11th International Research/Expert Conference
”Trends in the Development of Machinery and Associated Technology”
TMT 2007, Hammamet, Tunisia, 05-09 September, 2007.

APPLICATION OF ARTIFICIAL IMMUNE SYSTEM FOR DATA CLASSIFICATION

Małgorzata Galek, Andrzej Sokołowski
Silesian University of Technology, Faculty of Mechanical Engineering
Konarskiego 18a, 44-100 Gliwice,
Poland

ABSTRACT
Artificial immune system can be applied for binary classification into two classes: self and non-self elements. For such an application we propose an algorithm based on B cells net, cloning and maturation of B lymphocytes for antigens entering the body. Here, the primary response is responsible for creating the memory cells. Based on the memory, the classification is performed. The classification is considered as secondary immune response. The suggested approach is tested with data reflecting problem of tool state diagnostics during drilling.
Keywords: Artificial immune system, cutting tool wear diagnostics, knowledge discovery

1. INTRODUCTION
Artificial immune system is metaphor of biological immune system, which can discover and grow its qualifications by learning and experience. Very simplify way of functioning of immune system is shown in Figure 1. **A** square represents potential structure with which the body could have a contact. Then, the self - structures **B** are removed from all structures (**A-B=C**). The attacking intrusion structures **D** are memorised and stay in this way in the immune memory **E** (**C+D=E**).

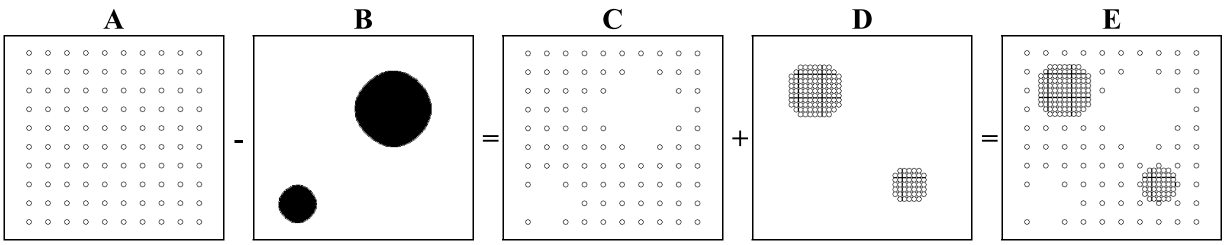

Figure 1. Idea of artificial immune system

The most important cells in immune system are as follows: lymphocytes B and T, cells representing antigen to lymphocytes T, immuglobine and receptors of T cells. Lymphocytes T respond for individual protection action, i.e. they recognize self and non-self elements. Lymphocytes B respond for creation of mechanism of protective cells in immunity system [6].
The whole immune system consists of two smaller systems by which the body identifies foreign material, i.e. the innate immune system and the adaptive immune system. The body is born with the ability to recognize certain microbes and immediately destroy them. It’s called the innate immune system. The adaptive immune system consists of immunology memory. This memory assures that the next answer for the same stimulating factor will be bigger and quicker [1]. Generally, in the first touch with antigen, a primary response emerges (Figure 2). Primary immune response is responsible for creating memory cells. Therefore, in the next contact with antigen, the body reacts more effectively. This is called secondary response.

The immune system can find its application in complex image recognition, automatic detection of data concentration or can be applied for data reduction, optimisation and anomaly detection. Also, data classification can be performed with artificial immune system. Such application is of our interest in this paper.

2. CLASSIFICATION SYSTEM

Immune system can be applied for binary classification into two classes, i.e. self and non-self elements. This application bases on B cell net, cloning and maturation of B lymphocytes for antigens entering the body. To initiate B cells, an input data set is chosen. The input data set is used for training and creating a network that consists of specific number of cells. Arc between two cells represents their similarity (Figure 5 b, c, d). Cells are connected by arc in the case when they are not identical and fit factor isn't higher than an assumed threshold value. Fit factor is Euclidian interval between cell a and p:

$$d(a,p) = \sqrt{\sum_{i=1}^{n}(a_i - p_i)^2} \qquad (1)$$

where: a – antigen (vector representation),
p – antibody (vector representation).

