Applied Soft Computing 36 (2015) 618–632

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

# Negative selection algorithm with constant detectors for anomaly detection

Dong Li [a,b], Shulin Liu [a,*], Hongli Zhang [a]

[a] School of Mechatronics Engineering and Automation, Shanghai University, Shanghai 200072, China
[b] School of Petroleum Engineering, Changzhou University, Changzhou 213164, China

ARTICLE INFO

*Article history:*
Received 20 October 2014
Received in revised form 6 July 2015
Accepted 3 August 2015
Available online 10 August 2015

*Keywords:*
Artificial immune system
Negative selection algorithm
Anomaly detection
Bearing

ABSTRACT

In the paper, two novel negative selection algorithms (NSAs) were proposed: FB-NSA and FFB-NSA. FB-NSA has two types of detectors: constant-sized detector (CFB-NSA) and variable-sized detector (VFB-NSA). The detectors of traditional NSA are generated randomly. Even for the same training samples, the position, size, and quantity of the detectors generated in each time are different. In order to eliminate the effect of training times on detectors, in the proposed approaches, detectors are generated in non-random ways. To determine the performances of the approaches, the experiments on 2-dimensional synthetic datasets, Iris dataset and ball bearing fault data were performed. Results show that FB-NSA and FFB-NSA outperforms the other anomaly detection methods in most cases. Besides, CFB-NSA can detect the abnormal degree of mechanical equipment. To determine the performances of CFB-NSA, the experiments on ball bearing fault data were performed. Results show that the abnormal degree based on the CFB-NSA can be used to diagnose the different fault types with the same fault degree, and the same fault type with the different fault degree.



## 1. Introduction

Anomaly detection problem can be stated as a one-class classification problem, because only normal samples are available at training stage [1]. The task of anomaly detection method is similar to that of the biological immune system because both of them aim to detect the abnormal information [2]. The negative selection algorithm (NSA) was proposed by Forrest et al. in 1994 under the inspiration of the mechanism of T-cell maturation in the thymus [3–5]. Many modified versions of NSA algorithms had provided more efficient solutions for problems of anomaly detection [6–8], fault diagnosis [9], computer security [10,11], and optimization [12].

The initial NSA used binary encoding to represent self and non-self samples [3]. Later, a real-valued NSA (RNSA) was presented, and the hypersphere detectors with constant radius were adopted [13,14]. Soon after, variable-sized detector [15,16], hypercube detector [2,14], hyper-ellipsoid detector [17,18], and multi-shaped detector [19] were proposed. Compared with these detectors, the hypersphere detector is more widely used, because it has a simple mathematic description.

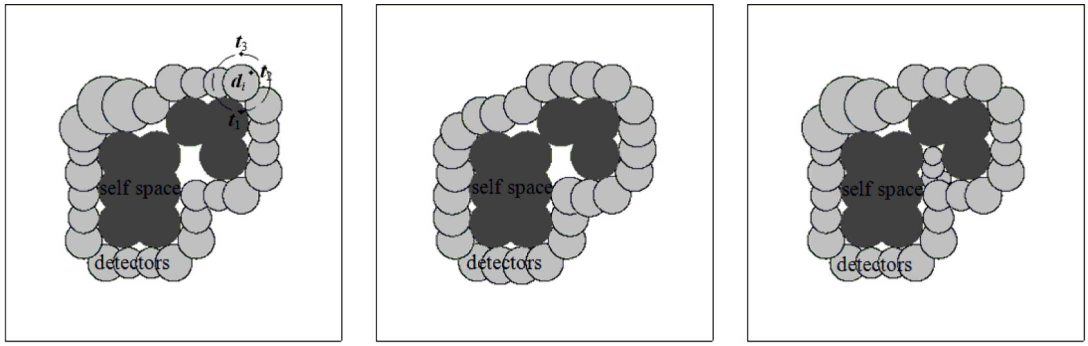
In order to achieve enough detector coverage and reduce the detector number, many methods were proposed [20–25]. Although the methods mentioned above can improve the detection rate and eliminate the holes, the detectors of these methods are generated at random. Even for the same training data, the position, size and quantity of detectors generated in each time are different. Little attention has been paid to the NSA with constant detectors. That is, the position, size and quantity of NSA detectors are constant. These detectors are only related to the training samples, and have nothing to do with training times.

The paper presents two negative selection algorithms with constant detectors. One is named as Boundary-Fixed Negative Selective Algorithm (FB-NSA). The other is an improved FB-NSA algorithm named Fine Boundary-Fixed Negative Selective Algorithm (FFB-NSA). FB-NSA and FFB-NSA generate a layer of detectors, which are around the self space. The FB-NSA detectors and FFB-NSA detectors seem to be the boundary of self space and non-self space. Self samples are on one side of the detectors, and non-self samples are on the other side of detectors or within detectors.

The remaining sections of the paper are structured as follows. The models of FB-NSA, FFB-NSA and CFB-NSA are presented in

* Corresponding author. Tel.: +86 21 56331523.
*E-mail address:* lsl346@shu.edu.cn (S. Liu).

http://dx.doi.org/10.1016/j.asoc.2015.08.011
1568-4946/




(a) VFB-NSA detectors ($m$=15). (b) CFB-NSA detectors ($m$=15). (c) FFB-NSA detectors (15-30).

**Fig. 1.** FB-NSA detectors and FFB-NSA detectors ($r_s$ = 0.07).

detail in Sections 2, 3 and 4, respectively. The experimental results are presented in Section 5. In Section 6, conclusions are provided.

