Intrusion Detection Based on Improved SOM with Optimized GA

ZHAO Jian-Hua[1, 2]
[1]College of computer, Northwestern Polytechnical University, Xi’an 710072, China
[2]Department of Computer Science, ShangLuo University, ShangLuo 726000, China
E-mail: zhaojh2009@yahoo.com.cn

LI Wei-Hua
College of computer, Northwestern Polytechnical University, Xi’an 710072, China

***Abstract*—In order to improve the effectiveness of supervised self-organizing map (SSOM) neural network, a kind of genetic algorithm is designed to optimize it. To improve its classification rate, a real number encoding genetic algorithm is provided and used to optimize the learning rate and neighbor radius of SSOM. To speed up the modeling speed, a binary encoding genetic algorithm is provided to optimize input variables of SSOM and reduce its dimension of input sample. Finally, intrusion detection data set KDD Cup 1999 is used to carry out experiment based on the proposed model. The results show that the optimized model has shorter modeling time and higher intrusion detection rate.**

***Index Terms*—SOM, intrusion detection, classification, dimension reduction, genetic algorithm**

I. INTRODUCTION

Nowadays, network communications become more and more important to the information society [1, 2]. Business, e-commerce, online shopping, Internet bank and other network transactions require more secured networks. As these operations increases, computer crimes and attacks become more frequents and dangerous, compromising the security and the trust of a computer system and causing costly financial losses [3, 4].

While a number of effective techniques exist for the prevention of attacks, it has been approved over and over again that attacks and intrusions will persist and always be there [5, 6]. Although intrusion prevention is still important, another aspect of network security, intrusion detection, is just as important [7, 8]. With trenchant intrusion detection techniques, network systems can make themselves less vulnerable by detecting the attacks and intrusions effectively so the damages can be minimized while keeping normal network activities unaffected [2, 9, 10].

The intrusion detection system (IDS) is used to detect intrusion action. Collecting and analyzing the information of a network or system [11, 12], IDS can find the actions of violating security policy and detect the traces of being attacked from the network or system. According to the network information, it classifies the network behavior normal behavior or abnormal behavior [13, 14].

The neural network has the function of pattern recognition, it may be used in the field of the classification of intrusion detection and get very good results [15, 16]. At the same time, neural network has self-learning and adaptive capacity. As long as the system audit data and the network data packet are provided, neural network can extract normal user or system feature model from it and detect the attack mode from the abnormal activity [25].

The self-organizing map (SOM) neural network constitutes an excellent tool for knowledge discovery in a data base, extraction of relevant information, detection of inherent structures in high-dimensional data and mapping these data into a two-dimensional representation space. It has been applied successfully in multiple areas. Many researcher has apply it in the field of intrusion detection and got the good test result.

However, the network architecture of SOM has to be established in advance and it requires knowledge about the problem domain. Moreover, the hierarchical relations among input data are difficult to represent and it is an unsupervised network and not easy to determine the classification type. Some researcher has improved SOM and they improved unsupervised SOM to supervised SOM (name SSOM) and obtain good results. However, there are still some problem exiting for SOM and SSOM. For example, it is difficult to determine the parameters of SOM and SSOM [17, 18].

In light of the disadvantage of SOM and SSOM, this paper uses genetic algorithm to optimize their parameters (including learning rate and neighborhood radius). New neural network model (GA-SOM and New-GA-SSOM) are proposed and applied in the field of intrusion detection.

The rest of the paper is organized as follows: In Section II we describe the basic definitions and characteristics of SOM neural network, SSOM neural network and genetic algorithm. Section III designs a

This work was sponsored by Scientific Research Program Funded by Shaanxi Provincial Education Department (No.12JK0748) and National ministries foundation in China.

doi:10.4304/jcp.8.6.1456-1463

genetic algorithm based on real number encoding, which is to optimize the learning rate and neighbor radius of SSOM and solve the random initialization problem of learning rate and neighbor radius. Section IV designs a genetic algorithm based on binary encoding to optimize the input variables and reduce the dimension for SSOM. In Section V, intrusion detection experiment is carried out based on KDD Cup 1999 data sets to verify the effectiveness of the provided model.

II. Proposed Scheme

A. SOM Neural Network

Self-organizing feature map network (SOM) is also known as Kohonen network, which is proposed by Holland scholar Teuvo Kohonen in 1981. The network is a no-teachers, self-organization and self-learning network consisting of fully connected neurons array.

