The Dual Negative Selection Algorithm Based on Pattern Recognition Receptor Theory and Its Application in Two-class Data Classification

Xufei Zheng[a], Yanhui Zhou[a], Yonghui Fang[b]
[a] Faculty of Computer and Information Science, Southwest University, Chongqing 400715, China
Email: {zxufei, xiaohui}@swu.edu.cn
[b] Faculty of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
Email: fyhui@swu.edu.cn

***Abstract*— Negative Selection Algorithm (NSA) is an important artificial immune data classifiers generation method in Artificial Immune System (AIS) research. However, with the increase of the data dimensions, the current data classification algorithms which based on NSA exist the problems of excessive number of generated classifiers and too low classifier generation efficiency. In this paper, the Dual Negative Selection Algorithm based on Pattern Recognition Receptor theory (PRR-2NSA) is proposed, which simulates the process of Antigen Presenting Cells (APC) recognized the Pathogen-Associated Molecular Patterns (PAMP) to trigger the immune response. The PRR-2NSA algorithm generates the APC classifier based on training set clustering firstly, and then generates the T-cell classifiers within the coverage of the APC classifier set with dual negative selection algorithm (2NSA) secondly. The 2NSA avoids the unnecessary and time-consuming self-tolerance process of candidate classifier within the coverage of existing mature classifiers, thus greatly reduces classifier set size, significantly improves classifier generation efficiency. The PRR-2NSA introduces the APC classifiers' co-stimulation to the T-Cell classifier, which reduce the occurrence of false classification on one hand, and accelerate the data classification efficiency on the other hand. Theoretical analysis and simulations show that the PRR-2NSA algorithm effectively improves classification efficiency and reduces the time cost of algorithm.**

***Index Terms*— artificial immune system, real-valued negative selection algorithm, variable-sized classifier, dual negative selection algorithm, PRR-2NSA**

I. Introduction

In data mining research, data classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. A large number of classification algorithms have been proposed, including Bayesian classifier [1], K-Nearest Neighbor algorithm (KNN) [2], Fuzzy C Mean (FCM) [3], the decision tree method [4], BP neural network algorithm [5] and Artificial Immune System (AIS) [6]. The data classification method has been widely applied to pattern recognition, data mining, artificial intelligence and network intrusion detection, etc.

In recent years, it is the research focus which introduced artificial immune theory to the field of data mining [6]. The researchers from different angles to simulate biological immune mechanism to data classification analysis, including data classification method based on negative selection algorithm (Ji et al. put forward the real-valued negative selection algorithm and applied to network intrusion detection [7], [8]). They regarded monitoring targets (such as legal user activities, legal application usage activities, etc.) as self and expected the NSA to discriminate them from others (such as illegal user activities, virus infected data, network worm, etc.). As well as data classification method based on the immune network theory (De Castro et al. put forward the aiNet model [9] and Timmis et al. put forward the RLAIS method [10] used for data classification).

In this paper, we analyze the two-class data classification problem based on real-valued negative selection algorithm, and propose the dual negative selection algorithm based on pattern recognition receptor theory (PRR-2NSA). The PRR-2NSA can be used in many two-class data classification applications, such as data classification, data mining, pattern recognition and network intrusion detection, etc.

II. Related Works

The Negative Selection Algorithm (NSA), first proposed by Forrest [11], simulates the immune tolerance process of $T-$cells in thymus to generate detectors which avoid self reaction. The mature detectors are subsequently used for the recognition of non-self and applied to many important researches, such as data classification, data mining, pattern recognition, and anomaly detection, etc. [6], [8], [12]–[16].

The early NSA [11], termed SNSA (String represented Negative Selection Algorithm), which encodes antibody (classifier) and antigen (samples) as binary strings and calculates the affinity (match degree) between them by the r-contiguous-bits matching rule. The inefficiency problem

Manuscript received December 01, 2012; revised January 2, 2012; accepted April 16, 2012. © 2005 IEEE.
Project supported by the doctoral fund of southwestern university, china (NO. SWU112038)
Corresponding author: Yanhui Zhou, xiaohui@swu.edu.cn

doi:10.4304/jcp.8.8.1951-1959

of SNSA was discussed in refs. [17], [18]: the probability of candidate classifiers matured by passing the negative selection process is $P = (1 - P_m)^{N_s}$, where P_m is the match probability of candidate classifier and antigen, N_s is the training set size; thus with the increase of N_s, P will tend to be 0 ultimately; moreover, $N_0 = \frac{-\ln(P_f)}{P_m \cdot (1-P_m)^{N_s}}$ candidate classifiers are needed to reach the given failure probability $P_f \approx e^{-P_m \cdot N_c}$, which means that the count of candidate classifiers N_0 is exponentially related to the count of training set N_s, and the time complexity of SNSA is $O(N_0 \cdot N_s) = O\left(\frac{-\ln(P_f) \cdot N_s}{P_m \cdot (1-P_m)^{N_s}}\right)$.

The RNSA (Real-valued Negative Selection Algorithm) uses a fixed classifier radius r_c, and sets the count of detectors as the condition of algorithm termination [13]. In RNSA, the candidate detector was randomly generated with center $X(x_1, x_2, ..., x_n)$ firstly, and then the shortest Euclidean distance dis_{min} between the candidate detector and all self elements in training set was calculated, and finally the mature detector was generated if $dis_{min} > r_s + r_c$, where r_c is the radius of detector, and r_s is the radius of self element.

The V-Detector (Real-valued Negative Selection Algorithm with Variable-size Detector) uses variable-sized detector radius, and sets the expected coverage as the condition of algorithm termination [7], [8]. In V-Detector, the candidate detector was randomly generated with center $X(x_1, x_2, ..., x_n)$ firstly, and then the shortest Euclidean distance dis_{min} between the candidate detector and all self elements in training set was calculated, and finally the mature detector was generated if $dis_{min} > r_s$, where the radius of detector is $r_c = dis_{min} - r_s$.

As indicated in refs. [17], [18], for pattern recognition algorithms based on distance calculation, the primary time consumption is the distance calculation. Stibor et al. [17] pointed out that the unacceptable high time cost of RNSAs is caused by the inefficiency of the classifier generation process, and which significantly limited the applications of AIS.

