

Clustering data using the modified artificial immune network

¹Lytvynenko V.I., ²Bidyuk P.I., ¹Bardachev Ju.M., ¹Didyk A.A., ¹Fefelov A.A., ¹Rogalskiy F.B.

¹*Kherson National Technical University, 73008 Kherson, 24 Beryslavske Shosse*
²*National Technical University of Ukraine "KPI", 03056 Kyiv, 37 Prosp. Peremogy*

johndoe2004@mail.ru, pbidyuke@gmail.com, fao1976@mail.ru, fbrdom@mail.ru

Abstract. *The new version of artificial immune system for solution of automatic data clustering is presented. The algorithm uses properties of self-organizing of immune system and creates a stable immune network.*

Keywords

Inductive modeling, self-organizing, artificial immune system, data clustering

1 Introduction

The natural immune systems possess a number of attractive properties from the point of view of information processing, one of which is the ability of self-organizing [1]. Self-organizing is a process in which the results, are obtained, reproduced or improved on the basis of immanent properties of a system structure and functions independently without external operating influences [2]. Self-organizing principles have been an object of research for many outstanding scientists: J. Neumann, N. Wiener, At. R. Ashbi, and many others. A substantial contribution to development of this scientific direction was brought by researchers of Ukrainian Cybernetic Center under supervision of Acad. G. Ivakhnenko [3, 4, 5]. They developed the whole class of adaptive self-organizing models, that could be classified as "intellectual generalization" of empirical-and-statistical methods. Self-organizing in immune system is understood as an irreversible process occurring under the influence of an environment [3], and resulting in co-operative interaction of a set of elements (subsystems) to form more effective (from a system point of view) structures. Today are known two types of self-organizing biological systems – nervous and immune which possess possibilities of "intellectual" information processing, including memory, ability to be trained, distinguish and make decisions concerning unknown situations. The potential of natural nervous system as a biological prototype of various computing schemes have already been well developed in the form of a paradigm of artificial neural networks and neurocomputing. The computing possibilities of a natural immune system have got a worthy estimation only recently that led to creation of a new direction – artificial immune systems (AIS). As the latest studies show AIS possess a high potential regarding computing and form a basis for immunocomputing – the new approach based on the principles of information processing by immune system.

2 The problem of data clustering by means of artificial immune networks (motivation)

At the heart of clustering algorithms lays the self-organizing feature of artificial immune networks. They do not usually contain explicitly expressed criterion function, and as a criterion for cluster recognition are used generated sets of detectors only. Today there are most known two systems of data analysis based on an immune metaphor: aiNet [6] and AINE [7]. Both systems use a training set as antigens, and its purpose is creation of a set of B-cells or the antibodies representing these antigens. According to N. Erne's hypothesis algorithm AINE performs synthesis of a network (copies of idiopathic networks) which cover the basic properties of elements of data from a training sample. This system uses almost all mechanisms of the idiopathic networks theory: a) each B-cell is capable to recognize some antigens; b) B-cells that are similar combine an immune network; c) all immune cells are exposed to a hypermutation and clonal selection; d) it is possible to present B-cells in the form of artificial recognizing spheres or ARB (artificial recognition balls). Application of ARB formalism is caused by aspiration to improve robustness of an immune network. In the

given kind of immune network are used only four operating parameters: 1) a threshold of network affinity; 2) mutation rate; 3) number of ARB, and 4) number of created clones that activate ARB. The influence of these parameters on the quality of resulting networks is investigated in work [8]. The immune network aiNet represents the system which uses more simplified representation. In a given AIS instead of B-cells or ARB a population of the antibodies presented by real-valued, integer, binary or symbolic vectors values is used. In the given work the algorithm aiNet [6] is taken as a basis that can be explained by two reasons: a) it is one of the most known methods in the literature on clustering artificial immune systems; b) its efficiency is shown by solving of a considerable quantity of challenges. In the given algorithm the clonal selection operates the dynamics and metadynamics of a network. Population of antibodies originally created in a random way, is changed by means of clonal selection operators, hypermutation and apoptosis (removal of not stimulated cells). A drawback of the algorithm is in the considerable number of parameters to be defined by user, and also the necessity of using the standard clustering and analysis tools such as algorithms of constructing *the minimal spanning tree* (MST). An advantage of the algorithm is in the possibility of compressing data which in some cases reaches 90 % [10].

