An Improved Grey Self-Organizing Map Based DOS Detection

Rui CHEN
Department of Management Science and Engineering
Tianjin University of Technology
Tianjin,China,300191
chenrui_ff@163.com

Jin-yu WEI
Department of Management Science and Engineering
Tianjin University of Technology
Tianjin,China,300191
weijinyu@tjut.edu.cn

He-feng YU
Department of Computer Science and Technology
Tianjin University of Technology
Tianjin,China,300191
emaif@163.com

Abstract-**In this paper, an improved Grey Self-Organizing Map (GSOM) model is proposed and applied in the detection of Deny of Service (DOS) attack. For detection of the DOS attacks, this improved model can consider the relativity of the data set of DOS attacks. Finally, the experiments on the DOS data set confirm their validities and feasibilities over this improved GSOM model.**

Keywords-**Deny of Service (DOS), SOM, Grey Theory, Gray Correlation Coefficient**

I. Introduction

Among most of the attacks at present, DOS is a popular method. This kind of attack disables the network servers using lots of messages which need response and consumes the bandwidth of the network or the resource of the system. According to the statistics of CERT (Computer Emergency Response Team), the incidence of Distributed Deny of Service increased obviously in recent years.

Most of the research institute proposed the model of intrusion detection, the introduction of neural network raised the detection precision of DOS and using high efficient neural network model becoming the key of DOS attack. The Kohonen's self-organizing map(SOM) is a typical artificial neural networks model and algorithm that implements a nonliner feature projection from the high-dimensional space of signal data into a low-dimensional array of neurons in an orderly fashion[1,2].The mapping tends to preserve the topological relationships of signal domains.SOM has the ability of clustering, self-organizing, self-learning, visualization, classification etc. and is widely applied in the field of Pattern Recognition, Data Mining, Incipient Diagnosis and Intrusion Detection[3-6]. But its weight adjustment is determined only by its learning rate and the difference between the input pattern and the winner neuron's weight. It seems that the SOM obviously ignores some correlation relationships during the learning, which actually exist between the input pattern and the weights of all the nodes that participate in competition. Grey relational coefficient (GRC), which characterize and stress the aforementioned correlation relationships, are explicitly introduced into the learning rule of the traditional SOM by Hu Yi-Chung[7]. But the so-defined GRC still ignore some whole measure relationship as a similarity between the input pattern and weights.

II. An Improved Model of Grey Self-Organizing Map

A. Model of SOM

SOM was proposed by professor Kohonen, the neural network expert in 1981. According to the internal affiliation, the SOM learning rule can do the clustering automatically. The SOM network is consist of two layers. The first layer is Input Layer, which is responsible for the input; the second layer is Output Layer, which is a node matrix with the input nodes lying below.

1) About SOM Algorithm

The algorithm of SOM is recursive. First, every neuron corresponds to a N-dimensions vector $W_i(k)=[w_{i1},w_{i2},\cdots,w_{iN}]^{\mathrm{T}}$. At every stage of training, sampling vector $X(k)=[x_1,x_2,\cdots,x_N]^{\mathrm{T}}$ is selected from the training set randomly, then calculate the distance between X_k and all the weight vectors. c is the BMU(best-matching unit), and the minimum distance between c and X_k is:

$$\| X_k - W_c \| = \min_{i=1}^{M} \| X_k - W_i \| \tag{1}$$

Next, update the weight vector of the neuron which is in neighbourhood zone of the winner cell's topology. The rule is as follow:

$$W_i(k+1)=\begin{cases} W_i(k)+\alpha(k)[X_k - W_i(k)], & i\in N_c(k) \\ W_i(k), & i\notin N_c(k) \end{cases}$$

(2)

In equation(2), N_c refers to neighbourhood zone of the centre neuron wc. In the process of learning, the initialization of $N_c(k)$ can be big, then contracts gradually, as follow:

$$N_c = INT(N_c(0)(1-k/L)),\ \ k=0,1,2,...L \tag{3}$$

In equation(3), $N_c(0)$ means the initial neighbourhood radius, L, the times of the iteration, $INT(.)$, the integral function. $N_c(k_1)$, $N_c(k_2)$, $N_c(k_3)$ stand for the topology neighbourhood zone of the winner cell whose iterative times are k_1, k_2, k_3 ($k_1 < k_2 < k_3$).