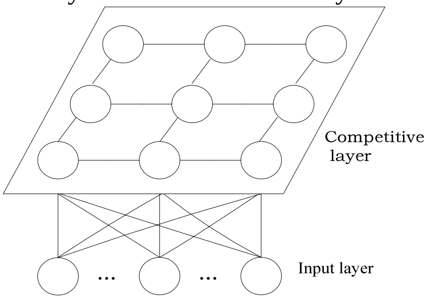

Figure 1. The structure diagram of SOM

SOM is an artificial neural network model and it is proved to be exceptionally successful for data visualization applications mapping from a usually very high-dimensional data space into a two-dimensional representation space. The remarkable benefit of SOM is that the similarity between the input data as measured in the input data space is preserved as faithfully as possible within the representation space. Thus, the similarity of the input data is mirrored to a very large extends in terms of geographical vicinity within the representation space [19, 20].

The structure of SOM neural network is shown in Figure 1, including two layers feed forward neural network structure which is an input layer and a competitive layer. The first layer is the input layer and its dimension is equal with the input vector dimension which is set to *m*. The second layer is a competitive layer and it generally shows a two-dimensional array distribution. A competitive layer node represents a neuron and the number of competitive layer node is set to *n*. The association between input layer and competitive layer is in the form of a full connection; its weight is indicated by ω_{ij} .

The basic working principle of SOM neural network is as follow: during the network train and learning the neurons on competitive layer get the response to the input model by competing with each other, the neuron having the minimum distance from input sample becomes the winning neuron.

Adjust the weights of the winning neuron and its adjacent neurons, so that the weights can reflect the relationship between the input samples. Through the repeated training and learning, the neurons are divided into different regions which have different response characteristics to input model and implement the clustering of input model. And it can realize the classification of the input samples and can be applied in various areas of the classification.

The steps of SOM neural network algorithm are as follow:

(1) Initialization. Initialize the weights and the neighbor radius etc.

(2) Distance calculation. Distance can reflect the similarity degree and closeness degree between samples. We calculate the distance d_j between input vector $x_i = (x_1, x_2, ..., x_n)$ and competitive layer neuron j , which is shown in equation (1).

$$d_j = \sqrt{\sum_{i=1}^{m} (x_i - \omega_{ij})^2} \qquad j = 1, 2...n \tag{1}$$

(3) The winning neuron selection on competitive layer.

Find out neuron *c* with the minimal distance from the winning neuron and calculate the neighborhood N$_c$ (t) of *c* in accordance with equation (2).

$$\mathrm{N}_c(t) = (t \,|\, find(norm(pos_t\, , pos_c) < r) \quad t = 1, 2, .., n \tag{2}$$

where pos_c represents the position of neuron *c* and pos_t represents the position of neuron *t*; *norm* represents the calculation of Euclidean distance between two neurons; *r* represents the neighborhood radius.

(4) Weight adjustment. Adjust the neuron weights of neuron *c* and others in its neighborhood N$_c$ (t) according to equation (3).

$$\omega_{ij} = \omega_{ij} + \eta(x_i - \omega_{ij}) \tag{3}$$

where ω represents the weight between input layer and competitive layer, η represents learning rate, η decreases with the increase of evolution number

(5) Judge whether the algorithm ends. If not end, return to (2).

B. SSOM Neural Network

SOM is an unsupervised neural network and it can effectively classify unlabeled data. However It cannot determine the classification types of labeled data more effectively in the help of data labels.. To facilitate the processing of classification problem and quickly get the classification type, some researchers improve the unsupervised SOM to supervised SOM which is named SSOM.

As shown in Figure 2, there are three-layer structures in SSOM instead of two layer structure in SOM. They are input layer, competitive layer and output layer. In this network, the number of output layer is equal with data classification category. Each output node represents a data category, and connection between the output layer node and the competitive layer node is also full connection way.

