Artificial immune system as a high parallel intelligent information processing and protection system

A I Alchinov¹, E V Volkova² and I M Daudov³

¹VATrapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia
²Irkutsk National Research Technical University, 83, Lermontov street, Irkutsk, 664074, Russia
³Chechen State University, 17a, Dudaev Boulevard, Groznyy, 364015, Russia

E-mail: kshakalov@mail.ru

Abstract. The article discusses the possibilities of applying the artificial immune system to ensure the protection of information. It is described that, according to studies, processing of the immune system of external signals is carried out using a highly parallel process.

1. Introduction
Methods based on the mechanisms of immunity are successfully used for solving problems of pattern recognition, diagnostics and fault detection, as well as to create computer security systems in some other areas. Studies of immune systems correlate with ideas about autopoietic systems for which certain isolation is a condition for adequate interaction with the environment [1, 2].

Concerning information processing, the following properties of immune systems are of interest:
• Recognition. The immune system can recognize and classify various molecular structures and selectively respond to them.
• Highlighting features. Each AIC acts as a “filter” that removes noise and a “magnifying glass” that draws the attention of lymphocyte receptors.
• Diversity. The immune system forms many different lymphocyte receptors using a combinatorial mechanism.
• Training. The immune system evaluates the structure of a particular antigen using its random contacts with the cells that make up this system. The training lies mainly in the mechanism of clone formation, leading to the formation of new immunocompetent cells, given the current state of the system.
• Memory. According to the short-term and long-term mechanisms of immune memory, the immune system provides an optimal balance between saving the body’s resources and fulfilling its function by preserving the minimum necessary, sufficient memory of previous contacts with the antigen.
• Distributed search. At its core, the immune system is a distributed system.
• Self-regulation. The immune system has the properties of self-regulation. The central body that controls the functions of the immune system does not exist.
• Joint stimulation. An additional stimulating signal tightly regulates B-lymphocyte activation. The
second signal helps ensure tolerance and distinguish between threat and “fake challenge”.

- Dynamic protection. The presence of dynamic protection gradually leads to the expansion of the surveillance zone controlled by the immune system.
- Probabilistic detection. Cross-reactions in the course of the immune response is a stochastic process. A lymphocyte can interact with several structurally similar antigens.

2. Artificial Immune Network Jisys

In the UK, an Isys research project is being funded. The Isys system tries to use the basic ideas of immunity, using other approaches if necessary.

Self-learning systems, built on the principles of immunity, have such properties as:

- Availability of capacious and addressable memory, which is accessed through a complex recognition mechanism
- Ability to recognize new substances after the first contact with them
- Construction in the form of a decentralized, self-organizing system without a central governing and control body
- Ability to limit specialization

The resulting model of the artificial immune system was close in its characteristics to the natural immune system. Now, this model is known as Pisys. Pisys includes a root object, a network of nodes, a dataset for training, and a dataset for testing. Each network node includes a recognition element with a randomly generated recognition pattern. These data are changed to recognize the input data in the first phase of the system – the training phase. During the recognition process in the second phase, Pisys tries to classify the new data with the training data.

In the training phase, the incoming data randomly interacts with the elements of the node. For evaluating the similarity between the pattern of the node and the data, an algorithm for sequential matching of plots is used. The algorithm is designed so that extended areas of similarity receive additional weights. As a result of this process, the "similarity value" of the node is evaluated. If the similarity value reaches a predetermined threshold, then the node is considered "stimulated" and begins to clone itself. In the process of cloning, a node pattern can stochastically mutate. Further, comparing the mutated patterns with the input data, the most affine ones are selected and added to the network.

Allocated n-nodes as close as possible to the new node and associated with it. If the selected node already has a predefined number of links m, the link with the least affine input node is given and the link with the new node is added. This leads to the formation of a region capable of recognizing similar signals.

The network, in turn, by the feedback mechanism enhances the stimulation of the best nodes and suppresses the stimulation of the worst [39]. This feedback provides a higher level of stimulation of nodes with a large number of neighboring nodes similar in their patterns, as compared to solitary nodes.

Nodes that do not reach the similarity threshold as a result of the presentation of melons are removed from the network and destroyed. As a result, the network changes over time and responds better to incoming data. In addition to nodes, the structure of bonds is constantly changing. Some disappear and others are created.