Aydin et al. [15] and Gao et al. [19] combine the genetic algorithm and chaos theory to optimize classifier generation process, which reduce the candidate classifiers' overlapping coverage. Bereta et al. [14] combine K-Means data clustering method to simplify negative selection process and applied to the data analysis. Gong. et al. [16] have two self-set training process in the self-tolerance stage on the basis of V-Detector, in order to improve the classifier generation efficiency.

Both RNSA and V-Detector employed only once negative selection process to eliminate the self-recognized invalid classifiers by matching candidate classifier with whole training set. In the negative selection process, there is only consideration of the relationship between candidate classifier and training set but without any consideration of repetitive coverage of candidate classifier with existing classifier set, which bring about the unnecessary self-tolerance of the candidate classifier which repetitive covered. Thus, the unnecessary self-tolerance of these candidate classifiers resulted in an excessive count of mature classifiers and extremely lowered classifier generation efficiency, and increased the computation time complexity of these RNSAs.

III. The basic definition of RNSA

Inspired by the self and non-self (SNS) theory [20], Forrest proposed the NSA to eliminate the self reactive detectors [11]. In this paper, the real-valued negative selection algorithm is discussed, and some basic conceptions of RNSA are defined as follows:

Def 1 **Antigen** $Ag = \{ag | ag =< x_1, x_2, ..., x_n >, x_i \in [0, 1]\}$, which represents all samples in the feature space, where n is the data dimension.

Def 2 **Self set** $Self \subset Ag$, which represents all normal samples in the antigen set Ag; **Non-self set** $Nonself \subset Ag$, which represents all abnormal samples in the antigen set Ag, and which satisfies formula 1.

$$Self \cup Nonself = Ag, Self \cap Nonself = \phi. \quad (1)$$

Def 3 **Training set** $Train \subset Self$, which represents the prior knowledge of detection, $r_s \in [0, 1]$ is the radius of self and N_s is the size of training set.

Def 4 **Classifier set** $CS = \{c | c =< y_1, y_2, ..., y_n, c_d >, y_i \in [0, 1], c_d \in [0, 1]\}$, which represents the mature classifier set generated by NSA based on the training set, where c_d is the radius of classifier and N_c is the size of the classifier set.

Def 5 **Estimated coverage rate** $C = \frac{Num_{covered}}{Num}$, which represents the ratio of samples fall in the coverage of classifier set CS and total samples in a sampling period, where Num is the count of total samples and $Num_{covered}$ is the count of samples fall in the coverage of classifier set CS in a sampling period.

$$C = \frac{Num_{covered}}{Num}. \quad (2)$$

Def 6 **Classification process** $f(Train) \rightarrow Nonself$, which represents the process to identify non-self set based on the self antigens training set.

IV. The implementation strategies of PRR-2NSA

In order to solve the problems of low classifier generation efficiency and high misclassification rate, the dual negative selection algorithm based on pattern recognition receptor theory (PRR-2NSA) is proposed in this paper which combines the Pattern Recognition Receptors theory (PRR) and the dual negative selection algorithm (2NSA).

A. The Pattern Recognition Receptor (PRR) Theory

In 1989, the famous immunologist Janeway first proposed the PRR theory [21]. In biology, the PRR model added additional layer of pathogen-associated molecular patterns (PAMP) to the self-nonself model [22]. The PRR model assumes that APC are quiescent until they are activated via encoded pattern recognition receptors

that recognize conserved PAMPs. To mirror this, T-cell classifiers in the proposed algorithm are first used to recognize the antigen according to negative selection. The co-stimulation of APC classifier will not be conducted until the so-called suspicious antigen is encountered in the system. The co-stimulation of APC classifier will not be conducted until the detection from T-cell classifiers becomes unsure, that is, the suspicious antigen is encountered in the system. Although this solution shows its strength in terms of algorithmic complexity, its performance relies on the application domain since the definition of suspicious antigen is not always in accordance with a specific application.

Inspired by this metaphor, we combine both PRR theory and T-cell negative selection process to achieve the data classification of two-class dataset, which had been proven to effectively reduce high false classification rate that often occurred in traditional NSAs.

B. The Implementation Strategies of PRR-2NSA

There are two separate stages in the traditional two-class data classification algorithm which based on negative selection process, respectively, the stage of antigen toleration process to generate data classifier stage and the stage of data classification process by using classifier set. The PRR-2NSA includes three separate stages, respectively: 1) the antigen clustering to generate APC classifier stage; 2) the negative selection process to generate T-cell classifier stage; and 3) using the generated APC and T-cell Classifier to execute data classification stage.

1) The Antigen Clustering to Generate the APC Classifier Stage: The APC classifier is generated by antigen training set hard clustering. The definition of hard clustering and dissimilarity measure are shown as Def.7 and Def.8, as well as the nearest neighbor metric of data vector x and cluster C is calculated by the nearest neighbor metric function.

Def 7 Hard Clustering Suppose data set $X = \{x_1, x_2, ..., x_n\}$, where x is data vector. The m clusters of data set X is the m subset $s_1, s_2, ..., s_m$ of m that satisfies the formula 3.

$$\begin{cases} \emptyset \subset c_i \subset X, i = 1, 2, \cdots, m \\ \bigcup_{i=1}^{m} c_i \\ c_i \cap c_j = \emptyset, i \neq j, i, j = 1, 2, \cdots, m \end{cases} \tag{3}$$

Def 8 Dissimilarity Measure Function $d : X \times X \to \Re$ is the dissimilarity measure, where $\Re$ is the set of real numbers, and d satisfies the formula 4.

$$\begin{cases} \exists d_0 \in \Re : -\infty < d_0 \leq d(x, y) < +\infty, \forall x, y \in X \\ d(x, x) = d_0, \forall x \in X \\ d(x, y) = d(y, x), \forall x, y \in X \end{cases} \tag{4}$$

There are three kinds of nearest neighbor metric function, respectively, maximum neighbor function (formula 5), minimum neighbor function (formula 6) and average neighbor function (formula 7).

$$\rho_{\max}(x, C) = \max_{y \in C} \rho(x, y). \tag{5}$$

$$\rho_{\min}(x, C) = \min_{y \in C} \rho(x, y). \tag{6}$$

$$\rho_{avg}(x, C) = \frac{1}{n_c} \sum_{y \in C} \rho(x, y). \tag{7}$$

After the training set data clustering, we get the APC classifier set, where the number of APC classifiers C_{num} is the number of clusters, and the center of each APC classifier $X = \langle x_1, x_2, \cdots, x_n \rangle$ is every cluster's center, that is $CS_{APC} = \{apc | apc = \langle x_1, x_2, \cdots, x_n, r_{apc} \rangle, x_i \in [0, 1]\}$. The radius of each APC classifier is the maximum distance between the classifier's center and every antigen's center, that is $r_{apc} = dis_{\max}(X, E_i)$, where E_i is the element in the cluster. There are two important purposes of APC classifier in the PRR-2NSA algorithm, respectively: 1) the rapid response of the data to be classified for the coverage of APC classifier set; and 2) the APC classifiers' co-simulation to the T-cell classifier can help to reduce the false classification rate. The complete algorithm process is shown as table I.