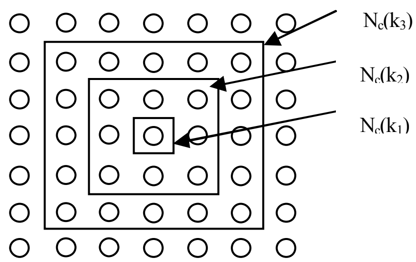

Fig1. Topology adjacent domains on two- dimension network

Usually, the learning rate $\alpha(k)(0<\alpha(k)<1)$ is a constant, which is close to 1.0 in the beginning, then lessens gradually. For example, $\alpha(k)$ can be 0.8(1-k/L). With the increase of the times of the iteration, $\alpha(k)$ tends to zero, which ensures the learning process to refrain from rash action.

2) The steps of the learning algorithm of SOM

The concrete steps of the learning algorithm of SOM are as follows[8-10]:

Step 1. Setting variables and parameters: Let $X(k)=[x_1,x_2,\cdots,x_N]^T$ be the input vector, or training sample, $W_i(k)=[w_{i1},w_{i2},\cdots,w_{iN}]$ be the weight vector, i=1,2, …,M, and the total times of iteration be L.

Step 2. Initialization: Initialize the weight vector W_i with a small random number in a certain interval. Let the neighbourhood radius be $N_c(0)$; the learning rate be $\alpha(k)$; and then normalize weight vector $W_i(0)$ and all the input vector X.

$$X^{'}=\frac{X}{\|X\|} \tag{4}$$

$$W_i^{'}(0)=\frac{W_i(0)}{\|W_i\|} \tag{5}$$

In the above formulas, $\|W_i(0)\|=\sum_{j=1}^{N}[w_{ij}(0)]^2$ and $\|X\|=\sum_{j=1}^{N}(x_i)^2$ are Euclidean norm of the weight vector and input vector.

Step 3. Data sampling. Select training samples $X^{'}$ from the input space.

Step 4. Approximate matching: According to the standard of the minimum Euclidean distance:

$$\|X^{'}-W_c^{'}\|=\min_i\|X^{'}-W_i^{'}\| \quad i=1,2,\ldots,M \tag{6}$$

select winner cell c, implement the competitive process of neurons.

Step 5. Updating. Update the weight vectors of the cordial neuron, who are in the topology neighbourhood zone of the winner cell $N_c(k)$ under the following rules:

$$W_i^{'}(k+1)=W_i^{'}(k)+\alpha(k)h_{i,j}(k)[X-W_i^{'}(k)] \tag{7}$$

$h_{i,j}(k)$ is a neighborhood updating function, called Maxican-Hat Function.

Step 6. Updating the learning rate $\alpha(k)$ and the topology neighborhood zone, and then normalize the weights after learning.

$$\alpha(k)=\alpha(0)(1-\frac{k}{L}) \tag{8}$$

$$N_c(n)=INT\left[N_c(0)\left(1-\frac{k}{L}\right)\right] \tag{9}$$

$$W_i^{'}(k+1)=\frac{W_i(k+1)}{\|W_i(k+1)\|} \tag{10}$$

Step 7. Judging whether the times of the iteration k exceeds L or not, if k<= L then turn to stage 3, otherwise end the process of iteration.

During the training, output neurons are sorted through adjusting their weight vector, in order to be close to the probability density. Though produced randomly at the beginning, the weight vector of the output neurons get closer and closer to the distribution of the input data after long time running, through updating the weight vector continuously.

B. *Grey System Theory*

Grey System Theory is put forward by professor Deng Ju-long in 1982[11]. Through nearly 20-year development, Grey System Theory has build up a new study structure.

