad (11)$$

The right part T_r contains the value of the output parameter T when performing an r -th set of conditions P_r (11).

When generating a set of rules PR to the rules P_r antecedently, we will include only the boundaries of

features for which they are not the limit values, i.e., $Ab_{kmin} \neq \min_{q=1,2,\dots,Q} (p_{qm})$ and $Ab_{kmax} \neq \max_{q=1,2,\dots,Q} (p_{qm})$. For example, the rule PR_k is created for detector $Ab_k = \{ \langle 5, 7 \rangle, \langle 8, p_{2max} \rangle, \langle p_{3min}, p_{3max} \rangle, \langle 4, 6 \rangle \}$. This rule has the following structure:

If $(p_1 > 5 \wedge p_1 < 7) \wedge (p_2 > 8) \wedge (p_4 > 4 \wedge p_4 < 6)$ then the instance corresponds to the class of "non-self" ($T \neq t_1$).

The upper limit of feature p_2 and feature p_3 has not been included in the rule explicitly. Because the corresponding values of the detector do not affect the quality of recognition. In addition, the exclusion of such values from the rule PR_k reduces its complexity. Interoperability and comprehensibility of the rule will be increased from the other side. So, classification rules based on each detector Ab_k are constructed. So, a set PR of N_{Ab} classification rules $PR_r : P_r \rightarrow T_r$ has been created with the using of proposed approach.

The proposed method for the classification rules synthesis is based on negative selection approach. This method is oriented to the case of uneven distribution of class instances of sample in generating a set of detectors. The proposed method uses known information about instances of all classes of the sample. It also takes into account information about the individual significance of features. A hypercube of maximum possible volume is used as a form of detector. It allows to exclude irrelevant and redundant features from the sample, thereby reducing the search space and time of the method implementation. As a result, a set of detectors with high approximation and generalization capability is formed.

The proposed method increases the generalizing properties of the synthesized model by reducing the number of detectors and conditions of antecedents. This method improves interoperability of the model, reduces its dimension (structural and parametric complexity) and volume of the used memory. All of these improvements increase the model performance with sequential computation.

4. Experiments and results

A computer program has been developed for implementation of the proposed method of classification rule synthesis based on negative selection. This software is oriented to the verification and analysis of different characteristics of this method. This software deals with a blade diagnosis of an aircraft engine gas turbine [23]. The blades of gas turbine were characterized by the values of the power spectra of damped oscillations after impact excitation. These values of the the power spectra are used as input features. Classes of blade quality were defined with the help of experts: undamaged and defective (potentially dangerous). Each blade was described by 10240 characteristics of the power spectrum of damped oscillations. Artificial features were constructed to reduce

the search space based on these characteristics. A set consisting of 80 artificial features was obtained based on this reducing.

The resulting sample $S = \langle P, T \rangle$ does not have statistical representativeness, because it does not display the actual frequency distribution of classes. Really, the number of undamaged blades is substantially greater in the general population than the number of defective blades. These defective blades ($t_q = t'_1$) in the sample represent typical cases of nonconformity, which provides a topological representation of defective blades in the sample. All the possible cases of the class of undamaged blades ($t_q = t'_0$) cannot be present in a sample from a practical point of view. Therefore, it is necessary to build a diagnostic model for aircraft engine blade class recognition based on the available sample $S = \langle P, T \rangle$ with uneven distribution of instances of classes.

The sample $S = \langle P, T \rangle$ contains 42 instances characterizing defective (potentially dangerous) blades and 72 instances representing undamaged blades. The proposed method for the synthesis of classification rules was compared with the existing methods of negative selection. These methods synthesized a set of detectors based on "self" instances $S_1 \subseteq S$ of the sample only. Therefore, the problem of blade diagnosis of gas turbine of aircraft engines was solved with the proposed method two times:

- using a sub-sample $S_1 \subseteq S$, which contains information about defective (potentially dangerous) instances ("self") only;
- using all the original samples $S = \langle P, T \rangle$.

Experimental investigation of characteristics of the proposed method has been compared with other methods of negative selection.