According to different prediction category of input samples, SSOM selects different weight adjustment formula to adjust weights and train network. SSOM not only adjusts the weight ω_{ij} between input layer and competitive layer, but also adjust the weight ω_{jk} between competitive layer and output layer. Finally, the classification results are generated by the combination of the two weights

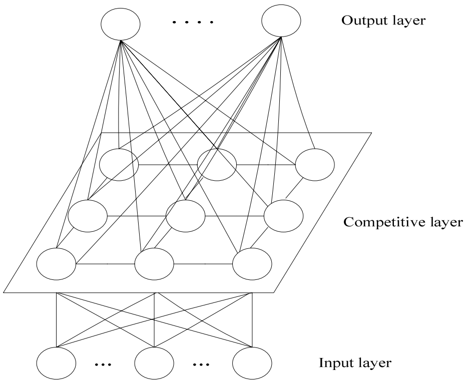

Figure 2. The structure diagram of SSOM

The learning and train step of SSOM is as follow:

(1) Initialization. Initialize the weight ω_{ij} between input layer and competitive layer, the weight ω_{jk} between competitive layer and output layer, and the neighbor radius r etc.

(2) The winning neuron selection on competitive layer. Compute the distance between input sample x_i and competitive layer neural *j* to get the neuron *c* with the minimal distance from the winning neuron. Assume d_i is the minimal distance, use c_i to marker output categories connected to it.

(3) Weight adjustment. Adjust the neuron weights ω_{ij} and ω_{jk} according to the predictive value of input sample x_i. Here we assume the actual output value of x_i is c_x. If $c_i = c_x$, adjust the weights in the neighbor area of *Nc (t)* according to equation (4) and (5).

$$\omega_{ij}^{new} = \omega_{ij}^{old} + \eta_1(x - \omega_{ij}^{old}) \tag{4}$$

$$\omega_{jk}^{new} = \omega_{jk}^{old} + \eta_2(x - \omega_{jk}^{old}) \tag{5}$$

If $c_i \neq c_x$, adjust the weights according to equations (6) and (7).

$$\omega_{ij}^{new} = \omega_{ij}^{old} - \eta_1(x - \omega_{ij}^{old}) \tag{6}$$

$$\omega_{jk}^{new} = \omega_{jk}^{old} - \eta_2(x - \omega_{jk}^{old}) \tag{7}$$

where η_1 and η_2 represent the learning rate, they decrease with the evolution number increasing.

(5) Judge whether the algorithm ends. If not end, return it to step (2).

During the train process of SSOM neural network, The initial parameters such as weight ω_{ij}, ω_{jk}, learning rate η_1, η_2 and neighbor radius r have much influence on testing result. These parameters randomly selected will have a negative effect on test result. In this paper, we use genetic algorithm to optimize the parameters (η_1, η_2 and neighbor radius r) of SSOM..

Genetic algorithm (GA) is a kind of parallel search optimization method, which simulates the natural genetic mechanisms of biological evolution and Darwinian natural selection. Genetic algorithm simulates the phenomenon of duplication, crossover and mutation that occur in natural selection and genetic replication.

Starting at a group which is a potential solution set of problem, it performs selection, crossover and mutation operation to generate a group of individuals better adapted to the environment. Then, group evolves into better and better areas in the search space and continues to evolve through the generations. Eventually they converge to a group of individuals best adapted to the environment and obtain the optimal solution.

In recent years, genetic algorithm has been successfully used in the fields of economic management, traffic transportation, and industrial design and resolved many technical problems successfully. For example, reliability optimization, flow shop scheduling, job shop scheduling, machine scheduling, equipment layout design, image processing and data mining etc.

The basic operation of optimization using genetic algorithm includes population initialization, fitness function calculation, selection, crossover and mutation operation.

III. Optimization of Weights and Thresholds

After train and learning SSOM network can quickly and easily achieves the classification of testing data. It can be used in a variety of classification field for labeled data such as text classification, intrusion detection, fault detection, etc. However during the train and learning process of SSOM neural network, the initialization of three parameters (learning rate η_1 and η_2, neighborhood radius r) have much influence on the experiment result. If the choice of these parameters is not very good or not very correct, it will have much negative effect on the test results and lead to lower correct classification rate.

The real number encoding method is an important encoding method of genetic algorithm, in which each individual gene value is real number. The real number encoding method has following advantages:

- Suitable for the scope of the larger number.
- Easy to expand the space of the genetic search.
- It can improve the accuracy requirements of the genetic algorithm.

- It can improve the computational complexity and efficiency of operations.
- Easy to use together with other classical optimization method.