1) Gray Correlation Analysis

Gray Correlation Analysis is used to quantitatively compare or depict the relative change between systems or every factors of the system, while the approximate extent of the size, direction and speed of the change is used to measure the relevance between them[12]. In this paper, Gray Correlation Analysis refers to the method of studying the correlation between a given reference pattern and another comparative pattern, both of them must be standardization[7]. Let input pattern X be the reference pattern, weight vector W_i(i=1,2,…M) be the comparative pattern, so as to analyze the grey correlation between X and W_i. And let the grey relational coefficient between the component of the input pattern x_j (j=1,…,N) and weight value w_{ij} be ξ_{ij} [7]:

$$\xi_{ij} = \frac{\Delta_{\min} + \lambda\Delta_{\max}}{\Delta_{ij} + \lambda\Delta_{\max}} \tag{11}$$

In equation(11), $\lambda(0 \le \lambda \le 1)$ is the discrimination coefficient, which usually takes 0.5. Further, $\Delta_{\min}, \Delta_{\max}, \Delta_{ij}$ can be defined as follows:

$$\Delta_{\min} = \min_{i=1,\cdots,M} \min_{j=1,\cdots,N} |x_j - w_{ij}| \tag{12}$$

$$\Delta_{\max} = \max_{i=0,1,\cdots,M} \max_{j=0,1,\cdots,N} |x_j - w_{ij}| \tag{13}$$

$$\Delta_{ij} = |x_j - w_{ij}| \tag{14}$$

From equation(11), we can see $0 < \xi_{ij} \le 1$, and if ξ_{ij} is close to 1, then Δ_{ij} will be close to $\Delta_{\min}$. That is to say, the closer the difference between the component of input pattern x_j and the weight value w_{ij}, the closer the difference between every component of the input pattern X and the weight vector W_i. So, the grey relational coefficient ξ_{ij} reflects the extent of similar between x_j and w_{ij}.

Considering this character of ξ_{ij}, ξ_{ij} can be used in learning rule of the traditional model of SOM, thereby let the learning rule be more reasonable and effective. In addition to the introduction of grey relational coefficient in the traditional model of SOM, an improved grey relational coefficient is proposed in this paper.

2) An improved Gray Correlation Coefficient

Under the analysis of gray correlation above, the grey relational coefficient ξ_{ij} only consider the every component of input pattern x_j and the weight value w_{ij}, but ignoring some important information between the reference pattern X and the comparative pattern W_i while they are seen as a whole. So, some amendments were made as follows:

$$\overline{\xi_{ij}} = \frac{\Delta_{\min} + \lambda\Delta_{\max}}{\Delta_{ij} + \lambda\Delta_{\max} + \eta \frac{1}{N}\left[\sum_{k=1}^{N} |x_k - W_{kj}|^p\right]^q} \quad \begin{pmatrix} 0 \le \lambda \le 1 \\ \eta \ge 0 \\ p \ge 1; 0 \le q \le 1 \end{pmatrix} \tag{15}$$

$\overline{\xi_{ij}}$ is defined as the improved GRC. λ is the discrimination coefficient, usually equals to 0.5; η is a non-negative real number , which is an important measurement of the whole character between X and W_i. If η is bigger, then the whole character is more important; if η=0, that is to say the whole character between X and W_i is ignored, meanwhile, $\overline{\xi_{ij}}$ become ξ_{ij}. So the improved GSOM has better depiction character than GRC. p , q are plus integers, usually $p \cdot q = 1$, which is Minkovski Distance being used to measure the close extent between X and W_i . Obviously, $0 < \overline{\xi_{ij}} \le 1$ and $\overline{\xi_{ij}}$ has the similar character as ξ_{ij} .

Defined $F\left(\overline{\xi_{ij}}\right)$ as the function of $\overline{\xi_{ij}}$, and it can improved the spread ability of GRC. Three kind of gray correlations are used to depict the function, they are index format, polynome format, tangent format, show as follows:

$$F\left(\overline{\xi_{ij}}\right) = \begin{cases} b^{\overline{\xi_{ij}}-1} & (b > 1) \\ \tanh(\beta \bullet \overline{\xi_{ij}} + \theta) & (\beta, \theta > 0) \\ \overline{\xi_{ij}}^{\,k} & (k > 1) \end{cases} \tag{16}$$

C. An Improved Model of GSOM

In the algorithm of SOM, the updating of the weight vector is decided by the formula bellow:

$$\Delta w_{ij} = \alpha(k) |x_i - w_{ij}(k)| \qquad i \in Nc(k) \tag{17}$$

$\alpha(k)$ is the learning rate, $|x_i - w_{ij}(k)|$ is the distance between x_j and $w_{ij}(k)$, and $N_c(k)$ is the neighbourhood zone.

