Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 19 (2009) 261–266
Progress in Natural Science
www.elsevier.com/locate/pnsc

Short communication

A self-adaptive negative selection algorithm used for anomaly detection

Jinquan Zeng, Xiaojie Liu *, Tao Li, Caiming Liu, Lingxi Peng, Feixian Sun

Department of Computer Science, Sichuan University, Chengdu 610065, China

Received 18 April 2008; received in revised form 30 May 2008; accepted 3 June 2008

Abstract

A novel negative selection algorithm (NSA), which is referred to as ANSA, is presented. In many actual anomaly detection systems, the training data are just partially composed of the normal elements, and the self/nonself space often varies over time. Therefore, anomaly detection system has to build the profile of the system based on a part of self elements and adjust itself to adapt those variables. However, previous NSAs need a large number of self elements to build the profile of the system, and lack adaptability. In order to overcome these limitations, the proposed approach uses a novel technique to adjust the self radius and evolve the nonself-covering detectors to build an appropriate profile of the system. To determine the performance of the approach, the experiments with the well-known dataset were performed. Results exhibited that our proposed approach outperforms the previous techniques.

Keywords: Artificial immune system; Anomaly detection; Negative selection algorithm

1. Introduction

The biological immune system (BIS) is of great interest to computer scientists, because it provides a unique and fascinating computational paradigm for solving complex problems. Inspired by BIS, artificial immune system (AIS) has become a vibrant and active research area [1–4]. Currently, major types of AIS methods include negative selection algorithm (NSA), clone selection and immune network model [5]. Forrest and Perelson proposed the initial NSA and used binary encoding to represent normal and abnormal space [6]. Later, a real-valued approach was presented [7]. In order to cover the abnormal space and generate good detectors, the genetic technique [8] was proposed. Among the latest studies on the NSAs, Zhou and Dipankar [9–11] presented a real-valued negative selection algorithm using V-detector, and Balachandran et al. [12] proposed a multi-shaped detector negative selection algorithm.

The NSA is one of the most successful methods in AIS, and its typical applications include change detection, fault detection, and network intrusion detection [5]. The NSA is believed to have distinct process from alternative methods and to be the most effective algorithm available [13]. Although there have been a lot of successful applications of the NSA, some problems still exist to prevent the AIS and the NSA from being applied extensively.

Firstly, for the scalability of the NSA, many researchers have found this problem, for example, Harmer et al. [14] found in their tests that CVIS (computer virus immune system) required approximately 1.45 years for generating antibodies (detectors) to scan an 8 GB hard disk drive. Kim and Bentley [15] studied the problem in-depth and concluded that the NSA produced poor performance due to scaling issues on real-world problems. Simultaneously, the cost for the detectors training is exponentially related to the size of self set [16], and so the profile of the system is only represented by part elements of the self space.

Secondly, the low-level representation of detectors prevents the extraction of meaningful domain knowledge. It is difficult to map back to problem space, e.g. binary

* Corresponding author. Tel./fax: +86 2885405568.
E-mail address: liuxiaojie8@126.com (X. Liu).

1002-0071/$ - see front matter
doi:10.1016/j.pnsc.2008.06.008

representation [9]. Balthrop et al. [17] also pointed out that the problem of scalability of the NSA could lie in the representation and the *r*-continuous match rule, not in the negative selection process itself.

Thirdly, self and nonself space often varies over time, e.g. a computer administrator often stops some network services and new network attacks always occur. So anomaly detection system should adjust the built profile of the system in real time and adapt these varieties.

However, almost in all NSAs [6,9–12], the self radius is in a constant size, so these methods cannot build an appropriate profile of the system and lack adaptability. While our proposed approach uses a novel technique to adjust the self radius and evolve the nonself-covering detectors, which appears to outperform the previous techniques in building the profile of the system and covering the abnormal space. It is applied to perform anomaly detection for the well-known dataset, and it is a general approach that can be applied to different anomaly detection problems.

