An Outlier Robust Negative Selection Algorithm Inspired by Immune Suppression

Guiyang Li
School of Computer Science, Sichuan University, Chengdu, 610065, China
Email: guiyang.li@gmail.com

Tao Li, Jie Zeng, Haibo li
School of Computer Science, Sichuan University, Chengdu, 610065, China
Email: litao@scu.edu.cn,

***Abstract*—The negative selection algorithm (NSA) is one of models in artificial immune systems. Traditional NSAs do not perform any differentiation for training self dataset and only use the mechanism of negative selection. They will generate excessive invalid detectors and have poor detection performance when the training selves contain noisy data. Inspired by immune suppression mechanism, an outlier robust NSA is proposed. The new algorithm will divide the training selves into internal selves, boundary selves and outlier selves. At the same time, the information hiding in different kind of selves is fully utilized. Furthermore, by combining negative selection mechanism with positive selection mechanism, the new algorithm can cover the non-self region more effectively. The experiment results show that no matter the training self data is clean or not, the new algorithm can obtain better detection performance with fewer detectors.**

***Index Terms*—negative selection algorithm; immune suppression; hypothesis testing; ROC;**

I. INTRODUCTION

Artificial Immune Systems are soft computing techniques inspired by the biological immune system[1-3]. Among various mechanisms in the biological immune system that are explored for Artificial Immune Systems, negative selection, clonal selection and immune network model are the most discussed models[4-7] .

The NSA is one of the most successful methods in Artificial Immune Systems, and its typical applications include change detection, fault detection, virus detection and network intrusion detection[8]. The idea of NSA comes from the T cells maturation process in the immune system: if T cells in thymus recognize any self proteins, they are eliminated before deploying for immune functionality. As a result, only the self-tolerant T cells survive in the negative selection process and are released into the blood stream. In the same way, the outline of NSA illustrated in figure 1 is applied in two stages[9]. In the training stage, the detector set is generated. In the detection stage, the detector set is used to detect anomaly in the incoming new data.

The NSA firstly proposed by Forrest et al[9] used binary representation for self and non-self space. Then, a real-valued representation was presented by Gonzalez et al[10]. Among the latest researches on the NSAs, in order to cover non-self space more effectively, Ji et al[11] raised and improved a real-valued NSA with variable-sized detectors (termed V-detector). Balachandran et al[12] proposed a multi-shaped detectors NSA.

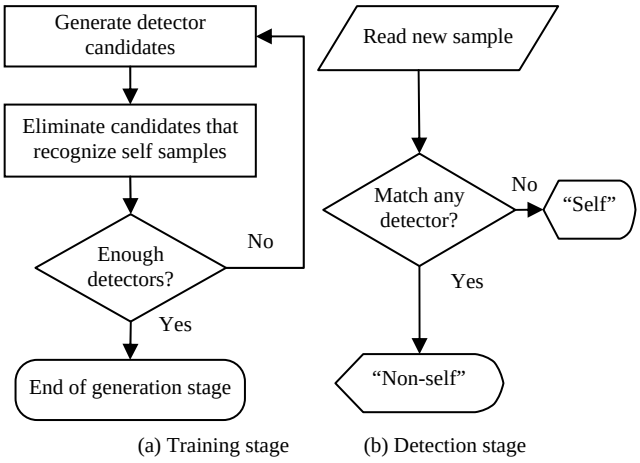


Figure 1. Outline of typical negative selection algorithm

After ten years of development, the outline of NSA shown in Figure 1 has never been changed. However, the outline of NSA does not make full use of information hiding in different kind of selves. When a large number of outliers existed in the self dataset, the NSAs will generate excessive invalid detectors and can not cover non-self region effectively. In addition, traditional NSAs are built on the basis that selves in training dataset are reliable, so, it can not adapt the training self dataset containing noisy data. The above-mentioned defects in traditional NSAs have greatly hindered its practical application.

The mechanism of biology immune system is very mysterious and complicated. Its associated theory is still in development and improvement stage. How to extract inspiration to improve and even build a novel immune algorithm is an important research topic in the theory and application of artificial immune system[6]. In biology immune system, T lymphocyte not only has helper T cells (T_h) to assist in the immune response, but also has suppressor T cells (T_s) to suppress immune response[13].

doi:10.4304/jcp.5.9.1348-1355

The two groups of cells interact with each other in normal body and maintain a balance to keep body healthy. The typical NSA just uses T_h cell metaphor and lacks of research and application for T_s cell metaphor. Inspired by the metaphor of T_s cells. this paper proposes a novel NSA which uses outlier and boundary detection technology to differentiation selves into outlier selves, boundary selves and internal selves. In detection stage, the detectors serve as T_h cells to assist in the immune response and the outlier and boundary selves serve as T_s cells to suppress auto-reactive T cells. Thereby, fewer detectors can be used to cover non-self space and the number of holes can be reduced. The experiment result shows that the improved algorithm is able to deal with clean data effectively and can handle the noisy data existed in selves. In the meantime, no matter whether outlier is noisy date or not, the algorithm can reduce the number of detectors under the circumstance of ensuring detection rate.

