PROCEEDINGS OF SCIENCE

Prediction Model of Network Security Situation based on Elman Neural Network

Huayu Fei[1]
National Information Academy
Wuhan,430000, China
E-mail: 457210430@qq.com

Jun He
National Information Academy
Wuhan,430000, China
E-mail: felix753@sina.com

To improve the prediction accuracy of time sequence data, this paper presents a model based on Elman neural network optimized by Genetic Algorithm. We initialized the parameters with GA, and then train the network. After training, we simulated the results with the ultimate model, and evaluate the performance between different algorithms. The experiment indicated that GA-Elman can improve the accuracy of prediction.

\

CENet2017
22-23 July 2017
Shanghai, China

[1] Speaker

1.Introduction

With the development of information technology, the internet security issues pop up one after another, the methods of attack change constantly and the negative impact grows. The network security administrators need to turn passive defense to active defense. As network situation awareness is one of the effective ways in active defense, the improvement of the network security situation awareness has become a consensus. Network situation awareness technology is composed of factor acquisition, situation perception, situation prediction and supporting the administrators to response network security events quickly by visualizing the situation directly.

Wang Hui Qiang proposed NSAS（Network Situation Awareness System）framework based on Endsley module and JDL module [1]. He provided valuable experience for network security situation research and also proposed a breakthrough point for further research at the same time. But the association analysis between events is not enough in the framework, and the further improvement in visulization of situation data is needed.

Literature combined ARIMA model and BPNN model. Though the method took into account the time characteristics of time series data, those two models were blanced after simply put together [2]. The models themselves were not modified which limited prediction effect.

Literature proposed a network security situation prediction method based on generalized Radial Basis Function (RBF) neural network [3]. The method adopted the K-means clustering algorithm to determine the data center and employed the least mean square algorithm to adjust the weights. Literature proposed a prediction model of optimized RBF neural networks based on modified Artificial Bee Colony (ABC) algorithm [4]. The modified algorithm was used to confirm center and unit numbers of the hidden layers of the RBF neural network. But the RBF neural network still belongs to static neural network structure, so this algorithm didn't explore the time series characteristics of network security situation data.

Chen Tao set up the BP neural network to learn and predict the network performance [5]. In the papaer, some data were randomly selected as training data with some as validation. But the information of squences in time dimension was not fully utilized.

You Mayan presented a method to predict network security situation based on Elman neural network [6], but the network weights were generated randomly on the initial stage which became an unstable factor to the training of the model.

This paper presents a method to predict network security situation based on Elman neural network optimized by genetic algorithm (GA) according to the characteristics of nonlinear time series. After integrating the advantages of Elman network in processing time series, the global search ability is improvedin the genetic algorithm optimization stage.

2.Elman Neural Network

Elman network is a dynamic feedback neural network. The network consists of input layer, hidden layer, delay layer and output layer. As the paper wants to deal with daily alarm messages, one sample contains a short sequence data put into the ENN at a time. After processed by the nonlinear function of the hidden layer, the data of the input layer are transmitted to the delay layer. At the same time, the delay layer can reflect the memory of historical data, which helps to fit the time series better.