The first part of experimental results is given in Table 1. This table contains next values. Column N_{it} describes a number of iterations of the method. Column t contains calculation time. Misclassification error on the training data $S = \langle P, T \rangle$ indicated at column E . Column E_t contains misclassification error of the test data. Columns $P_{t,t_q=i'_1/t_q=i'_0}$ and $P_{t,t_q=i'_0/t_q=i'_1}$ indicated probability of misclassification. So, value $P_{t,t_q=i'_1/t_q=i'_0}$ is error probability of assignment to class "self" ($t_q = t'_1$) when the instance actually belongs to a class of «non-self» ($t_q = t'_0$). Similarly, value $P_{t,t_q=i'_0/t_q=i'_1}$ is error probability of assignment to class "non-self" ($t_q = t'_0$) when an instance actually belongs to a class of «self» ($t_q = t'_1$). These probabilities are calculated using (12) and (13) respectively:

$$P_{t,t_q=i'_1/t_q=i'_0} = \frac{N_{t,t_q=i'_1/t_q=i'_0}}{N_{t,t_q=i'_0}}, \quad (12)$$

$$P_{t,t_q=i'_0/t_q=i'_1} = \frac{N_{t,t_q=i'_0/t_q=i'_1}}{N_{t,t_q=i'_1}}. \quad (13)$$

These expressions contain next variables. Variable $N_{t,t_q=i'_1/t_q=i'_0}$ is a number of instances of the test sample recognized as "self" ($t_q = t'_1$), but actually belonging to the class of «non-self»

($t_q = t'_0$). Variable $N_{t_q=t'_0}$ is a number of instances of the test sample belonging to the class of “non-self” ($t_q = t'_0$). Variable $N_{t_q=t'_0/t_q=t'_1}$ is a number of instances of the test sample recognized as “non-self” ($t_q = t'_0$), but actually belonging to the class of «self» ($t_q = t'_1$). Variable $N_{t_q=t'_1}$ is a number of instances of the test sample belonging to the class of “self” ($t_q = t'_1$). Test sample contains 273 instances. The subset of 261 instances from them belongs to the class $t_q = t'_0$ (undamaged products) and the subset of next 12 instances belongs to the class $t_q = t'_1$ (defective instances).

The second part of experimental investigation compared the proposed method with other classification methods. The above-described problem of blade diagnosis of a gas turbine of aircraft engines [23] was solved by these methods. Several different models of this problem were obtained based on these methods. We analyzed the next classification models:

- a model in form of classification rules synthesized by the proposed method MPRSBNS;
- a feed-forward neural network which has been trained by using error back-propagation. This network consists of three layers of neurons. The first layer of the neural network has five neurons, the second has three neurons and the third layer has one neuron only. Neurons in the first and the second layers have got a logistic sigmoid activation function. In the third layer a single neuron has got a threshold activation function;
- a model in form a set of detectors constructed by the method MMD [16].

The whole training set of 114 instances was used for experiment investigation of the first and the second models. This training set included (a) a subset of 42 instances characterizing defective blades and (b) a subset of 72 instances representing

undamaged blades. Part of the sample ($S_1 \subseteq S$) was used for constructing the third model based on a set of detectors using the method MMD. We had to do it because this method deals with instances of one class only.

Some criteria of these models have been calculated for analysis. The results of calculation are given in Table 2. We analyzed the next criteria of such classification models:

- N_{param} is a criterion determining the parametric complexity of the model. This criterion is calculated as the number of model parameters. In particular, the total number of parameters $Ab_{km\ min}$ and $Ab_{km\ max}$ for the first and the third models; and the total number of adjustable parameters (weight coefficients) for the second model;
- criteria E , E_t , $P_{t_q=t'_1/t_q=t'_0}$ and $P_{t_q=t'_0/t_q=t'_1}$ were described in Table 1.

5. Discussion

Table 1 shows that the misclassification error values E produced by the method MMD [16] ($E=0.018$) and by the proposed method MPRSBNS ($E = 0.026$ and $E = 0.009$) are acceptable. The low recognition errors of these methods were provided by the wide coverage of field of “self” instances $S_1 \subseteq S$ by synthesized detectors. The proposed method MPRSBNS synthesized a set of detectors based on instances of all classes of the sample $S = \langle P, T \rangle$. This method provided more acceptable results ($E = 0.009$) compared to a set of detectors synthesized using “self” instances $S_1 \subseteq S$ ($E = 0.026$) only. Method RNS [13] and model V-Detector [14 and 15] had less acceptable misclassification error value E ($E = 0.070$ and $E = 0.035$, respectively). This fact indicates that synthesized detectors do not cover the area of “self” instances