## 2. Boundary-Fixed Negative Selection Algorithm (FB-NSA)

### 2.1. The implementation of FB-NSA

The detectors of traditional RNSA are generated at random. The detectors which are around the self space can be generated easily, but they can not be directly used for anomaly detection. It is easy to calculate the distance between a detector and a testing sample, but it is difficult to confirm the position relationship between a detector and a testing sample. Therefore, these detectors cannot recognize what side of detectors the testing sample is on.

As shown in $[0,1]^2$ in Fig. 1a, $\boldsymbol{t}_1$ belongs to self space and it is a self sample, marked as $\boldsymbol{t}_1 \in \boldsymbol{S}$; $\boldsymbol{t}_2$ and $\boldsymbol{t}_3$ belong to non-self space, and they are non-self samples, marked as $\boldsymbol{t}_2 \in \boldsymbol{N}$, $\boldsymbol{t}_3 \in \boldsymbol{N}$. When the testing algorithm of the traditional NSA is used, $\boldsymbol{t}_1 \in \boldsymbol{S}$ (right), $\boldsymbol{t}_2 \in \boldsymbol{N}$ (right), $\boldsymbol{t}_3 \in \boldsymbol{S}$ (wrong).

To sum up, the FB-NSA detector can be defined as:

**Definition 1.** *FB-NSA detector*,

$$D = \{< \boldsymbol{d}_i, r_i, p_i > | \boldsymbol{d}_i \in \boldsymbol{R}^n, \ r_i \in \boldsymbol{R}\}$$

where $\boldsymbol{d}_i$ is the center of FB-NSA detector; $r_i$ is the radius of $\boldsymbol{d}_i$, $r_i = d - r_s$; $d$ is the distance between $\boldsymbol{d}_i$ and its the nearest training sample; $r_s$ is the radius of training samples; $p_i$ is the position information of $\boldsymbol{d}_i$.

When the radii of FB-NSA detectors are the same, this approach is named Boundary-Fixed Negative Selective Algorithm with Constant-Sized Detectors (CFB-NSA). When the radii of FB-NSA detectors are different, this approach is named Boundary-Fixed Negative Selective Algorithm with Variable-Sized Detectors (VFB-NSA). The FB-NSA and FFB-NSA detectors are described in $[0,1]^2$ shown in Fig. 1. There are 26 FB-NSA detectors and 29 FFB-NSA detectors around the self space. $\boldsymbol{t}_1 \in \boldsymbol{S}$, and it is on one side of the FB-NSA detectors; $\boldsymbol{t}_3 \in \boldsymbol{N}$, and it is on the other side of the FB-NSA detectors; $\boldsymbol{t}_2 \in \boldsymbol{N}$, and it is in the detector $d_i$, as shown in Fig. 1a.

It is difficult to confirm the position relationship between two hyperspheres, but it is easy to confirm the position relationship between two hypercubes. If the non-self space is filled with the same hypercubes, the non-self space can be approximated with Eq. (1):

$$V_{Nonself} = \lim_{V_{hypercube} \to 0} \sum_{i=1}^{\infty} (V_{hypercube})_i \tag{1}$$

Fig. 2 shows the approximation process in 2-dimensional space. There are $20^2$, $40^2$, $80^2$, and $160^2$ squares in Fig. 2a–d, respectively. It is clear that the boundary of squares in the non-self space can be approximated to the boundary of non-self space.

Therefore, the FB-NSA detectors can be generated with the hypercubes which are close to the self space. The center of the hypercube is the center of the detector, and the position relationship between the two hypercubes is the position relationship between two detectors. These hypercubes are defined as boundary hypercubes, and every boundary hypercube generates a FB-NSA detector.

The key to generate FB-NSA detector is to obtain the boundary hypercubes. To obtain boundary hypercubes, the state space $\boldsymbol{T}$ should be evenly divided into $m^n$ hypercubes:

$$\boldsymbol{T} = \bigcup_{i=1}^{m^n} h_i, \tag{2}$$

where $m$ is the number of the segments of each dimension, and $n$ is the number of space dimension.

**Definition 2.** *Empty hypercube* and *non-empty hypercube*, when a hypercube $h_i$ is covered by a self sample or a self sample located in $h_i$, it is non-empty, marked as $h_i = \Theta$. Otherwise it is empty, marked as $h_i = \text{O}$.

It is complicated to determine whether a hypercube is covered by self sample through calculation. To simplify the algorithm, a definition, recognition radius $\delta$, is proposed.

**Definition 3.** *Recognition radius* $\delta$, $\delta$ is a pre-set distance.

$$f(h_i) = \begin{cases} \Theta & d \le \delta \\ \text{O} & d > \delta \end{cases} \tag{3}$$

where $d$ is the distance between $\boldsymbol{c}_i$ and training sample $\boldsymbol{s}_j$, and $\boldsymbol{c}_i$ is the center of hypercube $h_i$.

Recognition radius $\delta$ is an important parameter to determine the hypercubes' property. According to Definition 3, whether a hypercube is empty is determined according to the distance between its center and self sample. It is clear that to obtain all the hypercubes which are covered by self sample $\delta$ should be larger than $r_s$.

The number of FB-NSA detectors increase with the increase of $\delta$, but the detection efficiency decrease with the increase of $\delta$. The hypercube which is less covered or not covered by self sample can be recognized as the boundary hypercube, when $\delta$ is relatively large. Besides, the generated FB-NSA detectors can not cover the holes near self space, when $\delta$ is relatively large.

The effects of $\delta$ on the quantities of FB-NSA detectors in $[0,1]^2$ is shown in Fig. 3. The state space is evenly divided into $15^2$ squares: