Research on Network Intrusion Detection Technology Based on Data Mining Technology

Lijun Zhou[1, a], Hong Lv[1, b] and Yuan Zhao[1, c]

[1] Naval Aeronautical and Astronautical University, Yantai 264001, China.

[a]jungle730@163.com, [b]lvhong1884@126.com, [c]ascendtop@126.com

Keywords: Data mining, BP neural network, Network intrusion detection, particle swarm optimization algorithm.

Abstract. In this paper, the technology of network intrusion detection based on data mining technology is studied. As the conventional BP neural network be used to establish the network intrusion detection techniques has some problems, because the BP neural network is easy to fall into minimum value and the accuracy is low, the paper uses particle swarm algorithm to optimize the BP neural network model, and uses dynamic inertia weight coefficient to determine the parameters of BP neural network. By using dynamic inertia weight coefficient to determine the parameters of BP neural network, by combining the network intrusion traffic characteristics and BP neural network parameters to encode to a particle, we achieved the parameters of the network intrusion traffic characteristics and BP neural network synchronization selection. By using the KDD CUP99 database of intrusion traffic data to train and test the model we proposed and the conventional model separately, the results show that the algorithm we proposed has better detection efficiency and detection accuracy.

Introduction

With the increasing complexity of computer Internet environment, as well as the increasingly common network technology, Internet intrusion detection methods are becoming more and more diverse, intelligent and complex, so the use of traditional firewall and operating system reinforcement can't resist the current intrusion attacks, and can't meet the requirements of network security. In this situation, intrusion detection system came into being [1]. The intrusion detection system can detect the network traffic data and detect the abnormal traffic. However, the network traffic data is huge, which leads to the need of data to support the decision of the system, which is often difficult to start with the mass of data, and can not quickly or efficiently extract the required information and valuable knowledge. To comply with this demand, data mining technology has developed rapidly in recent years. As a first-class data analysis tool, data mining technology has strong ability of data processing and analysis, and the current merger of communication technology, computer technology and network technology makes the amount of data to further expand. So the status of data mining technology has been promoted in the field of information management[2].

Machine learning algorithms such as BP neural network, are commonly used in data mining technology, they can automatically find the operation need parameters and patterns in a large number of training samples and after learning, they have excellent data processing ability and self learning ability, and can accurately identify. But the network intrusion detection technology based on conventional BP neural network has some problems such as: the BP neural network is easy to fall into the minimum, which leads detection efficiency and the accuracy to be low[3]. In this paper, the BP neural network model was optimized by using particle swarm algorithm, we used dynamic inertia weight coefficient to determine the parameters of BP neural network, combined the network intrusion traffic characteristics with BP neural network parameters, then encoded it into a particle in order to achieve the synchronous selection of the network intrusion traffic characteristics and parameters of BP neural network.

Improved BP neural network

BP neural network

Set the training sample as (x_k, y_k), x_k is the input vector, y_k is the desired output vector， namely, actual type. The output of the BP neural network and the error of the actual type y_k is expressed as:

$$E = \sum_{k=1}^{m} E_k \tag{1}$$

In Formula (1)， E_k is expressed as:

$$E_k = \sum_{i=1}^{n} \phi(e_{ik}) = \frac{1}{2}\sum_{i=1}^{n}(y_{ik} - \hat{y})^2 = \frac{1}{2}\sum_{i=1}^{n} {e_{ik}}^2 \tag{2}$$

The output of BP neural network and the error of the actual type y_k are changed by the adjustment of the weights. Set the output (which is numbered i) of BP neural network hidden layer as follows:

$$\begin{cases} \hat{y}_{ik} = \sigma_0(\bar{y}_{ik}), \\ \bar{y}_{ik} = {W_{ik}}^{(o)^T} \hat{H}_k^{(l)} = \sum_{j=1}^{nl} {w^{(0)}}_{ij} {h^{(l)}}_{jk} \end{cases} \tag{3}$$

The method of adjusting the weight of BP neural network is as follows:

$$\frac{\partial E_K}{\partial {W_{pk}}^{(0)}} = \frac{\partial E_K}{\partial e_{pk}} \cdot \frac{\partial e_{pk}}{\partial \hat{y}_{pk}} \cdot \frac{\partial \hat{y}_{pk}}{\partial \bar{y}_{pk}} \cdot \frac{\partial \bar{y}_{pk}}{\partial {\hat{y}_{pk}}^{(0)}} \tag{4}$$

$$= -e_{pk} \cdot \sigma_0'(\bar{y}_{pk}) \cdot \hat{H}_{pk}^{(1)}$$

$$\Delta W_{pk}^{(0)} = W_{pk}^{(0)} - W_{pk-1}^{(0)} = -\alpha \frac{\partial E_k}{\partial W_{pk}^{(0)}} \tag{5}$$