2. The proposed approach

2.1. Self generalization

Building the profile of the system on part of the self elements in a two-dimensional space is illustrated in Figs. 1–3. The self radius of the self element specifies the capability of its generalization (the elements within the self radius of the self element is considered to be the self elements). The bigger the self radius of the self element is, the more generalization the self element will be. Fig. 1 shows that the self radius is too small, part of the self elements cannot cover enough of the self space and high false positive rate occurs. In order to cover the self space enough and decrease the false positive rate, a large number of self elements are needed, but the cost for training the detectors increases. Fig. 2 illustrates that the self radius is too large, part of the self elements covers the nonself space and results in false negative errors. Fig. 3 illustrates that the variable self radius can appropriately cover the self space and builds the profile of the system. In conventional NSAs, the self radius is in a constant size, and so the appropriate profile of the system cannot be built. While our proposed approach adopts variable self radius and the appropriate profile of the system can be built, the appropriate profile of the system can increase the true positive rate and decrease the false positive rate.

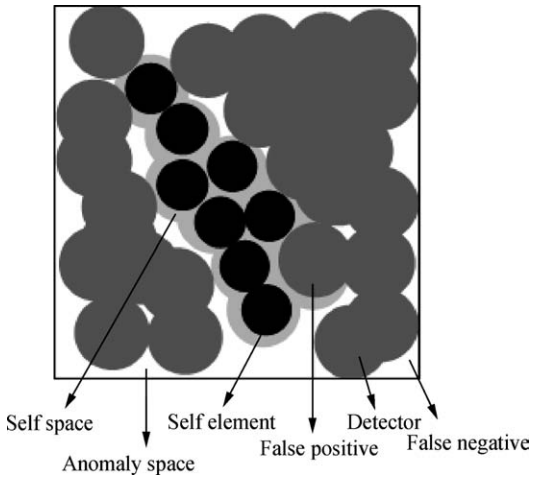

Fig. 1. The generalization of small self radius.

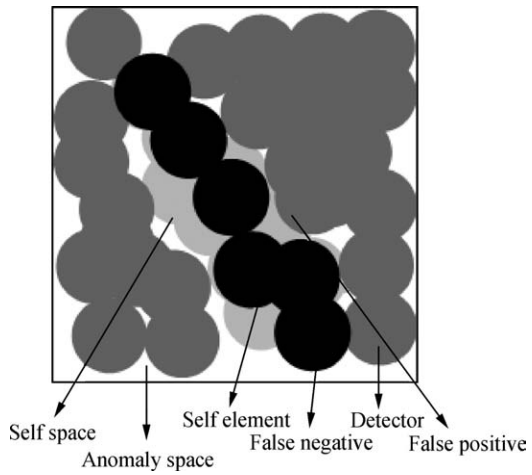

Fig. 2. The generalization of large self radius.

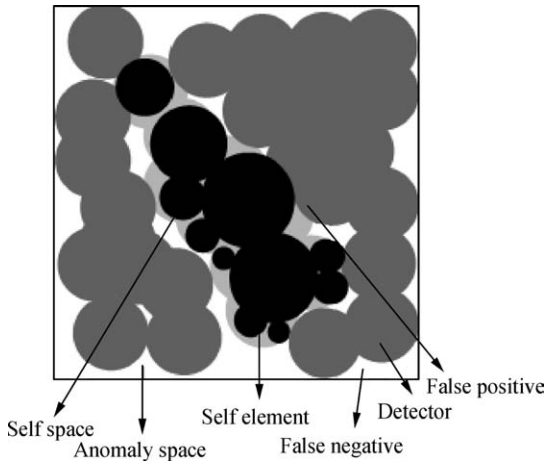

Fig. 3. The generalization of variable self radius.

2.2. Anomaly detection problem definition

Anomaly detection aims at building an appropriate profile of the system that reflects the normal behavior, and at classifying a given state of the system into the normal or abnormal state. The states of the system can be represented by a set of features.