There is a certain class of problems where ability of immune networks to information compression could be a negative feature. For these problems it is possible to allocate the following general indications: a) the clusters are located too close to each other; b) the density of data points arrangement varies from cluster to cluster; c) the borders clusters have fuzzy or non-uniform character. They are blocked and consequently distort the data structure, caused by their compression; creation of "an internal image" networks can deform the key properties of data necessary for correct identification clusters. As a result of solving the problems the relative distances between clusters in an internal image will not correspond to relative distances between starting points of data that can complicate considerably an application of clustering algorithm. Our research experiments touching the given class of problems by means of Kohonen Clustering Networks [10], have shown that there is the same problem. In work [10] it is shown, that U-matrix in which the criterion of distances for distribution of images over clusters is used, it is not capable to solve similar problems. In the given work it has been shown, that use of information representation of density of data in a combination with U-matrix considerably improves the quality of clustering. In [14] the similar practical problems are described in details. At the moment only one work [11] is known where the problem described is solved by the means of immune networks. The authors have offered an algorithm of an immune network with adaptive selection of coverage radius which is capable to keep as much as possible information on density, selecting values of radius in inverse proportion to local density of data. Therefore compression on the basis of given values of immune network with high density of data become covered by antibodies of smaller radius that are located closer to each other. The areas with low density of antibodies have larger radius and are not supposed to be close to each other within compression procedure. The goal of this research is creation of a new adaptive clustering algorithm for analyzing the data based on metaphors and principles of artificial immune system which uses mechanisms the clonal expansions and suppression of immune network, and is capable to create an internal image with as much as possible full preservation of the information on density of data that will allow to solve effectively the class of earlier mentioned difficult clustering problems.

3 The proposed modified algorithm for solving the clustering problem, based on the artificial immune network

We will define an artificial immune network as a full connected graph consisting of set of nodes – antibodies (vectors of numerical values), antigens (vectors of numerical values of data) and set of the weighed edges (values of affinities) which establish communications between antigens and antibodies [7,8]. Thus affinity is defined as similarity or distinction degree between corresponding attributes of symbol lines, such that $S^P \times S^P \rightarrow \mathfrak{R}^+$.

Conceptually it is possible to present any algorithm of an artificial immune network in the form of the scheme (Fig. 1):

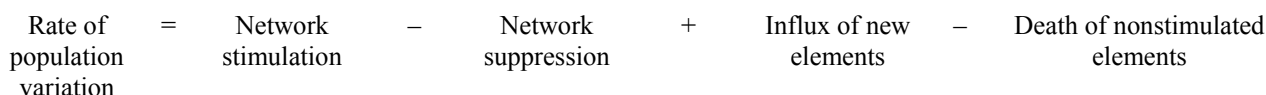


Fig. 1. The generalized algorithm scheme of an immune network

The developed immune algorithm clustering uses material coding of antibodies (fig. 2), where Ab_{ij} are coordinates of an antibody center in Euclidean space. Unlike classical algorithm of an immune network, in this case in a

line of an antibody the parameter r - the radius of an antibody is used, which is carrying out cross-reactive function of a threshold which that is adaptive steals up in the process of algorithm work.

r	Ab_{i1}	Ab_{i2}	\dots	Ab_{il}
-----	-----------	-----------	---------	-----------

Fig. 2. Representation of an antibody of immune algorithm clustering

Formal representation of the offered algorithm of an immune network

$$Modif_{immNET}(Ab, Ag, N_C, Gen, D, d, d_M, S, n, m_R, n_C, \zeta, \sigma_S, r, f_{Ab-Ag}, f_{Ab-Ab}, nAg_{rec}, l, H, D_E, k_r) \quad (5)$$

where Ab is a population of antibodies; Ag is a dataset consisting N_p vectors of p dimension; N_C – total number of clones created by stimulated cells in each generation (at network activation); Gen – Quantity of generations; D – matrix of elements d_{ij} of $Ag-Ab$ affinities; d is replacement level (quantity of antibodies with the minimum affinity, selected for replacement by new antibodies); d_M is threshold value for clone deleting; S is a matrix of elements s_{ij} with $Ab-Ab$ affinities; n is selection level (quantity of antibodies Ab with the maximum affinity, selected for cloning); m_R , is mutation level; n_C is a level of selection from population of clones; ζ is a percent of improved cells selected from population of clones for the subsequent processing; σ_S is a threshold value suppression; r is antibody radius; f_{Ab-Ag} is affinity between an antibody and an antigen; f_{Ab-Ab} is affinity between antibodies; nAg_{rec} is a number of recognitions given by an antibody; l is a dimension of data space; H is a rate of distribution dimension of antigens in the space limited by a radius of an antibody; D_E is Euclidean distance between the centre of an antibody and the average centre of all antigens recognized by this antibody; k_r is a parameter of the importance of radius of an antibody.