In order to illustrate the process of antigen clustering to generate APC classifier, we have an experiment to generate APC classifiers through the antigen training set clustering with 25 "$Iris - Setosa$" instances in the "Iris" dataset (4-dimension). In order to display the result in 2-dimensional graphics, only the "$sepallength$" and "$petallength$" properties are selected. All data normalized to the real value $[0, 1]$ space, the radius of self is $r_s = 0.05$. As shown as Figure 1, the "$Iris - Setosa$" training set are clustered as 3 APC classifiers, where the filling small rounds are the "$Setosa$" antigen training set elements and the dash circles are the APC classifiers.

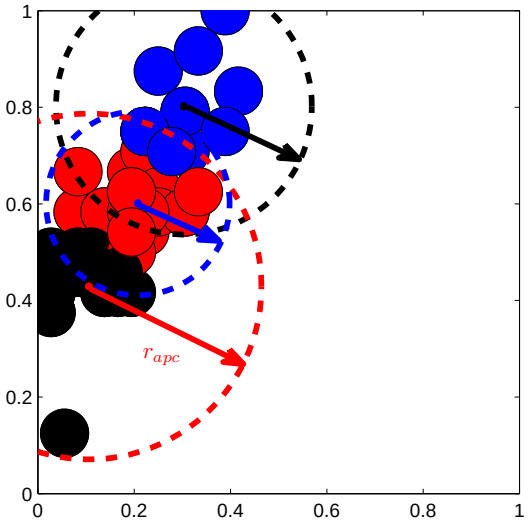

Figure 1. The antigen clustering to generate APC classifier set, where training antigen set is "Iris-Setosa", $r_s = 0.05$, the filling small rounds are training antigen elements, the dash circles are APC classifiers.

TABLE I.
THE PSEUDO CODE OF THE ANTIGEN CLUSTERING TO GENERATE APC CLASSIFIER STAGE.

$Train$: antigen training set	CS_{APC} : APC classifier set r_{apc} : the radius of APC classifier
Step 1	Initialization of the antigen training set $Train$ and APC classifier set $CS_{APC} = \emptyset$.
Step 2	Set the initial number of clusters, and using K-Means hard clustering method to execute data clustering of $Train$.
Step 3	Generation of C_{num} APC classifiers according to the clustering results.
Step 4	Determining appropriateness of each APC classifier radius, if $r_{apc} > 5r_s$ then $C_{num} \leftarrow C_{num} + 1$, else if $r_{apc} < 2r_s$, then $C_{num} \leftarrow C_{num} - 1$.
Step 5	If each $2r_s \leq r_{apc} \leq 5r_s$, then $CS_{APC} \leftarrow CS_{APC} \bigcup \{APC_{classifier}\}$, and stop.

2) The Implementation Strategies of 2NSA: In order to overcome the shortcomings of the current RNSAs, we propose the dual negative selection algorithm (2NSA). The 2NSA uses variable-sized classifier radius, and sets the expected coverage as the condition of algorithm termination. In the 2NSA, the randomly generated candidate classifier tolerates with classifier set to generate semi-mature classifier firstly (the 1st negative selection); and then the semi-mature classifier tolerates with training set to generate mature classifier (the 2nd negative selection).

The 1st negative selection process: every randomly generated candidate classifier c_{new} tolerates with mature classifier set and becomes semi-mature classifier when it does not match any existing mature classifier. The candidate classifier c_{new} was randomly generated with center $X(x_1, x_2, ..., x_n)$ firstly, and then calculated the Euclidean distance $dis(c_{new}, c_i)$ between the candidate classifier and every mature classifier c_i in the classifier set CS. The candidate classifier successfully tolerated with the classifier set and becomes a semi-mature classifier c_{semi} if the c_{new} satisfies the formula (8). Otherwise, the candidate classifier c_{new} will be eliminated if it been recognized by any mature classifier, that is the termination of the 1st negative selection process, and a new candidate classifier c_{new} will be randomly generated and the 1st negative selection process be restarted again.

$$dis(c_{new}, c_i) > r_{c_i} \qquad i = 1, 2, ..., N_d. \tag{8}$$

The 2nd negative selection process: the semi-mature classifier c_{semi} tolerates with self set and becomes mature classifier when it does not match any self element. The shortest distance $dis_{min}(c_{semi}, s_j)$ between the center $Y(y_1, y_2, ..., y_n)$ of semi-mature classifier c_{semi} and every self element of training set was calculated according to the Euclidean distance. The c_{semi} successfully tolerated with the training set and becomes a mature classifier c_{mat} if the formula (9) be satisfied, the mature classifier c_{mat} joins to the classifier set $CS, CS \leftarrow CS \cup \{c_{mat}\}$, the radius of c_{mat} is $r_c = dis_{min}(c_{semi}, s_j) - r_s$. Otherwise, the semi-mature classifier c_{semi} will be eliminated if it been recognized by any self element, and the 1st negative selection process will be restarted again.

$$dis_{min}(c_{semi}, s_j) > r_s \qquad j = 1, 2, ..., N_s. \tag{9}$$

The 2NSA algorithm avoids the unnecessary and time-consuming self-tolerance process of candidate classifier within the coverage of existing classifier set. The candidate classifier that repetitively covered with existing mature classifiers will be eliminated in the 1st negative selection process, and thus decreases classifier set size and improves classifier generation efficiency. The pseudo code of 2NSA is shown as table II.

