A Novel Self Suppression Operator Used in TMA

Jungan Chen, ShaoZhong Zhang, and Yutian Liu

Electronic Information Department
Zhejiang Wanli University,
No.8 South Qian Hu Road,
Ningbo, Zhejiang, 315100, China
friendcen21@hotmail.com, {dlut_z88,lyt808}@163.com

Abstract. In V-detector or TMA-OR, the parameters self radius rs or Omin are required to be set by experience. To solve the problem, a novel self suppression operator based on self radius learning mechanism is proposed. The results of experiment show that the proposed algorithm is more effective than V-detector or TMA-OR when KDD and 2-dimensional synthetic data are as the data set.

Keywords: artificial immune system, negative selection, self suppression operator.

1 Introduction

Nowadays, Artificial Immune System (AIS) has been applied to many areas such as computer security, classification, learning and optimization [1]. Negative Selection Algorithm, Clonal Selection Algorithm, Immune Network Algorithm and Danger Theory Algorithm are the main algorithms in AIS [2][3].

A real-valued negative selection algorithm with variable-sized detectors (V-detector Algorithm) applied in abnormal detection is proposed to generate detectors with variable r. A statistical method (naïve estimate) is used to estimate detect coverage in V-detector algorithm[4]. But as reported in Stiboret later work, the performance of V-detector on the KDD Cup 1999 data is unacceptably poor[5]. So a new statistical approach (hypothesis testing) is used to analyze the detector coverage and achieve better performance [6]. In the statistical approach, p is defined as the proportion of covered nonself points, n is defined as the number of detectors.The assumption, np>5, n(1-p)>5 and n>10, are required to be satisfied.When p is set to 90%, n must be set to at least 50. Sometimes the number of detectors do not have to be more than 50, so it is unreasonable and the performance of the algorithm is less effective because the number of detectors affect the detect performance.

Naïve estimate and hypothesis testing are two methods discussed above. When the number of detctors required is less than 50 , hypothesis testing is not the ideal solution. Actually in naïve estimate method, the candidate detector is added to valid detector set only when it is not detected by any of valid detectors, which means that the distance between candidate detector and any of valid detectors is bigger than the match threshold of the related valid detector. This process can maximize the distance

E. Corchado et al. (Eds.): IDEAL 2014, LNCS 8669, pp. 303–308, 2014.

among valid detectors. But it is difficult to find valid detector with the number of valid detectors increasing. At worst, it is possible that there is no other valid detector generated after one valid detector is generated, which leads that naïve estimate method shows unacceptable result on the KDD data. So the distance among valid detectors chosen in naïve estimate method can affect the number of detectors generated.

To choose the appropriate distance among valid detectors and achieve the optimized number of detectors, a parameter overlap rate (Omin) in T-detector Maturation Algorithm (TMA) is proposed to control the distance among detectors [7].But the optimized Omin is required to be set by experience. To solve this problem, a suppression operator called Negative Selection operator (NS operator) proposed in refrence[8] is used in TMA. and there is no parameter Omin in TMA with NS operator.

Later, a self radius learning mechanism is proposed to achieve the adaptive self radius. By combining NS operator and self radius learning mechanism, an augmented TMA called TMA with Adaptive Capability (TMA-AC) is proposed[9]. But TMA-AC shows less effective than TMA-OR when 2-dimensional synthetic data is as the data set.

In this paper, a novel self suppression operator based on the self radius learning mechanism is proposed to suppress the number of detectors and the TMA algorithm with Self Suppression operator (TMA-SS) is put forward.

2 Algorithm

2.1 Match Range Model

$U=\{0,1\}^n$,n is the number of dimensions. The normal set is defined as selves and abnormal set is defined as nonselves. selves∪nonselves=U. selves∩nonselves =Φ.There are two points $x=x_1x_2\ldots x_n$, $y=y_1y_2\ldots y_n$. The Euclidean distance between x and y is:

$$d(x, y) = \sum_{i=1}^{n} (x_i - y_i)^2 \tag{1}$$

The detector is defined as dct = {<center, selfmin, selfmax > | center ∈ U, selfmin, selfmax∈N}. center is one point in U. selfmax is the maximized distance between dct.center and selves. selfmin is the minimized distance. The detector set is defined as DCTS. Selfmax and selfmin are calculated by setMatchRange(dct, selves), dct.center∈U, i∈[1, |selves|], $self_i$∈selves :

$$\text{setMatchRange} = \begin{cases} \text{selfmin} = \min(\{d(\text{self}_i, \text{dct.center})\}) \\ \text{selfmax} = \max(\{d(\text{self}_i, \text{dct.center})\}) \end{cases} \tag{2}$$

[selfmin,selfmax] is defined as self area. Others is as nonself area. Suppose there is one point $x \in U$ and one detector dct ∈ DCTS. When d(x,dct) $\notin$ [dct.selfmin, dct.selfmin], x is detected as abnormal.

2.2 Self Radius Learning Mechanism

To learning the appropriate self radius, a property minselfList is added to the detector, which has four properties include center, selfmin, selfmax and minselfList. The property minselfList has three element $\{self_0, self_1, self_2\} \in selves$, which have the minimized distance with the center of a given detector.