II. Proposed Outline of New NSA

A. Classification of self

In the view of machine learning, NSA is a kind of one-class classification algorithm. That is, only one class (normal or self data) can be used when learning. Traditional NSAs do not perform analysis and differentiation for selves. The algorithm is based on an assumption that the selves are reliable and same important. However, in the use of negative selection algorithm to solve the specific problems, it is impossible to guarantee training self data reliable completely. For example, when the NSA is applied in intrusion detection, virus detection and etc, training self data usually contains noisy data. In these circumstances, the traditional NSA can not adapt well.

Different from traditional NSAs, the new NSA proposed in this paper analyzes the training self dataset and distinguishes different kind of selves. Through analysis of different information hiding in self data, we believe that each self is not same important. In figure 2, the training selves can be divided into internal selves, boundary selves and outlier selves. Obviously, Outlier selves and boundary selves are obviously different from internal selves. To the date, the information and function of outlier selves and boundary selves has not been studied deeply in the current study of artificial immune systems.

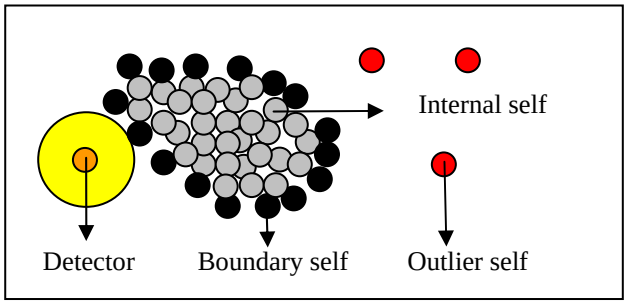


Figure 2. Different class of self

Obviously, different class of self has different amount of information. Outlier self is likely to be caused as a result of noisy data. Since the majority of miss and false alarm occur in the boundary between self region and non-self region, the boundary self contains more accurate boundary information and the information is important for better balancing detection rate and false alarm rate. In order to make full use of different types of information hiding in selves, the algorithm proposed in this paper uses different processing approach for different selves.

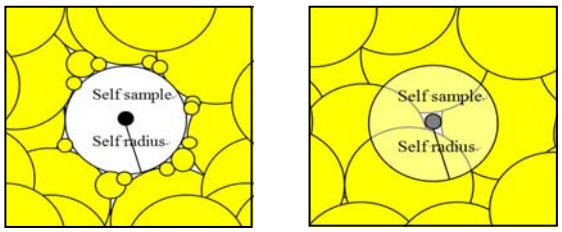
B. Negative selection and positive selection

Since the beginning research of negative selection, there are many discussions about the advantages and disadvantages of negative selection and positive selection. Stibor et al[14] considered that positive selection had better detection performance than negative selection had. But, Ji et al[15] considered that when the number of detectors generated by negative selection was less than the number of self samples, negative selection was better than positive selection. Further more, positive selection lacked of inductive ability.

The above discussions considered the positive mechanism and negative mechanism contradictory and mutually exclusive. In fact, in biology immune system, positive mechanism and negative mechanism are put into good use. For example, T cells not only have T_h cells to assist in the immune response, but also have T_s cells to suppress immune response. Inspired by the immune mechanism, the algorithm proposed in this paper combines both positive and negative mechanisms. Specifically, the detectors are used to cover non-self region by using negative selection mechanism. In the meantime, the boundary selves and outlier selves are regarded as reverse detectors (r-detectors) cover part of self region by using positive selection mechanism. Table I shows the relationship between theory of immunity and novel NSA framework.

TABLE I. Relationship Between Theory of Immunity and Novel NSA Framework

Theory of immunity	Novel NSA framework
Self	Self data
T_h cell	Detector
T_s cell	R-detector
Thymic tolerance	Negative selection Positive selection
Antigen	Test data
Affinity	Pattern match



(a) Only Negative (b) Negative and Positive

Figure 3. Basic idea of two NSA

Figure 3(a) shows the schematic with only negative selection mechanism. In this figure, the yellow circle represents detector, white circle represents self and the no-covered party represents the well-known "hole" in negative selection[9]. We can see the mechanism had two main problems: 1) holes in the boundary can't be covered effectively; 2) many invalid detectors (small yellow circle) without detection ability will be created in self and non-self boundary.

Figure 3(b) shows the schematic with negative and positive selection mechanism. From the figure, we may see that using both negative and positive selection mechanisms is able to cover non-self space completely. In the mean time, the invalid detector is reduced.

C. Outline of outlier robust NSA

The data flow diagram of the outlier robust NSA (ORNSA) proposed in this paper is shown in Figure 4. The ORNSA firstly uses outlier detection algorithm to recognize and filter outlier selves which are in the training self dataset. Then, the detectors are generated by using the filtered self dataset in the training stage, at the same time, boundary selves are also recognized. After recognizing, the boundary selves are recorded completely and will be used as r-detectors in detection stage. Whether the outlier selves need to be transformed to r-detectors, it depends on the specific problems solved by the algorithm. If training self dataset is reliable, these outliers will be transformed to r-detectors. Whereas, if the training self dataset is dirty, these outlier selves can be recognized as noisy data and no longer appear in the detection stage.