In order to illustrate the 2NSA algorithm, we have an experiment to generate classifiers through the self-tolerance with 25 $Setosa$ instances in the Iris dataset. That is, the 25 $Setosa$ instances composed the self training set, and the other two kinds of flower ($Versicolour$ and $Virginica$) composed the non-self set. In order to display the result in 2-dimensional graphics, only the $sepallength$ and $petallength$ 2 properties are selected. All data normalized to the real value $[0, 1]$ space, the radius of self is $r_s = 0.05$, and the classifier in RNSA with fixed radius $r_c = 0.10$.

The implementation strategy of 2NSA as well as the difference from RNSA and V-Detector can be illustrated by Figure 2. As shown as Figure 2, the classifiers generated with fixed radius in RNSA and with variable-sized radius in V-Detector, and many classifiers repetitively cover with existing classifiers in both RNSA and V-Detector, thus results in many candidate classifiers undergone the unnecessary and time-consuming self-tolerance process. However, in the 2NSA, the candidate classifiers tolerate with mature classifier set CS to generate semi-mature classifier firstly, which avoids the repetitive coverage with existing classifier set CS and guarantees the center of semi-mature classifier locates outside the coverage of the classifier set in meanwhile. The additional negative selection process avoids the unnecessary and time-consuming self-tolerance process and ensures that the new generated mature classifier covers more uncovered non-self space. As shown as Figure 2, the classifier set size of 2NSA is dramatically reduced compare to RNSA and V-Detector.

SNSA, RNSA and V-Detector are major classifier generation algorithms which widely used in data classification and pattern recognition fields. The 2NSA eliminates the unnecessary and time-consuming self-tolerance process of candidate classifiers locate outside the coverage of existing classifier set CS through the 1st negative selection process, thus dramatically reduces the classifier set size and the time complexity of current NSAs, greatly improves the classifier generation efficiency, and reduces the system false classification rate.

TABLE II.
THE PSEUDO CODE OF THE DUAL NEGATIVE SELECTION ALGORITHM (2NSA).

$Train$: self training set	CS : classifier set $r(c_i)$: the radius of classifier c_i C : estimated coverage C_{exp} : expected coverage
Step 1	Initialization of the training set $Train$ and classifier set $CS \leftarrow \phi$.
Step 2	Randomly generate a candidate classifier c_{new}, and calculate the distance between c_{new} and every classifier c_i in classifier set CS, goto Step 4 if $dis(c_{new}, c_i) \leq r(c_i)$.
Step 3	The candidate classifier c_{new} that successfully tolerated with classifier set CS changes to semi-mature classifier $c_{new} \rightarrow c_{semi}$, and then calculate the shortest distance dis_{min} between the c_{semi} and every self element in $Train$, If $dis_{min} > r_s$, then $c_{semi} \rightarrow c_{mat}$, $r(c_{mat}) = dis_{min} - r_s, CS \leftarrow CS \cup \{c_{mat}\}$.
Step 4	Calculate current estimated coverage C, goto Step 2 if $C < C_{exp}$, else stop and return the classifier set CS.

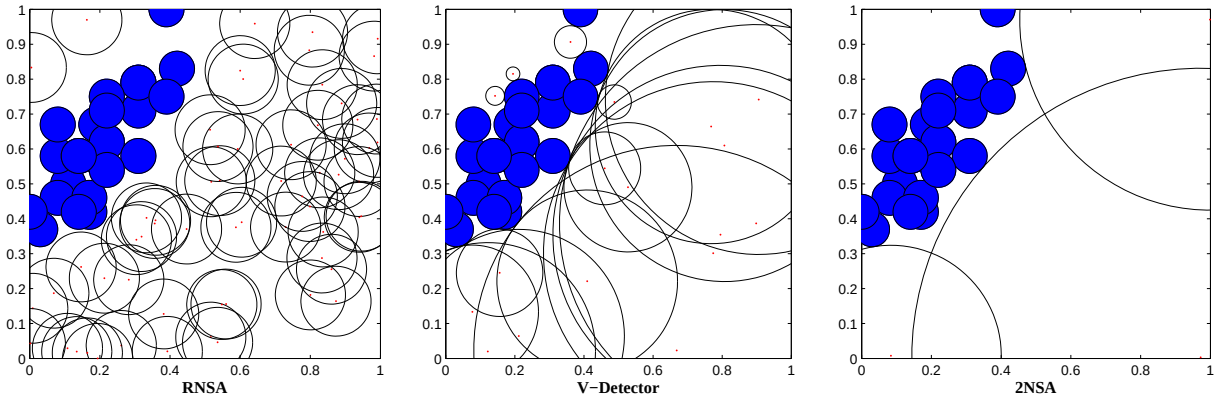

Figure 2. When the expected coverage reaches $C_{exp} = 90\%$, the classifier set size of RNSA, V-Detector and 2NSA are 62, 18, and 3 respectively, where the blue circles are the 25 $Setosa$ self elements in the training set and the white circles are the generated classifiers.

3) Using 2NSA to Generate T-cell Classifier Stage: After the introduce of PRR model, since the PAMP recognition of APC classifier, then the coverage of the APC classifier set defines the PAMP type data, and thus outside the coverage area of the APC classifier set can be judged as an alternative space. Therefore, the real-valued negative selection algorithm which based on PRR theory (PRR-2NSA) conducts antigen self-tolerance to generate T-cell classifier within the coverage of the APC classifier set, but not the whole real value space $[0, 1]^n$.

The complete negative selection process of the PRR-2NSA is roughly the same as the 2NSA algorithm. The difference of the two algorithms lies in the range to generate T-cell classifier, in which 2NSA in the whole real value space $[0, 1]^n$ but PRR-2NSA in the coverage of the APC classifier set. The PRR-2NSA algorithm used variable-sized classifier radius, and set the expected coverage as the condition of algorithm termination. In PRR-2NSA, the candidate classifier was randomly generated with center $X(x_1, x_2, ..., x_n)$ firstly, and then the shortest Euclidean distance dismin between the candidate classifier and all antigen elements in $Train$ was calculated, and finally the mature classifier was generated if $dis_{min} > r_s$, where the radius of T-cell classifier is $r_c = dis_{min} - r_s$.

In order to illustrate the negative selection process to generate T-cell classifier, we have an experiment to generate T-cell classifier through the antigen training set self-tolerance with "$Iris-Setosa$" instances in the "Iris" dataset. As shown as Figure 3, the 16 T-cell classifiers are generated within the coverage of 3 APC classifiers by dual negative selection process, where the dash circles are APC classifiers, and the green circles are T-cell classifiers.

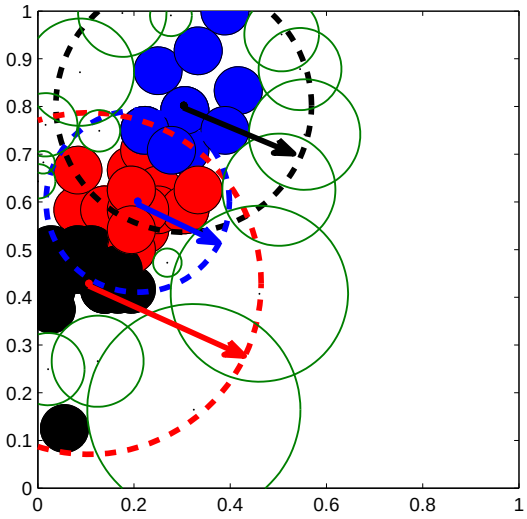

Figure 3. The negative selection process to generate T-cell classifier within the coverage of APC classifier set, where training antigen set is 'Iris-Setosa", $r_s = 0.05$, the points are training antigen elements, the dash circles are APC classifiers, and the green circles are T-cell classifiers.

4) Using the generated APC and T-cell classifier set to classify data: There are some differences between the classification process of PRR-2NSA and traditional

RNSAs. On one hand, T-Cell classifier need the co-stimulation of APC classifier to achieve the classification of abnormal data in the PRR-2NSA, thus effectively reducing the false classification rate. On the other hand, for the clustering normal data of APC classifiers, thus the coverage outside of the APC classifier set can be judged as abnormal data directly, thereby effectively enhancing the data classification efficiency.

V. Performance Evaluation

In this section, the effectiveness and performance of the PRR-2NSA is verified by a group of comparative experiments. The experimental data are the classical UCI standard datasets: "Iris" [23], "Breast Cancer Wisconsin Diagnostic" (BCW) [24] and "Chess" [25], which have been widely used for the performance test and generation effective analysis of classifiers [6], [8], [13]–[15], [19]. As pointed out in ref. [7], [8], the real-valued negative selection algorithm with fixed radius (RNSA) has poor classifiers' generation efficiency and performance. Therefore, we just compare the effectiveness and performance of the V-Detector, 2NSA and PRR-2NSA in the paper.

The Metrics the effectiveness of these algorithms include the classifier set size (CSS), the true classification rate (TCR), the false classification rate (FCR) and the classification time (CT). The experimental parameters are shown as table III, and all data is preprocessed to real-value and normalized in range $[0, 1]$. All experiment were repeated 50 times and averaged.

Def 9 True Classification Rate TCR, which represents the ratio of true positive count and the total non-self samples identified by classifier set, where TP and FN are the counts of true positive and false negative.

$$TCR = TP/(TP + FN). \tag{10}$$

Def 10 False Classification Rate FCR, which represents the ratio of false positive count and the total self samples identified by classifier set, where FP and TN are the counts of false positive and true negative.

$$FCR = FP/(FP + TN). \tag{11}$$

In order to verify the detection capability (Metric with TCR and FCR) of PRR-2NSA, we conduct the comparative experiments with the datasets of "Iris", "BCW" and "Chess" in comparison with V-Detector and 2NSA.

The results are shown as table IV and table V. As shown as table IV, the comparison of these 3 RNSAs' true classification rate in low-dimensional dataset ("Iris" dataset is 4-dimension and "Chess" dataset is 5-dimension) and high-dimensional dataset ("BCW" dataset is 30-dimension), in which TCR is roughly the same in the same dataset. As shown as table V, the comparison of these 4 RNSAs' false positive rate in different dataset, in which FCR is very different. For example, when estimated coverage rise up to $C = 99\%$, the FCR of RNSA, V-Detector, 2NSA and FPR-2NSA are $FCR_{V-Detector} = 45.76\%$, $FCR_{2NSA} = 22.49\%$ and $FCR_{PRR-2NSA} = 8.22\%$ respectively in "Iris" dataset, there is 82.03% and 63.45% drop of PRR-2NSA in comparison with V-Detector and 2NSA. Accordingly, when estimated coverage rise up to $C = 99\%$, there is 69.79% and 55.36% drop of PRR-2NSA in comparison with V-Detector and 2NSA in "Chess" dataset, as well as there is 76.15% and 53.21% drop of PRR-2NSA in comparison with V-Detector and 2NSA in "BCW" dataset.

In sum up, the PRR-2NSA algorithm significantly reduce the false classification rate FCR in different dimensions and scale of training set, and keep the almost the same true classification rate TCR in comparison with V-Detector and 2NSA.

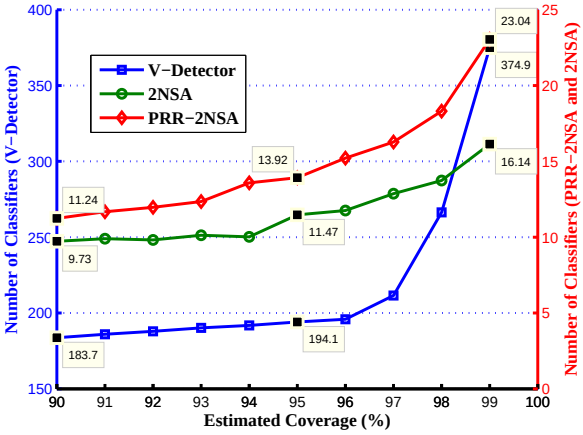

Figure 4. The comparison of classifier set size of V-Detector, 2NSA and PRR-2NSA in low-dimensional dataset, where dataset is "Iris", the training set is "Iris-Setosa", $r_s = 0.05, Num = 200$.

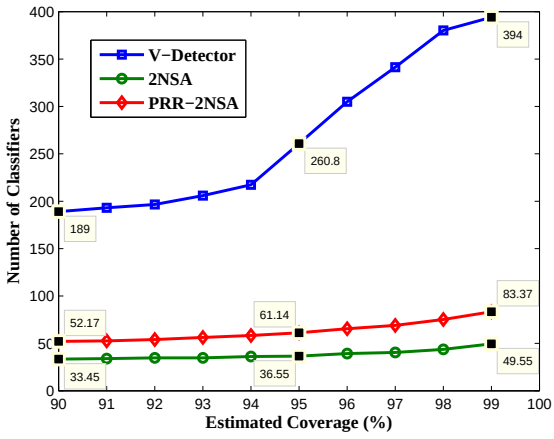

Figure 5. The comparison of classifier set size of V-Detector, 2NSA and PRR-2NSA in high-dimensional dataset, where dataset are "BCW", the training set is "BCW-Benign", $r_s = 0.05, Num = 200$.

