Evolutionary Negative Selection Algorithms for Anomaly Detection

Wenjian Luo[1] Jiying Wang[2] Xufa Wang[1]

[1] Department of Computer Science and Technology, University of Science and Technology of China, HeFei, Anhui 230027, China

[2]Department of Electronic Science and Technology, University of Science and Technology of China, Hefei, Anhui, 230026, China

Abstract

Biological immune system has a strong self-protection function. On the one hand, biological immune system is an anomaly detection system because it has a very good immune detection mechanism based on discrimination between self and non-self. On the other hand, biological immune system is a self-adaptive complex system because it has a very good immune learning mechanism based on evolution. By combining the immune detection mechanism and immune learning mechanism, two evolutionary negative selection algorithms (ENSAs) are proposed in this paper, which are called as simple ENSA and basic ENSA. Test experiments are given and the results prove that these two evolutionary negative selection algorithms are effective.

Keywords: Artificial Immune System; Evolutionary Negative Selection Algorithm; Anomaly Detection

1. Introduction

Artificial Immune System (AIS) is a novel intelligent computing research field after Artificial Neural Network and Evolutionary Computation, and it is an emergent interdisciplinary field derived from life science and computer science [1-3]. The purpose of Artificial Immune System is to extract special information processing mechanisms contained in biological immune system, and then to study and design the corresponding models and algorithms, which could be used to solve many kinds of complex problems [3].

The basic function of biological immune system is to recognize self and non-self, and then to classify and eliminate non-self. Biological immune system has immune recognition, immune memory, immune regulation, immune tolerance, immune surveillance and other characteristics. It is a self-adaptive, self-learning, self-organization, parallel and distributed complex system [3-4]. Meanwhile, biological immune system can be regarded as an anomaly detection system because it recognizes pathogens by antibody, which represents many non-self patterns [3]. So the evolutionary mechanism and anomaly detection mechanism are two important aspects of biological immune system.

This paper proposes two evolutionary negative selection algorithms, which are called as simple ENSA and basic ENSA, by introducing immune non-self learning operators based on affinity mutation, random generating (bone marrow and receptor editing) mechanism in succession. Both simple ENSA and basic ENSA are designed to effectively solve the detector adaptively generating problem and fast detect abnormal changes. The corresponding test experiments are given to prove their universal validity.

2. An Evolutionary Anomaly Detection System

Biological immune system has good learning ability based on evolution and can detect abnormal external pathogens and internal changes based on negative selection. Consequently, to some extent, biological immune system can be regarded as an evolutionary anomaly detection system. In other words, the main characteristics of biological immune system can be summarized into two major aspects:

(1) Anomaly detection mechanism, namely, discrimination mechanism between self and non-self based on negative selection. The biological immune system recognizes non-self by matching antibody with antigen. Antibody is a representative of some non-self patterns. And all mature antibodies should be negatively selected before they enter lymphoid circulation and are used to detect abnormal antigens.

Based on the negative selection mechanism in the course of the generation of T-Cells of biological immune system, Negative Selection Algorithm (NSA) is proposed [5]. S. Forrest and her colleagues also proposed detectors greedy generating algorithm [6]. NSA has several excellent features. It has been paid

much attention to by many researchers from different fields, especially anomaly detection [2, 3, 7].

(2) Evolutionary mechanism, i.e., immune learning mechanism based on evolution. The antigen pattern space is very large. In order to protect biology itself at all circumstances, biological immune system needs to generate a best antibody set to fast recognize and eliminate current antigens. Therefore, the immune learning mechanism based on evolution plays an important role in biological immune response.

The immune learning mechanism based on evolution is very complex in biological immune system, including affinity mutation, random generating (bone marrow and receptor editing), gene library evolution, immune memory, immune network regulation, and so on [3, 4, 7-14]. And some of them have been modeled and used in some applications [2-3, 7, 9-12, 14].

It is noted that negative selection mechanism and immune evolution mechanism are tightly united in the course of antibody growing and maturating, in order to recognize and eliminate antigens fast. So we can combine these two major aspects to propose some new algorithms, namely, evolutionary negative selection algorithms (ENSAs).