The adjustment method of weights of layer r in BP neural network is as follows:

$$\begin{aligned} &\Delta W_{pk}^{(r)} = W_{pk}^{(r)} - W_{pk-1}^{(r)} = \alpha \cdot \varepsilon_{pk}^{(r)} \cdot \hat{H}_{pk}^{(r+1)} \\ &\varepsilon_{pk}^{(r)} = \sigma_r'(\bar{h}_{pk}^{(r)}) \cdot \sum_{i=1}^{n_r - 1} {\varepsilon_{ik}}^{(r-1)} w_{ip}^{(r-1)} \end{aligned} \tag{6}$$

In the use of BP neural network to establish the network intrusion detection model in practical application, because of the gradient descent algorithm of BP neural network to optimize the weights can lead to local optimal solution, the error of the network intrusion detection model is gradually increased, and the detection accuracy is reduced. Therefore, the particle swarm optimization algorithm is used to optimize the weights of BP neural network in this paper[4].

Improved particle swarm optimization algorithm.

Particle swarm optimization algorithm is inspired and evolved by birds' foraging behavior. Set the location of No. i particle in the particle swarm optimization algorithm as $X_i = \{x_{i1}, x_{i2}, \cdots, x_{in}\}$, the flight speed of No. i particle as $V_i = \{v_{i1}, v_{i2}, \cdots, v_{in}\}$, the optimal position vector of the particle is expressed as $P_i = \{p_{i1}, p_{i2}, \cdots, p_{in}\}$, the optimal position vector of the whole particle swarm is expressed as $G_i = \{g_{i1}, g_{i2}, \cdots, g_{in}\}$. Then the particle swarm update speed and update location method are expressed as:

$$\begin{aligned} &{V^{k+1}}_{id} = wV_{id}^k + c_1 r_1 (pb_{id}^k - x_{id}^k) + c_2 r_2 (gb_d^k - x_{id}^k) \\ &X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \end{aligned} \tag{7}$$

Because the search efficiency and accuracy of the particle swarm optimization algorithm are determined by the inertia weight in particle swarm optimization algorithm, the dynamic inertia weight coefficient is used to determine the parameters of BP neural network:

$$w = \begin{cases} w_{min} - \dfrac{(w_{max} - w_{min}) \times (f - f_{min})}{(f_{avg} - f_{min})}, f \geq f_{avg} \\ w_{max} \qquad\qquad\qquad\qquad\quad , f < f_{avg} \end{cases} \tag{8}$$

In the formula (8), $w_{\max}$ represents the maximum value of inertia weight; $w_{\min}$ represents the minimum value of inertia weight; f_{avg} represents mean value of the fitness function; $f_{\min}$ indicates the minimum value of fitness function[3].

In this paper, the characteristics of network intrusion traffic flow and the parameters of BP neural network are combined to form a particle in order to realize the synchronized selection of the network traffic characteristics and the parameters of BP neural network. The position vector of the particle is represented as: (1) the feature of the intrusion, "1" is selected, "0" is not selected; (2) BP neural network parameters. The fitness function of the particle is expressed as:

$$f = w \times precison + (1 - w)\left(\sum_{i=1}^{N_f} f_i\right)^{-1} \tag{9}$$

Among them, f_i is the characteristic state; w is the weight value of the detection rate.

In this paper, the network intrusion detection method is as follows:

Step 1: Collect the data of the network intrusion traffic and select the characteristics, and normalize the characteristics of the intrusion;

Step 2: Set up the model of BP neural network and particle swarm optimization algorithm, initialize the particle swarm, the particle swarm is an encoded combination of network intrusion traffic characteristics and the parameters of the BP neural network;

Step 3: Calculate the fitness value of each particle, update the optimal position of particle history and particle swarm optimization;

Step 4: Inertia weight of the particle swarm optimization algorithm and update the speed and position of the particles;

Step 5: If the update iteration satisfies the termination condition, the optimal parameters of the BP neural network are obtained by the particle swarm optimization. If the termination condition is not satisfied, then the iterative update will be performed from step 3;

Step 6: Establish the network intrusion detection model on the base of the optimal parameters of BP neural network.

In this paper, the method of network intrusion detection is shown in Figure 1.