If a node entering the network checks for similarity with certain data, then all nodes associated with it are also mapped to this data. The matching process is carried out with a decreasing window. First, 75% of all related nodes are matched, then 50%, and finally, before completing the matching process, only 25% of the connected nodes are considered. After the analysis of the next batch of data is completed, the threshold levels are refined for a possible improvement in the operation of the entire system. After the network training phase, test data is offered. The system tries to find common characteristics with training data in them.

With a relatively small number of test nodes, this approach works well. However, when considering more complex tasks in the framework of this implementation, problems and limitations
appear. To get rid of them, the system was improved and was called Jisys.

Jisys was developed in the Java programming language using a relational database to store information about nodes and the relationships between them. The functioning of Jisys is divided into four stages. First, in the first stage, immune memory is initiated. On the second, the memory is converted into the immune network, then the generalization of the immune memory is performed using other data than on the first stage. The last stage is the classification of new unfamiliar data.

The included Jisys network is considered a host. In addition to the proposed fields, there is also a "stimulation level" field, which is modified by the Jisys system itself.

When initializing Jisys system memory, new data is presented as a linear memory node. A feature of a linear memory node is that it does not impose any organizational structure on the node. Each node is potentially as significant as the others. This fact increases performance compared to the original Pisys system, in which the network is restructured after each new node is turned on. In this case, the network is restructured only after the inclusion of all new nodes and the completion of initialization.

The algorithm for including new nodes is based on the algorithm described in work. In the Jisys system, it is used only once, after all nodes have been received that are included in the final version of the network. The scheme of the algorithm consists in sorting n the most similar nodes with an estimate of the number of links. If a node has m links, then this link is deleted, and the nodes are "reconnected".

As a result of the execution of the algorithm in the network, areas are formed that combine node designed to solve similar problems.

The Jisys system uses three methods to evaluate nodes:
• Sequential circular comparison method
• Trigram method
• Number comparison method

These methods allow getting different estimates of proximity. The first estimate is obtained by comparing with successive shifts of looped lines, matching groups of three characters for all text fields, and comparing numbers for numeric fields. For each field, a weight is entered, which is used for calculating the measure of proximity of the structures as a whole.

The second estimate is obtained by combining all the text fields of the node in one line and applying the methods of sequential circular comparison and trigrams to this line. For numerical fields, a comparison of numbers is used.

Both ratings are expressed as a percentage of the possible maximum similarity. Then the highest estimates of the first and second type are used as estimates of the similarity of the entire node.

After building the network, the generalization stage is used. This stage aims to generalize the memory of the node, as a result of which it would be able not only to process the data but also to identify trends and patterns. For assessing the similarity of the source data with the new, all nodes are scanned. The system adds value for assessing the similarity of a new node to the stimulation levels of existing nodes, obtaining n best nodes using the following algorithm:

For existing Node in best N Nodes do
   match Value := match (new Node, existing Node);
   existing Node. stimulation Level := existing Node. stimulation Level + matchValue;

When the selection threshold is reached, the node is cloned. The number of clones is determined by the difference between the average threshold of stimulation of the network and the threshold of stimulation of a given node. Clone selection is implemented as a random selection of one of three outcomes:
• Preservation. The clone remains an exact copy of the original.
• Generalization. Clones and their inducing data are compared. All matching fields remain in the clone, and the remaining elements are filled with wildcards.
• Mutation. The clone mutates so that the new node is slightly different from the original.

The clone is integrated into the network structure and receives a corresponding set of stimulation
levels in the refinement phase. In this way, nodes that better match the new data clone themselves and provide a high level of similarity for the new node.

At the refinement stage, the Jisys system scans all nodes of the network and finds those whose stimulation level is below the stimulation threshold, and removes them from memory. As a result, the nodes that are most often involved in the classification of the presented data remain in the memory. Then the memory of the nodes is transformed into the immune network.

The constructed immune network can be used for classifying new data. The matching algorithm compares the incoming data with all nodes of the network. If the similarity level is higher than the selection threshold, then the data is provided to the user as containing close ones to the input.