Step-by-step implementation of the algorithm is presented in a Fig. 3. On **Step 1** the initial set of population of antibodies is generated. On **the Step 2** the value of affinity of communication "antigen-antibody" is calculated using (6):

$$f_{Ab-Ag} = \frac{k_r}{r} + \frac{r^2}{H \cdot nAg_{rec}} \cdot D_{E(Ab-Ag_{rec})}, \quad (6)$$

where k_r is parameter of radius importance of an antibody, i.e. the value of cross-reaction of a threshold (the increase in value of this parameter forces an immune network to support antibodies large radius, that gives more rough clustering); r is antibody radius (the antigens located in the given area are considered as recognized as an antibody); nAg_{rec} - Quantity of the antigens distinguished by a given antibody; $D_{E(Ab-Ag_{rec})}$ is Euclidean distance between the centre of an antibody and the average centre of all antigens recognized by the antibody; H is an indicator of uniformity of distribution of antigens in the field of the space, limited by the antibody radius.

The indicator is calculated H as follows:

$$H = \sum_{i=1}^p H_i, \quad (7)$$

where p is a dimension of space of data (length of a receptor of an antibody); H_i is an indicator of uniformity of distribution of the recognized antigens along i -th coordinate axis of data space.

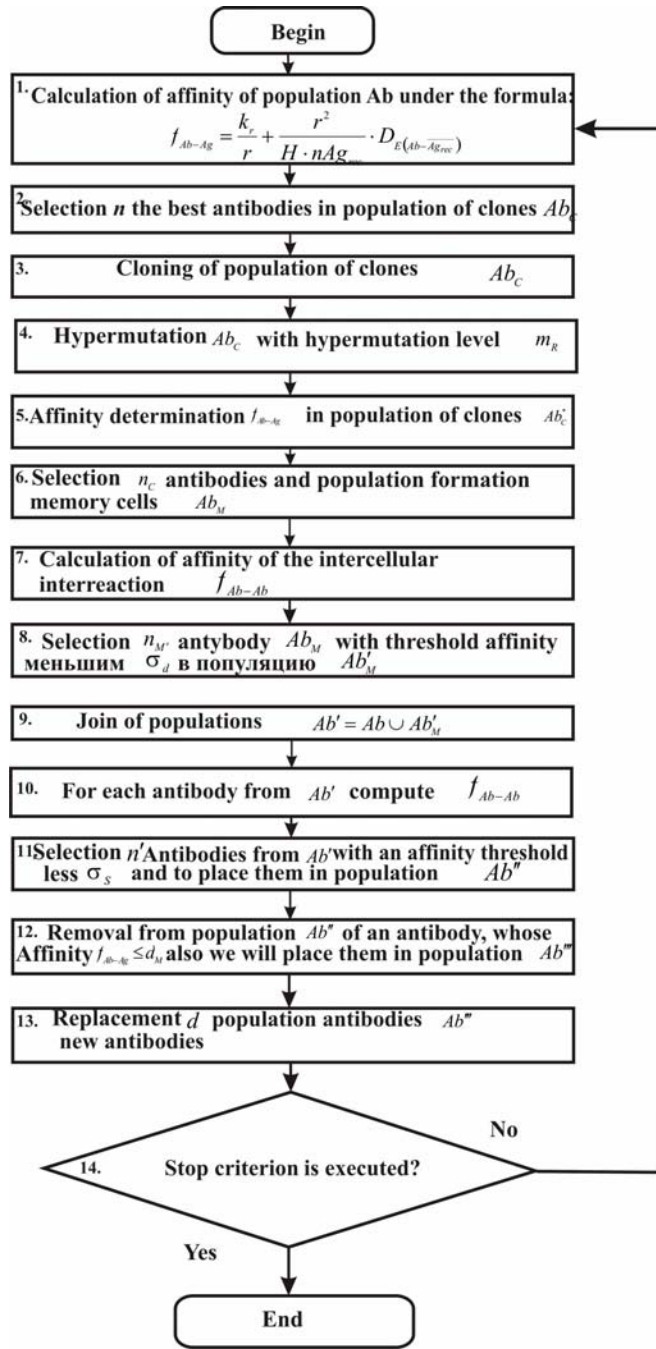


Fig. 3. Step-by-step implementation of the algorithm

$$H_i = - \sum_{j=1}^{nAg_{rec}} p_j^{Ag_{rec}} \cdot \log p_j^{Ag_{rec}}, \quad (8)$$

where $p_j^{Ag_{rec}}$ is a probability of joining of the recognized antigen into j -th interval of i -th coordinate axis of data space (within the limits of the area limited by the radius of an antibody); $p_j^{Ag_{rec}} = \frac{nAg_{rec}^j}{nAg_{rec}}$ nAg_{rec}^j is a quantity of the recognized antigens which have got into j -th interval of i -th coordinate axis of data space.

