International Journal of Security and Its Applications
Vol. 10, No. 7 (2016) pp.41-50
http://dx.doi.org/10.14257/ijsia.2016.10.7.04

Robust Anomaly Detection Using Supervised Relevance Neural Gas with Discriminant Analysis

Jia Weifeng* and Chen Weijun

School of Software Engineering, Anyang Normal University, Anyang, 455000, China;
E-mail: jiaweifeng@163.com

Abstract

Neural Network (NN) models employed in lots of researches about system call intrusion detection, however, most probably suffer from problems, such as sensitivity to random initialization, local optimum, outlier of training data, etc. A robust supervised relevance neural gas with discriminant analysis, called RSRNG-DA, is firstly proposed. By incorporating several robust strategies, e.g. outlier resistant scheme and building discriminant analysis on dissimilarity spaces of prototypes, into supervised relevance neural gas (SRNG) framework, RSRNG-DA possesses better robust properties. RSRNG-DA can tolerate the influence cased by outlier and random ordering of training data, and present a significant improvement on classification efficiency. Moreover, the relevant degree of each system call as feature that contributes most to classification performance can be determined, such that the relevance of all features can help to prune irrelevant pattern dimensions. Our technique is evaluated on the system call database maintained by NMU. Experimental results are compared with other existing methods in the literature, and have shown the superior performance on detection rate, false positive and computation time aspects.

Keywords: *neural network, anomaly detection, discriminant analysis*

1. Introduction

With the widespread use of network computers and the rapid increase in connectivity and accessibility of computer systems, computer security has been attracting more and more attention , since intrusion detection have become a significant threat in recent years. The techniques for intrusion detection are generally classified into two categories, which are anomaly detection and misuse detection respectively. Unlike misuse detection, which attempts to model attacks as specific signatures, anomaly detection identifies activities that deviate from the normal behavior of the monitored system and thus has the potential to detect novel attack [1]. Thus, anomaly detection draws more preference than misuse detection from research community with the reason of its' intelligent characteristic.

Over the past decade, many anomaly detection techniques by using neural networks (NNs) [2][3] have been reported. The goal of these approaches is to be able to recognize known attacks and detect novel attacks in the future. Regardless of the approaches used, some unavoidable drawbacks from neural networks techniques are obvious. First, NN is sensitive to random initialization of weighing vector, and the random ordering of training data fed to NN model also degrades the result stability. Second, local optimum but not overall optimum has a great deal of opportunity to be achieved. Third, Euclidian distance measurement is the most widespread metric in NN. However, it is not always suitable for all cases, especially for heterogeneous data and high dimensional data. Supervised Relevance Neural Gas (SRNG) [4] is one of most recently proposed NN model that can

[*] Corresponding Author

solve the second and third disadvantages mentioned above to some extent, yet still suffers from the outlier and random ordering of training data. Despite of these drawbacks, SRNG still paves a promising road to solve the unanswerable problems encountered by anomaly detection in computer security research field.

In this paper, we present a new approach, which is named by Robust Supervised Relevance Neural Gas with Discriminant Analysis (RSRNG-DA), based on Supervised Relevance Neural Gas (SRNG), and firstly introduce this NN model into system call anomaly detection. Our novel approach, RSRNG-DA, makes three contributions as follows. First, RSRNG-DA exploits an outlier resistance strategy to weaken the negative influence of random ordering of training phase and outlier presentation of training data. Second, RSRNG-DA builds linear discriminant analysis (LDA) [5] on dissimilarity space created by prototypes of RSRNG, which can achieve a great improvement on the classification performance. Third, our experiments show the superiority of RSRNG-DA not only in terms of high intrusion detection accuracy and low false positives but also in terms of its preponderance on running time over other techniques. More importantly, RSRNG-DA can utilize the product of RSRNG, say, feature relevance vector, to inform user which system calls as features contribute most to classification performance and which ones can be considered as irrelevant so that they could be ignored. At last, the experiment for RSRNG-DA on system call database maintained by New Mexico University (NMU) shows us very interesting, yet uncovered finding results.. Since then, various anomaly detection approaches have been implemented for users, program or network behavior. In modeling process behavior, system call data are one of the most common types of data used.In order to detect the deviation of anomalous system call sequences from the normal set of sequences, Forrest *et. al.*, introduced a simple anomaly detection method based on monitoring the system calls invoked by active and privileged processes [6][7]. Anup K. Ghosh et al. exploited Back Propagation Neural Network to predict the novel abnormal process [2][3].

