Network Security Situation Prediction Based on Adaptive Clustering RBF Network

Fangwei Li[1,a] , Bo Zheng[1,b] , Jiang Zhu[1,c] and Zhuxun Peng[1,d]

[1] Chongqing Key Lab of Mobile Communications Technology，Chongqing University of Posts and Telecommunications（CQUPT), Chongqing 400065, China

[a]lifw@cqupt.edu.cn, [b]zhbook@163.com, [c]zhujiang@cqupt.edu.cn, [d]pengzhux@163.com

Keywords: Network Security Situation Prediction(NSSP); Radical Basis Function (RBF) Neural Network; Adaptive Clustering

Abstract. Proposed is an algorithm for network security situation prediction (NSSP) based on adaptive clustering radical basis function (RBF) neural network. Experiment results show that, the proposed method not only reflects the general trend of network security situation, but also improves the prediction accuracy.

Introduction

Network security situation awareness（NSSA）is a technique that aims at providing the network entity with the ability to access, understand, and predict its security elements. It offers strategies to deal with network security threats and provides new ideas to build a seamless network security system [1, 2].

To solve such technical problems of NSSA, and achieve a higher accuracy on predicting the network security situation awareness, this paper proposes an adaptive clustering RBF network security situation prediction method.

RBF Network Model

A typical structure of an RBF network is shown in Fig.1

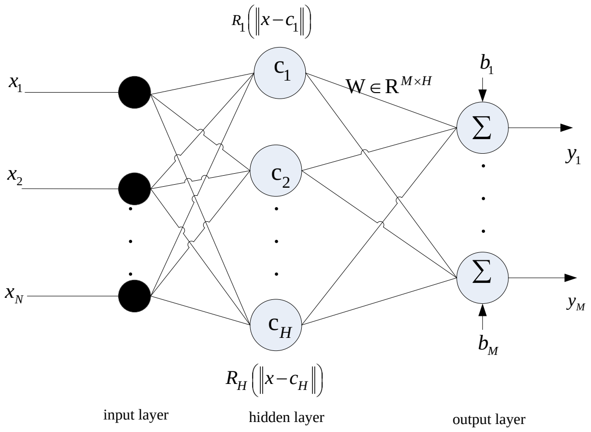

Fig.1 RBF neural network structure

The RBF network output of the ith hidden node is given by:

$$\phi_i = R_i\left(\|x-c_i\|\right) \qquad i=1,2,...,H \tag{1}$$

where x is the input vector of the RBF network, c_i is the center vector of the ith node, H is the number of hidden layer node, $\|x-c_i\|$ is the Euclidean norm of the difference between the input vector and the node center. $R(\cdot)$ denotes the base function having radial symmetry properties.

The final output of jth hidden node is produced by a linear combination of the hidden nodes responses as

$$y_j = \sum_{i=1}^{H} w_{ji}\phi_i + b_j \quad j = 1,2,...,M \tag{2}$$

where w_{ji} is the weight coefficient between the jth output node and the ith hidden node. b_j is the threshold of the jth node in output layer.

Online Adaptive Mechanism

This paper presents a novel online adaptive clustering algorithm by taking the characteristics of network security situation into consideration. Algorithm takes the following steps:

Step1: The network security situation input sample is described by

$$X = (x_1, x_2, ..., x_K) \tag{3}$$

We define two collections $A(L)$ and $B(L)$, where $A(L)$ is used to accumulate all kinds of sample vector of network security situation, $B(L)$ is a counter to storage the number of samples category, and L is the number of the adaptive Clustering categories, respectively.

Step2: From the first sample of network security situation on, each sample point is regarded as a possible data center. Calculate the density index of sample points according to (4), select the largest density index sample points as the first initial data center, and assume that $A(1) = x_i$, $B(1) = 1$.

$$D_i = \sum_{j=1}^{K} \exp(-\frac{\| x_i - x_j \|^2}{(d_1/2)^2}) \quad i = 1,2,...,K; i \neq j \tag{4}$$

Step3: For each of the remaining sample points x_j, calculate the Euclidean distance r between x_j and the determined cluster center in Step2. If $r \leq d_2$, the sample point will be attached to the category which correspond to the center of this cluster, and execute $A(1) = A(1) + x_j$, $B(1) = B(1) + 1$.

Step4: Find out the rest network security situation samples that are not belong to any category, collect them as a new input sample collection X, and re-execute Step1, Step2, and Step3 in order. Loop above process until $B(L) < Q$, where Q is a predefined threshold.

Step5: For each network security posture category $A(i)$, calculate its center of gravity according to (5).

$$c_i = \frac{A(i)}{B(i)} \quad i = 1,2,...,L \tag{5}$$

Then, c_i is the initial data center of hidden layer neurons.

The data center of each hidden node is adaptively determined in section A. Then initialize the extended width of radial basis functions according to (6)

$$\sigma_i = \frac{c_{\max}}{\sqrt{2L}} \quad i = 1,2,...,H \tag{6}$$

where $c_{\max}$ is the maximum distance among the initial data centers in the hidden layer .

In this paper, we use the Gaussian function as the activation function. For a N dimension network security situation sample $x = (x_1, x_2, ..., x_N)^T$, the output of the ith hidden node is:

$$\phi_i(x) = R(\| x - c_i \|) = \exp\left(-\frac{(x_1 - c_{i1})^2 + ... + (x_n - c_{in})^2}{2\sigma_i^2} \right) \tag{7}$$

The data center of hidden layer radial basis functions c_i, the expanded width σ_i, and the output layer weight coefficient w_{ji} can be learned by Gradient descent method.

The objective function of RBF neural network learning is defined as:

$$E = \frac{1}{2}\sum_{k=1}^{K}\sum_{j=1}^{M}\left(y_{kj} - \hat{y}_{kj}\right) \tag{8}$$

In the formula, y_{kj} and $\hat{y}_{kj}$ is the expected output and actual output of the kth network security posture samples in the jth node of the output layer, respectively.

The system considers all training samples of network security situation. The derivative of (8) with respect to the data center c_i , expansion width σ_i and output layer weights w_{ji} is

$$\frac{\partial E}{\partial c_i}=\sum_{k=1}^{K}\sum_{j=1}^{M}\frac{\partial E}{\partial y_{kj}}\frac{\partial y_{kj}}{\partial \phi_i\left(x_k\right)}\frac{\partial \phi_i\left(x_k\right)}{\partial c_i}=\sum_{k=1}^{K}\sum_{j=1}^{M}\left(2\left(y_{kj}-\hat{y}_{kj}\right)w_{ji}\times\exp\left(-\frac{\left\|x_k-c_i\right\|^2}{2\sigma_i^2}\right)\times\frac{x_k-c_i}{\sigma_i^2}\right) \tag{9}$$

$$\frac{\partial E}{\partial \sigma_i}=\sum_{k=1}^{K}\sum_{j=1}^{M}\frac{\partial E}{\partial y_{kj}}\frac{\partial y_{kj}}{\partial \phi_i\left(x_k\right)}\frac{\partial \phi_i\left(x_k\right)}{\partial \sigma_i}=\sum_{k=1}^{K}\sum_{j=1}^{M}\left(2\left(y_{kj}-\hat{y}_{kj}\right)w_{ji}\times\exp\left(-\frac{\left\|x_k-c_i\right\|^2}{2\sigma_i^2}\right)\times\frac{\left\|x_k-c_i\right\|^2}{\sigma_i^3}\right) \tag{10}$$

$$\frac{\partial E}{\partial w_{ji}}=\sum_{k=1}^{K}\sum_{j=1}^{M}\frac{\partial E}{\partial y_{kj}}\frac{\partial y_{kj}}{\partial w_{ji}}=\sum_{k=1}^{K}\sum_{j=1}^{M}\left(2\left(y_{kj}-\hat{y}_{kj}\right)\times\exp\left(-\frac{\left\|x_k-c_i\right\|^2}{2\sigma_i^2}\right)\right) \tag{11}$$

, respectively. The correction method of hidden layer neurons data center c_i , expansion width σ_i and the output layer weight coefficient w_{ji} is:

$$c_i\left(t+1\right)=c_i\left(t\right)-\eta\frac{\partial E}{\partial c_i} \tag{12}$$

$$\sigma_i\left(t+1\right)=\sigma_i\left(t\right)-\eta\frac{\partial E}{\partial \sigma_i} \tag{13}$$

$$w_{ji}\left(t+1\right)=w_{ji}\left(t\right)-\eta\frac{\partial E}{\partial w_{ji}} \tag{14}$$

where t is the training times, and η is the learning rate.

Simulation and Analysis

Network security situation calculation method is shown in Fig.2.

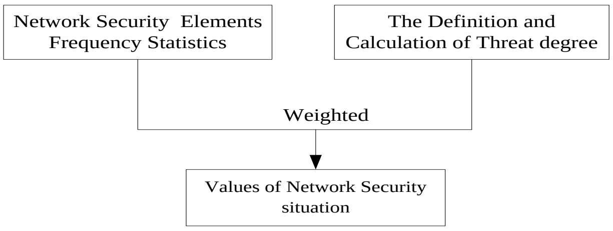
Network Security Elements Frequency Statistics
The Definition and Calculation of Threat degree
Weighted
Values of Network Security situation

Fig.2 Calculated values of network security situation

Simulated data sets adopt the hacker attack data captured by the network security department of a corporate. We set all kinds of network security incidents to the united level of threat. Then we can use security incident warning number as the situation values of the day [3]. According to the data provided by the company's security department, we found that a attack period is about three days, so the RBF network parameters of input layer and output layer can be set as follows:

(1) The RBF network input dimension N =3, the output dimension M =1.

(2) Network security situation training sample size of RBF neural network K =90. The partitioning method of network security situation training sample is shown in Tab.1.

Tab.1 Divide the training sample of network security situation

Input vector	Output vector
x_1,x_2,x_3	x_4
x_2,x_3,x_4	x_5
$\cdots$	$\cdots$
x_{90},x_{91},x_{92}	x_{93}

In order to avoid the negative impact of the large difference between the orders of the trend value, we use formula (15) to normalize the values which come from September 1, 2013 to December 1, 2013 (90 sample values). The normalized situation value is shown in Fig.3.

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{15}$$

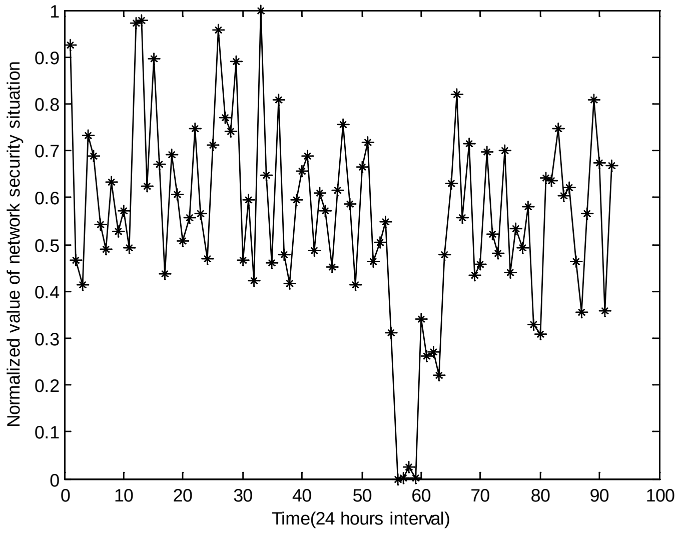

Fig.3 The normalize of network security situation values

According to the normalized network security situation samples that have been adaptively clustered, the number of nodes in hidden layer can be derived as 6. Based on the data base from December 2, 2013 to December 15, 2013, we predict network security situation value using the trained RBF neural network. Proposed algorithm is compared with the K-means RBF neural network model [3] and the support vector machine (SVM) forecasting model [6]. The results are shown in Fig.4.

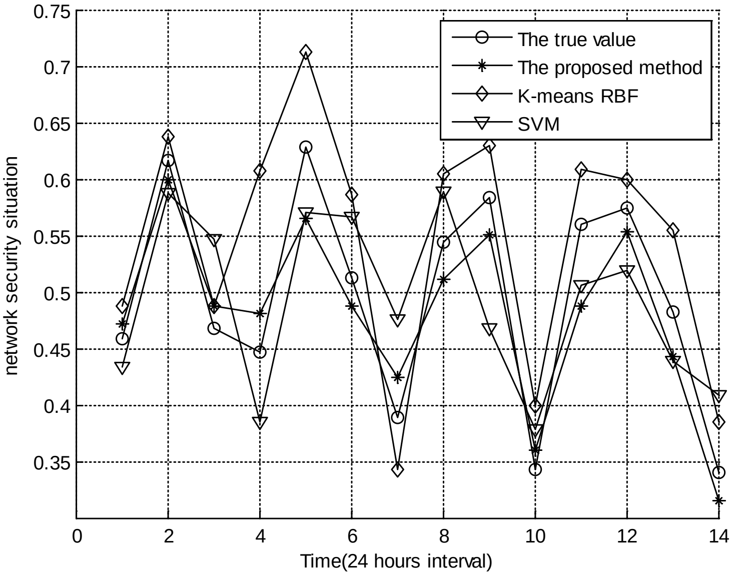

Fig.4 Prediction curve compare of every method

Compared with the other two forecasting methods, Figure 4 shows that the proposed method is more consistent with the real trend of network security situation. The proposed RBF network prediction model has a stronger extrapolated performance and improved prediction accuracy.

Next, we quantitatively compare the prediction performance of three prediction methods by using the absolute error indicators.

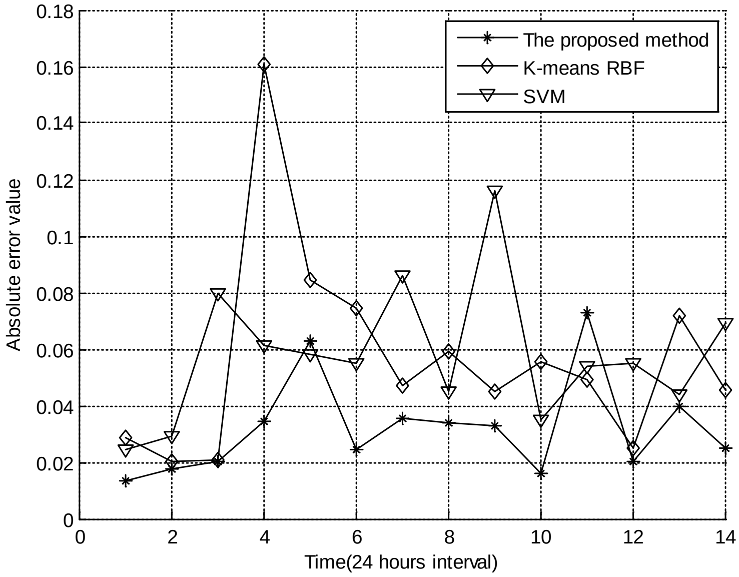

Fig.5 Absolute error curve compare of every method

From Figure 5, the prediction error of the proposed method is slightly larger than the other two methods in some individual points. But from the error curve, the absolute error is less than 0.04 in major prediction time. The prediction accuracy is significantly higher than the K-means RBF and SVM prediction model.

Conclusion

This paper presents a network security situation prediction method based on adaptive clustering RBF neural network. Simulation results show that proposed method can achieve high prediction accuracy and fit the network security situation well. Compared with the K-means RBF and the SVM prediction methods, proposed method can adaptively determine the neural network structure based on the real network environment. The superior forecasting performance of the proposed algorithm over existing methods is verified by simulations and calculation results. So it is more suitable for large-scale network environment network security situation prediction.

Acknowledgement

The National Nature Science Foundation Projects of China (NO.61102062, 61301122),The Science and Technology Research Project of Chongqing Education Commission (NO.KJ120530)

References

[1] ZHANG Y, TAN X B, CUI X L, et al. Network security situation awareness approach based on Markov game model[J]. Journal of Software, 2011, 3: 009.

[2] Rongzhen F, Mingkuai Z. Network Security Awareness and Tracking Method by GT[J]. Journal of Computational Information Systems, 2013, 9(3): 1043-1050.

[3] Ren W, JIANG X, SUN T. RBFNN-based prediction of networks security situation [J]. Computer Engineering and Applications, 2006, 31: 136-139.

[4] Chen H, Gong Y, Hong X. Online Modeling with Tunable RBF Network [J]. IEEE Transactions on Cybernetics, 2013, 43(3): 935-947.

[5] Jajodia S, Liu P, Swarup V, et al. Cyber situational awareness [M]. Springer, 2010.

[6] ZENG B, ZHONG P. Simulation Study on Network Security Situation Forecast [J]. Computer Simulation, 2012, 5: 041.