In order to verify the classifier set generation capability (Metric with CSS) of PRR-2NSA, we conduct the comparative experiments with the datasets of "Iris", "BCW" and "Chess" in comparison with the V-Detector and 2NSA.

The results are shown as Figure 4 $\sim$ 6. As shown as Figure 4, the comparison of the 3 RNSAs' classi-

TABLE III.
EXPERIMENT PARAMETERS OF RNSA, V-DETECTOR, 2NSA AND PRR-2NSA.

Dataset	Instances	Dimension	Type	Self Set	Non-self Set	Training Set and Size	Detection Set and Size
Iris	150	4	real	Setosa: 50	Versicolour:50 Virginica: 50	Setosa: 25	Setosa: 25 Versicolour: 25 Virginica: 25
BCW	569	30	real	benign: 357	malicious: 212	benign: 150	benign: 150 malicious: 150
Chess	28056	6	Categorical Integer	draw: 2796	others: 25260	draw: 1000	draw: 1000 others: 2000

TABLE IV.
THE COMPARISON OF THE TCR AMONG V-DETECTOR, 2NSA AND PRR-2NSA IN DIFFERENT DATASET

		True Classification Rate in Different Estimated Coverage (%)									
Dataset	Algorithm	50%	60%	70%	80%	90%	92%	94%	96%	98%	99%
Iris	V-Detect	92.63	94.27	95.90	96.64	97.26	97.82	98.15	98.40	98.81	99.47
	2NSA	92.56	94.12	95.23	95.96	96.64	97.34	97.86	98.22	98.69	99.38
	PRR-2NSA	93.40	94.71	96.18	97.32	97.75	98.23	98.54	98.90	99.26	99.57
Chess	V-Detect	67.02	76.84	83.92	89.53	93.27	94.62	95.93	96.60	98.01	98.72
	2NSA	66.58	75.36	82.54	89.05	93.03	94.47	95.76	96.53	97.89	98.46
	PRR-2NSA	68.24	77.08	84.27	89.71	93.60	94.96	96.11	96.72	98.19	99.01
BCW	V-Detect	78.57	83.97	87.31	88.96	90.34	90.72	91.29	91.74	92.65	94.05
	2NSA	77.94	83.03	86.79	88.32	89.25	89.96	91.18	91.50	92.39	93.87
	PRR-2NSA	80.21	84.15	87.62	89.14	90.81	91.10	91.51	92.06	93.17	94.68

TABLE V.
THE COMPARISON OF THE FCR AMONG V-DETECTOR, 2NSA AND PRR-2NSA IN DIFFERENT DATASET

		False Classification Rate in Different Estimated Coverage (%)									
Dataset	Algorithm	50%	60%	70%	80%	90%	92%	94%	96%	98%	99%
Iris	V-Detect	22.16	25.07	29.40	32.08	34.42	35.86	37.65	39.51	42.61	45.76
	2NSA	9.52	11.64	13.22	14.78	16.05	16.53	17.60	18.32	20.35	22.49
	PRR-2NSA	4.27	5.20	5.87	6.34	6.71	6.94	7.16	7.42	7.69	8.22
Chess	V-Detect	25.97	31.33	35.68	40.85	45.96	49.54	54.78	63.82	67.27	69.85
	2NSA	15.77	18.52	21.70	24.59	27.18	31.12	33.57	39.32	44.51	47.27
	PRR-2NSA	8.05	9.61	11.84	13.73	15.29	16.50	17.12	18.44	19.67	21.10
BCW	V-Detect	5.18	5.38	5.45	5.83	6.17	6.38	6.53	6.78	7.19	8.26
	2NSA	2.37	2.68	3.05	3.31	3.56	3.73	3.86	4.02	4.15	4.21
	PRR-2NSA	1.04	1.15	1.28	1.42	1.54	1.63	1.70	1.78	1.92	1.97

fier set size in low-dimensional dataset ("Iris" dataset is 4-dimension), in which classifier set size increased dramatically in V-Detector but only increased slowly in 2NSA and PRR-2NSA with the grow of estimated coverage. When estimated non-self coverage rise up to $C = 99\%$, the classifier set size of RNSA, V-Detector and 2NSA are $CSS_{V-Detector} = 374.9$, $CSS_{2NSA} = 16.14$ and $CSS_{PRR-2NSA} = 23.04$ respectively, there is 93.85% drop 29.94% rise of PRR-2NSA in comparison with V-Detector and 2-NSA. As shown as Figure 5, the comparison of classifier set size of V-Detector and 2NSA in high-dimensional dataset ("BCW" dataset is 30-dimension), in which the classifier set size of 2NSA and PRR-2NSA are significantly reduced. As shown as Figure 6, the comparison of the 3 RNSAs' classifier set size with big training set ($N_s = 1000$ in "Chess" dataset). When estimated non-self coverage rise up to $C = 99\%$, the classifier set size of RNSA, V-Detector and 2NSA are $CSS_{V-Detector} = 1860$, $CSS_{2NSA} = 567.5$ and $CSS_{PRR-2NSA} = 949.9$ respectively, the classifier set size of 2NSA and PRR-2NSA also significantly reduced in comparison with V-Detector.

In sum up, the 2NSA and PRR-2NSA algorithm significantly reduce the classifier set size in different data dimensions and scale of training set. For the introduce of APC classifier in the PRR-2NSA, the T-Cell classifier need to be generated within the coverage of APC classifier set, which resulting in the more smaller classifier's radius in comparison with 2NSA, and thus resulting in the bigger PRR-2NSA's T-Cell classifier set size than 2NSA.

In order to verify the detection generation efficiency (including classifier set generation time and classification time, Metric with CT) of PRR-2NSA, we conduct the comparative experiments with the datasets of "Iris", "BCW" and "Chess" in comparison with the V-Detector and 2NSA.

