Optik 142 (2017) 621–643

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

A novel radius adaptive hybrid detector generation algorithm

Chen Jinyin*, Su Mengmeng, Zheng Haibin

College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023 China

ARTICLE INFO

Article history:
Received 13 March 2017
Accepted 7 June 2017

Keywords:
Radius-adaptive
V-detector
Self-detector
Non-self-detector
Hybrid detect
NSA

ABSTRACT

Artificial immune algorithms have been widely used in anomaly detection. Negative selection algorithm (NSA) is one of the most popular detector generation algorithms. However NSA has problems such as large detector size, high overlapping rate and low detection efficiency etc. In order to reduce its overlap rate and detector size in guarantee of high detection efficiency, a novel radius adaptive hybrid detector generation algorithm is proposed, abbreviated as RAH-NSA. In order to reduce the number of self-detectors, the number of self-samples in different directions are evaluated of various radius to make sure the generated detector could cover each direction as possible. Based on the principle that self-set edge will make the number of self-set less in a certain direction, the radius of self-detector is self-adaptive. In this way, the number of self-detectors and the overlapping rate could be reduced sharply. For each non-self-detector, distance from sample self is calculated as its radius threshold to reduce the number of self-samples and false alarm rate. And non-self-detector centers are automatically generated by normalized endpoints. Shortest distance from the initially detector is used to generate two new negative detectors, whose radius are bigger to reduce the overlapping rate and the number of detector. Finally both self-detector and negative detector are applied as hybrid detectors for data sets detection. When the data sample belongs to self-detector means it's normal, while either detector includes test sample or the test sample belongs to the nearest one. Simulation results testify that proposed RAH-NSA has higher detector accuracy while reducing the negative detector size and overlapping rate compared with other classic detector generation algorithms without obvious execution time increase.

1. Introduction

Artificial Immune System is the simulation of biological immune system which could learn, remember and processing information [1]. Negative selection algorithm (NSA) is one of the most important algorithms of artificial immune system which can identify it and anomalies. However because of fixed radius, the traditional NSA detectors may have problems including too many loopholes, high overlapping ratio, large detector set size and poor detection rate. Recent researches pay much attention to its limitation for further applications [2]. NSA was proved to be a very effective detection method, and several revised version of NSA were come up to overcome its disadvantages. V-detector NSA [4] is brought up with variable radius for detectors to improve detection rate. The detector radius is determined based on the distance between the center of detector samples and self-sample.

* Corresponding author.
E-mail address: chenjinyin@163.com (C. Jinyin).

http://dx.doi.org/10.1016/j.ijleo.2017.06.034
0030-4026/© 2017 Published by Elsevier GmbH.

Further training NSA [5] is put forward based on V-detector. According to the distance between sample point, self-detector and non-self-detector, the excepted sample point of to improve detection rate and reduce false alarm rate. By eliminating the non-self-detector which is covered with another non-self-detector, Dual negative selection algorithm [6] solves the problem that non-self-detectors cover each other, reduces the number of non-self-detectors. By clustering hierarchical of self-set, real-valued negative selection algorithm based on hierarchical clustering of self-set (CB-RNSA) [7] improves the rate of detector generation efficiency in self-sample space. By introducing red-black-tree to create Index, fast negative selection algorithm [8] reduces the comparison between negative algorithms to improve the generation efficiency. By changing the conditions of negative hypothesis testing, improved V-detector [9] reduces the number of void detector. Worm detection with improved V-detector algorithm [10] according to non-self-space distribution to produce a few large coverage detectors. Optimized negative selection algorithm on research of fault diagnosis of network [11] proposed a new optimizing negative selection algorithm (DE-NSA) based on the differential evolution. The algorithm employed the negative selection algorithm and the differential evolution algorithm to generate and optimize the distribution of the detectors. The proposed method also adopted the local outlier factor (LOF)as the fitness function to optimize the distance between the detectors to avoid the overlapping area of them.

Since most negative selection algorithms have to carry out simulations for many times to find out the best detector radius, the radius searching is quite time consuming. Moreover, V-detector [4] reduces the number of negative detector to a certain extent. Further training NSA [5] reduces the loophole influence on test. However, they are both need a large number of detectors and have a high overlap rate. Aiming at these problems, a novel radius adaptive hybrid detector generation algorithm is proposed in this dissertation, abbreviated as RAH-NSA. The detector radiuses are self-adaptive according to the distribution of self-radius. A few of negative detectors are randomly generated from far and near from self-sets. And all self-detectors and abnormal detectors are grouped into hybrid detectors for testing. Abundant simulations are carried out to testify the high performances of RAH-NSA. The results prove that RAH-NSA improves detection rate and reduces false alarm rate and misjudgment rate, with fewer detectors compared with other classic detector generation algorithms.

2. Related algorithm

V-detector algorithm and FTNSA algorithm are two classic detector generation algorithm based on NSA. V-detector is a variable radius detector with its radius self-adaptive based on the distance between the detector and nearest self-samples. It improved detector overlap rate caused by high fixed radius, and solve the problem of loophole. FTNSA uses the way of generating self-detectors after generating negative detectors to reduce the influence of loophole for detection rate and false alarm rate.

In order to solve the problems of fixed radius NSA that there are too many detectors. The literature [4] proposed V-detector algorithm, using the distance between the detector and the nearest self-samples as negative detector radius. In this way, we can generate a smaller radius at the edges of the detector, to achieve the purpose of increase the detection rate. Farther more, it can generate a larger radius negative detector to make the negative detector and self-set suit better.

Definition 1. All the feature of the samples from the feature space strings antigen set $U = \left\{ g | g = (f_1, f_2, f_3, \ldots, f_n), f_i \in [0, 1] \right\}$ where n is the dimension of data, f_i represents the i_{th} teature value of the sample after normalization.

Definition 2. $Self \subseteq U$ represents the feature of the normal sample, r_s represents self-radius, $Non-self \subseteq U$ represents the feature of the abnormal sample, $Self \cup Nonself = U$, $Self \cap Nonself = 0$.

Definition 3. Detector $d(c, r)$ is used to identify the self-antigen, and $c \in Nonself$ representative of the position of detector's vector, r_d is detector's radius. The antigen whose distance with $d(c, r)$ is closer than r_d are recognized as non-self-elements.

Theorem 1. *If the detector d's distance with any self-set satisfies: . The detector i n recognition of its range avoid autoimmune.*

Proof. Assuming that $d \in D$, $d = \langle c, r \rangle$,$\forall s \in Self$ the distance $dis(d, s) > r + r_s$, and there exist p in attachment c and s on the edge of the self-set, $dis(d, p) = dis(d, s) - r_s$. According to the assumption we can fund $dis(d, p) > r + r_s - r_s = r$. By Definition 3 we can know that p will be recognized as self-element. Because within the radius of s, any element with d's distance is more than $dis(d, p)$, so that the detector in recognition of its range avoid autoimmune.

Definition 4. Detector abnormal coverage is the ratio of the detector which means the areas can be identified by detector and non-self which means non-self-set. P is calculated by the formula (1).

$$p = \frac{V_{Detector}}{V_{Nonself}} = \frac{\int_{Detector} dx}{\int_{Nonself} dy} \tag{1}$$

As defined in the formula, the actual coverage of the detector can estimate by sampling fixed abnormal samples. For example, we use k abnormal samples, there are m abnormal samples are activated, the remaining are not activated by detector. Therefore the coverage P can calculate by the formula (2).

$$p = \frac{m}{k} \tag{2}$$