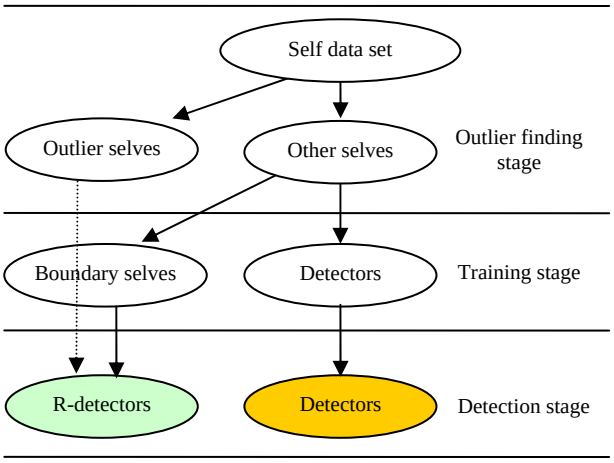


Figure 4. Data flow chart of ORNSA

Figure 5 shows the training stage and detection stage of ORNSA. In Figure 5(a), the training stage of ORNSA is similar to the traditional NSA, but the mechanism of identifying and recording boundary selves is added. After the end of the training stage, detector set is obtained to meet the requirement. In the meantime, a set of boundary selves is also obtained. Figure 5(b) shows the detection stage of ORNSA. In this stage, new samples are detected using detectors and r-detectors. Once the sample is recognized by any detector, it needs to match with r-detectors to check if the sample is a self. Since positive and negative mechanisms complement each other, this stage improves the issues of holes and invalid detectors encountered by traditional NSA effectively.

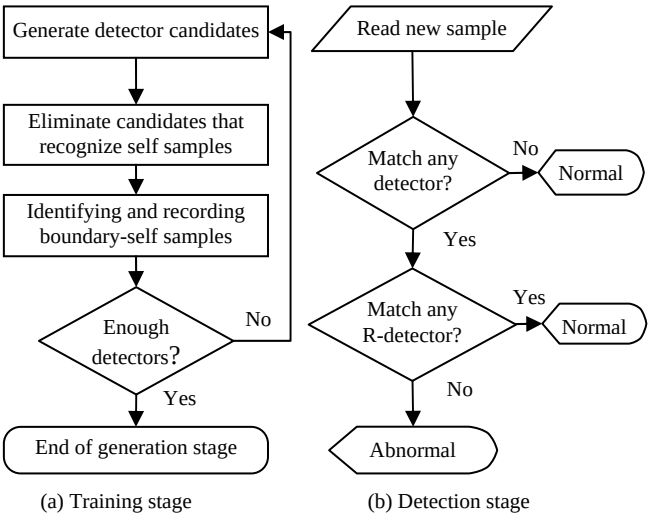


Figure 5. Training stage and detection stage of ORNSA

Figure II shows the comparison of three typical selection algorithms: positive selection algorithm (PSA)[14], NSA[9] and ORNSA. Table 2 indicates that ORNSA is the only selection algorithm based on immune suppressive theory and combines two kinds of mechanisms for negative and positive. The ORNSA also can classify self and deal with outlier.

TABLE II. COMPARISON OF DIFFERENT SELECTION ALGORITHM

Algorithm elements	**PSA**	**NSA**	**ORNSA**
Immune cells	No	T_h	T_h, T_s
Algorithm stage	Detection	● Training ● Detection	● Outlier finding ● Training ● Detection
Negative or Positive	Positive	Negative	● Negative ● Positive
Self-Classification	Internal self	Internal self	● Interior self ● Boundary self ● Outlier self
Whether it can handle noisy data or not?	No	No	Yes

III. AN ACTUAL ORNSA BASED ON REAL-VALUED NSA

In order to demonstrate the effectiveness of the new algorithm, this paper introduces an actual ORNSA based on real-valued NSA proposed by Gonzalez et al[10].

A. Basic Definition

Definition 1: system state space. The actual values of varaible could be scaled or normalized to fit a defined range [0.0,1.0]. The n-dimensional state space U can be expressed as hypercube $[0.0,1.0]^n$. then a state of the system is represented by a vector of features, $x^i = \left(x_1^i, x_2^i, ..., x_n^i\right) \in U$. A set of feature vectors, $Self \subset U$ represents the normal states of the system. Its

complement is called Non_self and is defined as $Non\_self = U - self$.

Definition 2: Mature detector set. A set of feature vectors $S' \subseteq Self$ is used for tolerance in the training stage. Any real-valued detector is defined as $d^i = \{< c, r > | c \in U, r \in R\}$, R represents a real value set, c represents the detector position in the system state space, r represents the detection radius of detector. The detector set $UD = \{d^1, d^2, ..., d^l \mid \exists s \in S', f(d^i.c, s) < d^i.r, i = 1, ..., l\}$ will be deleted in training stage, the undeleted detector set $MD = \{d^1, d^2, ..., d^k \mid \forall s \in S', f(d^i.c, s) \geq d^i.r, i = 1, ..., k\}$ will become mature detector set, f represents a distance measuring function.

Definition 3: Reverse Detector set. Suppose we get outlier-self set $OS = \{s^1, s^2, ..., s^m \mid s^i \in S', i = 1, ..., m\}$ in outlier finding stage and get boundary-self set $BS = \{s^1, s^2, ..., s^n \mid s^i \in S', i = 1, ..., n\}$ in training stage. If outlier selves are all real self, we get reservse detector set $RD = \{d^1, d^2, ..., d^k \mid d^i.c \in OS \cup BS, d^i.r = r_{rd}, i = 1, ..., k\}$, r_{rd} is the r-detector detection radius. If the outlier is noise data, have $RD = \{d^1, d^2, ..., d^k \mid d^i.c \in BS, d^i.r = r_{rd}, i = 1, ..., k\}$.