Definition 1 (*System states space*)**.** A state of the system can be represented by a vector of features $x = (x_1, x_2, \ldots, x_n)$, and it is assumed that each feature is normalized to $[0, 1]$. The state space of the system, denoted by $U = [0, 1]^n$, is an n-dimensional space. Define the set Ag as

$$Ag = \{ag | ag \in U, |ag| = n, n \in N\} \quad (1)$$

where n is the dimension and N is the set of natural numbers. The normal set is denoted by $Self \subset Ag$, and let $Nonself \subset Ag$ be the complementary space of $Self$, such that

$$Self \cap Nonself = \Phi, Self \cup Nonself = Ag \quad (2)$$

Definition 2 (*Anomaly detection problem*)**.** Given a problem space $S' \subseteq Self$, which represents the normal behavior of the system, build a good estimate of the normal space characteristic function f: $[0, 1]^n \to \{0, 1\}$

$$f(x) = \begin{cases} 1, & \text{if } x \in Self \\ 0, & \text{if } x \in Nonself \end{cases} \quad (3)$$

Given a state of the system, the function f can distinguish between normalcy and abnormality.

2.3. Self definition

Let S denote the self set given by

$$S = \{\langle ab, rd\rangle | ab \in U, rd \in R\}, \quad (4)$$

where ab is the antibody (antibody gene), rd is the self radius and R is the set of real numbers. In a dynamic environment, as time goes on, self elements are variables. The evolution of the self elements is expressed as

$$S(t) = \begin{cases} \{x | x.ab \in U, x.rd = r_0\}, & t = 0, \\ S_r(t-1) \cup S_v(t) \cup S_n(t), & t \geqslant 1, \end{cases} \quad (5)$$

where $S(t)$ evolves from $S(t-1)$, the fit self elements, $S_r(t-1)$, are retained at time $t-1$, the self elements, $S_v(t)$, that cover some nonself spaces are updated, and the newly defined self elements $S_n(t)$ at time t are added. The fit self elements, $S_r(t-1)$, are

$$S_r(t-1) = \{x | x \in S(t-1), \forall n \in Ag(t-1), \quad f_c(n) = 0, \\ f_d(n, x) > x.rd\} \quad (6)$$

$$f_c(x) = \begin{cases} 1, & x \text{ is the anomaly element confirmed by external signal,} \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

where $f_c(x)$ ($x \in Ag$) simulates the co-stimulation in the human immune system and further confirms whether the element is an anomaly element, and $f_d(x, y)$ is the distance computing formula between x and y. If a nonself element lies within the radius of the self elements, the radius of the self elements decreases, which is expressed as

$$S_v(t) = \{x | x \in S(t-1), \exists n \in Ag(t-1), f_c(n) = 1, \\ f_d(n, x) \leqslant x.rd, x.rd = \frac{1}{\mu} \times x.rd\Big\}, \quad (8)$$

where μ is the decreasing coefficient of the self radius. Therefore, there is no detector that can detect the nonself element, and the radius of the self elements is decreased to make the detectors increase their detection radius to detect the nonself elements. The new self elements are

$$S_n(t) = \{x | x.ab \in U, x.rd = \lambda \times r_0\} \quad (9)$$

Eq. (9) shows that the self radius of the new self elements ($t > 0$) is bigger than that of the initial elements ($t = 0$), where λ ($\lambda > 1$) is the increasing coefficient of the self radius. If a new self element is defined at time t ($t > 0$), it is hopeful that more self elements will be defined around the self element, and so a bigger self radius value is set.

2.4. Detectors

Like immunocytes in human immune systems, detectors are similarly used to detect the nonself elements in our proposed approach. The set of detectors D is defined as

$$D = \{\langle ab, rd\rangle | ab \in U, rd \in R\} \quad (10)$$

where ab is the antibody (antibody gene), rd is the detection radius and R is the set of real numbers.

Like self space, the nonself space is also dynamic and so the detectors have to respond to the variations. The evolution of the detectors is expressed as

$$D(t) = \begin{cases} \{x | x.ad \in U, \exists s \in S(0), \forall s' \in S(0), \\ f_d(x, s') > f_d(x, s), x.rd = f_d(x, s) - s.rd\} & t = 0 \\ D_n(t) \cup D_r(t-1) \cup D_v(t) - D_d(t) & t > 0 \end{cases} \quad (11)$$

where $D(t)$ evolves from $D(t-1)$, the new detectors, $D_n(t)$, are added; the detectors, $D_r(t-1)$, that do not cover the self space are retained; the detectors, $D_v(t)$, that have to increase the covering space are updated; and the detectors, $D_d(t)$, that cover the self space are removed. The new detectors are shown as

$$D_n(t) = \{x | \exists n \in Ag(t), f_c(n) = 1, \forall y \in D(t), f_d(n, y) > y.rd, \\ x.ab = n.ab, \exists s \in S(t), \forall s' \in S(t), f_d(s, x) < f_d(s', x), \\ x.rd = f_d(s, x) - s.rd\} \quad (12)$$