Here, we design a real encoding genetic algorithm to optimize the parameters η_1 , η_2 and r of SSOM to obtain the optimal parameters. Using these optimal parameters we create a new SSOM network model named GA-SSOM and perform intrusion classification based on KDD Cup 1999 data set.

To complete the optimization using GA, we firstly should initialize the individual population composed of parameters η_1 , η_2 and r in real coding, then design a proper fitness function to perform selection operation, crossover operation and mutation operation. After many times repeated iteration, the optimal individual including the optimal parameters is obtained. It is what we needed to create the GA-SSOM model.

The implementation step is shown in Figure 3 and the detailed process is as follows:

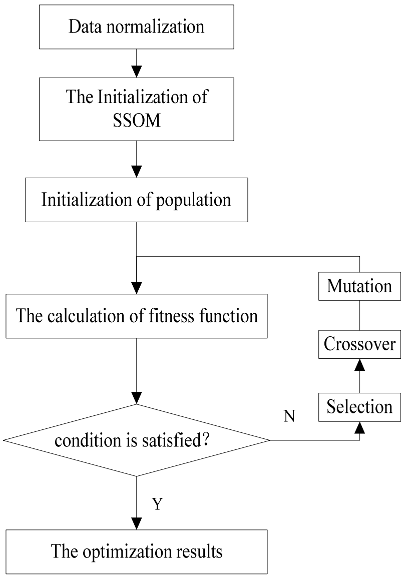

Figure 3. The optimization of parameter

(1) Data normalization

Data normalization is a data preprocessing procedure before training network; it is accomplished by data normalized function. Data normalization function is used to cancel the orders of magnitude difference between the dimensions of data and avoid large prediction error caused by differences in input and output. In this paper, the input feature value is normalized to [0, 1] by data normalization function in equation (8).

$$x_k = (x_k - x_{\min}) / (x_{\max} - x_{min}) \tag{8}$$

where x_k represents the data sequence, x_{min} and x_{max} represents the minimum value and maximum value in data sequence.

(2) The initialization of population

A population is formed by N individuals generated randomly and genetic algorithm starts the iteration from this population as the initial point.

In this part, individual coding adopts real coding and each individual is a real number series, which consists of 3 components: learning rate η_1 , learning rate η_2 and neighborhood radius r . And N is set to 20 in our work.

(3) Fitness function calculation

Fitness value is to measure the excellent degree that each individual approach or reach in optimization calculation. The higher fitness value the individual has, the larger probability it is genetic to next generation than others. Fitness value is usually calculated through a fitness function.

Here, we choose the reciprocal of the square of the absolute error between forecast output and the desired output data as the fitness function to judge the quality level of individual. The individual with greater fitness value will have more opportunity to be selected and inherited to the next generation. The fitness function is shown in equation (9).

$$F = \frac{1}{\sum_{i=1}^{m} (\mathrm{c}_i - \mathrm{c}_x)^2} \tag{9}$$

Where F represents the fitness value, c_i represents the forecast output and c_x represent the desired output of the first i node, m is the number of output node.

(4) Selection operation

The task of select operation is to select body from the parent group to inherit to the next group. The genetic algorithm uses selection operator (or copy operation) to achieve the group individual survival of the fittest operation. The probability that high fitness individual is inherited to the next generation of group is great, and the probability that small fitness individual is inherited to the next generation of group is small.

During the process of selection operation, proportional selection method is used. The basic idea of proportional selection method is as follow: the probability that individuals are selected is proportional to the size of its fitness.

The selection probability p_i which represents the first i individual is shown in equation (10).

$$p_i = \frac{F_i}{\sum_{j=1}^{N} F_j} \tag{10}$$

where Fi is fitness value of the first i individual, N is the population size.

(5) The crossover operation

The crossover operation is a process in which the two paired chromosomes exchange some of its genes in a certain way to form two new individuals. The crossover operation is an important feature that the genetic algorithm is different from other evolutionary algorithms. It plays a key role in the genetic algorithm and is the main method to generate new individual.

Because we use real number encoding GA, the crossover operation uses arithmetic crossover which is a linear combination of the two individuals to produce a new individual. The process is shown in equation (11).