Results of the first part of experiments

Table 1

Method	N_{Ab}	N_g	t , ms	E	E_t	$P_{t_q=t'_1/t_q=t'_0}$	$P_{t_q=t'_0/t_q=t'_1}$
Real-Valued Negative Selection (RNS) [13]	207	50	27.3	0.070	0.136	0.126	0.333
Model V-Detector (MVD) [14, 15]	41	50	24.1	0.035	0.077	0.069	0.250
Method with masking of detectors (MMD) [16]	19	14	13.2	0.018	0.055	0.054	0.083
Method of classification rules synthesis on the basis of negative selection MPRSBNS (using sample $S_1 \subseteq S$)	20	12	12.1	0.026	0.037	0.038	0
Method of classification rules synthesis on the basis of negative selection MPRSBNS (using sample $S = \langle P, T \rangle$)	31	19	13.7	0.009	0.011	0.011	0

Results of the second experimental part. Comparison of different classification models

Table 2

Model	N_{param}	E	E_t	$P_{t_q=t'_1/t_q=t'_0}$	$P_{t_q=t'_0/t_q=t'_1}$
Model as a set of classification rules synthesized by the proposed method MPRSBNS	652	0.009	0.011	0.011	0
Feed-forward neural network	427	0.018	0.070	0.065	0.167
Model of a set of detectors constructed by the method with masking of detectors MMD [16]	804	0.018	0.055	0.054	0.083

$S_1 \subseteq S$. The experimental results show that the method RNS [13] and the model V-Detector [14 and 15] generate the largest number of detectors ($N_{Ab} = 207$ and $N_{Ab} = 41$, respectively). It increases the learning time t and computational costs of computer resources. The time of calculations is not so important criterion in the experiment. Accuracy of prediction (criteria E_p , $P_{t,t_q=t'_1/t_q=t'_0}$ and $P_{t,t_q=t'_0/t_q=t'_1}$) is more important. Method MMD [16] and the proposed method MPRSBNS (using sample $S_1 \subseteq S$) generated significantly fewer number of detectors ($N_{Ab} = 19$ and $N_{Ab} = 20$, respectively). It indicates a more efficient operation of these methods. In particular, the method MPRSBNS uses a priori information about the significance of features at the initial stage. This method eliminates from further consideration irrelevant and redundant features that can reduce the search space and create a set of a small number of detectors based on highly informative features with a high approximation and generalization capability.

Criteria E_p , $P_{t,t_q=t'_1/t_q=t'_0}$ and $P_{t,t_q=t'_0/t_q=t'_1}$ were used for the analyzing of properties and characteristics of the investigated methods. These criteria describe misclassification error and the probability of making a wrong decision based on test data. Misclassification errors of models synthesized by the proposed method MPRSBNS and methods [13-16] are shown in Table 1. These errors have been calculated on test data E_t . Misclassification error of the proposed method MPRSBNS is significantly lower than the error of other known methods ($E_t = 0.136$, $E_t = 0.077$ and $E_t = 0.055$ for the methods [13, 15 and 16], respectively). It can be explained by using the characteristics $G(Ab_k)$. These characteristics allow to evaluate the ability of the detector for the generalization of data. The proposed method MPRSBNS allowed to reach misclassification error $E_t = 0.037$ (using a part of the sample $S_1 \subseteq S$) and $E_t = 0.011$ (using the full sample $S = \langle P, T \rangle$).

It is important to note the specificity of the solved problem of blade diagnosis. An error of assignment to “non-self” class ($t_q = t'_0$) has a very high cost when the instance actually belongs to a «self» class ($t_q = t'_1$). This error has been evaluated by criterion $P_{t,t_q=t'_0/t_q=t'_1}$. This is due to the fact that the classification of defective blades to the class of undamaged can cost human lives. The test data has zero error probability $P_{t,t_q=t'_0/t_q=t'_1}$ for the proposed method MPRSBNS (see Table 1). This fact indicates high efficiency of the proposed method for solving such problems. The zero level of error probability $P_{t,t_q=t'_0/t_q=t'_1}$ by using the proposed method is explained by:

- a high level of coverage of typical instances of class $t_q = t'_1$. This coverage was made by generated set of detectors $AB = \{Ab_1, Ab_2, \dots, Ab_{N_{Ab}}\}$. Note that this set of detectors was obtained using a priori information about the importance of features;
- a high generalizing ability of synthesized set of detectors. That is caused by the use of the criteria (6) and (7). These criteria allow to estimate ability of the detector to the data generalization.