3. Algorithm of negative selection
A feature of the vertebrate immune system is the ability to distinguish its cells from virtually any foreign cells and molecules. The detection of foreign antigens is a function of receptors located on the surface of T and B lymphocytes. These receptors appear during pseudo-random permutation of the genome during maturation of lymphocytes. Then the lymphocytes undergo a negative selection, during which cells that react against their antigens are eliminated. As a result, only those lymphocytes that do not respond to their antigens enter the body. This process is essential for separation into a friend or foe.

In 1994, Stefanie Forrest proposed a negative selection algorithm. It is based on the following components:

- It is defined as a normal pattern of activity or sustainable behaviour or process. In particular, a database describing the normal dynamics of processes is represented as a set of S lines of equal length l, composed of letters of a finite alphabet.
- A set of detectors R is created, none of which matches any of the lines of the set S. A partial matching rule is used, according to which two lines coincide if and only if they are identical in r-adjacent positions, where r is selected depending on the problem being solved.
- Data is monitored by continuously matching detectors with new arrivals in S. Detection of coincidence with the detector is considered as a change in the behaviour of the monitored system.

Candidates for a set of detectors in the initial version of the algorithm are randomly generated. Then they are checked for coincidence with their lines, and the candidate is rejected if a match is found. The process is repeated until the required number of detectors is generated. The main limitation in the case of random generation of detectors is an exponential increase in computational costs with an increase in its size. To generate a large number of detectors, S. Forrest and P. Helman developed an improved algorithm, the operating time of which grows linearly with an increase in its size. The operation of the algorithm is divided into two stages. At the first stage, all non-repeating lines are recalculated, i.e. All valid detectors. At the second stage, the required number of detector lines is formed from the set of valid detectors. In other words, at the first stage, from the set of its lines S and the correspondence threshold r, the algorithm determines the total number of lines that do not coincide with the lines of the set S. Then, at the second stage, part of these lines is used for forming a detector that controls the patterns of incoming data.

4. Clonal algorithm
An artificial immune network can be represented in the form of a graph, which consists of many nodes – cells of the immune network, and edges, the weight coefficients of which determine the affinity of the network nodes to each other. There are two types of affinity:

- Antigen-Antibody (Ag-Ab) Affinity – Degree of Difference
- Antibody-antibody affinity (Ab-Ab) – degree of similarity

Formally, the immune network algorithm can be represented as:

\[ \text{immNET} = (P, G^k, l, k, m_{Ab}, \delta, f, l, r, AG, AB, S, C, M, n, d, H, R) \]

where \( P \) is the search space; \( G^k \) is the representation of space; \( l \) is the length of the attribute vector; \( k \) is the cell receptor length; \( m_{Ab} \) — cell population size; \( \delta \) is the expression function; \( f \) is the affinity
function, \( I \) is the initialization function of the initial population of network cells; \( \tau \) is the termination condition of the algorithm; \( AG \) is a subset of antigens; \( AB \) – population of network cells; \( S \) – selection operator; \( C \) is the cloning operator; \( M \) is the mutation operator; \( n \) is the number of the best cells selected for cloning; \( d \) is the number of worst cells to be replaced with new ones; \( H \) is the clonal deletion operator; \( R \) is the network compression operator.

Figure A.1 shows a block diagram of the immune network algorithm.

Step 1. Initialization.
- Step 1.1. Creation of an initial population of memory cells (\( M_R \)).
- Step 1.2. Creating a receptor population (\( AB \)).

Step 2. Antigenic presence. At this step, the algorithm performs one pass for each antigen.
- Step 2.1. The calculation of affinity. The affinity of all memory cells \( m_j, m_j \in M_R \) for the next antigen \( Ag_i, Ag_i \in AG \) is calculated, and one best cell \( mb \) is selected.
- Step 2.2. Cloning. Proportional to its affinity, the selected memory cell is cloned to form a population of \( M_C \) clones.
- Step 2.3. Maturation of affinity. Mutation of clones from \( M_C \) is performed. Altered clones are added to the antibody population, i.e., \( AB \leftarrow AB \cup M_C \). The affinity of the \( AB \) antibody population with the \( Ag \) antigen is calculated.
- Step 2.4. Metadynamics Unstimulated cells are removed according to the threshold \( \sigma_d \).
- Step 2.5. Some antibodies from the \( AB \) population are re-cloned to form a population of \( M_C \) clones.
  - Go to step 2.3 if the average affinity of the \( AB \) population is below a given threshold value.
- Step 2.6. A candidate cell (the best antibody) is selected from the \( AB \) population for the \( Ab_b \) memory cell population.
- Step 2.7. Go to step 3 if \( f(AB_{b}, Ag_i) < f(mb, Ag_i) \).
- Step 2.8. Adding \( Ab \) antibodies to the \( M_R \) population.
- Step 2.9. Intercellular interaction. The affinity of the interaction of all cells of the \( M_R \) population with a friend is determined, i.e., \( f(m_i, m_j), m_i, m_j \in M_R \).
- Step 2.10. Network compression. The cells of the \( M_R \) population recognizing each other are deleted following a given threshold \( \sigma_c \).
- Step 3. Verify that the stop condition for the algorithm is satisfied and go to step 2 if the stop condition is not met.