3. ENSAs

In this section, evolutionary negative selection algorithms are proposed by integrating immune evolution models with negative selection model.

3.1 Immune metaphors

The growing and maturing process of antibody can be approximately divided into several phases [3-4, 8-9]:

(1) Bone marrow generates many immature antibodies. Immature antibodies may result in autoimmunity.

(2) Immature antibodies grow into mature antibodies after negative selection. Here mature antibodies will not recognize any self individual, and all of them denote some non-self patterns. The antibody that we often talk about means mature antibody.

(3) Mature antibodies are released, cruising in body and stimulated by antigens. If the affinity between an antibody and an antigen reaches a certain threshold, the antibody can recognize the antigen effectively, and then eliminate it.

(4) To recognize and eliminate a great deal of antigens, biological immune system maintains a dynamically updated antibody set [8]. According to clone selection principle proposed by Burnet, the antibody with higher affinity will be activated and rapidly differentiates and proliferates. Therefore, a lot of similar antibodies are generated. The higher the affinity between the antibody and the antigen, the more antibodies are generated. Contrariwise, the antibody with low affinity will decrease and be replaced by new antibodies. This is called as affinity maturation. Meanwhile, Bone marrow continually recruits new antibodies to renew the antibody set.

By exploring the process of antibody's growing and maturing, we can know that there are three important steps: (1) negative selection; (2) affinity maturation; (3) generating new antibodies by bone marrow. Since mutation is the main means of affinity maturation, we would like to simply regard affinity maturation as affinity mutation.

3.2 Affinity mutation

Affinity means the match degree between the antibody and the antigen, i.e., the binding degree between them. The affinity decides whether an antibody will be activated or not, and how it is activated [3, 9]. The higher the binding degree is, the higher the activated degree is, and the more rapidly it proliferates. Meanwhile, the antibody with lower affinity will not be activated and decrease, and disappears finally. The antibody with middle affinity may be activated, but its proliferating speed is lower and the proliferated amount is smaller.

In fact, it is an Activating-and-Proliferating process. The following simple ENSA is proposed according to this idea.

3.3 Simple ENSA

Figure 1 shows the flowchart of a simple ENSA. This simple ENSA is proposed as a normal algorithm in this paper.

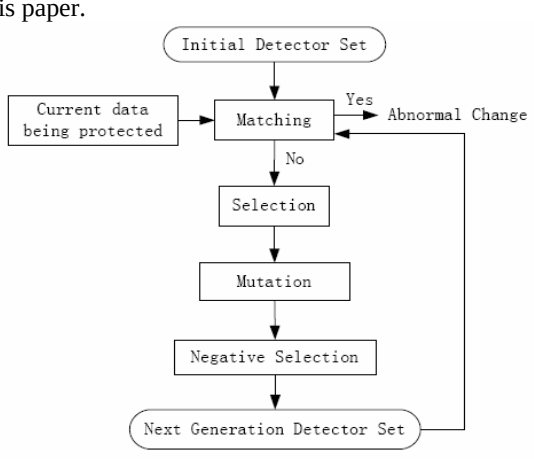


Fig. 1: Simple Evolutionary Negative Selection Algorithm

This simple ENSA is inspired by the Activating-and-Proliferating process and negative selection model.

Simple ENSA can be summarized into the following several steps:

(1) Generating an initial detector set. Since detector set is often a small one in ENSA, the initial detector set can be generated at random.

(2) Matching. Match the current data being protected with all detectors in the detector set. The match degree is also called as affinity. If the detector matches the data successfully, an abnormal change has occurred. Matching rule can be r-contiguous-bits matching, Hamming distance matching, and so on.

(3) Selection. The higher the affinity is, the higher selection probability is. The selection strategy can be roulette wheel method, fitness proportional method, rank-based method, and so on.

(4) Mutation. Then immature detectors are generated according to the antibodies' affinity. The higher affinity is, the more immature detectors are generated by mutation.

