An immune approach to classifying the high-dimensional datasets

Andrzej Chmielewski
Faculty of Computer Science
Białystok Technical University
Wiejska 45a, 15-351 Białystok, Poland
Email: achmielewski@wi.pb.edu.pl

Sławomir T. Wierzchoń
Institute of Computer Science
Polish Academy of Sciences
and
Faculty of Mathematics, Physics and Informatics
Gdańsk University, Poland
Email: stw@ipipan.waw.pl

Abstract—**This paper presents an immune-based approach to problem of binary classification and novelty detection in high-dimensional datasets. It is inspired by the negative selection mechanism, which discriminates between self and nonself elements using only partial information. Our approach incorporates two types of detectors: binary and real-valued. Relatively short binary receptors are used for primary detection, while the real valued detectors are used to resolve eventual doubts. Such a hybrid solution is much more economical in comparison with "pure" approaches. The binary detectors are more faster than real-valued ones, what allows minimize computationally and timely complex operations on real values. Additionally, regardless of type of encoding, the process of sample's censoring is conducted with relatively small part of its attributes.**

I. Introduction

NATURAL Immune System (NIS) prevents organism against intruders called *pathogens*. It consists of a number of cells, tissues, and organs that work together to protect the body. The main agents responsible for the adaptive and learning capabilities of the NIS are white blood cells called lymphocytes and produced by the thymus, spleen and bone marrow.

There are two kinds of lymphocytes: B- and T-lymphocytes called also B- and T-cells for brevity. Lymphocytes start out in the bone marrow and either stay there and mature into B-cells (this process is called *affinity maturation*), or they leave for the thymus gland, where they mature into T-cells. Both the types of lymphocytes have separate jobs to do: T-lymphocytes are like the body's military intelligence system, seeking out their targets and sending defences to lock onto them. B-cells are like the soldiers, destroying the invaders that the intelligence system has identified. It is obvious, that lymphocytes have to tolerate own cells of the organism; otherwise the organism could be self-destroyed.

To better imagine, how difficult and complex is the task of proper classification of different cells and microorganisms, we can mention that the total number of various types of pathogens is far greater then 10^{16}, whereas there are "only" about 10^6 own cells types. Despite of such huge diversity of infectious organisms, we require, that efficiency of immune systems has to be at the very high level as every intrusion can case disease or even can lead to die. Moreover, what is a characteristic feature of NIS, T-lymphocytes are "learned" only on the own cells, without any examples of pathogens. It means, that even never seen intruders can be properly recognized as they are different from own cells. This process called *negative selection* is only a small part of complex learning process performed by NIS. A reader interested in detailed description of immune system is referred to e.g. [12], [18].

Mentioned features of NIS are very desirable in many domains such as: intrusion detection systems (IDS), computer viruses detection, novelty detection and other binary classification purposes. Especially, a defending of computer networks against various types of attackers seems to be a natural application domain for immune based algorithms as there is a relatively easy to find out similarities between them. Both systems should effectively, with very high efficiency, recognize undesirable objects (binary classification) which constantly harass them. The huge diversity of intruders, additionally described by significant (from the computational complexity point of view) number of various types of attributes (labels, real and integer values, etc.), makes this problems relatively difficult to solve.

The approach presented in this paper was designed mainly for high-dimensional datasets. It is an effect of many experiments performed with the use of various types of detectors [9], [10], [4] and various affinity measures [3]. At the cost of lengthen learning stage (what is not crucial from e.g. IDS point of view), there is possible to increase the efficiency and significantly speed up the classification process.

This paper is organized as follows. Section II presents current state in this domain. Section III presents the proposed approach to classifying the high-dimensional datasets. Conducted experiments and its results are described in Section IV. Section V concludes the results and discusses a possibility of application of presented approach for other domains.

II. Background

A. Negative Selection

The *negative selection* algorithm, proposed by Forrest et. al, [6], is inspired by the process of thymocytes, i.e. young T-lymphocytes, maturation. It is designed to discriminate between own cells (called *self*) and others (called *nonself*).