Intrigued by employing the frequencies of system calls used by a process, Liao conducted research work on KNN that was used for anomaly detection by using cosine similarity metric usually adopted by text categorization [8]. Great inspirations of improving performance of this work, such as increasing detection rate while decreasing false positive, toleration of noisy training data, *etc*., had been drawn. As a successive study, Hu established RSVM, considering that KNN has low capacity to resist the negative influence of outlier training data [9]. Rawat proposed another efficient anomaly detection algorithm, applying a new similarity metric, say Binary Cosine Weight, into KNN classifier [10]. This metric not only considered the frequency of all system call in one process, but also the binary based common degree the two processes have. The experiment had shown this new technique owns the same noisy resistant capacity as RSVM, and outperforms those previous works. Sharma continued to expand Liao's work, and incorporated the kernel trick into similarity measurement [11]. The experiment argued that the best performance was proved against previous techniques using the same experiment dataset. However, the importance degree of each system call as feature is ignored by all previous works.

This paper introduces a novel neural network model, RSRNG-DA, to be utilized in system call anomaly detection, and compares the intrusion detection performance of Liao and Sharma's work. More importantly, we deliberately derive an interesting finding by using feature relevance vector produced by RSRNG-DA.

2. Review of Supervised Relevance Neural Gas

A brief mathematical description of Supervised Relevance Neural Gas (SRNG) is presented in the appendix. More elaborate illustration can be found in the literature [4]. Below, we try to explain why SRNG can pave a promising road to solve the problems faced by anomaly detection.

SRNG is a most recent proposed model in the Local Vector Quantization (LVQ) [12] based neural network model series, which is essentially prototype based classifier. As we know, LVQ has been successfully applied to clustering, pattern recognition, *etc*. However, some drawbacks of LVQ and it's variants prohibit achieving optimum results. Firstly, LVQ is sensitive to initialization, thereby gets easily stuck in local optima. Second, Euclidian metric usually adopted by LVQ is not always appropriate for heterogeneous data. Third, the underlying cost function causes instable behavior of LVQ.

Many research works has been aspired by LVQ's drawbacks. Generalized Learning Vector Quantization (GLVQ) [13] was proposed to develop underlying cost function to satisfy the convergence condition, such that reference vector doesn't diverge and improve the recognition ability. Neural Gas (NG) [14] brought the neighborhood cooperation scheme into updating the reference vectors, which is determined by 'neighborhood ranking' list, such that the random initialization of prototypes is much less critical to achieving global optima. Generalized Relevance Learning Vector Quantization (GRLVQ) [15] was generated, considering some dimensions are irrelevant and can be pruned. This method introduced weighting factors to the data dimensions that are adapted automatically when tuning the reference vectors.

SRNG incorporates several advantages of GLVQ, NG and GRLVQ to conquer above-mentioned problems of LVQ, and has been utilized in classification task, time series analysis, and bioinformatics. Naturally, SRNG provides us with the opportunity to solve the problems mentioned in Section 1.

3. Robust Supervised Relevance Neural Gas Algorithm with Discriminant Analysis

In this section, We exploit the advantages of SRNG and establish a stronger model, RSRNG-DA, by incorporating outlier resistant strategy and building linear discriminant classifier on dissimilarity space of prototypes generated by SRNG.

a) Outlier Resistant Strategy

Although SRNG utilizes advantages of GLVQ, NG and GRLVQ, the influence of outlier and random ordering of training data can not be overcome. The emergence of outlier data in different order can significantly degrade the performance of SRNG, since the outlier data possesses greater amplitude than normal ones. Ideally, the outlier should have little impact on the prototypes' updating procedure. Thus, we refer SRNG as RSRNG, when outlier resistant strategy is built into SRNG.