In this equation, it shows that the first *k* chromosome named a_k performs crossover operation with the first *l* chromosome named a_l, and the crossover bit is at first *j* bit. After crossover operation, a new pair of individual with good genes is generated.

$$\begin{cases} a_{kj} = a_{kj}b + a_{lj}(1-b) \\ a_{lj} = a_{lj}b + a_{kj}(1-b) \end{cases} \tag{11}$$

where *b* represents a random number between 0 and 1.

(6) Mutation operation

The so-called mutation operation is a process in which the value of certain genes in the individual encoded string is replaced by other genetic value to form a new individual. The mutation operation is a helper method to generate new individual, but it is essential to a computing step. Mutation operation determines the local search ability of genetic algorithms.

Equation (12) shows the process of mutation operation in this part. Select a_{ij}, which is the first *j* gene of the first *i* individual to perform mutation operation. The mutation operation is as follows:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{\max}) \times r & r > 0.5 \\ a_{ij} + (a_{\min} - a_{ij}) \times r & r < 0.5 \end{cases} \tag{12}$$

where a_{max} is the upper bound of a_{ij}, a_{min} is the lower bound of a_{ij}, *r* is the random value between 0 and 1.

Through the above steps, we get the optimal chromosome which is composed of the optimal learning rate η_1 and η_2, the optimal neighborhood radius r. Use these optimal variables to create SSOM neural network model named GA-BP. Then we use this model to carry out intrusion detection experiment based on KDD Cup 1999 data set.

IV. Optimization of Input Variable for Dimensionality Reduction

Using SSOM neural network to establish the model, the excessive input variable is easy to over fitting, which leads to the low precision, low rates of detection and excessive time. So it is necessary to optimize the selection of input variables, remove the redundancy variables and retain the variables which can most reflect the relationship between input and output variables in the model [21].

Binary encoding is one of the most commonly coding of genetic algorithm. It has the following advantages:

- Encoding, decoding operation is simple.
- The cross and mutation operation is easy to realize.
- Meeting the minimum character set encoding principle.
- Easy to use schema theorem theoretical to analyze the algorithm.

In this part we use binary encoding genetic algorithm to optimize input variable and reduce its dimensionality. To complete it, first encode the individual components, initialize the number of populations and the evolution, and design the fitness function. Then perform selection operation, crossover operation and mutation operation to generate the best individual which is the optimal combination of independent variables. The workflow is shown in Figure 4, each part functions as follows:

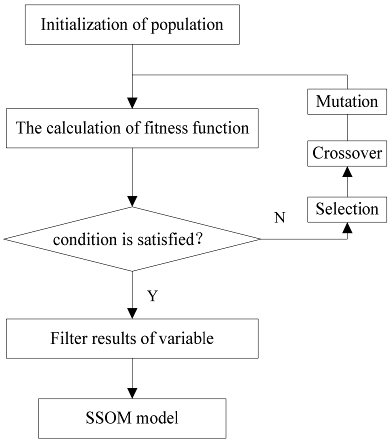
Initialization of population
The calculation of fitness function
Mutation
Crossover
Selection
N
condition is satisfied?
Y
Filter results of variable
SSOM model

Figure 4. The optimization of input variables

(1) Data normalization

Data normalization is a data preprocessing procedure before the experiment, it is also important for variable dimension reduction. Here, the input feature value is also normalized to [0, 1]. Data normalization function also uses equation (8).

(2) The initialization of population

In this optimization process, the individual coding adopts the binary coding mode. As the intrusion detection data has 41 features, the length of coding is designed to 41 and every individual is a binary string composed of 41 binary bits. Every chromosome corresponds to an input feature and every gene can only be 1 and 0. If the value of a particular chromosome is 1, it means that the input variable corresponding to this bit takes part in the final model, otherwise not.

A population is formed by *N* individuals generated randomly and genetic algorithm starts the iteration from this population as the initial point.

(3) Fitness function calculation

Here, the reciprocal of the absolute error between forecast output and the desired output data is chose as the fitness function and it is shown in equation (13).

In the process of calculating the fitness function, the learning rate η_1 and η_2, neighborhood radius r of every individual is optimized by the genetic algorithm in Section III, avoiding the impact of its random on fitness function calculation.

$$F = \frac{1}{\sum_{i=1}^{m}(C_i - C_x)} \tag{13}$$

where F represents the fitness function, m is the number of output node, c_i and c_x represent respectively forecast output and the desired output of the first i node.