The results are shown as Figure $7 \sim 9$. As shown as Figure 7, the comparison of the 3 RNSAs' classification efficiency in low-dimensional dataset ("Iris" dataset is 4-dimension), in which classification efficiency of PRR-2NSA has improved dramatically. When estimated coverage rise up to $C = 99\%$, the time cost of V-Detector, 2N-

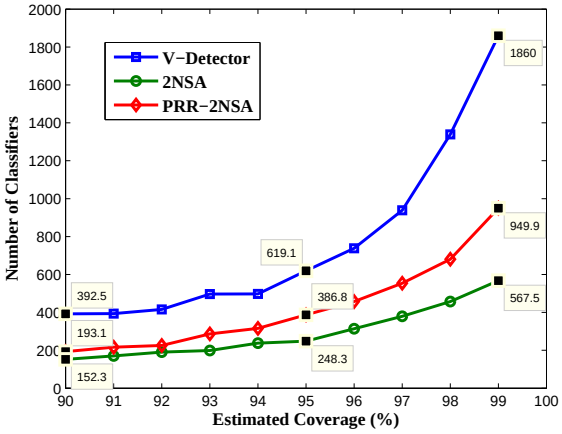

Figure 6. The comparison of classifier set size of V-Detector, 2NSA and PRR-2NSA with big training set, where dataset is "Chess", the training set is "Chess-Draw", $r_s = 0.05, Num = 200$.

SA and PRR-2NSA are $CT_{V-Detector} = 0.9234$ seconds, $CT_{2NSA} = 0.5286$ seconds and $CT_{PRR-2NSA} = 0.1652$ seconds respectively, there is 82.11% and 68.75% drop of PRR-2NSA in comparison with V-Detector and 2NSA. As shown as Figure 8, the comparison of the classifier generation efficiency of V-Detector and 2NSA in high-dimensional dataset ("BCW" dataset is 30-dimension), in which classification efficiency in high-dimension is also significantly improved of PRR-2NSA in comparison with V-Detector and 2NSA. As shown as Figure 9, the comparison of the 3 RNSAs' classification efficiency with big training set ($N_s = 1000$ in Chess dataset). When estimated non-self coverage rise up to $C = 99\%$, the time cost of V-Detector, 2NSA and PRR-2NSA are $CT_{V-Detector} = 159.4$ seconds, $CT_{2NSA} = 80.93$ seconds and $CT_{PRR-2NSA} = 28.45$ seconds respectively, there is 82.15% and 64.83% drop of PRR-2NSA in comparison with V-Detector and 2NSA.

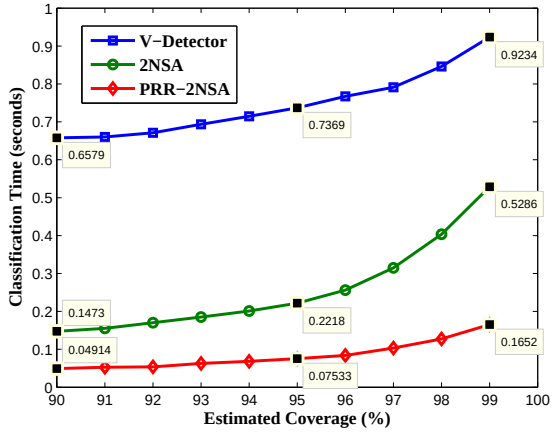

Figure 7. The comparison of classification efficiency of V-Detector, 2NSA and PRR-2NSA in low-dimensional dataset, where dataset is "Iris", the training set is "Iris-Setosa", $r_s = 0.05, Num = 200$.

In sum up, the 2NSA and PRR-2NSA algorithm signif-

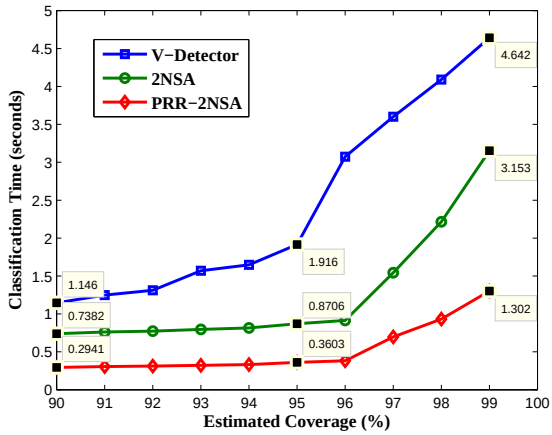

Figure 8. The comparison of classification efficiency of V-Detector, 2NSA and PRR-2NSA in high-dimensional dataset, where dataset is "BCW", the training set is "BCW-Benign", $r_s = 0.05, Num = 200$.

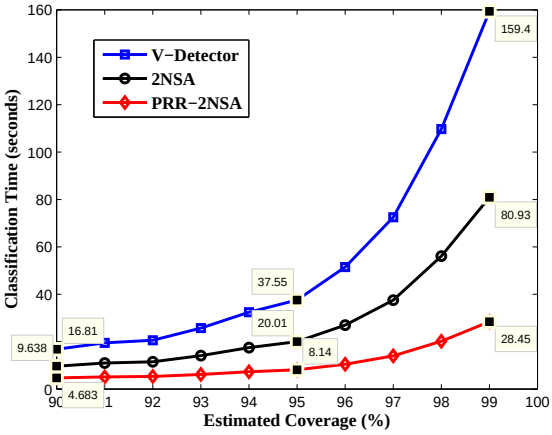

Figure 9. The comparison of classification efficiency of V-Detector, 2NSA and PRR-2NSA with big training set, where dataset is "Chess", the training set is "Chess-Draw", $r_s = 0.05, Num = 200$.

icantly reduce the classifier generation and classification time cost in different data dimensions and scale of training set in comparison with V-Detector. For the introduce of APC classifier in the PRR-2NSA, the data lied outside of the coverage of the APC classifier set can be judged directly as abnormal data which effectively enhance the data classification efficiency.

VI. Conclusions

The NSA is an important two-class data classifier generation algorithm in data mining field. The RNSA encodes antigens and antibodies using fixed classifier radius. The V-Detector algorithm uses variable-size classifier radius and achieves better classification results than RNSA in the simulation experiments. However, both RNSA and V-Detector algorithm employ one negative selection process to conduct the self-tolerance of antigen training set in the whole real value $[0, 1]^n$ space, which bring about the high false classification rate and low classifier generation efficiency.