(5) Negative selection. All new immature detectors are filtered by negative selection. Some of them are removed because of matching self, and others remain in the next generation detector set.

(6) The next generation detector set. The next generation detector set is generated by step (3) ~ step (5). Then, go to step (2) and continue to match and monitor.

3.4 Random generating mechanism

In reference [13], Tonegawa pointed out that somatic recombination and mutation are the most important mechanisms to maintain the diversity of antibodies. Bone marrow continually and randomly generates new antibodies to add into the current antibody set, while the antibody with low affinity in current antibody set will die gradually. It is estimated that every day about 5~8 percent (or 10 percent) antibodies in current antibody set die and are replaced by new antibodies generated by bone marrow [8].

Moreover, there is another similar mechanism named as receptor editing. B-cells in biological immune system can delete their low-affinity receptors and replace them by new receptors [9]. In other words, the B-cell changes its V-region (Variable Region) by recombination. Here V-region is the binding area with antigen. This is the meaning of receptor editing. From the point of view of modeling and designing algorithms, because B cells use receptors to recognize antigens, receptor editing can be regarded as a new antibody generating mechanism which is similar to bone marrow's function. Both of them belong to a mechanism that maintains the diversity of antibodies by generating new random antibodies. So they can avoid running into local optimizations.

3.5 Basic ENSA

Based on negative selection, affinity mutation and random generating mechanism of biological immune system, a more reasonable evolutionary negative selection algorithm is proposed, which is called as basic ENSA.

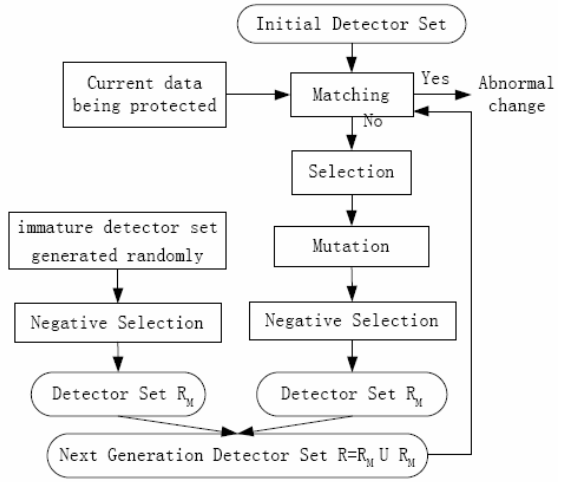


Fig. 2: Basic Evolutionary Negative Selection Algorithm

Figure 2 shows the structure of basic evolutionary negative selection algorithms.

Basic ENSA can be summarized into the following several steps:

(1) Generating an initial detector set.

(2) Matching.

(3) Selection.

(4) Mutation.

(5) Negative selection. This new detector set is denoted as R_M. The size of R_M accounts for about α percent of the size of R (total detector set).

(6) Generating new detectors randomly. The immature detectors are generated at random and become mature detectors after negative selection. This new detector set is denoted as R_R. The size of R_R accounts for about β percent of the size of R, and $\alpha + \beta = 1$.

(7) The next generation detector set R ($R = R_M \bigcup R_R$).The next generation detector set is generated by step (3) ~ step (6). Then, go to step (2) and continue to match and monitor.

Since basic ENSA uses affinity mutation and random generating mechanism for reference, it can search in the global space and avoid sinking into local optimizations.

In the practical implementation of basic ENSA, we can adjust the ratio of the size of R_M and R_R. According to the principle of biological immune system, we set $\beta \approx 5\% \sim 10\%$. Thus $\alpha \approx 90\% \sim 95\%$.

4. Experiments

Generally, we take 0-1 strings for example. String length is denoted by l. The partial match rule is r-contiguous-bits matching rule [5]. A detector's affinity f is defined as:

For two string X and Y, their affinity is $l - HD(X,Y)$. Here $HD(X,Y)$ means the Hamming distance and l means the string length.