(4) Selection operation

During this process, we adapt proportion selection operator and calculate the probability of each individual's fitness in accordance with the equation (10). The individuals with larger probability is selected as the best individual to the next generation of genetic population, the one with smaller probability not.

(5) The crossover operation

During crossover operation, two individual are selected randomly from population to generate a new and outstanding individual.

As the optimization of this part adopts binary coding, one-point crossover operator is used during the crossover operation. For a matched pair of individual, select randomly the cross-point and swap the other bit from cross-point. The operating diagram is shown in Figure 5. The binary string 1001 in individual A exchanges data information with binary string 0011 in individual B. After crossover operation, it generates two new individual and increases the diversity of individual.

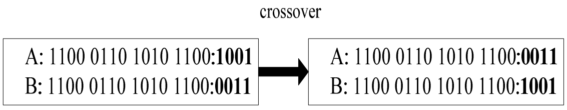
crossover
A: 1100 0110 1010 1100:**1001**
B: 1100 0110 1010 1100:**0011**
A: 1100 0110 1010 1100:**0011**
B: 1100 0110 1010 1100:**1001**

Figure 5. Crossover operation

(6) Mutation operation

Mutation operation can also increase the diversity of individual. Here, select a single point mutation operator and random mutation point, then 0 and 1 is exchanged. The principle is shown in Figure 6. Two new individual generate after this operation.

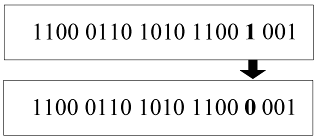

Figure 6. Mutation operation

(7) The establishment of New-GA-SSOM network model

After many times evolution, when meeting the iteration condition, the output of the population is the optimal solution of the problem. They are the handsome and the most representative input variable combination.

Through the above steps, we get the optimal chromosome which is composed of the optimal feature. Extract a set of variables from the best chromosome gene as the final input variables to achieve the dimension reduction of independent variables. That is the new neural network model, named New-GA-SSOM. Then we use this model to train network, and carry out intrusion detection data based on KDD Cup 1999 data set.

V. EXPERIMENT

KDD Cup 1999 data set is a standard data set for intrusion detection, including the training data set and test data set. The training data set includes 494 021 records and testing data set includes 311 029 records. In the KDD99 data set, each data example represents attribute values of a class in the network data flow, and each class is labeled either as normal or as an attack with exactly one specific attack type. There are 22 types of attacks in the training data set and an increase of new 14 kinds of attacks in the testing data set. All the attack types can be divided into four major categories: Probing, Denial of Service (DoS), User-to-Root (U2R) and Remote-to-Local (R2L). Each complete TCP (transmission control protocol) connection is considered as a record, including four types of attributes collection: time-based traffic features, host-based traffic features, content features and basic features [22, 23, 24].

Our experiment is based on the KDD Cup 1999 intrusion detection data set. Training data set is composed of 3 000 data of normal type and 3 000 data of attack type, selected randomly from KDD Cup99 of "10% KDD" dataset. Testing data set is composed of 2 000 data of normal type and 2 000 data of attack type, selected randomly from KDD Cup99 of the "Corrected KDD" dataset. The selected data set is shown in Table I.

Each data has 41 different attributes (32 continuous attributes and 9 discrete attributes) used as SSOM input value and 1 attack type label used as output value of SSOM. Some of them are the numerical types, and some are character types, but SSOM can only deal with numerical data. Therefore, before training we must make the input data numerical and normalized. This study used simple substitution symbols with numerical data types. The protocol-type, service and flag are replaced by digital attributes. For example, three kinds of protocol-type (tcp, udp and icmp) will be expressed with 1, 2, 3. Also, 70 kinds of services are substituted with 1, 2… 70. The attack types are also numbered with 1, 2, 3 and so on.

Experimental platform is the PC with Intel Core2 Duo CPU 2.0GHz, memory 2.0GB, Windows XP operating system and MATLAB 7.8.0 (R2009.0a) programming environment.

Based on the experiment data in Table I, training and test are carried out respectively using SSOM (its parameters are selected randomly), GA-SSOM and New-GA-SSOM neural network. According to the different classification number of attack type, experiment is carried out as following two cases.