Eq. (12) depicts that if a nonself element cannot be detected, a detector is generated and the detection radius of the detector is decided by the nearest self element to avoid covering the self space. The detectors that do not cover the self space are retained, which are shown as

$$D_r(t-1) = \{x | x \in D(t-1), \forall s \in Ag(t-1), f_c(s) = 0, \\ f_d(x, s) \geqslant x.rd + s.rd\} \quad (13)$$

If the self radius of self elements is big and self elements cover wrongly nonself space, Eq. (8), the detection radius of the detectors is increased to cover more nonself space, depicted by

$$D_v(t) = \{x | x \in D(t-1), \exists s \in Ag(t-1), f_c(s) = 0, \\ f_d(x, s) < x.rd + s.rd, x.rd = f_d(s, x) - s.rd\} \quad (14)$$

Simultaneously, if a detector detects a self element, the detector is removed, expressed by

$$D_{\mathrm{d}}(t) = \{x | x \in D(t-1), \exists s \in Ag(t-1), f_c(s) = 0, \\ f_d(x,s) \leqslant x.rd\} \quad (15)$$

3. Experiments and results

In order to determine the performance and possible advantages of our proposed approach, we performed the experiments with a classical dataset used extensively in the pattern recognition literature, the Fisher's Iris dataset [18], which includes three different classes of flowers: setosa, virginica and versicolor. In the dataset, each element is described by four attributes and each class is different from the others. We compared the results obtained by using the constant self radius [9], namely V-detector. V-detector combines real-valued negative selection algorithm and variable-sized detectors, and enhances the negative selection algorithm in efficiency [9]. Nevertheless, the self radius is constantly sized in V-detector, and cannot adapt the variety of self and nonself space, nor can it build an appropriate profile of the system.

Table 1 illustrates the comparison using the Fisher's Iris dataset. Different experiments were performed in each case using one of the three classes as the normal and the other two as the abnormal. The training data were either partially or completely composed of the elements of the normal class. The distance function was based on the Euclidean distance. The increasing coefficient of the self radius λ is 1.2. The decreasing coefficient of the self radius μ is 10. All the reported results were the average of 100 different runs.

There are two elements that define the cost function of an anomaly detection system: the detection rate of the detection system that detects an anomaly, and the false alarm rate of the detection system that produces an alarm in normal conditions. A good detection system has a high detection rate and low false alarm rate. In order to compare the performance of our proposed approach, we generated the average detection rate and false alarm rate curves for the ANSA and V-detector in Fig. 4. The average number of detectors was 50 in V-detector and 45.24 in ANSA. The training data were a half of the verginica, and another half of the verginica and the other two classes' data were the testing data. The distance function was based on the Euclidean distance. The increasing coefficient of the self radius λ is 1.2. The decreasing coefficient of the self radius μ is 10. All the reported results were the average of 100 different runs. Fig. 4 and Table 1 show that the self radius is an important parameter to control the detection rate and false alarm rate. The smaller the self radius is, the higher the detection rate and the false alarm will be. V-detector has a worse performance when the self radius increases and much anomaly cannot be detected. Our proposed approach, ANSA, has a higher detection rate but lower false alarm rate. In actual applications, we cannot predict all the anomalies, e.g. network attacks, in which novel and unknown attacks often occur. Simultaneously, normal conditions also vary, e.g. the computer administrator stops some network services, and so the detection system has to adjust itself to the changed environment. According to the detection feedback of the detection system, ANSA can adaptively adjust the profile of the system to adapt the vari-

Table 1
Comparison between ANSA and V-detector using Fisher's Iris dataset.