Definition 4: Distance measure. Distance measure is an important element in negative selection algorithms. Distance measure is used to calculate the distance between any two points in state space, the smaller the distance between two points, the higher affinity. In ORNSA, the Minkowski distance is used for distance measure. For any two points $x \in U, y \in U$, its m-dimensional Minkowski distance f is defined as

$$f(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^m \right)^{\frac{1}{m}} \tag{1}$$

Minkowski distance is also referred to as *m-norm* distance. The *1-norm* distance is also known as Manhattan distance and *2-norm* distance is the most commonly used Euclidean distance. For different norm, the detector (or recognition region) will take different geometric shapes and have different covering area[15].

Definition 5: Detection range. The detector detection radius is usually used to restrict detection range of detector. In self set *S'*, the nearest self to detector *d* is defined as

$$Nearest(d, S') = \{s | s \in S', \forall s \in S', f(d.c, s) \leq f(d.c, s)\} \tag{2}$$

The detection radius for detector *d* is defined as

$$d.r = f(d.c, Nearest(d, S')) \tag{3}$$

Definition 6: Detection. The set of reverse detector set *RD* and mature detector set *MD* are both used during the detection stage. In the stage, *MD* is used as T_h cells to assist in the immune response; *RD* is used as T_s cells to suppress immune response. Suppose antigen $x \in U$, the detection stage will determine whether the antigen is a self or not. the detection process is defined as

$$match(x) = \begin{cases} 1, if\ (\exists d \in MD, f(d.c, x) < d.r)\ and \\ \quad (\forall rd \in RD, f(rd.c, x) \geq d.r) \\ 0, otherwise \end{cases} \tag{4}$$

The value 1 returned by function *match* represents that the antigen is a non-self and the value 0 denotes that the antigen is a self. In different applications, the self and non-self has different meaning. For example, the non-self represents network attack and the self represents normal network access for the network intrusion detection. But, for virus detection, the non-self represents virus and the self represents legal code.

B. Aglorithm

The ORNSA has three stages: Outlier finding stage、Training stage、Detection stage. The following gives a specific implementation for each stage.

1) Outlier finding stage

As one of the important research area for knowledge discovery, there are many efficient outlier detection algorithms for issues of outlier detection[16, 17]. The algorithm proposed in this paper is based on the premise that selves are similar and so distance-based algorithm proposed by Knorr is used for recognizing outlier self[17]. The specific definition for outlier self is:

Definition 7: Outlier self. A self o in a self dataset S is a DB(p, D)-outlier self if at least fraction p of the selves in S lies greater than distance D from o.

2) Training stage

We may see that the space covered by detectors and selves are complementary in figure 2. Therefore, it is feasible to recognize boundary self by using detectors. The specific definition for boundary self is:

Definition 8: Boundary self. A self *o* in a self dataset S is a boundary self if this self is not outlier self and this self is at the edge of a detector.

In the training stage, there are several implementations for real-valued negative selection algorithm currently. One of the most famous algorithm is V-detector algorithm proposed by Ji and Dasgupta et al [18]. So far, there are several versions of the V-detector. The algorithm given by [14] is the latest and most mature version. It took the most of advantages from other versions. Different from other negative selection algorithms, the V-detector algorithm does not compute the number of the needed detectors in the training stage, but detectors are generated by using statistical estimate and hypothesis testing to meet the coverage requirement. There are several key running parameters exist in the V-detector: p is the target coverage of the non-self region by all the existing detectors for hypothesis testing, α is the significant level for hypothesis testing, n is the sample size, r_s is the self radius, z is the standard score for z-score using "Central Limit Theorem", and z_α is the z score for a confidence level of 1-α.

The flow chart of the V-detector algorithm is shown in Figure 6[18].

