A Suppression Operator Used in TMA

Jungan Chen, Qiaowen Zhang, and Zhaoxi Fang

Electronic Information Department, Zhejiang Wanli University, No.8 South Qian Hu Road
Ningbo, Zhejiang, 315100, China
friendcen21@hotmail.com, cn_hnzqw@yahoo.com.cn,
zhaoxifang@gmail.com

Abstract. In T-detector Maturation Algorithm with Overlap Rate, the parameter Omin is proposed to control the distance among detectors. But Omin is required to be set by experience. To solve the problem, T-detector Maturation Algorithm with NS operator is proposed. The results of experiment show that the proposed algorithm can achieve the same effect with TMA-OR when 2-dimensional synthetic data and iris data are used as the data set.

Keywords: Artificial immune system, suppression operator, TMA.

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) is proposed to generate detectors with variable r, which are applied in abnormal detection. A statistical method (naïve estimate) is used to estimate detect coverage [4]. But as reported in paper[5], the performance of V-detector algorithm on the KDD Cup(1999) data is unacceptably poor. So a new statistical approach (hypothesis testing) is used to analyze the detector coverage [6]. But hypothesis testing requires np>5, n(1-p)>5 and n>10. When p is set to 90%, n must be set to at least 50. While the number of detectors have an important effect on the detect performance, the hypothesis has poor performance. Actually in naïve estimate method, V-detector algorithm tries to maximize the distance among valid detectors. With the number of valid detectors increasing, it is difficult to find valid detector. To choose the appropriate distance among valid detectors and achieve less number of detectors generated, 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) is used in TMA. NS operator is first proposed to eliminate those network cells which are recognized by others in optaiNet [8]. So there is no parameter Omin in TMA with NS operator(TMA-NS).

Y. Wang and T. Li (Eds.): Knowledge Engineering and Management, AISC 123, pp. 47–52.
springerlink.com

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 the abnormal set is defined as nonselves. selves ∪ nonselves=U. selves∩nonselves=Φ.There is two point $x=x_1x_2\ldots x_n$, $y=y_1y_2\ldots y_n$. 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 definined as DCTS. Selfmax and selfmin is calculated by setMatchRange(dct, selves), dct.center ∈ U, i ∈ [1, |selves|], $self_i \in$ selfves:

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

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

2.2 NS Operator

NS operator is first proposed to eliminate those network cells which are recognized by others in optaiNet [8]. In this work, it is defined as following:

$$IsValidAnd = \begin{cases} false, NSMatchAnd(dctx, dct_k) = true, \exists dct_k \in DCTS \\ true, \quad others \end{cases} \tag{3}$$

$$NSMatchAnd = \begin{cases} true, d < dctx.self\min \wedge d < dctk.self\min \\ false, \quad others \end{cases} \tag{4}$$

$$d = d(dctx.center, dctk.center) \tag{5}$$

$$IsValidOR = \begin{cases} false, NSMatchOR(dctx, dct_k) = true, \exists dct_k \in DCTS \\ true, \quad others \end{cases} \tag{6}$$

$$NSMatchOR = \begin{cases} true, d < dctx.self\min \vee d < dctk.self\min \\ false, \quad others \end{cases} \tag{7}$$

As there are two logic operators including AND and OR, two type NS operators in equation 3 and 6 are provided.