By introducing $F\left(\overline{\xi_{ij}}\right)$ to the traditional model of SOM, another new model of SOM is produced, which is called an improved model of Grey Self-Organizing Map. Then, the updating of the weight vector will be as follows:

$$\Delta w_{ij} = \alpha(k) * F\left(\overline{\xi_{ij}}\right) * (x_j - W_{ij}(k)) \tag{18}$$

If η=0 in equation(15), then $\overline{\xi_{ij}} = \xi_{ij}$. That is to say, if $F\left(\overline{\xi_{ij}}\right) = \overline{\xi_{ij}}^k$, then the improved GSOM becomes GSOM. Because of the introduction of gray correlations, the analysis ability of the improved GSOM is improved.

Like SOM, the improved GSOM also includes two phases: sort phase and convergence phase. Sort phase is the updating phase of weight vector topology, the initial learning rate is 0.9, finally becomes 0.05, and the neighbourhood zone becomes from 8 to 1. When it comes to the convergence phase, the statistic behavior of the input space is needed, so the feature mapping can be updated further. The initial learning rate is 0.05, finally becomes 0.01, and the neighbourhood zone is 0 from beginning to end. The learning time of sort phase should be more than 2000; but when it comes to convergence phase, in order to save the expenses of algorithm, the learning time should be less than 1000.

III. AN IMPROVED GSOM BASED DOS DETECTION

In order to train the improved GSOM and test its ability to identify the DOS attacks, choosing some dataset from the DARPA'98 training data and testing data. Concretely as the TABLE I shows below:

TABLE I
THE DATA OF DIFFERENT DOS ATTACKS CHOOSE FROM DARPA'98

Types of DOS attack / Data	Normal	Neptune	Smurf	Pod	Land	Teardrop
training data	4863	10720	280791	264	21	980
testing data	60593	58001	164091	87	9	12

The total number of training data samples is 297639, and testing data samples is 282793. In order to deal with the neural network, data preprocess and standardization is necessary. But the data sample is too larger, so how to choose the network topology of the improved GSOM becomes the key of the problem.

A. *Choice of large sample SOM topology*

First of all, the number of SOM neurons should be determined. Define the whole training data is dlen, number of SOM neurons is munits. So, $munits = 1.25 * dlen \wedge 0.54321$

This makes enough neurons project on the large primitive data. When the number of neurons is determined, we can use PCA to determine the length and wide proportion of SOM.

Suppose the primitive dimension of vector is n, namely $X = (x_1, x_2, \cdots, x_n)^T$. Another n new character: $y_1, y_2 \cdots y_n$ should be constructed, and let them satisfy the conditions showing below:

1: Every new character is the linear combination of primitive characters, namely

$$y_i = u_i^T X \quad u_i = (u_{i1}, u_{i2}, \cdots, u_{in})^T, i = 1, 2, \cdots n \tag{19}$$

2: Every new character is irrelevant, namely the correlation coefficient is 0:

$$r\left(y_i, y_j\right) = 0 \quad i, j = 1, 2, \cdots n, i \neq j \tag{20}$$

3: u_i make the covariance of y_i maximize, $i = 1, 2, \cdots, n$

The new characters $y_1, y_2 \cdots y_n$ are the principal components of node $1, 2, \cdots, n$. In order to calculate every principal component, we only need to calculate every u_i, which can be proved as the feature vector of the sample covariance matrix S_x.

$$S_x = E\left[(X - \mu)(X - \mu)^T\right] \tag{21}$$

Let $U = (u_1, u_2, \cdots, u_n)^T$, content with orthogonal normalization, $UU^T = I$, then

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = UX \tag{22}$$

$S_x U^T = U^T \Lambda$, in which $\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$ is the characteristic value matrix. λ_i is mapping to the characteristic vector u_i.