To be more formal, denote U the problem space, or Universe of discourse and let $S \subset U$ be a subset of elements (e.g. measurements represented as binary strings or real-values) representing typical behavior of a system under considerations (*self*). Then the set of elements characterizing anomalous behaviour, N can be viewed as the set-theoretical complement of S (*nonself*):

$$N = U \backslash S \qquad (1)$$

The negative selection algorithm relies upon generation of so-called *detectors*, or *receptors*, being a counterpart of T-lymphocytes, in such a way, that a freshly generated detector d is added to the set D of valid detectors only if it does not recognize any self element. In the simplest case the detectors are generated randomly, but smart techniques are requested in general [6].

To mimic the process of self/nonself recognition we must designate an affinity measure, $match(d, u)$, specifying the degree with which a detector d bonds a given element u, see e.g. [18] for details. Usually, $match(d, u)$ is modeled by a distance metric or a similarity measure [7]. Majority of detection rules induce so-called *holes* (H), i.e. regions of N which are not covered by any detector and therefore samples from this region will be classified as self.

In real-life applications, it is not possible to gather all elements of S, because typically only its subset is observed. Therefore, it is assumed, that S is composed of S_{seen} and S_{unseen}

$$S = S_{seen} \cup S_{unseen}, \qquad (2)$$

and only S_{seen} is taken into the training phase (i.e. the phase of detector's generation). As a result, the detectors from the set D can recognize not only nonselfs from $N' \subset N$, but also some elements from S_{unseen} and it can result in wrong classification of not seen self samples.

In Fig. 1, there are presented possible combination of S, N and H regions. In Fig. 1(a) there is no holes as the self samples fully covers a space S and a detectors fully covers N region. This is only theoretical situation, not met in real applications. The desirable situation is in Fig. 1(b); detectors not covers S_{unseen} region. In this case, holes plays an important role as they are *"necessary to generalize beyond the training set"* (Stibor [15]). Problems of overfitting and underfitting are showed in Fig. 1(c) and Fig. 1(d), respectively. In both cases, a detectors covers either too much or too low of space, what leads to decreasing the accuracy of negative selection algorithms.

B. Optimal repertoire of detectors

In negative selection, the time complexity of classification as well as space complexity is relevant to the number of generated detectors. Every censored sample, in pessimistic case, have to be matched with all detectors from D. Thus, to decrease the complexity of classification process, it is desired, to generate minimal number of detectors covering a maximal part of N. A perfect set of detectors should cover
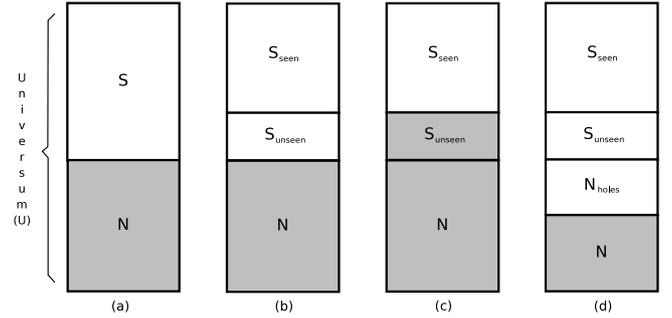

Fig. 1. Examples of possible combination of Self(S), Nonself(N) subsets, holes and regions covered by detectors (greyed regions).

themselves in minimal way and then each of them can be able to recognize different subsets of N (optimal repertoire of detectors).

This problem is especially important in on-line systems where huge dataset have to be classified without significant delays (i.e. IDS, virus detection, etc.).

C. Representing samples

To model the interactions (matching) between self and nonself samples we need three elements:

- proper samples encoding,
- appropriate affinity function (matching rule), and
- appropriate algorithm enabling generation of the receptors.

The choice of appropriate representation for samples gathered in the datasets seems to be a key issue. Typically, most of the samples are encoded as vectors of real-valued numbers. Unfortunately, this type of representation can involve the use of many time consuming operations (like multiplication and division), e.g. when classification process have to compute a distance between two samples to measure the degree of similarity between them. Therefore, in the case of huge dataset containing tens of attributes, the choice of this type of representation seems to be far from optimal, especially when the time of classification is crucial.

Since there, two types of samples representation were regarded in negative selection algorithms: binary and real-valued.

1) Binary representation: Primary, this type of representation was applied by Forrest et al. [5] to capture anomaly sequences of system calls in UNIX systems and next to model the system for monitoring TCP SYN packets to detect network traffic anomalies (called LISYS) [8].