2.3 The Model of Algorithm

The algorithm, called TMA-NS (TMA with Negative Select operator), is shown in Fig.1. Step 2~4 is used to generate candidate detector which does not covered by self with rs. Step 10 is used to estimate the detect coverage. Step 5 is used to decide whether candidate detector is a valid detector according equation 5 or equation 8. As AND or OR operator is used in step 5, there are two algorithms called TMA-NS-AND or TMA-NS-OR.

```
1.  Set the desired coverage pc, Self radius rs
2.  Generate one candidate detector dctx randomly
3.  setMatchRange(dctx,selves) // equation 2
4.  if dctx.selfmin< rs^2 then Go to 2;
5.  if isvalidAnd(dctx,DCTS) then// equation 3
//  isvalidOR(dctx,DCTS)// equation 6
6.       dctx 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-NS algorithm model

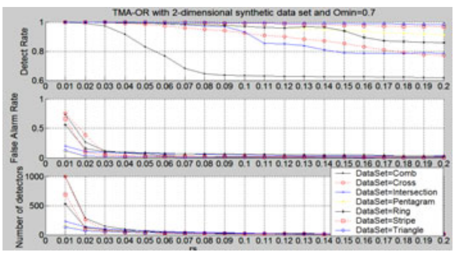
3 Experiments

For the purpose of comparison, experiments are carried out using every data set list in table 1.In table 1, 2-dimensional synthetic data is described in Zhou's paper[9]. Over the unit square $[0,1]^2$,various shapes 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. In the famous benchmark Fisher's Iris Data, one of the three types of iris is considered as normal data, while the other two are considered abnormal [4]. As for KDD data, 20 subsets were extracted from the enormous KDD data using a process described in [5]. Self radius from 0.01 up to 0.2 and Omin used in TMA-OR from 0 up to 0.7 is conducted in these experiments. All the results shown in these figures are average of 100 or 20 (see table 1) repeated experiment with coverage rate 99%.

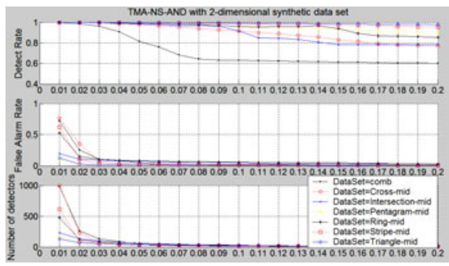
Table 1. Data set and parameters used in experiments

Data set		Parameters		
		r_s	*Omin*	Repeated times
2-dimensional synthetic data	Comb	0.01 ~ 0.2	0 ~ 0.7	100
	Cross			
	Ring			
	Triangle			
	Stripe			
	Intersection			
	Pentagram			
Iris data	Setosa as self data			
	Versicolor as self data			
	Virginica as self data			
KDD data		0.05~0.2		20

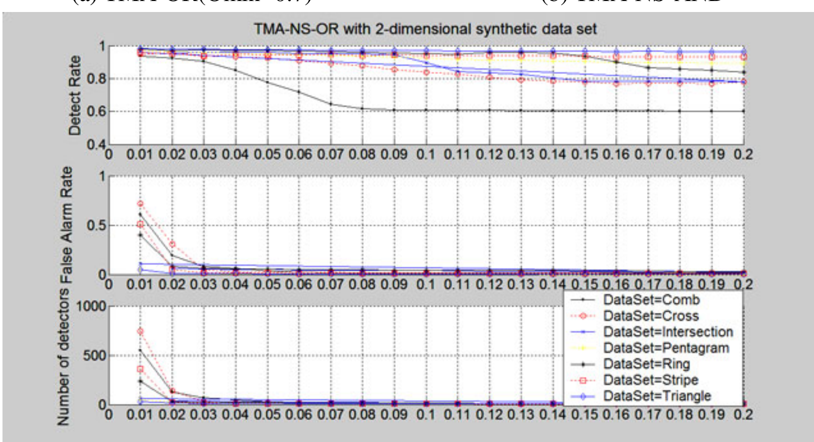