TABLE I.
TRAINING SET AND TEST SETS

Attack class	Attack type	Training set	Test set
Normal	normal	6000	3000
DOS	back	700	400
	neptune	2700	1200
	smurf	1600	800
R2L	guess_passwd	53	40
U2R	buffer_overflow	30	22
Probe	ipsweep	350	180
	portsweep	350	200
	satan	217	158

TABLE II.
DETECTION RATE AND MODEL TIME (TWO CLASSIFICATION)

Type / Model	Time	Detection rate (%)	
		normal	abnormal
SSOM	38.1s	93.2	90.2
GA-SSOM		98.5	95.3
New-GA-SSOM	13.5s	97.5	95.5

TABLE III.
DETECTION RATE AND MODEL TIME (FIVE CLASSIFICATION)

Model / Type		Model		
		SSOM	GA-SSOM	New-GA-SSOM
detection rate (%)	Normal	92.1%	98.4%	96.5%
	DOS	89.8%	94.4%	94%
	R2L	6.7%	7.7%	7.1%
	U2R	19.2%	23.4%	22.4%
	Probe	89.3%	95.1%	96.1%
time (s)		45.4.s		16.5s

Experiment 1: the attack types of selected experiment data are divided into normal data and attack data, the normal data is numbered with 1 and the attack data is numbered with 2. It is a two classification problem and the experiment result is shown in Table II.

Experiment 2: The attack types are classified into Normal data, DOS, R2L, U2L, Probe. The Normal label data is numbered with 1, the other four types are numbered with 2, 3, 4 and 5. It is a multiple classification problem and the experiment result is shown in Table III.

From Table II and Table III, we can know that the proposed GA-SSOM and New-GA-SSOM have higher detection than SSOM whose parameters are selected randomly. Although there is little difference in detection rate between GA-SSOM and New-GA-SSOM, New-GA-SSOM spends less time than SSOM and GA-SOM in modeling. So it shows that GA-SSOM has rather higher intrusion detection rate than SSOM, and New-GA-SSOM has higher intrusion detection rate and much shorter modeling time than SSOM.

VI. Conclusion

In this paper, we use genetic algorithm to optimize the SSOM which is an improved and a supervised SOM neural network. A real encoding genetic algorithm is applied to optimize the learning rate η_1 and η_2 , neighborhood radius r of SSOM neural network to improve detection rate. And a binary encoding genetic algorithm is used to reduce the dimension of input variable of SSOM neural network to improve the efficiency of modeling.

Through optimization, it can quickly and effectively establish SSOM network model and improve speed of training and learning. Classification experiments based on KDD Cup 1999 data set was carried out and results showed that the optimized model has shorter modeling time and higher intrusion detection rate.

In the future, we plan to propose a semi-supervised intrusion detection classifier based SOM, and use genetic algorithm to optimize the input parameters of this semi-supervised classifier.

Acknowledgment

The authors wish to thank the support of ShangLuo University and Northwestern Polytechnical University. This work was sponsored by Scientific Research Program Funded by Shaanxi Provincial Education Department (No.12JK0748) and National ministries foundation in China.

References

[1] E. J. Palomo, E. Domínguez, R. M. Luque and J. Muñoz, “An Intrusion Detection System Based on Hierarchical Self-Organization”, Advances in Soft Computing, 2009, Volume 53, 139-146. [Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS’08].

[2] Qinglei Zhang, Gongzhu Hu and Wenying Feng, “Design and Performance Evaluation of a Machine Learning-Based Method for Intrusion Detection”, Studies in Computational Intelligence, Volume 295, pp.69-83, 2010.

[3] Zhang yirong, Xiao ShunPing, Xian Ming, “An overview of intrusion detection techniques based on machine learning”, Computer Enginneering and Application, vol. 42 (2): 7-10, 2006.

[4] Zhou Honggang, Yang Dechun, “Anomaly detection approach based on immune algorithm and support vector machine”, Computer Application, vol. 26, no. 9, pp. 2145-2147, 2006.

[5] Jason Shifflet, “A Technique Independent Fusion Model for Network Intrusion Detection”. Proceedings of the Midstates Conference on Undergraduate Research in Computer Science and Mathematics, University of Denison, America, pp. 13-19, 2004.

[6] ZANG Weihua, GUO Rui, “The Application of Neural Network based on Evolutionary Strategy in Network Security Quantification Analysis”, AISS: Advances in Information Sciences and Service Sciences, vol. 4, no. 2, pp. 151 ~ 159, 2012.