If the number of current strings being detecting is $m(m \geq 1)$, and the corresponding affinity is $f_1, f_2, \cdots, f_m$, then a proper definition of the detector's affinity f is:

$$f = \max(f_1, f_2, \cdots, f_m)$$

For each selected detector d, kf new mature detectors are generated by mutation. Therefore, these kf detectors are similar to d but not the same as d. Here f is the detector's affinity, and k means a proliferation coefficient and it is an integer larger than zero. In this paper, we set $k = 1$.

In basic ENSA, the next generation detector set consists of two subsets. One is generated randomly; another is generated by affinity mutation. The next generation detector set consists of two parts. In this paper, we set $\beta \approx 10\%$ and $\alpha \approx 90\%$.

In the following experiments, we set the size of self set is 1000, the size of detector set is 200, the maximum evolutionary generation is 250.

4.1 Single abnormal string

We firstly test these two algorithms with the simplest experiments.

In these experiments, string length $l = 64$, partial matching parameter $r = 62$.

Figure 3 shows the detection results of 20 independent runs of simple ENSA. From figure 3, the evolutionary generation varies from 4 to 13, and the average evolutionary generation is about 9.

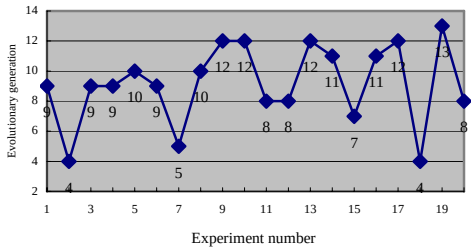


Figure 3: Detecting single abnormal string (Simple ENSA)

Figure 4 shows the detection effect of 20 independent runs of basic ENSA. From figure 4, the evolutionary generation varies from 4 to 12, and the average evolutionary generation is 8.65.

Compared figure 3 with figure 4, it is obvious that simple ENSA and basic ENAS has a almost equivalent detection effect in these experiments.

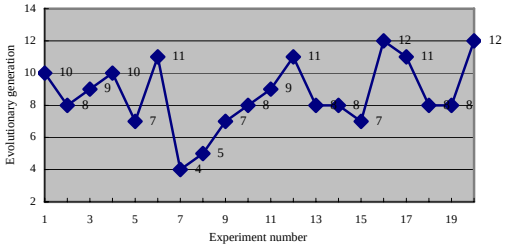


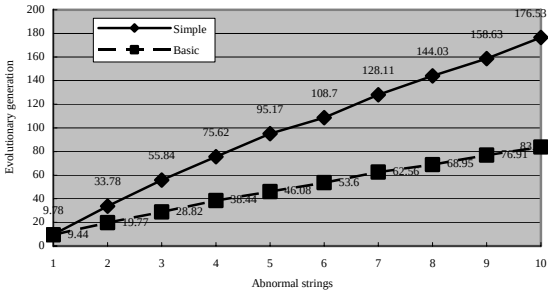
Figure 4: Detecting single abnormal string (Basic ENSA)

4.2 Several abnormal strings

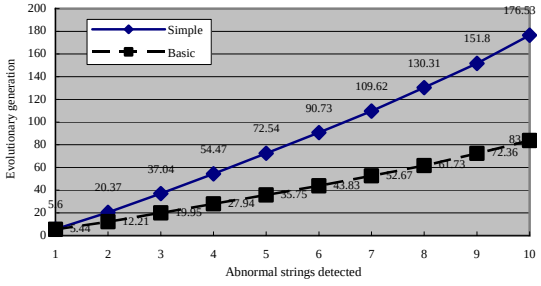
When some abnormal strings are detected together in these experiments, two different cases should be respectively considered. One is that some abnormal strings occur simultaneously, and the other is that some abnormal strings occur at intervals.

In these experiments, string length $l = 56$, partial matching parameter $r = 55$.

(1) Firstly, we consider that several abnormal strings occur simultaneously.



(a) Multi abnormal strings occurs simultaneously (1-10)



(b) The detecting process when ten abnormal strings occur simultaneously

Figure 5: Detecting multi abnormal strings

Figure 5(a) shows the detection effect when the abnormal strings number varies from 1 to 10. The evolutionary generation is an average of 100 independent runs.