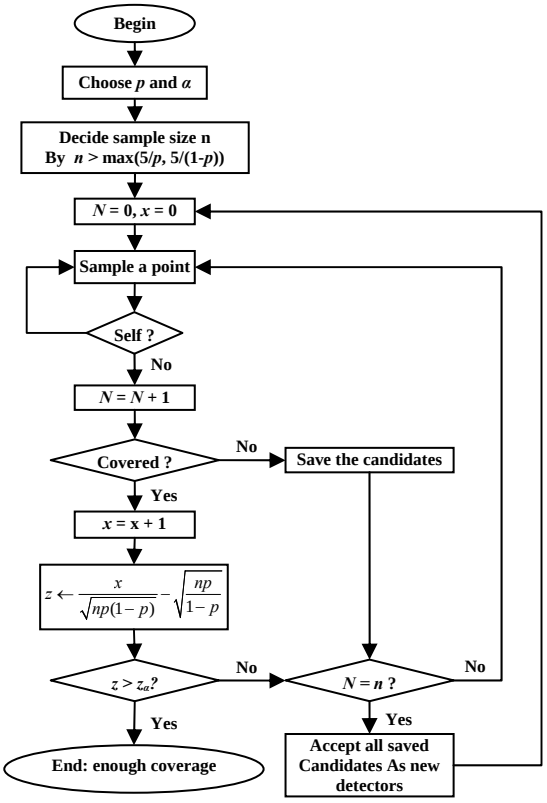


Figure 6. The training stage of V-detector algorithm

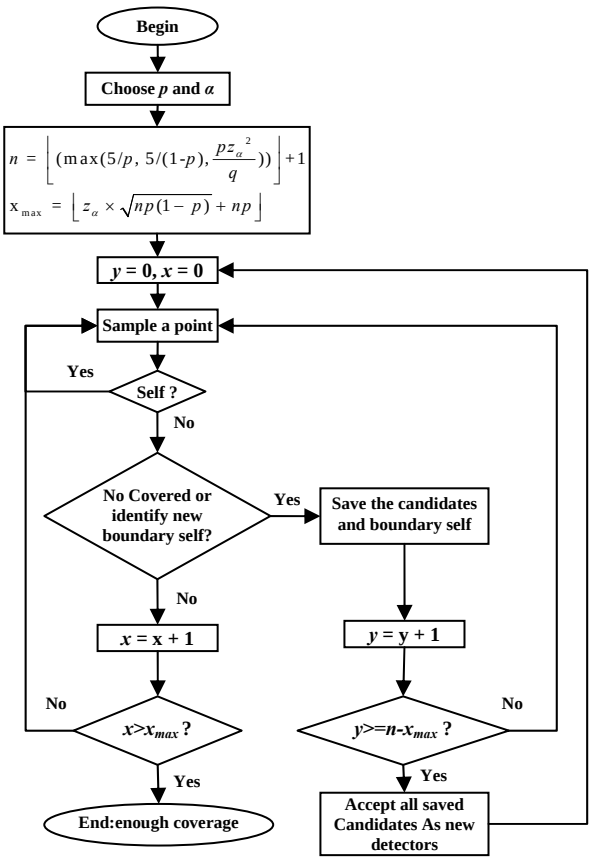


Figure 7. The training stage of improved V-detector algorithm

The V-detector algorithm shown in Figure 6 can not identify boundary selves and exists several problems [19, 20]. Therefore, it can not be used in the training stage of ORNSA. Figure 7 shows an improved algorithm which has a function of identifying boundary selves. In this improved algorithm, p is the target coverage of the non-self region by all the existing detectors for hypothesis testing, α is the significant level for hypothesis testing, n is the sample size, x is the number of covered points, y is the number of uncovered points, x_{max} is the upper limit of the hypothesis testing and z_α is the z score for a confidence level of 1-α.

In this improved algorithm, after initializing the value of p and α in the algorithm, the target coverage p and the impact of significant level α were fully considered for the determination of sample size n. Its value is defined as

$$n = \left\lfloor (\max(5/p,\ 5/(1\text{-}p), \frac{pz_\alpha^{\ 2}}{q})) \right\rfloor + 1 \qquad (5)$$

The Equation (5) is different from the algorithm given by Ji et al.[18] and can make this algorithm to simultaneously satisfy the requirements of "Central Limit Theorem" and hypothesis testing for sample n[20]. For the algorithm shown in Figure 7, no matter how much target coverage p and significant level α is, as long as the value of n in the algorithm is calculated using Equation(5), the algorithm will certainly be able to perform successfully.

The null hypothesis in the algorithm is "The coverage of the non-self region by all the existing detectors is below percentage p" [18]. The value of x_{max} is defined as

$$\mathrm{x}_{\max} = \left\lfloor z_\alpha \times \sqrt{np(1-p)} + np \right\rfloor \qquad (6)$$

In each round of the iteration for testing of n points, if y is equal to or greater than n-x_{max}, the null hypothesis will be accepted and the detector set will be updated. After several iterations, if x is greater than x_{max}, the null hypothesis will be rejected and the algorithm will be terminated successfully, the detectors generated at this time will have confidence level of 1-α to meet the requirement of target coverage p of non-self region.

After algorithm is finished, the detectors, which satisfy the coverage p, are generated. In the meantime, a number of boundary selves are generated too. These boundary selves can be used as r-detector in detection stage. The detailed improved V-detector algorithm refers to [19, 20].

3) Detection stage

The biggest difference between ORNSA and traditional NSA occurs in the detection stage. In this stage, the outlier selves and boundary selves are used as r-detectors. Then, it combines detectors and r-detectors to detect the antigen. The Figure 3(a) shows the algorithm process.

The generalization ability for traditional NSA depends on some parameters setting before training stage. For NSA that uses r-continuous match regulation, the parameter is r; for V-detector algorithm, the parameter is self radius. All of these parameters are set in the training stage. Once the training is completed, the relevant

parameter will no longer be adjusted. If these parameters need to be adjusted, the detectors will be re-trained.

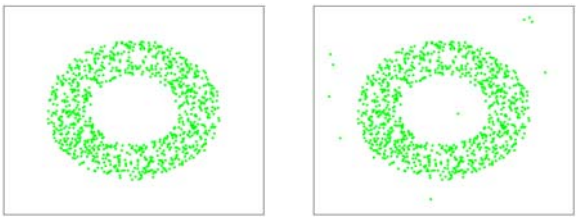
For ORNSA algorithm, the detection radius of r-detector can adjust the generalization ability during detection stage. Its value can be adjusted according to actual situation. The detector does not need to be trained when changing the detection radius of r-detector. This is because that the r-detector detection radius affects only the self and non-self boundary without affecting coverage of other non-self space.