Self radius r_s can be achieved by the equation 4.$\{m,n\} \in \{0,1,2\}$ and $m \neq n$

$$d_{mn} = d(self_m, self_n) \tag{3}$$

$$r_s = \frac{\sum d_{mn}}{3} \tag{4}$$

The average self radius is defined as avgrs. The equation 5 is used to calculate the value of avgrs.

$$avgrs = \begin{cases} r_s, & avgrs = 0 \\ \dfrac{avgrs + r_s}{2}, & others \end{cases} \tag{5}$$

2.3 Self Suppression Operator

The novel self suppression operator is proposed to eliminate those detectors which are recognized by others. The equation 6 is used to decided whehter a given detector dct_x is valid. dct_x will be removed if it is not valid.

$$\mathrm{IsValid}(dct_x, DCTS) = \begin{cases} \text{false}, & \mathrm{UnValid} > 1 \cap dct_x.self\min > avgrs \\ \text{true}, & \text{others} \end{cases} \tag{6}$$

$$\begin{aligned} &\exists dct_k \in DCTS \\ &\text{if}(\mathrm{NSMatchAnd}(dct_x, dct_k)) \qquad \mathrm{UnValid} = \mathrm{UnValid} + 1 \end{aligned} \tag{7}$$

$$\mathrm{NSMatchAnd} = \begin{cases} \text{true}, & \forall d_{ab} = 0 \\ \text{false}, & \text{others} \end{cases} \tag{8}$$

$$d_{ab} = d(dct_x.\mathrm{minselfList}.self_a, dct_k.\mathrm{minselfList}.self_b), \{a, b\} \in \{0,1,2\} \tag{9}$$

2.4 The Model of Algorithm

The algorithm, called TMA-SS (TMA with Self Suppression operator), is shown in Fig.1. Step 5 is used to decide whether candidate detector is a valid detector according equation 6.Step 10 is used to estimate the detect coverage.

```
1.  Set the desired coverage pc
2.  Generate one candidate detector dct_x randomly
3.  setMatchRange(dct_x,selves)
4.  compute the self radius r_s and average self raddius
avgrs according equation 4,5
5.  if isvalid(dct_x,DCTS) then // equation 6
6.     dct_x is added to detector set DCTS
7.      covered=0
8.  Else
9.      covered ++
10. If covered <1/(1- pc) then goto 2
```

Fig. 1. TMA-SS algorithm model

3 Experiments

For the purpose of comparison, experiments are carried out using KDD and 2-dimensional synthetic data in table 1,which is described in Zhou's paper[10]. In 2-dimensional synthetic data,various shapes over the unit square $[0,1]^2$ are used as the self region. In every shape, there are training data (self data) of 1000 points and test data of 1000 points including both self points and nonself points. As for KDD data, 20 subsets were extracted from the enormous KDD data using a process described in [5]. Self radius and Omin used in TMA-OR are given in table.1. All the results shown in these figures are average of 100 or 20 (see table 1) repeated experiment with coverage rate 99%.

Table 1. Data set and parameters used in experiments

<table>
<tr><td colspan="2" rowspan="2">Data set</td><td colspan="3">Parameters</td></tr>
<tr><td>r_s</td><td>Omin</td><td>Repeated times</td></tr>
<tr><td rowspan="7">2-dimensional synthetic data</td><td>Comb</td><td rowspan="7">0.03</td><td rowspan="8">0, 0.7</td><td rowspan="7">100</td></tr>
<tr><td>Cross</td></tr>
<tr><td>Intersection</td></tr>
<tr><td>Pentagram</td></tr>
<tr><td>Ring</td></tr>
<tr><td>Stripe</td></tr>
<tr><td>Triangle</td></tr>
<tr><td colspan="2">KDD data</td><td>0.05</td><td>20</td></tr>
</table>