[7] Wei Xiong, "Anomaly-based detection using synergetic neural network", JDCTA: International Journal of Digital Content Technology and its Applications, vol. 6, no. 4, pp. 188-196, 2012.

[8] Patcha, A., & Park, J. M, “Network anomaly detection with incomplete audit data”, Computer Networks, vol. 51, no. 13, pp. 3935–3955, 2007.

[9] SWARUP K S, CORTHIS P B, “ANN approach assesses system security”, Computer Applications in Power, vol.15, no.3, pp.32-38, 2002.

[10] Xiaomei YI, Peng WU, Dan DAI, Lijuan LIU, Xiong HE, “Intrusion Detection Using BP Optimized by PSO”, IJACT: International Journal of Advancements in Computing Technology, vol. 4, no. 2, pp. 268 -274, 2012.

[11] Liu Hui, CAO Yonghui, “The Research of machine learning algorithm for intrusion detection techniques”, JDCTA: International Journal of Digital Content Technology and its Applications, vol. 6, no. 1, pp. 343-347, 2012.

[12] Jie Ma, Zhi Tang Li, Bing Bing Wang, “Application of Singular Spectrum Analysis to the Noise Reduction of Intrusion Detection Alarms”. Journal of Computers, vol. 6, no. 8, pp. 1715-1722, 2011.

[13] Rauber A., Merkl D., Dittenbach M., “The growing hierarchical self-organizing map: Exploratory analysis of high-dimensional data”, IEEE Transactions on Neural Networks, Vol. 13, no. 6, pp.1331-1341, 2002.

[14] E. J. Palomo, E. Domínguez, R. M. Luque and J. Muñoz, "An Intrusion Detection System Based on Hierarchical Self-Organization, " Advances in Soft Computing, vol. 53, pp. 139-146, 2009.

[15] Jian Wu, Jie Xia, Jian-ming Chen, Zhi-ming Cui, "Moving Object Classification Method Based on SOM and K-means. Journal of Computers", vol.6, no.8, pp.1654-1661, 2011.

[16] YANG Ya-hui, JIANG Dian-bo, SHEN Qing-ni, XIA Min, "Research on intrusion detection based on an improved GHSOM", Journal on Communications, vol.32, no. 1, pp. 121-126. 2011

[17] Zhao Jianhua, LI Weihua, Application of Supervised SOM Neural Network in Intrusion Detection, Computer Engineering, vol. 38, no. 12, pp. 1-3, 2012.

[18] MIYOSHI Tsutomu, "Initial Node Exchange and Convergence of SOM Learning", Proceedings of The 6th International Symposium on Advanced Intelligent Systems (ISIS2005), pp. 316-319, 2005.

[19] Kohonen.T, "Self-organized formation of topologically correct feature maps", Biological cybernetics, vol. 43, no. 1, pp. 59-69, 1982.

[20] Chao Shao, Yongqiang Yang, "Distance-Preserving SOM: A New Data Visualization Algorithm", Journal of Software, vol. 7, no. 1, pp. 196-203, Jan 2012.

[21] SHI F, WANG S C, YU L, "Matlab neural network 30 cases analysis", Beijing University of Aeronautics and Astronautics Press, China, 2010.

[22] Mukkamala S, Sung AH, and Abraham A, "Intrusion dection using an ensemble of intelligent paradigms", Proceedings of Journal of Network and Computer Applications, vol. 2, no. 8, pp. 167-182, 2005.

[23] WANG Hui, ZHANG Guiling, E Mingjie, SUN Na, "A Novel Intrusion Detection Method Based on Improved SVM by Combining PCA and PSO", Wuhan University Journal of Natural Sciences, vol. 16, no. 5, pp. 409-413, 2011.

[24] Jimin Li, Wei Zhang, KunLun Li, "A Novel Semi-supervised SVM based on Tri-training for Intrusion Detection", Journal of Computers, vol. 5, no. 4, pp. 638-645, 2010.

[25] Hettich S, Bay S D.The UCI KDD Archive [EB/OL]. http: //kdd.ics.uci.edu/ databases/kddcup99.]

Zhao Jianhua was born in 1982. He is currently a lecturer and seeking for his doctor's degree. His research interests include machine learning, network security.

Li Weihua was born in 1951. He is currently a professor. His research interests include network security and intelligent decision.