In the above algorithm, the operator \( H \) reduces the size of the network by removing unstimulated cells, using the threshold death rate as a control parameter. Also, the network size is reduced using the compression operator \( R \), which uses the threshold compression coefficient \( \sigma_c \), due to the removal of self-recognized cells. During initialization, a large number of cells are created that are processed by this algorithm. During suppression, those cells that are affinity for each other, as well as those that are too far from the antigens, are destroyed.

Along with this, the study of the laws of the functioning of the system is becoming increasingly important in connection with the formulation of the problem of production and modernization of the Russian economy as a whole [3, 4]. The simplest software analogues working according to the general principles of the immune system are used in some areas of industry [5, 8-10].

5. Conclusions

The immune system is an excellent example of a system that implements adaptive processes at the global level based on local interactions. An example of the application of such systems can be intelligent energy systems, systems for managing the lifecycle of production processes and equipment, security systems, etc. [6, 7]. For solving recognition and classification problems, the immune system uses the mechanisms of learning, memory and associative search. Some successful products are being developed in this area.

References
[1] Shutaleva A V 2019 The fundamental problem of subjectivity in Francisco Valera’s
neurophenomenology Tomsk State Univer. J. of Philos. Sociol. and Political Sci. 48 84–90
[2] Shutaleva A V 2020 Consciousness as a problem of Charles D. Laughlin’s biogenetic structuralist neurophenomenology Tomsk State Univer. J. of Philos. Sociol. and Political Sci. 53 141–147
[3] Youn R B, Klyuev R V, Bosikov I I and Dzeranov B V 2017 The petroleum potential estimation of the north Caucasus and Kazakhstan territories with the help of the structural geodynamic prerequisites Sustainable Development of Mountain Territories 9(2) 172–178
[4] Tran V A, Klyuev R V, Bosikov I I and Tsidaev B S 2018 Estimation of the reliability of the ventilation control system at the mining sites Sustainable Development of Mountain Territories 10(1) 117–124
[5] Shishlyannikov D I, Romanov V A, Zvonarev I E 2019 Determination of the operating time and residual life of self-propelled mine cars of potassium mines on the basis of integrated monitoring data Journal of Mining Institute 237 336-343
[6] Zhukovskiy Y L, Starshaia V V, Batueva D E, Buldysko A D 2019 Analysis of technological changes in integrated intelligent power supply systems Innovation-Based Development of the Mineral Resources Sector: Challenges and Prospects - 11th conference of the Russian-German Raw Materials 249-258
[7] Zhukovskiy Y, Batueva D, Buldysko A, & Shabalov M 2019 Motivation towards energy saving by means of IoT personal energy manager platform. Journal of Physics: Conference Series 1333(6) 062033
[8] Maksarov V V, Gabov V V, Zadkov D A, Martyushev N V 2018 Preface IOP Conference Series: Earth and Environmental Science 194 012001 DOI:10.1088/1755-1315/194/1/012001
[9] Maksarov V V, Gabov V V, Zadkov D A, Martyushev N V 2019 Preface IOP Conference Series: Earth and Environmental Science 378 011001 doi:10.1088/1755-1315/378/1/011001
[10] Fetisov V G, Nikolaev A K, Lykov Y V, Duchnevich L N 2018 Mathematical modeling of non-stationary gas flow in gas pipeline. IOP Conference Series: Materials Science and Engineering 327(2) 022034