The antibodies produced are ranked and at **STEP 2** are selected the best n antibodies from which the population Ab_C is created. Further, on **the Step 3** the cloning of antibodies, Ab_C , in population of clones is carried out in conformity to their affinity using (9):

$$N_c = \sum_{i=1}^n \text{round}\left(\frac{\beta * N}{i}\right), \quad (9)$$

where N_c is a total number of clones in the population of clones; β is multiplication factor; N is a total number of antibodies in population; *round* – the operator that approximates argument to the nearest integer.

Composed sum corresponds to the size of a clone of each selected antibody, for example, for $N = 100$ and $\beta = 1$, the antibody with the highest affinity ($i = 1$) creates 100 clones, while the second antibody on affinity creates 50 clones, etc. On **STEP 4** the mutation of antibodies is performed proportionally to their affinity; thus the set is created, Ab_C , exposed to the directed process of maturing of affinity (an operated mutation) generating modified set, where each antibody k from Ab_C^* is exposed to a mutation with level in inverse proportion to m_{Rk} , the value of affinity of its $f_{Ab-Ag_{i,j}}$ parental antibody:

$$Ab_C^* = Ab_{Ck} + m_{Rk} (Ag_j - Ab_{Ck}); \quad m_{Rk} \sim 1/f_{Ab-Ag_{i,j}}; \quad k = 1, \dots, N_c; \quad i = 1, \dots, N. \quad (10)$$

On **STEP 5** the affinity of antibodies f_{Ab-Ag} from the population Ab_C^* is calculated according to (6) and creation of new population of clones is performed $f_{Ab-Ag} Ab_C^* Ab_C^{*'}$. On **STEP 6** ranging of antibodies on level of their affinity from the created population of clones and $Ab_C^{*'}$ selection of the best n_c antibodies for forming of them the population of memory cells is carried out $Ab_C^{*'}$ n_c Ab_M . On **STEP 7** the affinity calculation of inter-cells interactions is carried out f_{Ab-Ab} according to (11):

$$f_{Ab-Ab} = -\frac{D_{E(Ab_1-Ab_2)} - (r_{Ab_1} + r_{Ab_2})}{2 \cdot \min(r_{Ab_1}, r_{Ab_2})}. \quad (11)$$

The following interpretation of values f_{Ab-Ab} is thus possible: if it is ≤ 0 , then recognizing areas of antibodies are not blocked; $(0, 1)$ the areas are blocked by coverings, and the value is overlapping degree; ≥ 1 – the area of smaller radius (r) completely is in the area of larger radius. On **step 8** the selection of antibodies n_M from population according to Ab_M affinity threshold is carried out σ_d , the selected antibodies are located in population Ab'_M .

At **Step 9** a combination of initial population of antibodies and the Ab population created on the previous step Ab'_M ($Ab' = Ab \cup Ab'_M$) is made: $Ab Ab'_M Ab' = Ab \cup Ab'_M$. On the **Step 10** using formula (11) the affinity between antibodies is calculated for each antibody from the set Ab' . Further, on the **Step 11** selection of antibodies n' from Ab' with an affinity threshold, smaller than σ_s , and creation of new population of cells of memory is carried out, Ab'' . The threshold factor of compression σ_s is as much as possible admissible degree of overlapping of distinguishing coverings of antibodies. On **STEP 12** the population Ab''' is created; for this purpose those antibodies leave it, whose affinity is: $f_{Ab-Ag} \leq d_M$. On **Step 13** the replacement d of antibodies from population Ab''' is made by new antibodies. In the given implementation the algorithm finishes its work after the number of iterations defined by the user. After termination of the work the formed population of memory cells is used for constructing of a tree of minimum coverage (the minimum spanning tree) which reduction (removal of the maximum communications) finishes the clustering process.

4 Results of computer experiment

4.1. Experiment 1

For estimation the efficiency of the developed clustering algorithm the following three problems from **Marburg University** have been used (*Philips-Universität Marburg*): <http://www.uni-marburg.de/fb12/datenbionik/data/>: - Lsun – volume of sample of 400 examples, two variables and three classes; target – volume of sample of 770 examples with two variables and six classes; Two Diamond – volume of sample of 800 examples with two variables and two classes.