To clearly explain why and how outlier resistant strategy is incorporated into SRNG, we reformulate expression (4) and (5) in appendix as

$$\Delta\mathbf{w}_j = \varepsilon^{+}\mathbf{\Psi}^{+}\lambda \|\mathbf{x}_i - \mathbf{w}_j\|_{\lambda} \frac{(\mathbf{x}_i - \mathbf{w}_j)}{\|\mathbf{x}_i - \mathbf{w}_j\|_{\lambda}} \tag{1}$$

$$\Delta\mathbf{w}_k = -\varepsilon^{-} \sum_{\mathbf{w}_j \in W^{\mathbf{x}_i}} \mathbf{\Psi}^{-}\lambda \|\mathbf{x}_i - \mathbf{w}_k\|_{\lambda} \frac{(\mathbf{x}_i - \mathbf{w}_k)}{\|\mathbf{x}_i - \mathbf{w}_k\|_{\lambda}} \tag{2}$$

According to formula (1) and (2), if an outlier $\mathbf{x}_o$ is presented to the SRNG network, the amplitude $\|\mathbf{x}_o - \mathbf{w}_j\|_{\lambda}$ generated along the direction $\frac{(\mathbf{x}_o - \mathbf{w}_j)}{\|\mathbf{x}_o - \mathbf{w}_j\|_{\lambda}}$ will become large. Hence, the outlier can significantly influence the updating of prototypes. Moreover, the updating strength from outliers will differ according to input orderings and thereby destabilizing the final result, if these outliers are randomly located at input sequence. In order to enhance the robustness of the updating rule (1) and (2), we further modify those rules to new forms:

$$\Delta\mathbf{w}_j = \varepsilon^{+}\boldsymbol{\Psi}^{+}\lambda\sigma_{iter}\frac{(\mathbf{x}_i-\mathbf{w}_j)}{\|\mathbf{x}_i-\mathbf{w}_j\|_{\lambda}} \tag{3}$$

$$\Delta\mathbf{w}_k = -\varepsilon^{-}\sum_{\mathbf{w}_j\in W^{\mathbf{x}_i}}\boldsymbol{\Psi}^{-}\lambda\sigma_{iter}\frac{(\mathbf{x}_i-\mathbf{w}_k)}{\|\mathbf{x}_i-\mathbf{w}_k\|_{\lambda}} \tag{4}$$

In new formulas, we make use of σ_{iter} to control the force amplitude caused by the distance between outlier and reference vector. Thus, the influence of outlier and random ordering problem can be weakened. In addition, σ_{iter} is a dynamical parameter, which changes during the training iteration step. More detailed information about how σ_{iter} dynamically changes can be found in [16] [17].

b) Linear Discriminant Analysis on Dissimilarity Space

Prototypes vectors $W=\{\mathbf{w}_1,...,\mathbf{w}_K\}$ are formed after training phase, where dissimilarity space can be built. Dissimilarity space can be constructed as follows. Given a dissimilarity measure $d=\|\mathbf{x}_i-\mathbf{w}_j\|_{\lambda}$, where $\mathbf{w}_j\in W$, every object $\mathbf{x}_i$ is then described by a vector of dissimilarities computed between $\mathbf{x}_i$ and prototypes from W. The new dissimilarity space is described as:

$$D(\mathbf{x}_i,W)=[d(\mathbf{x}_i,\mathbf{w}_1),d(\mathbf{x}_i,\mathbf{w}_2),...,d(\mathbf{x}_i,\mathbf{w}_K)] \tag{5}$$

Traditionally, the 1-NN rule assigns an unknown object to the class of its nearest neighbor in the prototype set. Since research works on dissimilarity space literature [18][19] show us that constructing a classifier on dissimilarity space will improve the classification performance. Considering that RSRNG has capability of behave very well, it is reasonable for us to believe building another classifier on dissimilarity space can bring out better classification performance, to which we refer this approach as RSRNG-DA.