From figure 5(a), we can see that basic ENSA is much better than simple ENSA when the abnormal strings are more than one. Moreover, the contrast between them is more obvious as the abnormal strings number increases. When the number of abnormal

strings is 2, simple ENSA needs an average 33.78 generations to detect these two abnormal strings, while basic ENSA only needs an average 19.77 generations. When the number of abnormal strings is 5, simple ENSA needs an average almost 100 generations, while basic ENSA only needs an average 46.08 generations.

To further explore the detection process of these two algorithms, when ten abnormal strings occur at the same time, figure 5(b) shows the average generations as the abnormal string number varied from one to ten. The evolutionary generation is an average of 100 independent runs.

From figure 5(b), we can conclude that these two algorithms detect the first abnormal string almost at the same time. Simple ENSA needs an average 5.6 generations, while basic ENSA only needs an average 5.44 generations. But after that, simple ENSA needs an average 15~18 generations to detect the second abnormal strings, while basic ENSA only needs an average 7 generations or so. The larger the number of abnormal strings is, the more obvious the advantage of basic ENSA is.

Table 1. Detecting some abnormal strings which occur at intervals

	Total ten abnormal strings (Two abnormal strings occur every 10 generations)									
ASD	1	2	3	4	5	6	7	8	9	10
Simple	8.46	28.98	47.25	64.45	82.29	101.03	120.39	140.23	160.81	185.55
Basic	8.38	18.90	28.17	37.08	45.83	53.07	62.86	72.14	81.96	94.06
	Total ten abnormal strings (One abnormal string occurs every 5 generations)									
ASD	1	2	3	4	5	6	7	8	9	10
Simple	9.30	19.61	29.08	38.30	47.02	55.40	64.23	73.79	84.08	95.34
Basic	8.91	19.83	29.33	38.57	47.09	55.72	64.41	73.67	83.87	95.29
	Total five abnormal strings (Two abnormal strings occur every 10 generations)									
ASD	1	2	3	4	5	—	—	—	—	—
Simple	7.94	18.48	27.87	38.24	50.02	—	—	—	—	—
Basic	8.22	18.88	29.02	39.14	50.81	—	—	—	—	—
	Total five abnormal strings (One abnormal string occurs every 5 generations)									
ASD	1	2	3	4	5	—	—	—	—	—
Simple	10.09	32.87	53.00	74.83	100.93	—	—	—	—	—
Basic	9.46	20.07	29.80	39.93	51.86	—	—	—	—	—

(2) Secondly, several abnormal strings occur at intervals.

Some abnormal strings occur now and then. It means that some new abnormal strings abruptly happen while the algorithms are detecting the current abnormal strings.

Table 1 shows the experiment results when some abnormal strings occur now and then. In table 1, “ASD” means the abnormal strings detected, “Simple” means the generation that simple ENSA needs to detect the corresponding number of abnormal strings, and “Basic” means the generation that basic ENSA needs to detect the corresponding number of abnormal strings.

From table 1, compared with simple ENSA, basic ENSA has a much better performance.

Therefore, from table 1 and figure 5(a), despite that the several abnormal strings occur at the same time or at intervals, basic ENSA has a much better performance than simple ENSA.

5. Related works

There are some related works about how to generate detectors. S. Forrest and her colleagues proposed an exhaustive detector generating algorithm and greedy detector generating algorithm [5-6]. Some preliminary works about negative selection with mutation are also discussed [12]. M. Ayara and J. Timmis have given a summary in reference [15].

However, biological immune system can be regarded as an evolutionary anomaly detection system. This opinion denotes two sides of biological immune system, including evolutionary mechanism and anomaly detection mechanism. The evolutionary mechanism includes affinity mutation, detector

randomly generating mechanism, and so on. And the anomaly detection mechanism is based on negative selection model. This paper focuses on the design of two versions of ENSAs and the performance comparison between them, which are inspired by viewing biological immune system as an evolutionary anomaly detection system. Experimental results have proved that ENSAs have a good performance.