IV. EXPERIMENTS AND DISCUSSION

To demonstrate the basic behavior of the new algorithm (ORNSA), the synthesized data and benchmark Fisher's Iris data are used and compare with V-detector algorithm presented in [18]. The V-detector algorithm with boundary-aware and point-aware are abbreviated as VDA(B) and VDA(P) respectively.

A. Synthetic data

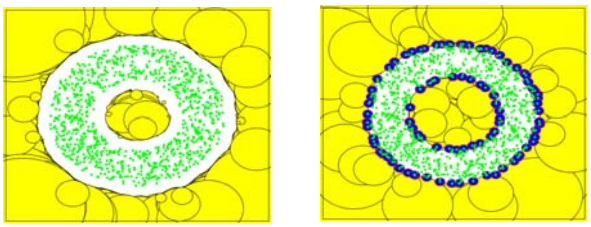
To demonstrate the basic behavior of the new algorithm (ORNSA), experiments were carried out using 2-dimensional synthetic data provided by [21]. Over the unit square $[0, 1]^2$, two training self datasets are used in these experiments. The shapes of training selves are shown in Figure 8. We may see that one dataset is clean, the other contains noisy data.



(a) Clean selves (b) Dirty Selves

Figure 8. Training self datasets of ring shape

Figure 9 shows a visualized experiments result of VDA(P) and ORNSA. The experiments were from the following setting: clean training selves, target coverage 99%, significant level 0.1 and self radius 0.05. Green dots represent training selves, yellow part represents detectors, blue part represents boundary selves. We may see that ORNSA generates 83 detectors and 82 r-detectors. This result is far less than 526 detectors generated by VDA(P). We may also see that the r-detectors may balance the detection rate and false alarm rate through adjusting its detection radius. Whatever VDA(P) or VDA(B) is used, the boundary is too strict and there is no flexible adjustment mechanism.



(a) 526 detectors (b) 83 detectors and 82 r-detectors

Figure 9. Regions Covered by selves and detectors (Clean)

Figure 10 shows the ROC curves using ORNSA and VDA. Other relevant parameters are as follows: cleanly training selves, target coverage 99% and significant level 0.1. Using the area under the ROC curve as the algorithm evaluation criteria, we may see that ORNSA has better performance compared with the VDA. For ORNSA, the determination of boundary depends on the radius of r-detectors and the radius can be adjusted according to specific application. Therefore, the better tradeoff can be obtained between detection rate and false alarm rate.

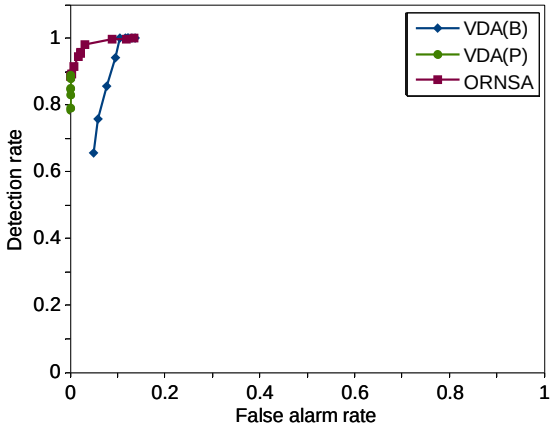
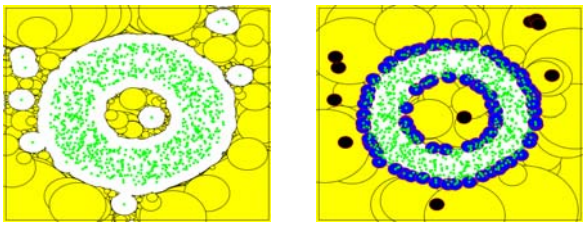


Figure 10. ROC comparison for clean training data

Figure 11 shows a visualized experiments result of VDA(P) and ORNSA The experiments were from the following setting: dirty training selves, target coverage 99%, significant level 0.1 and self radius 0.05. Green dots represent training selves, yellow part represents detectors generated, blue part represents boundary selves, and black part represents outlier selves. We may see that ORNSA algorithm can recognize boundary selves and outlier selves accurately. It generates 84 detectors, 81 boundary selves and 10 outlier selves. This result is far less than 865 detectors generated by VDA. Comparing Figure 9 and Figure 11, we may find that, for dirty training selves, ORNSA can ignore these noisy data in the training stage and detection stage. However, VDA can not deal with noisy data. The presence of the noisy data will increase the number of generated detector and decrease the detection performance.



(a) 865 detectors (b) 84 detectors and 81 r-detectors

Figure 11. Regions Covered by selves and detectors(Dirty)

Figure 12 shows the ROC curves using ORNSA and VDA with boundary-aware and point-aware algorithm respectively. Other relevant parameters are as follows: dirty training selves, 1000 self sample points, target coverage 99% and significant level 0.1. We may see that

ORNSA gets better performance than VDA does by using the area under ROC curve as evaluation criteria.