In RSRNG-DA, we exploit Fisher Linear Discriminant (FLD) analysis as new classifier on dissimilarity space, as in our experiment only two classes, normal or abnormal process, need to be examined. FLD analysis function can be illustrated as:

$$\begin{aligned} f(D(\mathbf{x}_i,W)) = (m_1-m_2)^T S_w^{-1}D(\mathbf{x}_i,W) - \\ \frac{1}{2}(m_1-m_2)^T S_w^{-1}(m_1-m_2) \end{aligned} \tag{6}$$

where m_1 and m_2 are the mean vectors, and S_w is the total within-class scatter matrix.

4. Experiments and Results

In this section, we use system call database of Computer Immune System at NMU to demonstrate the performance of our RSRNG-DA in three aspects, which are detection rate, false positive, time consuming and result stability.

Prior to experiments, we need to organize the experiment benchmark and set several common parameters. As system call datasets provided by NMU contain system call sequence from a variety of programs and computer operation systems, we choose synthetic sendmail, synthetic lpr , and live lpr programs as the data source. Each process trace is the list of system calls issued by a single process from the beginning of its execution to the end. Note that system call traces from those three programs collected by SunOS 4.1.4 are recorded by 182 different system calls. After thorough observation, we find that the system calls behind the 50 most frequency have very small frequency. Thus, it's reasonable to extract the 50 most frequent system calls from 182 items with loss much information contained by all different system calls. We also extract 2398 unique normal process traces and 83 unique abnormal ones from these programs.

We then prepare the dataset for experiments illustrated in Table I. It also should be declared that all the following experiments are conducted on WINDOW XP computer platform with Intel 2.4 GHz T8300 processor and 2G Main Memory. Note that without specific declaration, several common parameters needed for experiments are set as follows: prototypes number $K=24$ (12 prototypes for each class), epochs are equal to 20, $\varepsilon^{+}=0.2$, $\varepsilon^{-}=0.1$, $\delta=0.01$, and neighborhood range $\gamma=K/2$.(All mathematic symbols can be referred in appendix). Note that all those common parameters are delicately chosen for optimum.

Table 1. Experiments Data Preparation

Class Label	Training Data	Testing Data
Normal	200	300
Abnormal	48	35
Total	248	335

a) Comparison between RSRNG and SRNG

Here, we evaluate the efficiency of RSRNG and SRNG and make a comparison between them. Since Akaike's Information Criterion (AIC) [20] has been serving as model selection algorithm, it also could be employed as evaluation criterion, where AIC scores are derived, to determine which one is better between RSRNG and SRNG. Generally, the smaller the AIC score is, the higher possibility the model is to be chosen as the better one. AIC score is computed as following formula, which is a little different from standard one.

$$Score(AIC)=\ln\left(\frac{\alpha}{m}\sum_{i=1}^{m}e1_i^2+\frac{\beta}{l}\sum_{i=1}^{l}e2_i^2\right)+\frac{2(k+1)}{m+l} \tag{7}$$

m and l denote the number of normal and abnormal processes respectively; $e1$ indicates the error deriving from training data and prototypes labeled by normal class, and $e2$ from training data and prototypes labeled by abnormal class; k denotes the number of parameter required by model. α is the coefficient for normal class, and β is for abnormal class. Because the detection rate is placed more emphasis on anomaly detection other than classification accuracy, α is set to be 1, and β to be 2 in this experiment as more punishment when trained model gets prototypes weight vectors far away from training data of abnormal class. We then change the prototypes number of two models simultaneously, and record the corresponding AIC scores as well. It is obvious from Figure 1 that RSRNG is always superior to SRNG when the prototypes' number changes.