6. Conclusions

By extracting the rich information processing mechanisms of biological immune system, this paper proposes two evolutionary negative selection algorithms (ENSAs). Firstly, a simple evolutionary negative selection algorithm is introduced, which is proposed by integrating negative selection model with affinity mutation mechanism. Secondly, a basic evolutionary negative selection algorithm is introduced by integrating negative selection model with affinity mutation mechanism and random generating mechanism. We call the first one as a simple ENSA because it has good local search ability. And the second one is entitled as a basic ENSA because it not only has good local search ability, but also considers global search methods. In fact, basic ENSA uses the random generating mechanism to jump off the local optimizations. The work is not only very important to design novel intrusion detection mechanism, but also very significant to drive the related application research on anomaly detection.

Acknowledgement: This work is supported by NSFC Foundation (60404004), Post-doctor Foundation (2003034433) and Nature Science Major Foundation from Anhui Education Bureau (2004kj360zd).

7. References

[1] A O Taraknaov, V A Skormin and S P Sokolova, "*Immunocomputing: Principles and Applications*", Springer-Verlag, 2003.

[2] L N de Castro, J Timmis, "*Artificial immune systems: a new computational intelligence approach*", Springer-Verlag, London, 2002.

[3] Wenjian Luo, "*Research on artificial immune models and algorithms for intrusion detection*", Ph.d thesis, Department of Computer Science and Technology, University of Science and Technology of China, Hefei, China, 2003.

[4] Qi Anshen, Du Chanying, "*Nonlinear models in immunity*", Shanghai Scientific and Technological Education Publishing House, Shanghai, 1998.

[5] S Forrest, A S Perelson, L Allen, R Cherukuri, "*Self-Nonself discrimination in a computer*", Proceedings of IEEE Symposium on Research in Security and Privacy, pp. 202-- 212, Oakland, CA, 16-18 May 1994.

[6] P D'haeseleer, S Forrest, P Helman, "*An immunological approach to change detection: algorithms, analysis and implications*", Proceedings of the IEEE Symposium on Security and Privacy, IEEE Computer Society Press, 1996.

[7] J Kim, P J Bentley, "*Negative selection and niching by an artificial immune system for network intrusion detection*", Proceedings of Genetic and Evolutionary Computation Conference (GECCO'99), Orlando, Florida, pp.149-158, July 12-17 1999.

[8] S A Hofmeyr, "*An interpretative introduction to the immune system. Design Principles for the Immune System and other Distributed Autonomous Systems*", Oxford University Press, I Cohen and L Segel (Eds), 2000.

[9] L N de Castro, F J Von Zuben, "*Learning and optimization using the clonal selection principle*", IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems, 6(2), pp. 229-251, 2002.

[10] J Timmis, M Neal, "*A resource limited artificial immune system for data analysis*", Knowledge Based Systems, 14(3-4): 121-130, June 2001.

[11] Wenjian Luo, Xianbin Cao, Xufa Wang, "*An immune genetic algorithm based on immune regulation*", Proceedings of the 2002 Congress on Evolutionary Computation, Honolulu, Hawaii, pp. 801-806, 2002.

[12] Wenjian Luo, Xianbin Cao, Xufa Wang, "*NIDS Research Based on Artificial Immunology*", International Conference on Information and Communications Security 2001 (ICICS'2001), XiAn, China, Lecture Notes in Computer Science 2229, Springer, pp. 371-375, 2001.

[13] Susumu Tonegawa, "*Somatic generation of antibody diversity*", Nature, 202(14): 575-581, 1982.

[14] L N de Castro, F J Von Zuben, "*The clonal selection algorithm with engineering applications*", Proceedings of GECCO'00, Workshop on Artificial Immune Systems and Their Applications, pp. 26-27, 2000.

[15] M Ayara, J Timmis, L N de Lemos, R de Castro, R Duncan, "*Negative selection: How to generate detectors*", Proceedings of the First International Conference on Artificial Immune Systems, J. Timmis and P.J. Bentley (editors), pp. 89-98, September 2002.