The number parameters N_{param} of the model synthesized by the proposed method ($N_{param} = 652$) is less than a similar model constructed by the method MMD [16] ($N_{param} = 804$) (see Table 2). This is due to the fact that when using the proposed method, the average size of the generated detector is less. Such reduction can be explained that a priori information of feature significance is used during the process of negative selection. It allows to exclude from further consideration irrelevant and redundant features that complicate the process of synthesis of diagnostic models and reduce their interpretability. Thus, the model synthesized by the proposed method MPRSBNS is more simple and straightforward compared to the model created by the method of [16]. Approximation and generalization capabilities of the model synthesized by the method MPRSBNS are also higher. This fact is confirmed by the values of the criteria E , E_t , $P_{t,t_q=t'_1/t_q=t'_0}$ and $P_{t,t_q=t'_0/t_q=t'_1}$.

The comparison of the model synthesized on the basis of MPRSBNS method and the neural network model allows next to conclude. The model constructed by the proposed method has higher generalizing and approximation abilities (criteria E , E_t , $P_{t,t_q=t'_1/t_q=t'_0}$ and $P_{t,t_q=t'_0/t_q=t'_1}$). However, the number of model parameters constructed by the method MPRSBNS ($N_{param} = 652$) is greater than in the neural network model ($N_{param} = 427$). It can be explained by representation of neural network as a set of neurons interconnected in a certain way and characterized by weighting coefficients as adjustable parameters. And each neuron corresponds to a function of many arguments. At the same time, this neural network model is difficult enough for human perception. The model as a set of classification rules synthesized by the proposed method is more intuitive in comparison with the neural network model. Really, classification rules of the form “if condition, then action” are much more understandable and human-readable than a set of coefficients that reflect the degree of neuronal connections in the neural network model.

Thus, the results of experiments showed that the proposed method due to the usage of a priori information and exclusion of irrelevant and redundant features of the sample makes it possible to reduce the search space and time of execution. Proposed method allows to synthesize classification models in a form of a set of detectors with high approximation and generalization capabilities. Also by reducing the number of detectors and the conditions in antecedents it increases interpretability of the model, reduces its dimension and, therefore, the size of the used memory.

6. Conclusions

In this paper we solve the problem of automation of classification rule synthesis based on negative selection for the case of uneven class distribution in the sample.

The developed method of classification rule synthesis based on negative selection uses a priori information about instances of all classes in the sample at detector set generation. It also takes into account information about the individual feature significance. A hypercube of maximum possible volume is used as a form of detector. It allows to exclude irrelevant and redundant features from the sample, thereby reducing the search space and time of execution of the method, as well as to generate a set of detectors with high approximation and generalization capability. The proposed method improves interoperability of the model,

reduces its dimension, size of used memory and improves the model performance for the sequential computation. It is obtained by increasing the generalizing properties of synthesized model by reducing the number of detectors and conditions of antecedents.

An experimental study of the proposed method and its comparison with the known analogues is performed. A practical task of diagnosing the vanes of gas turbine of aircraft engines has been solved. The mathematical approach proposed at [24 and 25] can be used for reliability analysis of the proposed solution.