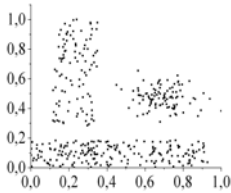


Fig. 3(a). Data: Lsun

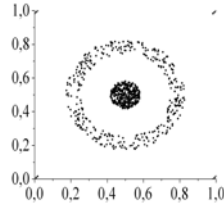


Fig.3(b). Data: Target

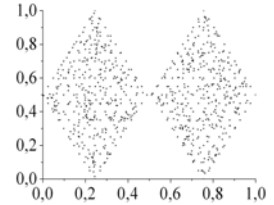


Fig. 3(c). Data: Two Diamond

In drawings 3 (a), 3 (b) and 3 (c) initial data, data sets Lsun, Target and Two Diamond are shown. In drawings 4 (a), 4 (b) and 4 (c) the immune networks constructed by data together with MST-tree are shown. In Fig.5 the dependence of the size of an immune network on quantity of iterations using investigated data is shown. In Fig. 6 the dependence of clustering quality on the size of a network is shown.

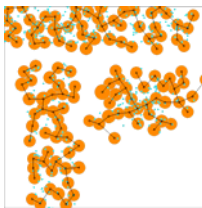


Fig.4(a). Definitive immune network for data Lsun

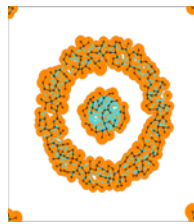


Fig.4(b). Definitive immune network for data Target

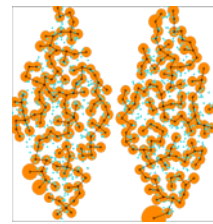


Fig.4(c). Definitive immune network for data Two Diamond

From the given drawing the influence of quantity of created antibodies on quality of the decision of a problem clustering is visible. In this connection, we have tried to investigate the influence of key parameters of adjustment on synthesis of antibodies.

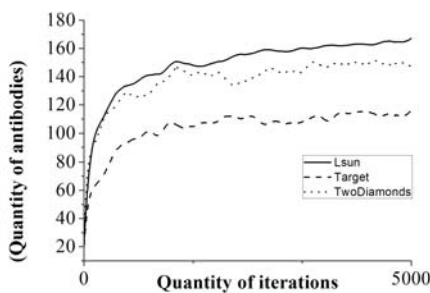


Fig.5. Dependence of a network size on the number of iterations

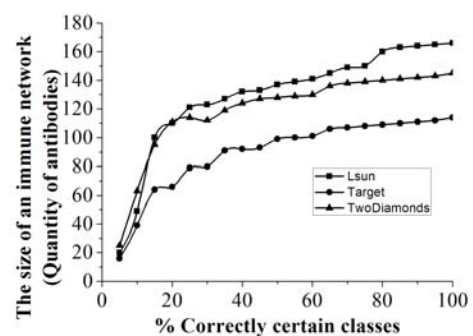


Fig.6. Dependence of correctness of classification on the size of an immune network

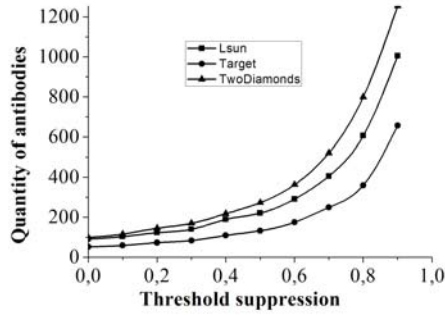


Fig. 7. Influence of threshold suppression value of the networks on the quantity of synthesized antibodies

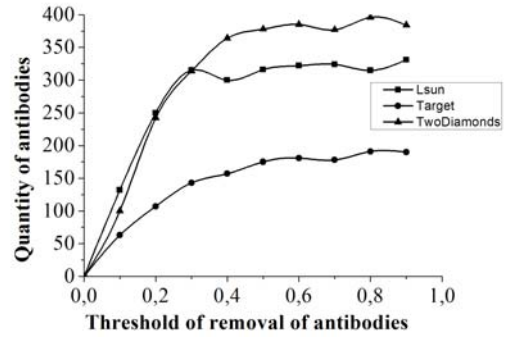


Fig. 8. Influence of value threshold clonal removals on the quantity of synthesized antibodies

In the first parameter, which has been investigated, the threshold suppression a step σ_s 2.9.2 algorithms (fig. 7) is σ_s . The given parameter displays an admissible level of overlapping coverings. It can be interpreted as follows: the less is the value the more strong is suppression, and the network will be especially discharged, i.e. antibodies will be located far from each other further. From the given drawing it can be seen that the threshold increase the suppression exponential influences the number of antibodies, and, hence, and the quality of clustering. In drawing 8 the influence of threshold clonal removals σ_d (a step 2.7.2) is shown σ_d . The parameter is interpreted as follows: the less is the value, there will be more removals. Despite some distinction, the general tendency in behavior of the given indicator is observed: the increase in value of the given indicator considerably increases synthesis of antibodies.