In case of binary encoding, we identify the Universe with l-dimensional Hamming space, $\mathbb{H}^l$ consisting of the binary strings of fixed length l:

$$\mathbb{H}^l = \{\underbrace{000...000}_{l}, \underbrace{000...001}_{l}, \ldots, \underbrace{111...111}_{l}\}$$

Hence the size of this space is 2^l. The most popular matching rules used in this case are:

(a) r-contiguous bit rule [11], or

(b) r-chunks [2].

Both the rules say that a detector bonds a sample only when both the strings contain the same substring of length r. To detect a sample in case (a), a window (part of detector) of length r ($0 \le r \le l$) is slided through censored samples of length l. In case (b) the detector specifies substring and its position within a string. In both the cases nonself samples are recognized based only on partial information.

Below an example of matching (r-contiguous rule) a sample by a detector for affinity threshold $r = 3$ is given:

$$\overbrace{1\ 0\ \mathbf{0\ 0\ 1}\ 1\ 1\ 0}^{l} \qquad \text{sample}$$
$$0\ 1\ \underbrace{\mathbf{0\ 0\ 1}}_{r}\ 0\ 0\ 1 \qquad \text{detector}$$

It is obvious, from optimal repertoire point of view (see Section II-B), that shortest detectors are more desirable as they are able to detect more samples. However, Stibor [16] showed the coherence between r and l values for various $|S|$, in the context of the probability of generating detectors. He distinguished three phases:

- Phase 1 (for lowest r) – the probability is very near to 0,
- Phase 2 (for middle r) – the probability rapidly increase from 0 to 1 (so called *Phase Transition Region*),
- Phase 3 (for highest r) – the probability is very near to 1.

Hence, we should be interested in generating receptors with medium length r (belonging to second region) and eventually with larger values of r if coverage of N is not sufficient. It is worth to emphasize, that detectors can not be too long, due to exponential increase in the duration of learning process, which should be finished in reasonable time.

Hofmeyr in [8] developed immune-based IDS, called LISYS, using binary representation and r-contiguous rule as a matching function. Detectors, were randomly generated and each network TCP SYN packet was encoded as a binary string of length 49 (32-bits external IP address, 8-bits local address, 8-bits type of service and 1 bit for indicating the server machine). However, he used only 4 attributes which, in the most cases, are not sufficient to detect intruders. For example, Snort [13], one of the most popular open source IDSs, involve the use several (or even tens) parameters at least, significantly lengthen the dimension of space l.

2) Real-valued representation: To overcome scaling problems inherent in Hamming space, Ji and Dasgupta [10] proposed real-valued negative selection algorithm, termed as *V-Detector*.

It operates on (normalized) vectors of real-valued attributes; each vector can be viewed as a point in the d-dimensional unit hypercube, $U = [0,1]^d$. Each self sample, $s_i \in S$, is represented as a hypersphere centered at $c_i \in U$ and constant radius r_s, i.e. $s_i = (c_i, r_s)$, $i = 1, \ldots, l$, where l is the number of self samples. Every point $u \in U$ which lies within any self hypersphere s_i is considered as a self element. Also, detectors d_j are represented as hyperspheres: $d_j = (c_j, r_j)$,
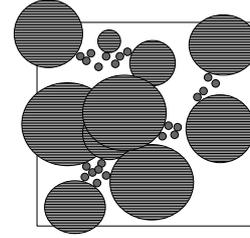

Fig. 2. Example of performance V-Detector algorithm for 2-dimensional problem. Grey circles denotes self samples, dashed circles denotes V-detectors, dashed area denotes detector which recognize all samples laying outside the space spanned over all self samples and white areas denotes holes.

$j = 1, \ldots, p$ where p is the number of detectors. In contrast to self elements, the radius r_j is not fixed but is computed as the Euclidean distance from a randomly chosen center c_j to the nearest self element (this distance must be greater than r_s, otherwise detector is not created). Formally, we define r_j as

$$r_j = \min_{1 \le i \le l} dist(c_j, c_i) - r_s \qquad (3)$$

The algorithm terminates if predefined number p_{max} of detectors is generated or the space $U \backslash S$ is sufficiently well covered by these detectors; the degree of coverage is measured by the parameter co – see [10] for the algorithm and its parameters description.