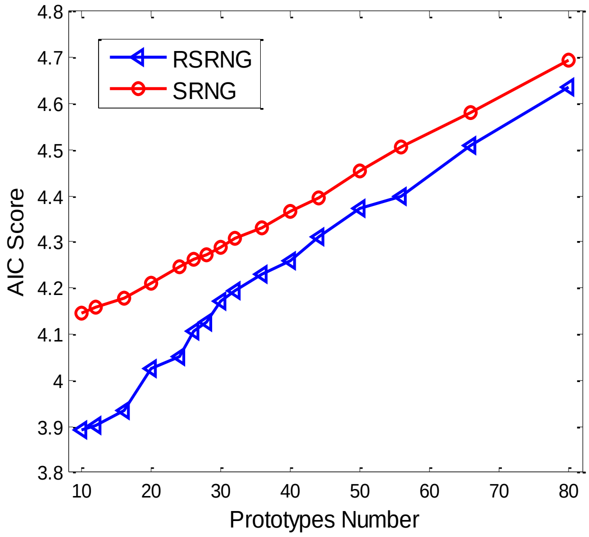

Figure 1. Comparison of AIC Scores between RSRNG and SRNG

Now, we continue to compare the detection rate and false positive performance of RSRNG and SRNG, where the performance of RSRNG-DA is deserved to get more attention. "-1NN" postfix indicates model utilizes 1-NN principle to determine the class of testing data, while "-DA" postfix denotes that the model utilizes linear discriminant analysis on dissimilarity space of prototypes to classify the class of testing data. Table II shows the average performance between two models, both of which are repeatedly conducted for ten times. We can realize from Table 1 that outlier resistant strategy exploited by RSRNG greatly improve the detection rate for both "-1NN" and "-DA" cases, for example the performance increases from 29.1% to 74.8% in "-1NN" case, and from 90.1% to 95.6% in "-DA" case. Interestingly, discriminant analysis helps RSRNG-DA keep the standard deviation of ten runs' results in 4%, which is much lower value than the corresponding value, e.g., 8.8%, of SRNG-DA. Moreover, RSRNG-DA remains the false positive to be a reasonable and acceptable low value, say 4%.

Table 2. Average Performance between RSRNG and SRNG

Algorithms	Detection Rate(%)	False Positive(%)
RSRNG-1NN	74.8	0.9
SRNG-1NN	29.1	3.8
RSRNG-DA	95.6 (4)	4 (3.9)
SRNG-DA	90.1 (8.8)	2.9 (2.7)

From above comparisons, we can arrive at a conclusion that Outlier Resistant Strategy significantly reduces the instability of results, overcoming the influence of outlier and the random ordering of training data. Linear Discriminant Analysis on dissimilarity space of prototypes assists RSRNG to greatly improve detection rate, keeping false positive an acceptable low value. All these characteristics of RSRNG-DA make it more accessible to practical utilization of anomaly detection.

b) Comparison between RSRNG-DA and Previous Work

In this section, we compare our proposed RSRNG-DA to previous works. Liao's research work referred as KNN and Alok Sharma's successive work referred as SBWRBF-KNN, where smooth radial basis function is used, are selected as two competitors against RSRNG-DA.

Figure 2 shows the ROC curve of these three algorithms, where the larger area ROC curve surrounds, the better performance one algorithm has. Note that k neighborhood is set to 5 for both KNN and SBWRBF-KNN algorithms. We could derive that RSRNG-DA is the most superior to other twos.

Although RSRNG-DA achieves the highest performance in detection rate and false positive aspects, RSRNG-DA also possesses the character of real-time computing, when time consuming is considered as a criterion of performance evaluation. Table 3 presents the time consuming of these three algorithms, when the best performances of detection rate are reached. We can see that RSRNG-DA has lower computation time than SBWRBF-KNN that considers both binary weighed matrix and kernel trick on similarity measurement. In addition, although DA cost the minimum computation time, it only obtains best 85.7% of detection rate without combination of RSRNG, while the 100% detection rate with the smallest false positive, say 4%, can be achieved by RSRNG-DA.