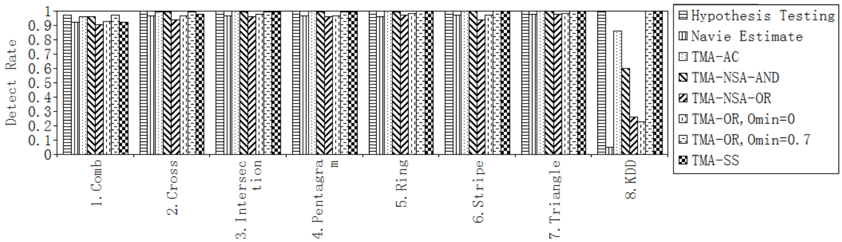

Fig. 2. Detect Rate

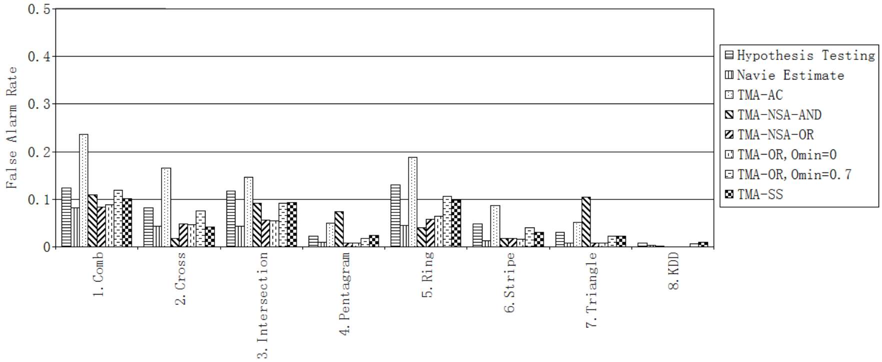

Fig. 3. False Alarm Rate

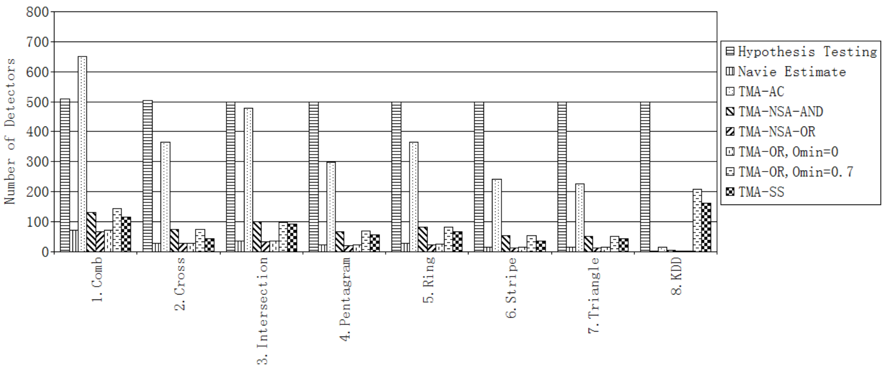

Fig. 4. Number of Detectors

As reference[7], TMA-OR(Omin=0.7) can achieve the best effect. By comparing with TMA-OR(Omin=0.7), TMA-SS get almost the same detect rate (Fig.2) with lower false alarm rate(Fig.3) and smaller number of detectors(Fig.4). So the result of experiment show that TMA-SS is more effective than other algorithm.

4 Conclusion

As the parameters Omin and self radius r_s in TMA-OR are required to be set by experience. To solve the adaptive problem, a novel self suppression operator based on the self radius learning mechanism is proposed, and then an augmented TMA called TMA-SS is proposed. The results of experiment show that the proposed algorithm is more effective than other algorithms.

Acknowledgement. This work is supported by National Natural Science Foundation of China 71071145, Zhejiang Public Welfare Technology Research Project Foundation 2013C31122,Zhejiang Provincial Nature Science Foundation Y1110200, Ningbo Nature Science Foundatio 2012A610010, 2013A610070 , The Science and Technology Innovation Team of Ningbo (Grant No.2013 B82009,2012B82006).Thanks for the assistance received by using KDD Cup 1999 data set [http://kdd.ics.uci.edu/databases / kddcup99/ kddcup99.html] , the 2-dimensional synthetic data set [https:// umdrive. memphis.edu/ zhouji/ www/ vdetector.html].

References

1. Hart, E., Timmis, J.: Application areas of AIS: The past, the present and the future. Journal of Applied Soft Computing 8(1), 191–201 (2008)
2. Timmis, J., et al.: An interdisciplinary perspective on artificial immune systems. Evolutionary Intelligence 1(1), 5–26 (2008)
3. Greensmith, J., Aickelin, U., et al.: Information Fusion for Anomaly Detection with the Dendritic Cell Algorithm. Information Fusion 11(1), 21–34 (2010)
4. Ji, Z., Dasgupta, D.: Real-Valued Negative Selection Algorithm with Variable-Sized Detectors. In: Deb, K., Tari, Z. (eds.) GECCO 2004. LNCS, vol. 3102, pp. 287–298. Springer, Heidelberg (2004)
5. Stibor, T., Timmis, J.I., Eckert, C.: A Comparative Study of Real-Valued Negative Selection to Statistical Anomaly Detection Techniques. In: Jacob, C., Pilat, M.L., Bentley, P.J., Timmis, J.I. (eds.) ICARIS 2005. LNCS, vol. 3627, pp. 262–275. Springer, Heidelberg (2005)
6. Ji, Z., Dasgupta, D.: Estimating the Detector Coverage in a Negative Selection Algorithm. In: Genetic and Evolutionary Computation Conference (2005)
7. Chen, J.: T-detector Maturation Algorithm with Overlap Rate. WSEAS Transactions on Computers 7(8), 1300–1308 (2008)
8. Chen, J.: A novel suppression operator used in optaiNet. BSBT 57, 17–23 (2009)
9. Chen, J., Zhang, Q., Fang, Z.: Improve the Adaptive Capability of TMA-OR. In: Omatu, S., Paz Santana, J.F., González, S.R., Molina, J.M., Bernardos, A.M., Rodríguez, J.M.C. (eds.) Distributed Computing and Artificial Intelligence. AISC, vol. 151, pp. 665–671. Springer, Heidelberg (2012)
10. Ji, Z.: Negative Selection Algorithms: from the Thymus to V-detector. PhD Dissertation, University of Memphis (2006)