In order to solve these problems of traditional RNSAs, the PRR-2NSA algorithm was proposed in this paper. The PRR-2NSA adopts the dual negative selection to avoid unnecessary and time-consuming antigen self-tolerance process, and introduce APC classifiers' co-stimulation for T-Cell classifier to reduce false classification rate. The PRR-2NSA avoid the unnecessarily and time-consuming self-tolerance of candidate classifiers which repetitive cover with existing classifier set, thus greatly reduces classifier set size, significantly improves classifier set generation efficiency, reduces the time cost and false positive rate of the algorithm. Theoretical analysis and simulations show that the PRR-2NSA has better classifier set generation efficiency and lower false classification rate in comparison with V-Detector and 2NSA.

REFERENCES

[1] F. Pernkopf, "Bayesian network classifiers versus selective k-nn classifier," Pattern Recogn., vol. 38, no. 1, pp. 1–10, Jan. 2005.

[2] F. Sebastiani, "Machine learning in automated text categorization," ACM Comput. Surv., vol. 34, no. 1, pp. 1–47, Mar. 2002.

[3] S. Ramathilagam and Y.-M. Huang, "Extended gaussian kernel version of fuzzy c-means in the problem of data analyzing," Expert Syst. Appl., vol. 38, no. 4, pp. 3793–3805, Apr. 2011.

[4] L. F. Mendonça, S. M. Vieira, and J. M. C. Sousa, "Decision tree search methods in fuzzy modeling and classification," Int. J. Approx. Reasoning, vol. 44, no. 2, pp. 106–123, Feb. 2007.

[5] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," J. Mach. Learn. Res., vol. 2, pp. 265–292, Mar. 2002.

[6] D. Dasgupta, S. Yu, and F. Nino, "Recent advances in artificial immune systems–models and applications," Applied Soft Computing, vol. 11, pp. 1574–1587, 2011.

[7] J. Zhou, "Negative selection algorithms: from the thymus to v-detector," Ph.D. dissertation, The University of Memphis, 2006.

[8] J. Zhou and D. Dasgupta, "V-detector: An efficient negative selection algorithm with "probably adequate" detector coverage," Information Science, vol. 19, no. 9, pp. 1390–1406, 2009.

[9] L. De Castro, "An evolutionary immune network for data clustering," in SBRN, 2000, pp. 84–89.

[10] J. Timmis and M. Neal, "A resource limited artificial immune system for data analysis," Knowledge Based Systems, vol. 14, no. 3/4, pp. 121–130, 2001.

[11] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri, "Self-nonself discrimination in a computer," in Proceedings of the 1994 IEEE Symposium on Security and Privacy, ser. SP '94. Washington, DC, USA: IEEE Computer Society, 1994, pp. 202–212.

[12] J. E. Hunt and D. E. Cooke, "Learning using an artificial immune system," J. Netw. Comput. Appl., vol. 19, no. 2, pp. 189–212, Apr. 1996.

[13] F. A. González and D. Dasgupta, "Anomaly detection using real-valued negative selection," Genetic Programming and Evolvable Machines, vol. 4, no. 4, pp. 383–403, 2003.

[14] M. Bereta and T. Burczyński, "Immune k-means and negative selection algorithms for data analysis," Inf. Sci., vol. 179, no. 10, pp. 1407–1425, Apr. 2009.

[15] I. Aydin, M. Karakose, and E. Akin, "Chaotic-based hybrid negative selection algorithm and its applications in fault and anomaly detection," Expert Syst. Appl., vol. 37, no. 7, pp. 5285–5294, July 2010.

[16] M. Gong, J. Zhang, J. Ma, and L. Jiao, "An efficient negative selection algorithm with further training for anomaly detection," Know.-Based Syst., vol. 30, pp. 185–191, 2012.

[17] T. Stibor, P. Mohr, J. Timmis, and C. Eckert, "Is negative selection appropriate for anomaly detection?" in Proceedings of the 2005 conference on Genetic and evolutionary computation, ser. GECCO '05. New York, NY, USA: ACM, 2005, pp. 321–328.

[18] M. Skala, "Measuring the difficulty of distance-based indexing," in Proceedings of the 12th international conference on String Processing and Information Retrieval, ser. SPIRE'05. Berlin, Heidelberg: Springer-Verlag, 2005, pp. 103–114.

[19] X. Gao, S. Ovaska, and X. Wang, "Genetic algorithms-based detector generation in negative selection algorithm," in 2006 IEEE Mountain Workshop on Adaptive and Learning System, 2006, pp. 133–137.

[20] P. Bretscher and M. Cohn, "A theory of self-nonself discrimination," Science, vol. 169, no. 3950, pp. 1042–1049, 1970.

[21] C. A. Janeway, "Approaching the asymptote? evolution and revolution in immunology," in Cold Spring Harbor symposia on quantitative biology, vol. 54, no. 1, march 1989, pp. 1–13.

[22] R. Medzhitov and C. Janeway Jr, "Decoding the patterns of self and nonself by the innate immune system." Science, vol. 296, no. 5566, pp. 298–300, 2002.

[23] "Uci iris dataset," http://archive.ics.uci.edu/ml/machine-learning-databases/iris/.

[24] "Uci breast cancer wisconsin dataset," http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/.

[25] "Uci chess (king-rook vs. king) dataset," http://archive.ics.uci.edu/ml/machine-learning-databases/chess/king-rook-vs-king/.

Xufei Zheng received his Ph.D degree in the College of Computer Science in Sichuan University, China, in 2012. From 2004 to now, he is a lecturer in the School of Computer and Information Science in SouthWest University, China. His current research interests include computer network security, artificial immune system and software testing theory.

Yanhui Zhou received Bachelor of Science and master degree in software engineering from Southwest University, China. He is currently an associate professor and a PhD student in the Faculty of Computer and Information Science at Southwest University. His research interests include Information Security and testing, artificial intelligence in software engineering, and automation technologies in software testing.

Yonghui Fang received her MS degree in Computer Science from SouthWest China Normal University in 2004. Currently, she is a Ph.D student in the College of Electronical Engineering in Chongqing University, China. From 2002 to now, she is a lecturer in the School of Electronic and Information Engineering at SouthWest University, China. Her current research interests include Intelligent Information Processing and data mining.