For initialising of memory, B cells are selected and they create embryo of memory. Part of input data is usually taken to perform such an operation. A complete algorithm of creating the memory is shown in Figure 4. The stimulation level is described with equation (2):

$$s(p) = sk(p) + wk(p) + ak(p) \qquad (2)$$

where: $sk(p)$ – level of cell stimulation by its neighbours,
$wk(p)$ – enmity level of neighbours,
$ak(p)$ – cell stimulation level by antigen.

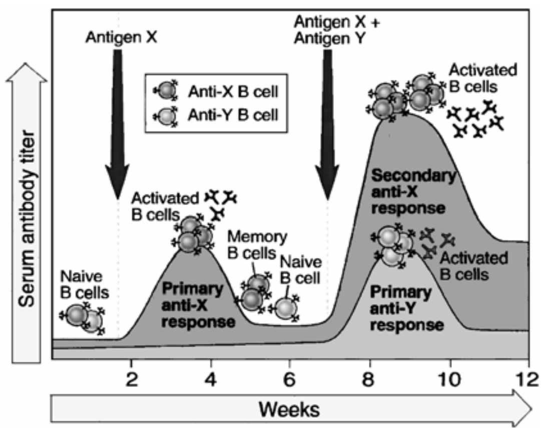

Figure 2. Primary and secondary immune response [1]

Creation of immune memory - *p set*
↓
Computation of fit factor of memory cells and antigens representing classification elements *d(a,p)*
↓
Comparison of fit factor with *NAT* parameters

$d(a,p) \geq NAT \downarrow$ | $\downarrow d(a,p) < NAT$

Self elements | Non-self elements

Figure 3. Classification algorithm

Set of antibody a
↓
Initiate of B cells (immune memory) set of *p* ←
↓
Calculation of affinity *d(a,p)*
↓
Calculation of stimulation level *s(p)*
↓
Cloning adequate to stimulation level
↓
Mutation
↓
Modification of memory
↓
Stopping criterion →
↓
END

Figure 4. Algorithm for learning network

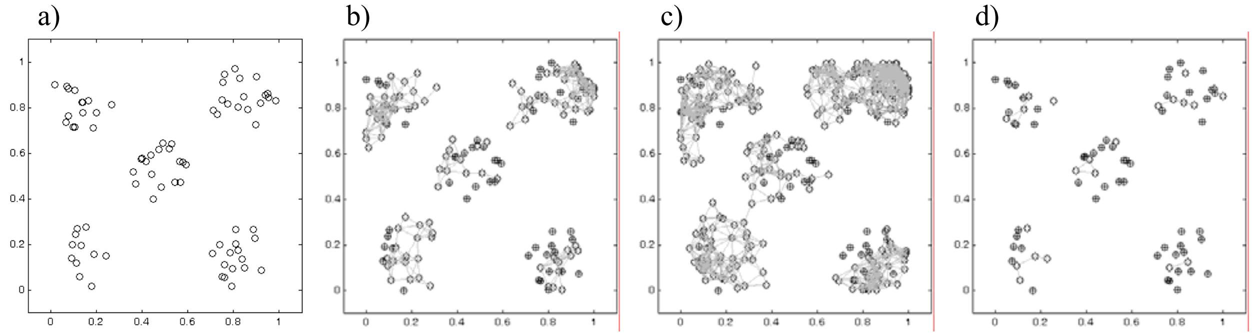

Figure 5. Data and memory cells:
a) input data; b) memory c) memory with high number of cells d) memory with low number of cells.