In its original version, the V-Detector algorithm employs Euclidean distance to measure proximity between each two samples. Therefore, self samples and the detectors are hyperspheres (see Figure 2). Formally, Euclidean distance is a special case of Minkowski norm (L_m, where $m \ge 1$), which is defined as:

$$L_m(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{d} |x_i - y_i|^m \right)^{\frac{1}{m}}, \qquad (4)$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_d)$ are points in $\Re^d$. Particularly, L_2-norm is Euclidean distance, L_1-norm is Manhattan distance, and L_∞ is Tchebyshev distance.

However, Aggarwal et. al [1] observed that L_m-norm loose its meaningfulness of proximity distance when the dimension d and the values of m increase. Thus, for example, Euclidean distance is the best (among L_m-norms) similarity metric when d is about 5. For higher d, metrics with lowers m (i.e. Manhattan distance) should be used.

Based on this observation, Aggarwal introduced *fractional distance metric* with $0 < m < 1$, arguing that such a choice is more appropriate for high-dimensional spaces. Experiments, reported in [3], partially confirmed efficiency of this proposition. For $1 < m < 0.5$, more samples were detected, in comparing to L_2 and L_1 norms. However, for $m < 0.5$ efficiency rapidly decreased and for $m = 0.2$, none samples were detected. Moreover, the same experiments showed also a trade-off between efficiency, time complexity and m. For fractional norms, the algorithm runs slower for lowers m;
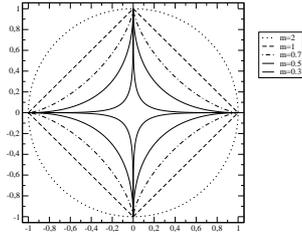
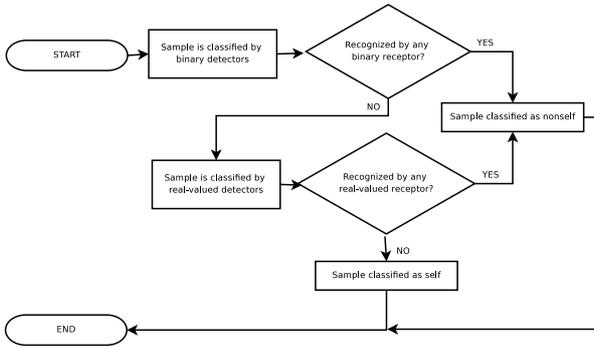Fig. 3.   Unit spheres for selected L_m norms in 2D.

Fig. 4.   Scheme of the classification process for dual representation receptors.

duration of learning processes for $L_{0.5}$ were even 2-3 times longer than for L_2.

One of the additional consequence of applying fractional metrics for V-Detector algorithm is the change of the shape of detectors. Figure 3 presents the unit spheres for selected fractional L_m-norms in 2D.

III. PROPOSED APPROACH

In [4] we proposed to use the dual representation of samples. The main motivation was to reduce duration of classification process as well as to improve the detection rate, especially in the case of high-dimensional datasets. When we attain this goal, then immune algorithms will be even more attractive alternative to traditional, i.e. *not immune* approaches in many domains i.e. computer security, spam detection, etc.

For example, IDSs which can represent on-line classification systems, should be able to analyze all network connections, regardless of intensity network traffic. The large number of various parameters which should be taken into consideration and still increasing the number of signatures of potential attackers, requires the use of computers with higher and higher computational power. Thus, signatures of normal (legal) behavior, offered by *negative selection*, seems to be reasonable solution, especially, when more effective models and algorithms will be developed.

The scheme of dual receptors is presented in Figure 4. It involves generating both types of detectors. Binary detectors, plays only the role of preliminary detection and those samples which are not recognized by them are censored by V-Detector algorithm. Thus, we do not expect that binary receptors covers the space N in very high degree, as it can takes too much time. More important aspect is its length. They should be relatively

short (with high generalization degree) to detect in the possible shortest time as much as possible nonself samples and then the profits of this approach will increasing. The optimal length of binary detectors r can be easily tuned. Usually, it is placed at the end of *phase transition region* (see Section II-C1). Then, both probability of detecting detectors as well as the corresponding to it coverage of N space are sufficiently high.