After the change, if X is mapping to the covariance matrix S_x, then the mapping covariance matrix of Y is:

$$S_y = E\left[YY^T\right] = US_X U^T = UU^T \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} \quad (23)$$

The covariance between every two new character $y_1, y_2 \cdots y_n$ is 0, namely they are irrelevant. The covariance reflects the incorporative information. Sort the characteristic value, let $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$, when the information of the former m new character is enough, namely $\sum_{i=1}^{m} \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i \ge 98\%$, the other new character can be discard.

So, $\psi = \lambda_1 / \lambda_2$ is defined as the length and wide proportion of SOM. The topology of SOM can be determined through lot of training data and characteristic value.

B. Detection results of SOM and Improved GSOM

1) Imbalance of training data

In order to solve the imbalance of training data and save the training time of the improved GSOM, the same data in the large sample is deleted; thereby the Normal data and the DOS sample data can keep balance.

2) Choice of experiment data and experiment results

In the experiment, the training data is input in the model of the improved GSOM. From the results of experiment 1 and experiment 3, Tangent is chosen to be the function format of grey relational coefficient. The choice of experiment parameters are as follows:

From the analysis of 3.1, the choice of SOM topology can be: 81*6;

Sorting phase: the largest time of learning is 50, the initial learning rate is 0.9, the last learning rate is 0.05, the initial value of neighbourhood zone is 8 and the last is 1;

Convergence phase: the largest time of learning is 40, the initial learning rate is 0.05, the last learning rate is 0.01, the initial value of neighbourhood zone is 0 and the last also is 0;

Tangent parameters are: $\beta = 1, \theta = \pi$;

After training, the testing data set is used to test the improved GSOM, experiment results can be seen in Tab3 and Tab4. Because of the mapping mechanism of SOM, the label of neurons on the class border can not be determined. So, let the label be 0, labeled as 'unkown' in TABLE III.

Something should be especially pointed out: the dimension of some data character is greater, but these dimensions reflect the main data character of this kind of attack. If normalizing these dimensions, the integrity of the primitive character will be broken.

$$M_j = \frac{1}{N} \sum_{i=1}^{N} x_{ij} \qquad i = 1, 2, \cdots N \quad j = 1, 2, \cdots 41 \quad (24)$$

$$C_j = \frac{1}{N-1} \left(\sum_{i=1}^{N} \left(x_{ij} - M_j \right)^2 \right)^{\frac{1}{2}} \qquad i = 1, 2, \cdots, N \quad j = 1, 2, \cdots, 41 \quad (25)$$

$$X_{ij} = \frac{x_{ij} - M_j}{C_j} \qquad i = 1, 2, \cdots, N \quad j = 1, 2, \cdots, 41 \quad (26)$$

M_j is the average value of column j, C_j is the covariance of column j, X_{ij} is the new value after normalization. The experiment results of TABLE III is the result after normalization, show in below.

TABLE II
THE IMPROVED GSOM BASED DETECTION RESULTS OF DOS ATTACK (NORMALIZATION)

	Normal	Neptune	Smurf	Pod	Teardrop	Unknow	Total	Correct（%）
Normal	20856	5	0	0	58	40023	60942	34.22
Neptune	2064	16386	0	0	0	38982	57432	28.53
Smurf	31819	0	130298	0	0	320	162437	80.21
Pod	5	0	0	0	78	0	83	0
Teardrop	0	0	0	0	0	11	11	0

From TABLE II, we can see that the normalization break the primitive character, this make the classification of DOS attacks and Normal data bad. TABLE III and TABLE IV are the classification results of SOM and the improved GSOM without normalization.

TABLE III
SOM BASED DETECTION RESULTS OF DOS ATTACK (NOT NORMALIZATION)

	Normal	Neptune	Smurf	Pod	Teardrop	Unknow	Total	Correct（%）
Normal	58302	76	668	256	852	723	60877	95.77
Neptune	58	55998	0	0	986	156	57198	97.90
Smurf	0	0	163987	0	0	20	164007	99.98
Pod	0	0	0	85	0	0	85	100
Teardrop	4	0	0	0	12	0	16	75

From TABLE III, we can see that SOM based detection rate of Normal is 95.77%, the detection rate of Neptune, Smurf, Pod and teardrop are separately 97.90%, 99.98%, 100% and 75%. The average detection rate is 93.73%. From the analysis above, the detection ability of the improved GSOM is very good.