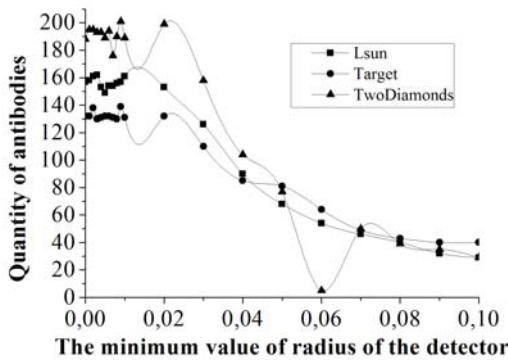


Fig. 9. Influence of the minimum value of radius of the detector on quantity of synthesized antibodies



Fig. 10. Influence of the importance of radius of the detector on quantity of synthesized antibodies

Considerably influences the minimum admissible value of radius of the detector (antibody) the synthesis of antibodies. In Fig. 9 it is shown how given value (is minimum admissible value of parameter in r the formula (2)) influences the quantity of synthesized detectors. The importance of radius of the detector (parameter in k_r the formula (2)) according to the experiments spent by us also considerably influences quality clustering (Fig. 10).

4.2 Experiment 2

For creation of a two-dimensional spiral from points, N parameters of the form of spirals for are calculated $1 \leq i \leq N$ in conformity of following expressions:

$$\Theta_i = \frac{i * \pi}{16 * \rho}, \quad (12)$$

$$r_i = \frac{r_{\max} * (104 * \rho) - i}{104 * \rho}, \quad (13)$$

$$x_i = r_i * \cos(\Theta_i), \quad (14)$$

$$y_i = r_i * \sin(\Theta_i), \quad (15)$$

where Θ_i is an angle; r_{\max} is maximum radius of a spiral; ρ is a density of points; i is time; $\pi=3.14$; y_i and x_i are generated points of the given spiral.

Parameters of radius and density are defined by the user. The fact of belonging to two various classes is shown with two different spirals. Changing these two parameters, it is possible to get the spirals of various radius and length. Operating the described parameters, it is possible to generate various spirals with the changing radius and length (Fig. 11, 12, 13). The main advantage of use of the equations (1)-(4) is that it is possible to model behavior of spirals to change complexity of the problem. The density variable ρ defines total number of points generated within curves which are defined by radius. The training set is presented by a vector $\{x, y\}$. The test set it is possible to consider as noisy spirals, i.e. it represents the level of additive white normal noise δ that is added to vectors of signals $\delta \{x, y\}$. For carrying out the experiments we used spirals with various density ρ equal 1, 2 and 3, various quantity of points of a spiral equal to 200, 402 and 600 (Fig. 11,12,13) or study the influence of noise level on quality of a problem clustering after contaminating the data white noise. Adding the white noise was carried out by means of function *awgn* from the **Communication Toolbox (Matlab)**:

$$y = \text{awgn}(x, \text{SNR}, \text{"measured"});$$

here x is a vector of input signal; SNR is a scalar - the signal/noise relation in decibels, the parameter "measured" measures the level of useful signal x to calculate the demanded noise level.

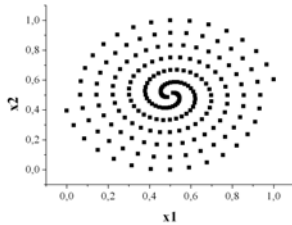


Fig. 11. Two spirals with the maximum radius $r_{\max}=1$ and density of points $\rho=1$, quantity of points in a spiral – 200.

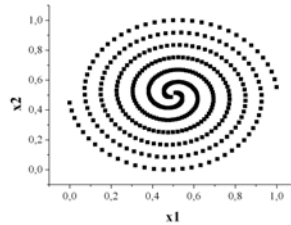


Fig. 12. Two spirals with the maximum radius $r_{\max}=1$ and density of points $\rho=2$, quantity of points in a spiral – 402.

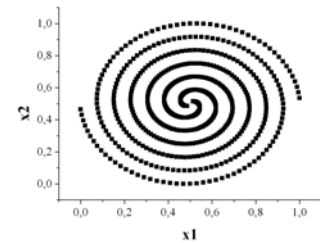


Fig. 13. Two spirals with the maximum radius $r_{\max}=1$ and density of points $\rho=3$, quantity of points in a spiral – 600.