To construct binary receptors, after normalization to unitary hypercube, self samples should be turned into binary representation. For every attribute x ($x \in [0, 1]$) the quantization function Q can be expressed as:

$$Q(x) = \lfloor M * x \rfloor, \tag{5}$$

where $\lfloor \cdot \rfloor$ is floor function and M is the number of ranges (clusters). In our case, the optimal number of clusters is $M = 2^{bpa}$, for $bpa = \{1, 2, \ldots\}$ (bpa - bits per attribute).

In [4] and [3] we conducted experiments on datasets containing about 30-40 attributes. For such dimensionality, Euclidean and Manhattan distance metrics provide too low detection rates, what was confirmed by performed experiments. As could be expected, fractional distances produced quite good results (about 70-80%), but we should keep in mind that, in comparison to L_2 and L_1 norms, they are much more computationally complex.

Unfortunately, there are many datasets which consist of 50 and more attributes (even up to 250), i.e. spam detection, handwritten recognition and even time series. In this case, probably none of known metrics is able to measure the similarity properly when the whole set of attributes is used. Thus, in this paper, we propose to incorporate the sliding window for both binary as well as real-valued detectors.

Let w ($w \le d$) be the window's length for real-valued detectors. Then, to censor a sample by only one detector, there is a need to calculate $d - w + 1$ distances (instead of 1 in original case), in the pessimistic case. Hence, for example, if $d = 100$, $w = 20$ and $|D| = 1000$, even about 81000 such operation can be needed to determine, if a sample belong to set S or N. Similarly, process of detector's maturation, which depends on d, w and S, will be also much more durable. Thus, especially for very high-dimensional datasets, binary receptors plays very important role in proposed model.

IV. EXPERIMENTS AND RESULTS

This section presents only a small part of conducted experiments with most valuable results. Several datasets (presented in I) from Machine Learning Repository and Keogh's were used. For all dataset relatively shorts binary receptors were generated with lengths $l = \{5, 7, 10, 12, 15, 17\}$

For all datasets, relatively shorts binary receptors were generated with lengths $r = \{5, 7, 10, 12, 15, 17\}$ and $bpa = \{1, 2, 3\}$ to examine its usefulness in our model. Selected results are presented in Table II.

One can see, that binary detectors allow to at least 30-40% detection of nonself samples. Moreover, for properly tuned values r and bpa, its efficiency increases even up to 80-100%

TABLE I
BRIEF DESCRIPTION OF USED DATASETS.

Dataset	Number of attr.	Samples count	Self count
Spambase	54	4601	2788
Madelon	250	2000	1000
Leaf	160	221	33
CBF_B	128	5000	1723
KDD_ICMP	41	11910	4427

TABLE II
PERCENT OF NONSELF SAMPLES RECOGNIZED BY BINARY RECEPTORS
FOR VARIOUS r AND bpa VALUES.

Dataset	bpa	Rcp. len.	Rcp. count	Det. [%]
Spambase	1	10	267	6.81
		12	2353	10.58
		15	29015	24.10
		17	126029	33.57
Madelon	1	10	50	2.50
		12	834	16.10
		15	18530	68.60
		17	100012	95.50
Leaf	1	5	10	21.27
		7	94	55.85
		10	974	75.00
		15	32620	79.25
	2	7	75	90.95
		10	875	96.80
		15	32355	98.93
	3	12	3540	100.00
CBF_B	1	12	259	34.26
		15	18804	54.28
		17	109255	58.22
	2	12	2754	72.26
		15	29078	81.84
		17	124673	88.98
KDD_ICMP	1	12	259	34.26
		15	32536	42.02

Fig. 5. Spambase. Detection rate of V-Detector algorithm.

Fig. 6. Madelon. Detection rate of V-Detector algorithm.

Fig. 7. Spambase. Number of detectors generated by V-Detector algorithm.

Moreover, one can see, for Spambase dataset Euclidean distance gives an optimal results when $w = 10$; for other, considered in this paper similarity metrics, higher efficiency is observed for $w = 15$. It is consistent with presented theoretical reflections [1] and experimental results [3]. On the other hand, for Madelon dataset, detection increase with w (up to 80%).

The duration of classification process, in the case of V-Detector algorithm is proportional to the cardinality of receptor's set. Also, in this case, obtained results are different for presented datasets. For Spambase, number of detectors decreases when w increase. However, for Madelon dataset, we can observe, that for $w = 15$ number of receptors is lowest. In contrast to Spambase, it does not decreases for high w values, what results in very high detection rates.