3.1 The Optimized rs Value



(a) TMA-OR(Omin=0.7)



(b) TMA-NS-AND

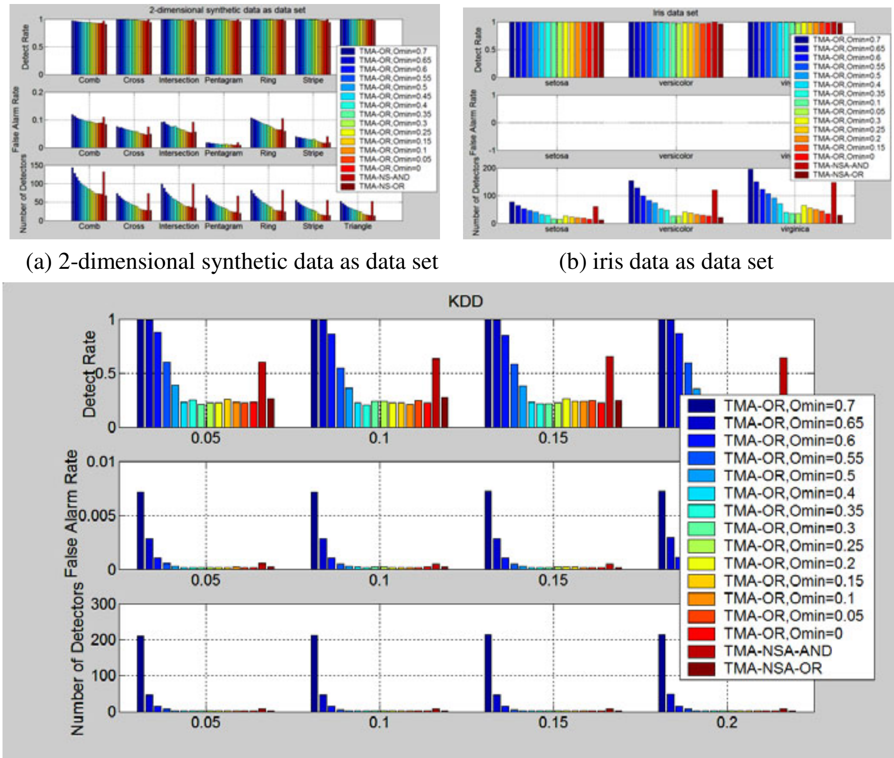


(c) TMA-NS-OR

Fig. 2. Results with 2-dimensional synthetic data and different rs

In Fig.2, it shows that these algorithms achieve the optimized value at rs=0.03. So 0.03 is taken the optimized value in following discussion.

3.2 Comparison



(a) 2-dimensional synthetic data as data set (b) iris data as data set

(c) KDD data as data set

Fig. 3. Results with rs=0.03

In fig.3(a)(b),the results of TMA-NSA-AND and TMA-OR(Omin=0.7) have almost the same effect So does between TMA-NSA-OR and TMA-OR(Omin=0). According the equation 4 and 7, TMA-NSA-AND requires AND operator and TMA-NSA-OR requires OR operator. So TMA-NSA-OR leads more detectors to be removed because OR operator is easier to become true than AND operator. As a result, TMA-NSA-OR generates less valid detectors in the third figure.

In Fig.3(c), when KDD data set is used, TMA-NSA-AND shows less effective than TMA-OR(Omin>=0.6) because TMA-NSA-AND generates less valid detectors in the third figure.

4 Conclusion

As the parameter Omin in TMA-OR is required to be set by experience. To solve the problem, T-detector Maturation Algorithm with NS operator (TMA-NS) is proposed. TMA-NS, where there is no Omin required to be set , can achieve the same effect in

2-dimensional synthetic data and iris data. But TMA-NS shows less effective than TMA-OR when KDD is as the data set.So further research is required to be done.

Acknowledgments. This work is supported by Zhejiang Provincial Nature Science Foundation Y1110200, Ningbo Nature Science Foundation 2010A610173, Ministry of Science and Technology project 2009GJC20045, Scientific Research Fund of Zhejiang Provincial Education Department Y201018538 . 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.: 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., et al. (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. Ji, Z.: Negative Selection Algorithms: from the Thymus to V-detector. PhD Dissertation. University of Memphis (2006)