To get the sets of boards with various levels of white noise to each of three sets of spirals with density ρ equal to 1, 2, and 3 the values of the spirals have been processed by function *awgn* using different values of parameter SNR : 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0. As a result, for each of three initial spirals we got 10 sets of noisy spirals. The data training was carried out on clean (no noise) data and were tested consistently on noisy data. The following immune networks have been as a result received:

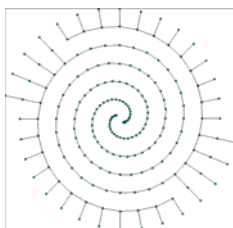


Fig. 7. An immune network with a MST-tree for a problem of two spirals with the maximum radius $r_{\max}=1$ and density of points $\rho=1$, on no noisy data; number of points in a spiral – 200.

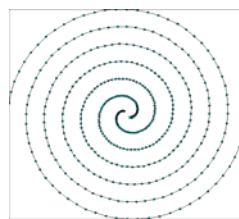


Fig. 8. An immune network with a MST-tree, for a problem of two spirals with the maximum radius $r_{\max}=1$ and density of points $\rho=2$, on no noisy data, number of points in a spiral – 402.

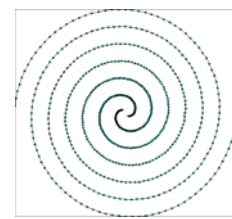


Fig. 9. An immune network with a MST-tree, for a problem of two spirals with the maximum radius $r_{\max}=1$ and density of points $\rho=3$, on no noisy data; number of points in a spiral – 600.

Table 1. Percent of correctly distinguished objects for a problem with two spirals with various noise level and various density of points

The signal/noise relation (value of SNR)	Level of the correct recognition in %, $\rho=1$, at $(x + \delta, y + \delta)$	Level of correct recognition in %, $\rho=2$, at $(x + \delta, y + \delta)$	Level of correct recognition of %, $\rho=3$, at $(x + \delta, y + \delta)$
Without noise	80	100	100
1	72	100	100
0.9	72	90	100
0.8	71	88	95
0.7	72	86	94
0.6	71	85	94
0.5	68	85	92
0.4	66	83	91
0.3	66	75	87
0.2	57	70	82
0.1	56	60	75

Conclusion and Future research

In the paper are presented the results of experimental research of the developed adaptive immune network for the problem of data clustering. It was considered the influence of structural information on the quality of clustering by means of the offered immune network. In particular the influence of density of an arrangement of points of the given spiral structures and level noise on the quality of clustering was investigated. The analysis of applicability of the developed adaptive immune network for clustering the spectral structures has shown the following results:

1. The clustering algorithm, based on the principles of functioning of an artificial immune network of the modified structure is developed.
2. The estimation of interrelations "antigen-antibody" and "antibody-antibody" are offered which have provided the possibility to reduce the number of the enclosed cycles of the algorithm that has essentially accelerated its work in comparison to other similar algorithms.
3. The experimental studies of influence of various parameters of algorithm on quality of solution of the clustering problems have been carried out.
4. For the considered standard test problems the convergence of the size of the network, that provides practically 100 % of recognitions is found out.
5. The revealed features of behavior of parameters related to the number of detectors using the minimum radius and the importance of radius of detectors are shown to have local minima that allows the possibility of further structure optimization of immune network.
6. The convergence of the procedure of increasing the size of a network has asymptotic character, i.e. converges to the final size of a network.
7. Testing of qualifiers only with one kind of density of points of spiral structures does not allow to get objective results allowing to state an objective estimation of quality of the algorithm.
9. Adding of a noise component to the test data allows to estimate more objectively the working capacity of the algorithm in real conditions.
10. The quality of clustering by means of an adaptive immune network substantially depends on the structural information about spirals.

The possible subject of further research is directed to improving the work of the offered algorithm assume the following:

1. Preliminary processing of data aiming to the use the auxiliary space which dimension will exceed initial dimension as it is carried out in nuclear methods.
2. The use of indistinctive sets for an estimation of affinity of communication "antigen-antibody" and affinity of communication "antibody-antibody" (similarity degree) with introducing the indistinct parameters of importance of the radius of an antibody (cross-reactive threshold) and indistinct radius of an antibody.