Training data	Self radius	Algorithm	Detection rate (%)	False alarm rate (%)	Number of detectors
Setosa (50%)	0.1	V-detector	99.43	3.68	10
	0.1	ANSA	99.53	2.40	9
	0.05	V-detector	99.62	7.96	10
	0.05	ANSA	99.66	5.84	9
Setosa (100%)	0.1	V-detector	99.14	0.00	10
	0.1	ANSA	99.50	0.00	9
	0.05	V-detector	99.45	0.00	10
	0.05	ANSA	99.56	0.00	8
Versicolor (50%)	0.1	V-detector	70.13	5.00	50
	0.1	ANSA	88.70	1.36	42
	0.05	V-detector	86.19	21.24	50
	0.05	ANSA	89.94	9.52	38
Versicolor (100%)	0.1	V-detector	60.63	0.00	50
	0.1	ANSA	88.56	0.00	40
	0.05	V-detector	79.48	0.00	50
	0.05	ANSA	89.98	0.00	41
Verginica (50%)	0.1	V-detector	87.39	16.96	40
	0.1	ANSA	92.16	8.16	34
	0.05	V-detector	93.55	29.92	40
	0.05	ANSA	94.83	16.68	37
Verginica (100%)	0.1	V-detector	80.53	0.00	40
	0.1	ANSA	92.00	0.00	35
	0.05	V-detector	92.21	0.00	40
	0.05	ANSA	93.23	0.00	32

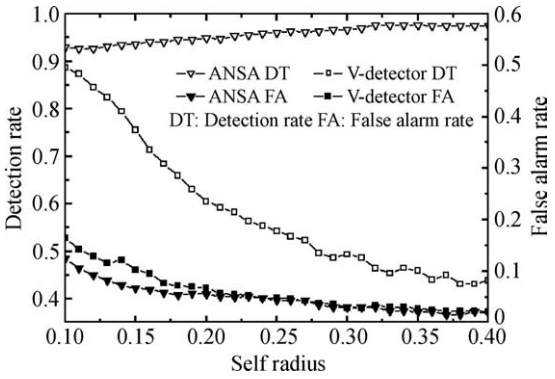

Fig. 4. Comparison curves generated by ANSA and V-detector with the verginica dataset.

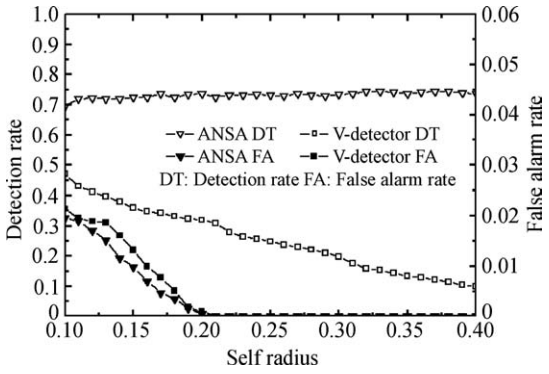

Fig. 5. Comparison curves generated by ANSA and V-detector using the biomedical dataset.

Table 2
Comparison between ANSA and V-detector using biomedical dataset.

Training data	Self radius	Algorithm	Detection rate (%)	False alarm rate (%)	Number of detectors
25%	0.2	V-detector	30.68	0.06	55
	0.2	ANSA	70.44	0.04	53
	0.1	V-detector	47.32	2.09	80
	0.1	ANSA	69.43	1.81	71
50%	0.2	V-detector	30.61	0.09	64
	0.2	ANSA	71.68	0.03	55
	0.1	V-detector	46.80	2.11	80
	0.1	ANSA	69.36	1.94	75
100%	0.2	V-detector	23.56	0.00	75
	0.2	ANSA	72.64	0.00	65
	0.1	V-detector	37.47	0.00	100
	0.1	ANSA	70.55	0.00	93

ables of the self and nonself space. Table 1 also shows that ANSA has another advantage, that is, ANSA needs less numbers of detectors and has a good coverage, e.g. when the training data are a half of the versicolor and the self radius r_0 is 0.1, V-detector needs 50 detectors, the detection rate is 70.13% and the false alarm rate is 5.00%; while ANSA only needs 42 detectors, the detection rate increases to 88.70% and the false alarm rate increases to 1.36%.