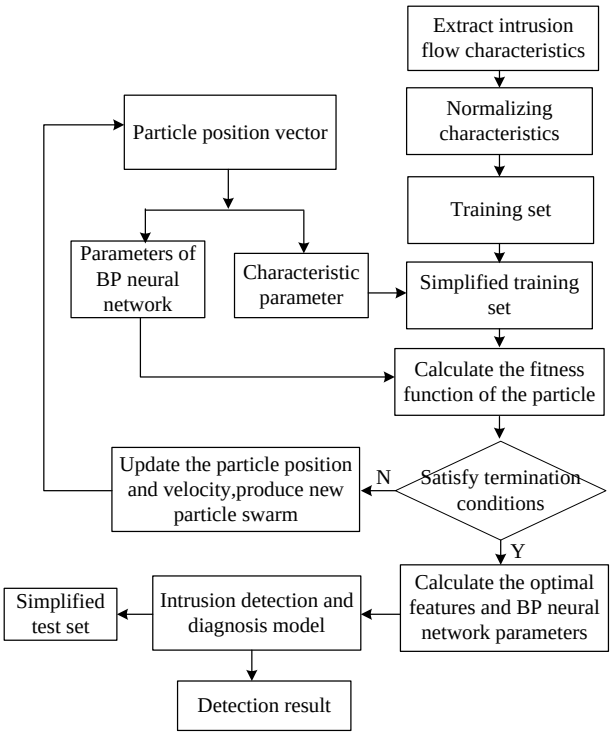

Fig. 1 Network intrusion detection process

Experimental study

In this paper, we use the database of KDD CUP99 to study the network intrusion detection technology, and use MATLAB to establish the intrusion detection model. KDD CUP99 database includes all kinds of intrusion traffic data, we select 4 kinds of classical intrusion models, they include Probe (scan attack), DOS (denial of service attack), U2L (unauthorized use of the local super privilege access attacks) and U2R (remote users not authorized to access). For the four types of intrusion, 200 data streams are selected, among them, 100 randomly selected data streams are used to train the detection model, and the other 100 are used to test the detection performance of the model test.

Normalized processing of data streams to simplify the model data processing :

$$S^* = \frac{0.9(S - S_{\min})}{S_{\max} - S_{\min}} + 0.05 \tag{10}$$

In the formula (10): s is the value before the normalized processing; $S_{\max}$ is the maximum of the normalized processing; $S_{\min}$ is the minimum value of the normalized processing.

In order to compare the detection performance of the network intrusion detection model based on the particle swarm optimization BP neural network algorithm, the conventional BP neural network is used to establish the same detection model, and the model training and performance tests are carried out using the same training and test data. The detection results of the detection model based on the two algorithms are shown in Figure 2 and Figure 3.

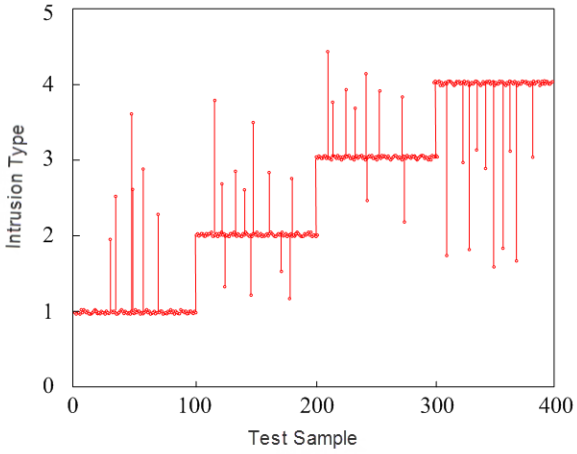

Fig. 2 Recognition results of BP neural network algorithm model based on particle swarm optimization

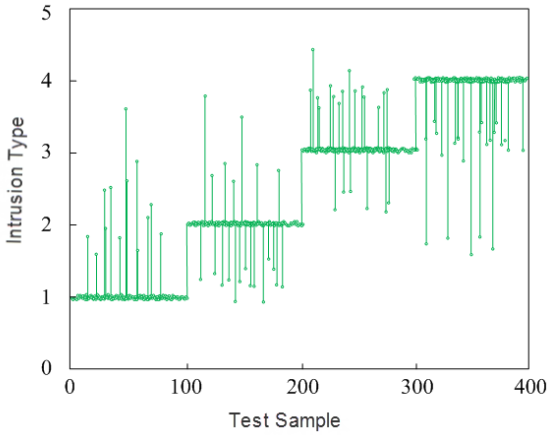

Fig. 3 Recognition results of traditional BP neural network algorithm model

In figure 2 and 3, the meaning of horizontal axis and vertical coordinate is shown in Table 1.