Finally, the best combinations of parameters for both types of detectors were chosen to achieve the most optimal solutions for our model, considered in this paper. Table III presents the most valuable results. For most testing datasets results are satisfactory. Only for Spambase, detection of *nonself* samples is only slightly higher than 60%.

V. CONCLUSION

The model presented in this paper base on the negative selection mechanism, developed within domain of Artificial

for almost all tested datasets. Thus, one can expected, that such results, should significantly lighten the real-valued detectors.

Next, there was investigated the possibility of applying real-valued receptors for testing datasets. V-Detector algorithm was executed with the following parameters (like in all the experiments described in this section):

- window's length $w = \{5, 10, 15, 20, 25, 30\}$,
- distance norms: L_2, L_1, $L_{0.7}$, $L_{0.5}$,
- detector's radius $d_r = 0.001$,
- estimated coverage $co = 99.99\%$,
- all self samples were used in learning stage to generate receptors. In consqequence, false positives (number of incorrectly classified self samples) is equal 0 in all cases.

Figures 5 and 6 shows the detection rates for Spambase and Madelon datasets, respectively. It is worth to emphasize, that in our case, detection rates denotes only the number of correctly recognized nonself samples, without correctly classified all selfs. Thus, is difficult to compare obtained results with other approaches, which incorporate all types of samples in at the learning stage.
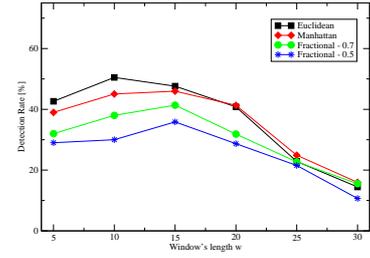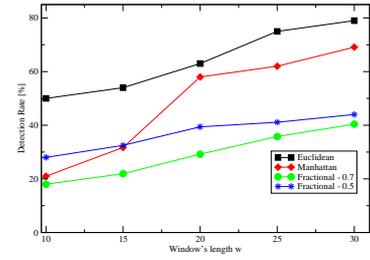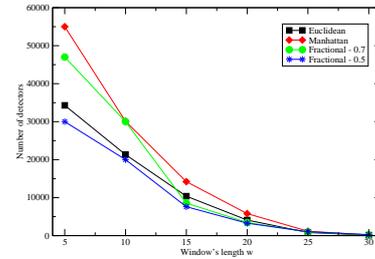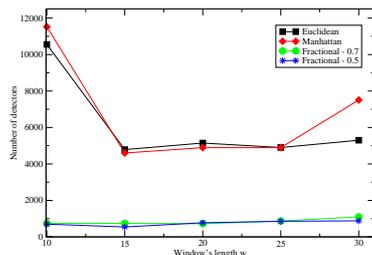
Fig. 8.   Madelon. Number of detectors generated by V-Detector algorithm.

TABLE III
THE MOST VALUABLE DETECTION RATES FOR SELECTED DATASETS.

Dataset	bpa	r	m	w	Det.[%]
Spambase	1	12	1	15	45.91
	1	15	1	15	49.10
	1	17	1	15	61.10
	2	12	1	15	45.15
	2	17	1	15	47.20
	2	12	2	10	50.28
	2	15	2	10	50.71
Madelon	1	12	2	20	71.45
	1	15	2	20	88.10
	1	17	2	20	98.50
Leaf	1	10	2	15	100.00
	2	10	2	15	100.00
CBF_B	2	12	1	15	100.00
	2	12	2	15	100.00
KDD_ICMP	1	10	2	10	86.61
	1	10	0.7	10	87.99

Immune Systems. It is designed for high-dimensional datasets, which are very difficult to classify, due to the lack of appropriate similarity metrics. Sliding window method applied for both types of detectors can be viewed as the one of the possibility to overcome the scaling problem. However, although the detection rates where quite satisfactory, duration of learning as well as classifying processes leave a lot to be desired. Fortunately, significant (and in some cases almost all) part of nonself samples can be relatively quickly detected by binary receptors, which are the power of this model.

Generally, applying only one type of receptors is not able to guarantee the sufficient coverage of nonself space N in reasonable time, especially, when dataset contains a thousands and more samples. Moreover, even slight increase in one of the following set of parameters: d, w and $|S|$, significantly influence on the performance of applied algorithms.