Similar comparison was performed with a biomedical dataset, which includes blood measurement of 209 patients [19]. The blood measurement was used to screen a rare genetic disorder. There are 134 normal patients and 75 carriers of the disease in the dataset. Table 2 and Fig. 5 show the comparison by using the dataset. The average number of the detectors in Fig. 5 was 75 in V-detector and 75.82 in ANSA. The training data were 25% of the normal patients, and 75% of the normal patients and all carriers of the disease were the testing data. The distance function was based on the Euclidean distance. The increasing coefficient of the self radius λ is 1.2. The decreasing coefficient of the self radius μ is 10. All the reported results were the average of 100 different runs. It further shows the capacity of ANSA in building an appropriate profile of the system, and shows that ANSA can increase the detection rate and decrease the false alarm rate.

4. Conclusions

This paper presents a self-adaptive negative selection algorithm, ANSA. The approach can build an appropriate profile of the system only by using a subset of normal elements, and can adapt the varieties of self/nonself space. It can also adaptively adjust the self radius, the detection radius and numbers of detectors to amend the built profile of the system. The experimental results show that ANSA is an efficient solution to anomaly detection and offers the characteristics of high detection rate, low false alarm rate, self-learning and adaptation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 60573130), and the National High Technology Research and Development Program (Grant No. 2006AA01Z435).

References

[1] Kim J, Bentley PJ, Aickelin U, et al. Immune system approaches to intrusion detection – a review. Nat Comput 2007;6(4):413–66.
[2] Dasgupta D. Advances in artificial immune systems. IEEE Comput Intel Mag 2006;1(4):40–9.

[3] Li T. An immunity based network security risk estimation. Sci Chin Ser F 2005;48(5):798–816.
[4] Timmis J, Andrews P, Owens N, et al. An interdisciplinary perspective on artificial immune systems. Evol Intel 2008;1:5–26.
[5] Li T. Computer immunology. Beijing Publishing House of Electronics Industry; 2004.
[6] Forrest S, Perelson AS. Self-nonself discrimination in a computer. In: Proceedings of IEEE symposium on security and privacy, Oakland; 1994. p. 202–13.
[7] Dasgupta D. An immunity-based technique to characterize intrusions in computer networks. IEEE Trans Evol Comput 2002;6(3):281–91.
[8] González F, Dasgupta D. An immunogenetic technique to detect anomalies in network traffic. In: Proceedings of the genetic and evolutionary computation conference, New York; 2002. p. 1081–8.
[9] Zhou J, Dipankar D. Real-valued negative selection algorithm with variable-sized detectors. In: Proceedings of the genetic and evolutionary computation conference, vol. 3102; 2004. p. 287–98.
[10] Zhou J, Dipankar D. Applicability issues of the real-valued negative selection algorithms. In: Proceedings of the genetic and evolutionary computation conference, Seattle, Washington, USA; 2006. p. 111–8.
[11] Zhou J. A boundary-aware negative selection algorithm. In: Proceedings of the international conference on artificial intelligence and soft computing, Benidorm, Spain; 2005.
[12] Balachandran S, Dasgupta D, Nino F, et al. A framework for evolving multi-shaped detectors in negative selection. In: Proceedings of the 2007 IEEE symposium on foundations of computational intelligence, Honolulu; 2007. p. 401–8.
[13] Garrett SM. How do we evaluate artificial immune systems? Evol Comput 2005;13(2):145–78.
[14] Harmer PK, Williams PD, Gunsch GH, et al. An artificial immune system architecture for computer security applications. IEEE Trans Evol Comput 2002;6(3):252–80.
[15] Kim J, Bentley P. Evaluating negative selection in an artificial immune system for network intrusion detection. In: Proceedings of the genetic and evolutionary computation conference, San Francisco; 2001. p. 1330–7.
[16] D'haseleer P, Forrest S, Helman P. An immunological approach to change detection: algorithms, analysis and implications. In: Proceedings of the 1996 IEEE symposium on computer security and privacy, Oakland; 1996. p. 110–9.
[17] Balthrop J, Forrest S, Glickman M. Revisiting lisys: parameters and normal behavior. In: Proceedings of the congress on evolutionary computation, Honolulu; 2002. p. 1045–50.
[18] Murphy PM, Aha DW. UCI repository of machine learning databases; 1992.
[19] StatLib-datasets archive. http://lib.stat.cmu.edu/dataset/.