References

- [1] Litvinenko V. I. Artificial immune systems as means of inductive construction of optimum models of difficult objects // *Problems of Control and Information Processing*. - 2008. № 3. - pp. 43-61.
- [2] *Philosophical Encyclopedic Dictionary* / Edition: L.F. Ilicheva, P.N. Fedoseev, S.M. Kovalev, V.G. Panov. – M.: The Encyclopaedia, 1983. – 840 p.
- [3] Ivakhnenko A.G. Self-training of recognition and automatic control system. – Kyiv: Technika, 1969.
- [4] Ivakhnenko A.G., Stepashko V. S. Noiseproof modeling. – Kyiv: Naukova. Dumka, 1985. – 300 p.
- [5] Ivachnenko A.G., Zajchenko J.P. Decision-making on the basis of self-organising. – M.: Sov. Radio, 1976. – 280 p.
- [6] Tarakanov A.O., Skormin V.A., Sokolova S.P. Immunocomputing: principles and applications. – New York: Springer-Verlag, 2003.
- [7] de Castro, L. N. & Von Zuben, F. J., aiNet: An artificial Immune Network for Data Analysis. In *Data Mining: A Heuristic Approach*, H. A. Abbass, R. A. Saker, and C. S. Newton (Eds.), Idea Group Publishing, USA, Chapter XII, pp. 231-259, 2001.
- [8] Timmis, J. & Neal, M., “A resource Limited Artificial Immune System for Data Analysis”, *Knowledge Based Systems*, 14(3-4): pp. 121-130, 2001.
- [9] Knight T., Timmis J.. AINE: An Immunological Approach to Data Mining. ICDM 2001: pp.297-304.
- [10] Wierzchoc S. T. Artificial Immune Systems. Theory and Applications (in Polish). Akademicka Oficynna Wydawnicza EXIT, Warszawa 2001. ISBN 83-87674-30-3, 282 + vii pp.
- [11] Bezerra G.B., Barra T.V., de Castro L.N., Zuben F.J.V.: Adaptive radius immune algorithm for data clustering. In: *Proceedings of 4th International Conference on Artificial Immune Systems – ICARIS*. Volume 3627 of Lecture Notes in Computer Science., Springer-Verlag (2005) 290–303
- [13] Lytvynenko V., Bidyuk P., Myrgorod V. Application of the method and combined algorithm on the basis of immune network and negative selection for identification of turbine engine surging /2nd International Conference on Inductive Modelling // September 15-19, 2008, Kyiv, Ukraine, pp.116-23.
- [14] Ultsch A. “U*-Matrix: a Tool to Visualize Clusters in High Dimensional Data”, in Technical Report No. 36, Department of Mathematics and Computer Science Philips-University Marburg, 2003.
- [15] Zahn C. T. Graph-theoretical methods for detecting and describing gestalt clusters // *IEEE Trans. Comput.*, 1971. C-20. P. 68–86.
- [16] Singh S. 2D spiral recognition using possibilistic measures, *Pattern Recognition Letters* 1998; 19(2): pp.141-147.
- [17] Fahlman S.E. Faster-learning variations on back-propagation: An empirical study. In *Proceedings of the 1988 Connectionist Models Summer School*, Morgan Kaufmann, 1988.
- [18] Fahlman S.E. and Lebiere C. The cascade-correlation learning architecture, In *Advances in neural information processing systems 2*, Touretzky D.S. (ed.), Morgan Kaufmann, 1990.
- [19] Lang K.J. and Witbrock M.J. Learning to tell two spirals apart, In *Proceedings of the 1988 Connectionist Models Summer School*, Morgan Kaufmann, 1988.
- [20] Tay L.P. and Evans D.J. Fast learning artificial neural network (FLANN II) using the nearest neighbor recall. *Neural, Parallel and Scientific Computations* 1994; 2(1):17-27.
- [21] Sun C.T. and Jang J.S. A neuro-fuzzy classifier and its applications. – In *Proceedings of the IEEE International conference on fuzzy systems*, 1993, vol. 1, pp. 94-98.
- [22] Chua H., Jia J., Chen L. and Gong Y. Solving the two-spiral problem through input data encoding, *Electronics letters* 1995; 31(10):813-14.
- [23] Jia J. and Chua H. Solving two-spiral problem through input data representation, In *Proceedings of the IEEE International conference on neural networks*, 1995, vol. 1, pp. 132-135.
- [24] Ulgen F., Akamatsu N. and Iwasa T. The hypercube separation algorithm: a fast and efficient algorithm for on-line handwritten character recognition, *Applied Intelligence* 1996; 6(2): pp. 101-116.
- [25] Singh S. A single nearest neighbour fuzzy approach for pattern recognition // *International Journal of Pattern Recognition and Artificial Intelligence*, 1999.
- [26] Singh S. 2D spiral recognition using possibilistic measures // *Pattern Recognition Letters* 1998; 19(2): pp. 141-147.
- [27] Singh S. Effect of noise on generalization in Massively Parallel Fuzzy Systems, *Pattern Recognition*, vol. 31, issue 9, 1998.