The performace of network intrusion detection method is evaluated by false positive rate(FPR), detection rate(DR) and detection time(DT)[5].

$$FPR = \frac{Number\,of\ normal\ samples\,being\ invaded\ by\ false\ positives}{Number\ of\ normal\ samples} \times 100\% \tag{11}$$

$$DR = \frac{Number\,of\ \text{intrusion}\ samples\ be\ \text{detected}}{Number\ of\ \text{intrusion}\ samples} \times 100\% \tag{12}$$

Table 1 Horizontal and vertical coordinates of figure 2 and figure 3

Horizontal axis	Meaning	Vertical axis	Meaning
1~100	Test Sample of Probe	1	Sample be judegd as Probe
101~200	Test Sample of DOS	2	Sample be judegd as DOS
201~300	Test Sample of U2L	3	Sample be judegd as U2L
301~400	Test Sample of U2R	4	Sample be judegd as U2R

The detection rate and the false positive rate of the two methods are shown in Table 2.

Table 2 Detection rate and false positive rate results of two methods

Intrusion Type	Improved method of this paper		Conventional BP neural network	
	Detection Rate	False Positive Rate	Detection Rate	False Positive Rate
Probe	95.6	4.4	82.3	17.4
DOS	88.1	11.9	76.5	23.5
U2L	92.3	7.7	81.6	18.4
U2R	91.6	8.4	83.5	16.5

The comparison of detection efficiency between Using the algorithm of this paper and the conventional BP neural network algorithm are shown in Table 3.

Table 3 Detection time comparison of two methods

Intrusion Type	Improved method of this paper		Conventional BP neural network	
	Training Time(s)	Detection Time(s)	Training Time(s)	Detection Time(s)
Probe	23.69	8.68	65.86	18.55
DOS	9.50	6.85	23.26	16.64
U2L	9.29	6.25	29.97	16.45
U2R	13.43	11.39	31.20	15.05

From the experimental data, we can see that the network intrusion detection model based on the improved particle swarm optimization BP neural network algorithm is significantly improved compared to the conventional BP neural network algorithm. For 4 types of intrusion detection, by using the improved algorithm, the average detection rate is 91.9%, the average false positive rate is 8.1%, while by using the conventional algorithm, the average detection rate is 80.975%, the average false positive rate is 18.95%. At the same time, the detection efficiency of this paper is higher than the conventional algorithm, for the four types of intrusion, the average training time is 13.98s, the detection time is 8.3s, while by using the conventional algorithm, the average training time is 37.6s, the detection time is 16.67s.

Summary

As the problem that by using the traditional firewall and strengthening the operation system can not resist the current intrusion or meet the requirements of network security[6], the paper researched on network intrusion detection technology which is based on data mining technology. Machine learning algorithms such as BP neural network, are commonly used in data mining technology, they can automatically find the operation need parameters and patterns in a large number of training samples after learning, they have excellent data processing ability and self learning ability, and can accurately identify. But the conventional BP neural network be used to establish the network intrusion detection techniques has some problems, because the BP neural network is easy to fall into minimum value and the accuracy is low, this paper used particle swarm algorithm to optimize the BP neural network model, by using dynamic inertia weight coefficient to determine the parameters of BP neural network,combining the network intrusion traffic characteristics and BP neural network parameters and encoding to a particle, to achieve the parameters of the network intrusion traffic characteristics and BP neural network synchronization selection. By using the KDD CUP99 database to train and test the intrusion traffic data using this method as well as the conventional BP neural network, the results show that the proposed algorithm has better detection efficiency and detection accuracy.

References

[1]. Shenzheng Zuo. Research on Machine Learning Approach For Network Anomaly Detection and Response[D]. Ph.D thesis of Beijing University of Posts and Telecommunications ,2010,p.5-25.

[2]. Lv Man. Research on Intrusion Detection Based on Data Mining. Master's thesis of Daqing Petroleum Institute,2007,p.10-28.

[3]. Guo Tong.Research on Network Anomaly Detection Technology based on Adaptive Flow Sampling Measurement[D].Ph.D thesis of Information Engineering University,2013,p.23-45.

[4]. Shah Bhavin, H Trivedi Bhushan.Artificial Neural Network based Intrusion Detection System: A survey[J]. International Journal of Computer Applications, Vol.39(2012),No.6,p.13-18.

[5]. Jiaolong Peng.Research and Application of Improved Ant Colony Clustering Algorithm to Optimize the RBF Neural Network[D].Master's thesis of Zhengzhou University,2013,p.8-33.

[6]. Jing G L，Du W T，Guo Y Y．Studies on prediction of separation percent in electrodialysis process via BP neural networks and improved BP algorithms[J].Desalination,2012,29(1):78-93．