TABLE IV
THE IMPROVED GSOM BASED DETECTION RESULTS OF DOS ATTACK (NOT NORMALIZATION)

	Normal	Neptune	Smurf	Pod	Teardrop	Unknow	Total	Correct（%）
Normal	60012	101	162	46	352	778	61451	97.65
Neptune	0	58212	0	0	192	0	58404	99.67
Smurf	0	0	163989	0	0	4	163993	99.99
Pod	5	0	3	88	0	0	96	91.66
Teardrop	0	0	0	0	15	0	15	100

From TABLE IV, we can see that the average detection rate is 97.794%.The attack of Pod mistakes 5 to be Normal, 3 to be Smurf. From the analysis above, the improved GSOM has better detection ability than SOM on the whole.

IV. CONCLUSION

This paper introduced the attack principle of DOS, analyzed the data character of DOS attack and proposed the improved GSOM model using grey relational coefficients and its correlative functions. The purpose is to make the input pattern and the weight value of the competitive neurons as a whole, express the inner relationship between them through general grey relational function, and melt the inner relationship into the learning rule of SOM, thus the capability of the algorithm is improved further. In contrast to SOM, the improved GSOM raise the detection precision of DOS attacks.

ACKNOWLEDGMENTS

The authors would like to sincerely acknowledge the valuable suggestions of the referees, which have immensely helped to enhance the quality of the paper.

REFERENCE

[1] T.Kohonen, Self-Organization and Associative Memory, Springer-Verlag Berlin Heidelberg New York,1989.

[2] T.Kohonen, The Self-Organizing Map, Proceedings of the IEEE,1990,78(9):1464-1480.

[3] J.Iivarinen, T.Kohonen, J.Kangas, S.Kaski, visualizing the Clusters on the Self-Organizing Map, in proceedings of Conference on Artificial Intelligence Research in Finland, Helsinki, 1994:122-126.

[4] D.Merkl, A.rauber, Altemative Ways for Cluster Visualization in Self-Organizing Map, in Proc. Workshop on Self-Maps(WSOM97), Helsinki, Finland,1997.

[5] J.Vesanto, E.Alhoniemi, Clustering of the Self Organizing Map, IEEE Transactions on Neural Networks, 2000,11(3):586-600.

[6] D.Alahakoon, S.K.Halgamuge, B.Srinivasan, Dynamic Self-Organizing Maps with Controlled Growth for Knowledge Discovery, IEEE Transaction on Neural Networks, 2000,11(3):601-614.

[7] Hu Yi-Chung, Chen Ruey-Shun, Hsu Yen-Tseng, Grey Self-Organizing Feature Maps, Neurocomputing,2002,48(4):863-877.

[8] Sahin Albayrak, Achim Muller, Christian Scheel, Dragan Milosevic, Combining Self-Organizing Map Algorithms for Robust and Scalable Intrusion Detection. In Proceedings of the 2005 International Conference on Computational Intelligence for Modeling, Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'05), 2005

[9] H. Giinev Kayacik, A. Nur Zincir-Heywood, Malcolm I. Heywood, On the Capability of an SOM based Intrusion Detection System. In IEEE International Joint Conference on Neural Network, July 20th-24th 2003. pp. 1808-1813

[10] Jianhong Gao, Lixin Xu, Yaping Dai, An intrusion detection system model based on self-organizing Map. In Proceedings of the 5th World Congress on intelligent Control and Automation, June 15-19, 2004, Hangzhou, P.R. China. pp. 4367-4369.

[11] J.L.Deng, Control Problems of Grey Systems, Systems Control Lett, 1982,1(5):288-294.

[12] Liu Si-feng, Guo Tian-bang, Dang Yao-guo, Grey System Theory and Application, Science Press,1999.