Table 3 .Comparison between RSRNG-DS and Other Algorithms

Algorithms	Detection Rate(%)	False Positive(%)	Time
KNN	100	20.6	1.89 s
SBWRBF-KNN	100	9.4	3.67 s
RSRNG-DA	100	4	2.76 s
DA	85.7	4.6	0.03 s

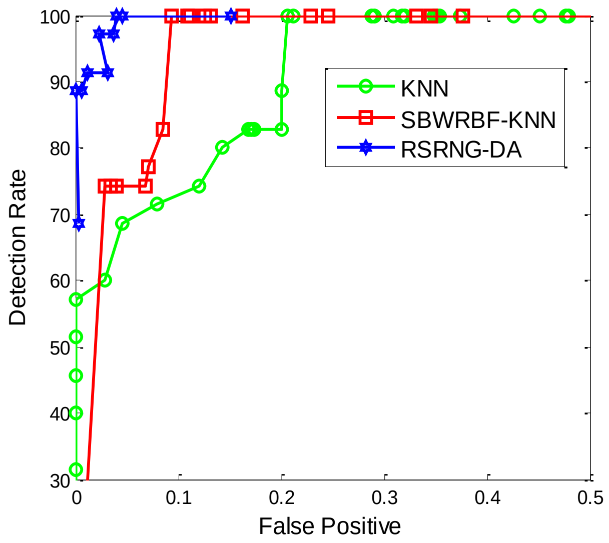

Figure 2. Comparison between KNN, SBWRBF-KNN and RSRNG-DA

5. Interesting Finding on Feature Relevance Vector

As the discussion illustrated above, a feature relevance vector is generated by RSRNG-DA that presents which features are of importance for classification performance and which features are less important. Table IV shows that the frequency rank of system call has no direct relationship with the rank of feature relevance coefficient. After delicate observation, we can realize the most contributing 10 system calls mainly focus in range [15,25] of frequency rank other than the most frequent system call. Since the appearance times of 50th system call in frequency rank is 9, we can generally ignore the influence to classification performance caused by the remanding rank system call.

Table 4. Comparison between Frequent Rank and Relevance Rank

Rele. Rank	System Call Name	Freq. Rank	Rele. Rank	System Call Name	Freq. Rank
1	phys	9	6	sendmsg	16
2	umount	27	7	setpgrp	42
3	ftruncate	23	8	stat	12
4	getdopt	24	9	fchown	25
5	creat	26	10	time	41

Rele. Rank denotes the relevance vector rank; Freq. Rank denotes the frequency rank.Importantly, it is worthy to declare that in Liao and his successive researches, the factor of performance effect caused by each system call did not receive much attention, such that 50 system calls were arbitrarily chosen for anomaly detection. However, our RSRNG-DA algorithm provides much possibility to determine which features are more important, helping us prune the irrelevance system calls to improve performance in time consuming aspect.

6. Conclusion

In this paper, we proposed new approach named RSRNG-DA for system call anomaly detection in computer security. Experiments with system call database maintained by NMU show that RSRNG-DA possesses better detection rate, lower false positive, and stability of result by utilizing outlier resistant strategy and building discrimnant analysis on dissimilarity space of prototypes. In addition, the running time of RSRNG-DA is also reasonable low to make it have more opportunity in real-time application. An interesting finding result is also presented that frequency rank of system calls isn't directly consistent with the rank of feature relevance coefficient. Thus, feature relevance vector can help us to prune the irrelevant system calls to improve performance in time consuming aspect.

References

[1] H. Debar, M. Dacier and A. Wespi, "Towards a taxonomy of intrusion detection systems", Computer Networks, vol. 31, no. 8, **(1999)** Apr., pp. 805-822.

[2] A. K. Ghosh and A. Schwartzbard, "A Study in Using Neural Networks for Anomaly and Misuse Detection", Proceedings of the 8th USENIX Security Symposium, Washington, D.C. US, **(1999)**, pp. 23-36.