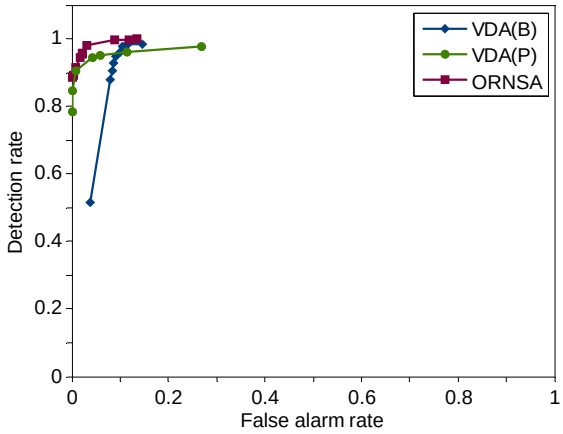


Figure 12. ROC comparison for dirty training data

Table III shows the comparison of VDA and ORNSA which detection radius of r-detectors is 0.01 and 0.02 respectively. The results were from the following setting: self radius 0.05 and target coverage 99%. The numbers are the mean of 100 repeated tests. The result shows that when the training self dataset is clean, comparing with VDA(B) and VDA(P), ORNSA generates fewer detectors and gets better tradeoff in detection rate and false alarm rate through the adjustment of detection radius. Furthermore, when training self dataset is dirty, comparing with VDA ORNSA can obtain better detection performance using fewer detectors.

TABLE III. COMPARISION BETWEEN VDA AND ORNSA

Ring shape	Algorithm	Detection Rate	False Alarm Rate	Number of detectors
Clean	VDA(B)	99.99%	12.7%	508.36
	VDA(P)	80.3%	0%	517.92
	ORNSA(0.01)	99.89%	7.26%	151.36
	ORNSA(0.02)	94.93%	1.32%	151.36
Dirty	VDA(B)	98.15%	11.04%	515.43
	VDA(P)	71.84%	0%	849.2
	ORNSA(0.01)	99.89%	7.26%	151.36
	ORNSA(0.02)	94.93%	1.32%	151.36

B. Real-world data

The benchmark Fisher's Iris dataset 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. One of the three types of iris is considered as normal data, while the other two are considered abnormal. The normal data are either completely or partially (50%) used to train the system.

Figure 13 shows the ROC curves using ORNSA and VDA Other parameters are as follows: versicolor (50%), target coverage 99% and significant level 0.1. We may see that ORNSA gets better performance than VDA does.

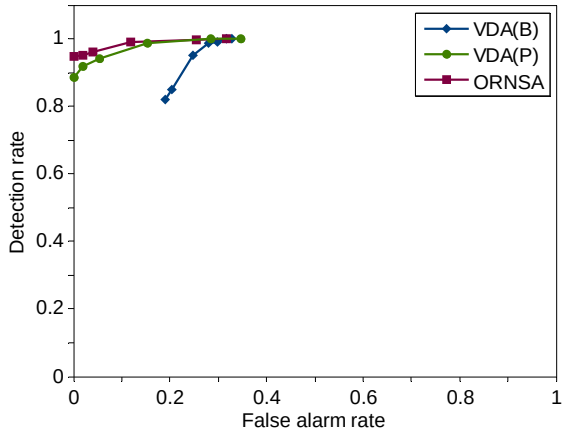


Figure 13. ROC for partial versicolor data

Table IV shows the comparison of the VDA and ORNSA using iris dataset. The results were from the following settings: self radius 0.1, target coverage 99%.When the training data is completely, the boundary self radius of ORNSA is set to 0. When the training data is partially (50%), the r-detector detection radius of ORNSA is set to 0.1. The numbers are the mean of 100 repeated tests. The result shows that ORNSA generates fewer detectors and can obtain better tradeoff between detection rate and false alarm rate than VDA does.

TABLE IV. COMPARISON USING FISHER'S IRIS DATA

Iris data	Algorithm	Detection Rate	False Alarm Rate	Number of detectors
Setosa (100%)	VDA(B)	100%	0%	500
	VDA(P)	100%	0%	592.23
	ORNSA	100%	0%	31.77
Setosa (50%)	VDA(B)	100%	36.88%	500
	VDA(P)	100%	4%	525.58
	ORNSA	100%	3.96%	24.54
Versicolor (100%)	VDA(B)	99.84%	0%	500
	VDA(P)	86.31%	0%	536.71
	ORNSA	99.88%	0%	47.67
Versicolor (50%)	VDA(B)	99.99%	35.3%	500
	VDA(P)	91.92%	0%	520.88
	ORNSA	96.99%	1.92%	38.22
Verginica (100%)	VDA(B)	99.44%	0%	500
	VDA(P)	91.93%	0%	1288.92
	ORNSA	100%	0%	88.14
Verginica (50%)	VDA(B)	100%	43.88%	500
	VDA(P)	92.06%	2.7%	903.45
	ORNSA	95%	5%	43.4

V. CONCLUSIONS

The traditional NSAs neither consider the mechanism of immune suppression, nor identify and use relevant mechanism of boundary selves and outlier selves. Inspired by immune suppression mechanism, this paper proposed an outlier robust negative selection algorithm which uses outlier and boundary detection technology to divide the selves into internal selves, boundary selves and outlier selves. Then, it combines the positive with negative mechanisms to cover non-self space more effectively. The experiment results show that the new algorithm have better adaptability and can obtain better detection performance by using fewer detectors. It also shows that the algorithm can deal with the training selves containing noisy data. The future research work will focus on how to integrate it into other NSAs and how to determine the value of several key parameters (self radius, r-detector detection radius) automatically.