After presenting antigen to the cells, each cell produces a certain number of clones, proportionally to standardized stimulation level. Created in this way cells succumb of mutation. If the stimulation level is not larger than NAT, the cell is added to memory cells. The NAT parameter represents fit level. 10 % cells with smallest stimulation level are removed from memory set. There is an additional parameter, which controls the number of cells in memory set. This parameter can be considered as a threshold for the number of cells in memory. Generally, created in this way memory is a base for performing classification, as shown in Figure 3. From other hand, creation of immune memory reflects the primary response and classification can be seen as a secondary immune response. It should be added that number of memory cells depends mostly on NAT parameter and number of iterations. Also, it is very important to control number of memory cells since this number can strongly affect the results achieved, as exemplary shown in Figure 5.

As it was mentioned, the type of classification performed with artificial immune systems is suitable for a task with two classes, only. For classification into more than two classes, it is necessary to build more complex system. In order to perform classification into three classes we propose application of two or three immune systems at the same time.

The proposed approach was tested with so-called academic data, first. Then, the classification of drill wear during multi - spindle drilling was analysed. In this case, classification system consists of three simple immune classifiers that classify data into two classes: “fresh” and “¬fresh”, “worn” and “¬worn”, “partly worn” and “¬partly worn”. During tests, the data was presented ten times. If the most of the conclusions pointed at a specific basic state (“fresh”, “worn” or “partly worn”), this state of the drill was assumed as a final result, e.g. drill is “worn”. The efficiency (performance) for that classification was, however, relatively low (on average about 30 %). Therefore, it was decided to introduce an additional rule that allows excluding some of the basic states of the drill. In other words, the classification was considered successful if we could decide, for example, that the drill is not worn, but we did not know whether the drill is fresh or partly worn. The described approach allowed reaching efficiency up to 100% in some cases.

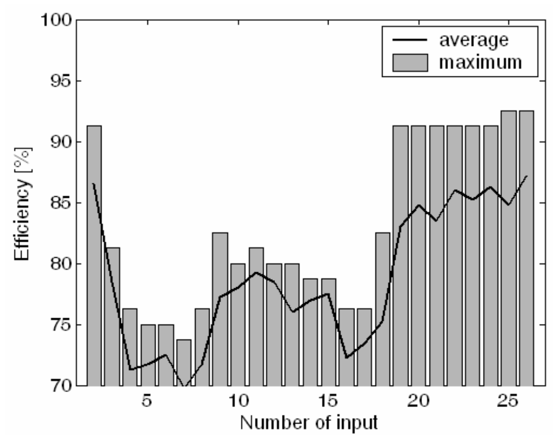

Figure 6. Efficiency for input sorted based on diminishing distance between input

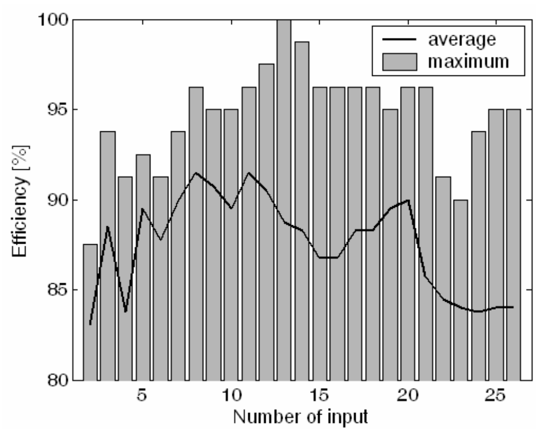

Figure 7. Efficiency for input sorted based on accumulative distance between input

The results expressed in the form of maximum and average classification efficiency are shown in Figure 6 and Figure 7. The tests were conducted with different numbers of inputs. We started with two inputs and kept adding inputs up to all available, i.e. 26 inputs. As it can be noticed, the number of inputs has substantial influence on the classification efficiency. Also, the way the inputs were selected (based on diminishing distance or accumulative distance) affects the results. At this point it is necessary to recall the influence of the NAT parameter, as well. This problem is graphically depicted in Figure 8. Figure 9, where the efficiency for different number of inputs in relation to the NAT value is presented.