References

- [1] SOBHANI-TEHRANI, E., KHORASANI, K.: *Fault Diagnosis of Nonlinear Systems using a Hybrid Approach*. Springer, 2009.
- [2] DENTON, T.: *Advanced Automotive Fault Diagnosis*. Elsevier, 2006.
- [3] GERTLER, J.: *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker Inc., 1998.
- [4] *Encyclopedia of Artificial Intelligence*. DOPICO, J. R., CALLE, J. D., SIERRA, A. P. (Eds). Information Science Reference, 2009.
- [5] *Encyclopedia of Machine Learning*. SAMMUT, C., WEBB, G.I. (Eds). Springer, 2011.
- [6] MUMFORD, C. L., JAIN, L. C.: *Fault Detection and Diagnosis in Engineering Systems*. Springer-Verlag, 2009.
- [7] MO, C. L.: *Handbook of Research on Artificial Immune Systems and Natural Computing: Applying Complex Adaptive Technologies*. Information Science Reference, 2009.
- [8] TARAKANOV, A., SKORMIN, V., SOKOLOVA, S.: *Immunocomputing: Principles and Applications*. Springer, 2013.
- [9] DASGUPTA D., NINO, F.: *Immunological Computation: Theory and Applications*. Auerbach Publications, 2008.
- [10] JI, Z., DASGUPTA, D.: Revisiting Negative Selection Algorithms. *Evolutionary Computation*, 2007, 15: 223-251.
- [11] ZENG, J., LI, T., LIU, X.: *A Feedback Negative Selection Algorithm to Anomaly Detection*. Proc. of Natural Computation, Haikou, 2007: 604-608.
- [12] ZHANG, P. T., WANG, W., TAN, Y.: A Malware Detection Model Based on a Negative Selection Algorithm with Penalty Factor. *Science China. Information sciences*, 2010, 53(12), 2461-2471.
- [13] OLIVEIRA, L. O., DRUMMOND, I. N.: Real-valued Negative Selection (RNS) for Classification Task. *Recognizing Patterns in Signals, Speech, Images and Videos*, 2010, 6388, 66-74.
- [14] JI, Z.: *A Boundary-aware Negative Selection Algorithm*. Proc. of Intern. Conference on Intelligence and Soft Computing, Benidorm, 2005, 49-55.
- [15] JI, Z.: *Negative Selection Algorithms: From the Thymus to V-detector*. doctor of philosophy in "Computer Science" thesis, Memphis, 2006.
- [16] ZAITSEV, S., SUBBOTIN, S.: *Negative Selection using Masked Detectors*. Proc. of seminar on Modelling of non-equilibrium systems, Krasnoyarsk, 2011, 95-98.
- [17] SUBBOTIN, S., OLEYNIK, A.: *Entropy Based Evolutionary Search for Feature Selection*. Proc. of Intern. Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, Lviv, 2007, 442-443.
- [18] OLIINYK, A., SUBBOTIN, S.: The Decision Tree Construction Based on a Stochastic Search for the Neuro-fuzzy Network Synthesis. *Optical Memory and Neural Networks (Information Optics)*, 2015, 24(1): 18-27.
- [19] OLIINYK, A., ZAIKO, T., SUBBOTIN, S.: Training Sample Reduction Based on Association Rules for Neuro-Fuzzy Networks Synthesis. *Optical Memory and Neural Networks (Information Optics)*, 2014, 23(2): 89-95.
- [20] OLIINYK, A., SKRUPSKY, S., SUBBOTIN, S.: Using Parallel Random Search to Train Fuzzy Neural Networks. *Automatic Control and Computer Sciences*, 2014, 48(6): 313-323.
- [21] OLIINYK, A., SKRUPSKY, S., SUBBOTIN, S.: Experimental Investigation with Analyzing the Training Method Complexity of Neuro-Fuzzy Networks Based on Parallel Random Search. *Automatic Control and Computer Sciences*, 2015, 49(1): 11-20.
- [22] BODYANSKIY, YE., VYNOKUROVA, O.: Hybrid Adaptive Wavelet-neuro-fuzzy System for Chaotic Time Series Identification. *Information Sciences*, 2013, 220: 170-179.
- [23] DUBROVIN, V., SUBBOTIN, S., BOGUSLAYEV, A., YATSENKO, V.: *Intelligent Diagnosis and Prediction of Reliability of Aircraft Engines (in Russian)*. Motor-Sich, 2003.

- [24] ZAITSEVA, E., KVASSAY, M., LEVASHENKO, V., KOSTOLNY, J.: Reliability Analysis of Logic Network with Multiple Outputs, *Communications - Scientific Letters of the University of Zilina*, 2015, vol. 1A: 44-50.
- [25] KVASSAY, M., KOSTOLNY, J., Evaluation of Algorithms for Identification of Minimal Cut Vectors and Minimal Path Vectors in Multi-State Systems, *Communications - Scientific Letters of the University of Zilina*, 2015, vol. 17, No. 4: 8-14.