ACKNOWLEDGMENT

This work was sponsored by the National Natural Science Foundation of China (Nos. 60573130, 60502011 and 60873246), the National High-Tech Research and Development Plan of China (No. 2006AA01Z435), and the National Research Foundation for Doctoral Program of Higher Education of China (No. 20070610032).

REFERENCES

[1] Hofmeyr S and F. S, "Architecture for an artificial immune system," Evolutionary Computation, vol. 8, pp. 443-473., 2000.
[2] L. N. de Castro and J. I. Timmis, "Artificial immune systems as a novel soft computing paradigm," Soft Computing, vol. 7, pp. 526-544, 2003.
[3] E. Hart and J. Timmis, "Application areas of AIS: The past, the present and the future," Applied Soft Computing Journal, vol. 8, pp. 191-201, 2008.
[4] D. Dasgupta, "Advances in artificial immune systems," Computational Intelligence Magazine, IEEE, vol. 1, pp. 40-49, 2006.
[5] S. M. Garrett, "How Do We Evaluate Artificial Immune Systems?," Evolutionary Computation, vol. 13, pp. 145-177, 2005.
[6] J. Timmis, "Artificial immune systems—today and tomorrow," Natural Computing, vol. 6, pp. 1-18, 2007.
[7] J. Timmis, A. Hone, T. Stibor, and E. Clark, "Theoretical advances in artificial immune systems," Theoretical Computer Science, 2008.
[8] T. Li, "Computer Immunology," Publishing House of Electronics Industry, Beijing, 2004.
[9] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-Nonself Discrimination in a Computer," in 1994 IEEE Computer Society Symposium on Research in Security and Privacy, Proceedings. Los Alamitos: IEEE, Computer Soc Press, 1994, pp. 202-212.
[10] F. Gonzalez, D. Dasgupta, and J. Gomez, "The effect of binary matching rules in negative selection," Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pp. 196-206, 2003.
[11] Z. Ji and D. Dasgupta, "Real-valued negative selection algorithm with variable-sized detectors," LNCS, vol. 3102, pp. 287–298, 2004.
[12] S. Balachandran, D. Dasgupta, F. Nino, and D. Garrett, "A Framework for Evolving Multi-Shaped Detectors in Negative Selection," 2007.
[13] E. M. Shevach, "CD4+ CD25+ suppressor T cells: more questions than answers," Nature Reviews Immunology, vol. 2, pp. 389-400, 2002.
[14] T. Stibor, J. I. Timmis, and C. Eckert, "A Comparative Study of Real-Valued Negative Selection to Statistical Anomaly Detection Techniques," LECTURE NOTES IN COMPUTER SCIENCE, vol. 3627, pp. 262, 2005.
[15] Z. Ji and D. Dasgupta, "Applicability issues of the real-valued negative selection algorithms," presented at Genetic and Evolutionary Computation Conference (GECCO), Seattle, Washington, 2006.
[16] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," 2000.
[17] E. M. Knorr, R. T. Ng, and V. Tucakov, "Distance-based outliers: algorithms and applications," The VLDB Journal, vol. 8, pp. 237-253, 2000.
[18] Z. Ji and D. Dasgupta, "Estimating the detector coverage in a negative selection algorithm," presented at Genetic and Evolutionary Computation Conference (GECCO), Washington, DC, 2005.
[19] G. Y. Li, T. Li, J. Zeng, and H. B. Li, "An improved V-detector algorithm of identifying boundary self," in 2009 International Conference on Machine Learning and Cybernetics, vol. 6. Baoding, China, 2009, pp. 3209-3214.
[20] G. Li, T. Li, J. Zeng, and H. Li, "Significant Level in V-detector Generation Algorithm," in 2009 Asia-Pacific Conference on Information Processing, vol. 2. Shenzhen, China 2009, pp. 513-516.
[21] Z. Ji, "V-detector java source code and 2DSyntheticData. http://www.zhouji.net/prof/vdetector.html," 2006.

Guiyang Li received the M.S. degree from computer science of Southwest Jiaotong University in 2002 and is currently working toward the Ph.D. degree in computer science at Sichuan University, China. He is also an instructor in the Department of Computer Science, Sichuan Normal University, China. He is actively involved in the development of network security. His current areas of interest include artificial immune computation, network security.Email:guiyang.li@gmail.com

Tao Li received the Ph.D. degree in Computer Science from University of Electronic Science and Technology of China in 1994. From 1994 to 1995, he was a visiting scholar in University of California at Berkeley for neural networks theory. Now, he is the professor in the Department of Computer Science at Sichuan University, China. His current research interests include network security, artificial immune theory and disaster recovery application. Email: litao@scu.edu.cn

Jie Zeng received the M.S. degree in computer science in 2006 and is currently working toward the Ph.D. degree in computer science at Sichuan University, China. His current areas of interest include artificial immune computation, network security. Email: jet.zengjie@gmail.com

Haibo Li received the M.S. degree in computer science in 2007 and is currently working toward the Ph.D. degree in computer science at Sichuan University, China. His current areas of interest include artificial immune computation, network security. Email: whatersky9908@sina.com