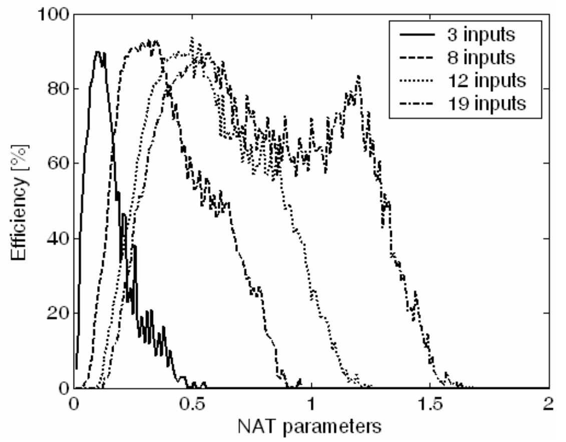

Figure 8. Efficiency for different NAT parameters value

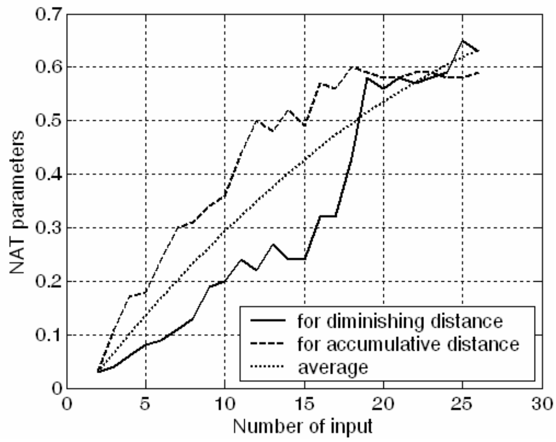

Figure 9. NAT parameters value for different numbers of inputs for maximum efficiency

3. SUMMARY

Immune classification system can have the efficiency (performance) of classification equal 100 %. However, efficiency of classification in average is about 90 %. There are few parameters and values that decide on possibility of achieving the high efficiency. Considering these parameters, one can point at proper data selection, i.e. selection of proper inputs (see Figure 6, Figure 7). This is a difficult task and some more sophisticated methods should be used in this case. Next, one must decide on NAT parameter value, which has the highest influence on efficiency of classification system (Figure 8). Authors propose to choose NAT parameter value based on the following equation:

$$NAT = -0{,}005 \cdot x^2 + 0.0388 \cdot x - 0{,}0463$$

where: x – number of inputs.

4. REFERENCES

[1] ABBAS A.K., LICHTMAN A.H.: *Basic Immunology. Fraction and Disorders of the Immune system.* SAUNDERS. An Imprint of Elsevier, Philadelphia 2004

[2] DE CASTRO L. N., VON ZUBEN F. J.: *Artificial Immune Systems: Part I – Basic Theory and Applications.* Technical Report TR – DCA 01/99, Campinas-SP 1999

[3] FORREST S., PERELSON A. S., ALLEN L., CHERUKURI R.: *Self-nonself discrimination in a computer*. Proceeding of 1994 IEEE Symposium on Research in Security and Privacy, Los Alamitos, CA: IEEE Computer Society Press 1994

[4] FORREST S., HOFMEYR S. A.,SOMAYAJI A., LONGSTAFF T. A.: *A sense for self for UNIX process*. Proceeding of 1996 IEEE Symposium on Research in Security and Privacy, Los Alamitos, CA: IEEE Computer Society Press 1996

[5] GALEK M.: *Sztuczne systemy immunologiczne.* Prace naukowe Katedry Budowy Maszyn Nr 4/2005. Zeszyt nr 22, Gliwice 2005

[6] WIERZCHOŃ S. T.: *Sztuczne systemy immunologiczne. Teoria i zastosowania*. Akademicka Oficyna Wydawnicza EXIT. Warszawa 2001