[3] V. N. P. Dao and V. R. Vemuri, "A Performance Comparison of Different Back Propagation Neural Networks Methods in Computer Network Intrusion Detection", Differential Equations and Dynamical Systems, vol. 10, no 1&2, **(2002)** Jan&April, pp. 201-214.

[4] B. Hammer, M. Strickert and T. Villmann, "Supervised neural gas with general similarity measure", Neural Processing Letters, vol. 21, no. 1, **(2005)**, pp. 21-44.

[5] R. O. Duda, P. E. Hart and D. G. Stork, "Pattern Classification (2nd Edition)", United States: Wiley-Interscience, **(2000)**, ch.3.

[6] J. P. Anderson, "Computer Security Threat Monitoring and Surveillance", Technical Report, James P. Anderson Co., Fort Washington, Pennsylvania, **(1980)** Apr.

[7] S. Forrest, S. A. Hofmeyr, A. Somayaji and T. A. Longstaff, "A Sense of Self for Unix Process", Proceedings of the 1996 IEEE Symposium on Securiry and Privacy, IEEE Computer Society Press, Los Alamitos, CA, **(1996)**, pp. 120-128.

[8] Y. Liao and V. R. Vemuri, "Use of K-Nearest Neighbor Classifier for Intrusion Detection", Computers & Security, vol. 21, no. 5, **(2002)** Oct., pp. 439-448.

[9] W. Hu, Y. Liao and V.R. Vemuri, "Robust Support Vector Machines for Anomaly Detection in Computer Security", Proceedings of International Conference on Machine Learning and Applications, Los Angeles, CA., **(2003)**.

[10] S. Rawat, V.P. Gulati and A.K. Pujari, "Intrusion Detection Using Text Processing Techniques With a Binary-Weighted Cosine Metric", Journal of Information Assurance and Security, vol. 2, **(2006)**, pp. 43–50.

[11] A. Sharma, A.K. Pujari and K.K. Paliwal, "Intrusion detection using text processing techniques with a kernel based similarity measure", Computers & Security, vol. 26, no. 7, **(2007)** Dec., pp.488-495.

[12] T. Kohonen, Self-Organizing Maps. Springer-Verlag, **(1997)**.

[13] A. Sato and K. Yamada, "Generalized learning vector quantization", in Advances in Neural Information Processing Systems, G. Tesauro, D. Touretzky, and T. Leen, Eds., vol. 7, MIT Press, **(1995)**, pp. 423–429.

[14] T. M. Martinetz, S. G. Berkovich and K. J. Schulten, "Neural-gas network for vector quantization and its application to time-series prediction", IEEE Trans. Neural Networks, vol. 4, **(1993)**, pp. 558–569.

[15] B. Hammer and T. Villmann, "Generalized relevance learning vector quantization", Neural Networks, vol. 15, **(2002)**, pp. 1059–1068.

[16] A. K. Qin and P. N. Suganthan, "Robust growing neural gas algorithm with application in cluster analysis", Neural Networks, vol. 17, no. 8, **(2004)** Oct., pp. 1135-1148.

[17] A. K. Qin and P. N. Suganthan, "A Robust neural gas algorithm for clustering analysis", Proceedings of 2004 International Conference on Intelligent sensing and information processing, Chennai, India, **(2004)**.

[18] E. PeRkalska, R. P. W. Duin and P. Paclik, "Prototype Selection for Dissimilarity-Based Classifiers", Pattern Recognition, vol. 39, no. 2, **(2006)**, pp. 189-208.

[19] E. PeRkalska and R.P.W. Duin, "The Dissimilarity Representation for Pattern Recognition", Foundations and Applications, World Scientific, Singapore, **(2005)**.

[20] H. Akaike, "Information theory as an extension of the maximum likelihood principle", In: B.N. Petrov,

F. Csaki.(Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary, **(1973)**, pp. 267-281.

Author

Weifeng Jia, assisted professor, Research direction: Software engineering