Obtained results are very preliminary and should be confirmed with other datasets in the future, regardless of its applications. Moreover, in our opinion, presented model can be, almost directly, applied to detect the various types of

anomalies in time series. And it should be a direction of our future investigations.

ACKNOWLEDGMENT

REFERENCES

[1] Aggarwal C. C., Hinneburgand A., Keim D. A., *On the Surprising Behavior of Distance Metrics in High Dimensional Space*, Lecture Notes in Computer Science, Vol. 1973, 2001, pp. 420–434.
[2] Balthrop J., Esponda F., Forrest S., Glickman M., *Coverage and generalization in an artificial immune system*, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2002), New York, 9-13 July 2002, pp. 3–10.
[3] Chmielewski A., Wierzchoń S. T., *On the distance norms for multidimensional dataset in the case of real-valued negative selection application*, Zeszyty Naukowe Politechniki Białostockiej, No. 2, 2007, pp. 39–50.
[4] Chmielewski A., Wierzchoń S. T., *Dual representation of samples for negative selection issues*, Computer Assisted Mechanics and Engineering Sciences, Vol. 14, No. 4, 2007, pp. 579–590.
[5] Forrest S., Hofmeyr S. A., Somayaji A., Longstaff T. A., *A Sense of Self for Unix Processes*, Proceedinges of the 1996 IEEE Symposium on Research in Security and Privacy, IEEE Computer Society Press, 1996, pp. 120–128.
[6] Forrest S., Perelson A., Allen L., Cherukuri R., *Self-nonself discrimination in a computer*, in Proceedings of the IEEE Symposium on Research in Security and Privacy, Los Alamitos, 1994, pp. 202–212.
[7] Harmer P. K., Wiliams P. D., Gunsch G. H., Lamont G. B., *Artificial immune system architecture for computer security applications*, IEEE Transactions on Evolutionary Computation, Vol. 6, 2002, pp. 252–280.
[8] Hofmeyr S. A., *An Immunological Model of Distributed Detection and its Application to Computer Security*, Ph.D. thesis, Department of Computer Sciences, University of New Mexico, 1999.
[9] Hofmeyr S. A., Forrest S., *Architecture for an artificial immune systems*, Evolutionary Computation, Vol. 8(4), 2000, pp. 443–473.
[10] Ji Z., Dasgupta D., *Real-valued negative selection algorithm with variable-sized detectors*, Genetic and Evolutionary Computation GECCO-2004, Part I, LNCS Vol. 3102, Seattle, WA, USA, Springer-Verlag, 2004, pp. 287–298.
[11] Percus J. K., Percus O. E., Perelson A. S., *Predictiong the size of the T-cell receptor and antibody combining region from consideration of efficient self-nonself discrimination*, Proceedings of National Academy of Sciences USA (90), 1993, pp. 1691–1695.
[12] Perelson A., Weisbuch D., *Immunology for physicists*, Reviews of Modern Physics, Vol. 69, 1977, pp. 1219–1265.
[13] Snort, Intrusion Detection System, http://www.snort.org.
[14] Stibor T., Mohr P., Timmis J., Eckert C., *Is Negative Selection Appropriate for Anomaly Detection?*, in Proceedings of the ACM SIGEVO Genetic and Evolutionary Computation Conference (GECCO 2005), Washington, D.C., June 25-29, 2005, pp. 321–328.
[15] Stibor T., Timmis J., Eckert C., *The Link between r-contiguous Detectors and k-CNF Satisfiability*, Part of the World Congress On Computational Intelligence, Canada, 2006, pp. 491–498.
[16] Stibor T., *Phase Transition and the Computational Complexity of Generating r-contiguous Detectors*, Proceedings of 6th International Conference on Artificial Immune Systems, LNCS (4628), 2007, pp. 142–155.
[17] Weber A., Schek H. J., Blott S., *A quanttitative analysis and performance study for similarity-search methods in high-dimensional spaces*, 24th International Conference Very Large Data Bases, 1998, pp. 194–205.
[18] Wierzchoń S. T., *Sztuczne systemy immunologiczne*, Teoria i zastosowania. Akademicka Oficyna Wydawnicza EXIT, Warszawa 2001.