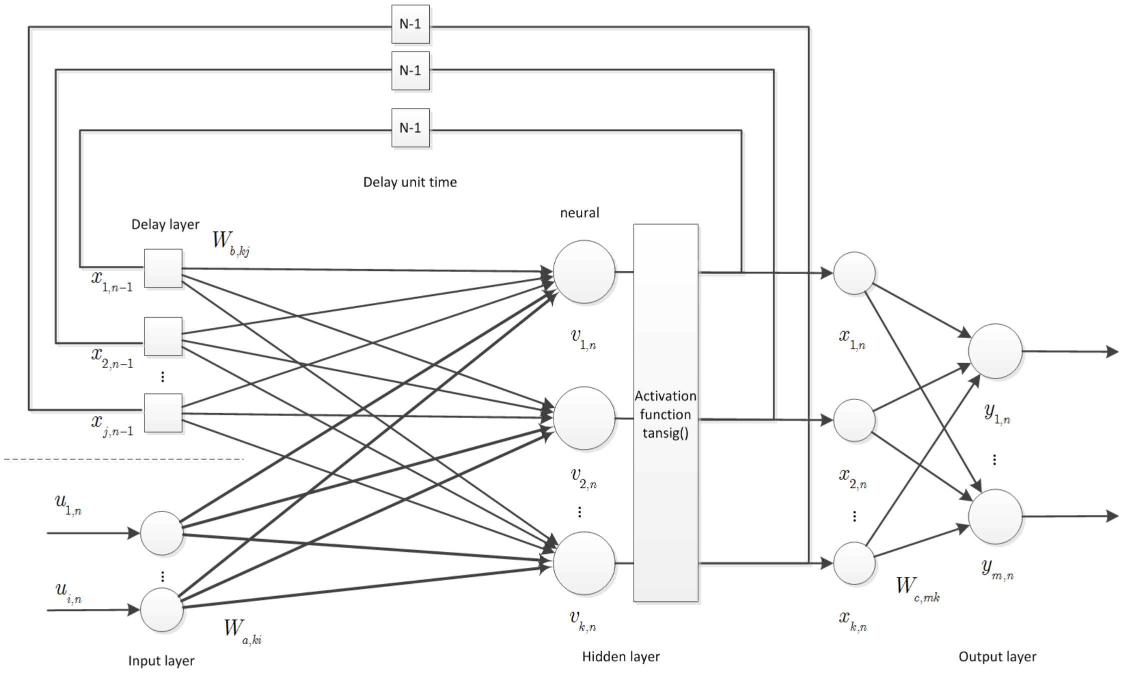

Figure 1: Elman Neural Network

$$v_{k,n}=\sum_{j} w_{a,kj} x_{j,n-1}+\sum_{i} w_{b,ki} u_{i,n} \tag{2.1}$$

$$x_{k,n}=tansig(v_{k,n})=\frac{1-e^{-2v_{k,n}}}{1+e^{-2v_{k,n}}} \tag{2.2}$$

$$y_{m,n}=\sum_{k} w_{a,mk} x_{k,n} \tag{2.3}$$

In Formula (2.1), $x_{j,n-1}$ is the j_{th} output of delay layer; $w_{a,kj}$ represents weight from the j_{th} neural in delay layer to the k_{th} neural in hidden layer; $u_{i,n}$ represents the i_{th} input in input layer; $w_{b,ki}$ represents the weights between i_{th} input and k_{th} neural; $v_{k,n}$ is the input of the k_{th} neural in hidden layer. In Formula (2.2), $tansig(v)$ is the activation function of hidden layer, and $x_{k,n}$ represents the kth output of hidden layer. Formula (2.3) represents the m_{th} output. For all the formulas, n or $n-1$ represents the literation of the algorithm.

3.Genetic Algorithm

The genetic algorithm is a randomly global search optimization method based on the law of biological evolution. It operates on the parameters itself directly and there is no need to require continuity and derivative of the function. It optimizes the parameters by probabilities, applied in a wide scope. The first step is the parameter coding, then the objective function is used as the search information during optimization. The selection, crossover and mutation of genetic algorithm are carried out in a certain probability to avoid falling into local minimum in conventional optimization algorithm. The combination of genetic algorithm and Elman neural network can speed up the convergence rate under the condition of global search.

In the initialization stage of Elman network, the method of using random weight matrix has some defects. On the one hand, a random way tends to make the results uncertain; on the other hand, the random matrix makes the training process uncertain. Particularly, the randomly generated initial maxtrix may cause the network training to fall into local minima.This paper uses the genetic algorithm to initialize the network, and then the network will be trained by back-propagation training algorithm with momentum and adaptive learning rate method.

3.1Initialize Weights with Genetic Algorithm

N is defined as the dimension of input layer; *K* is the number of hidden layer; *M* is the dimension of output layer.

(1) Coding

The coding is to transform the parameters into the search space that genetic algorithm operator can handle. As population *C* is set up, *L* is defined as the individual number and *D* is dimension of the chromosome. Finally, the population matrix comes out.

$$C_{l,d}=\begin{pmatrix} c_{1,1} & \cdots & c_{1,D} \\ \vdots & \ddots & \vdots \\ c_{L,1} & \cdots & c_{L,D} \end{pmatrix}=\left[C_{l,K\cdot N},C_{l,K\cdot K},C_{l,M\cdot K}\right] \tag{3.1}$$

The genetic algorithm uses fixed length binary string to represent the individual in the group. This paper considers gray code encoding principle to improve the local search ability of genetic algorithm.

(2) Fitness function

The fitness is an important factor to measure the probability of being eliminated. This paper chooses objective function as the fitness function. The error between expected output, which is defined as $y_{\exp}$, and predicted the output, which is y_{test} , will be optimized to minimum. is defined as the total error of each individual.

$$E_l=\sum_p^P\sum_m^M(y_{\exp,m}(p)-y_{test,m}(p))^2 \tag{3.2}$$

P is defined as the number of total samples, and p represents the p_{th} sample. $y_{\exp,m}(p)-y_{text,m}(p)$ is error between the p_{th} sample's expected result and test result. As the roulette algorithm is usually used in the selection process, which prefers to the individual whose fitness valueis bigger, so the new fitness function is as follows:

$$f_l=\frac{1}{E_l} \tag{3.3}$$

(3)Selection

The classical selection operators usually use roulette algorithm, that is to say, individuals with the smaller error have a greater probability of being retained. Considering the randomness of crossover, mutation and other operations, the best individuals in the existing group may be destroyed. As a result, the efficiency of the algorithm will be reduced and the overall quality of the population will be affected. Thus the modified process is as follows:

Step1: find the better initial fitness individual and preserve them to the next generationwithout crossover or mutation.

Step 2: screenthe best individual in the offspring, compare with the record and update the record with the better one.

The range of betterrecords can be adjusted appropriately depending on the quality of the population.If the quality is poor, the records should be reduced, so the high-quality individuals can expand rapidly；otherwise, expand the range.

(4) Crossover

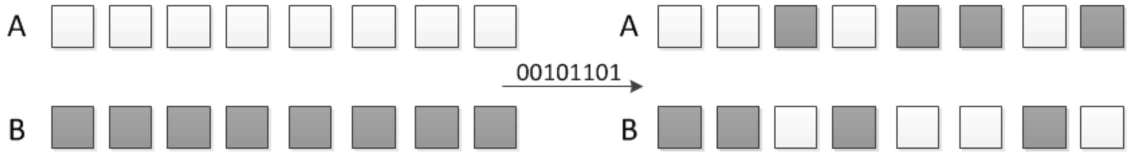

Figure 2: Crossover Operator

The crossover operator is the process of crossing the genetic information in a specific position with certain probability. As shown in Fig. 2, the crossover is the main form to generate new individuals in genetic algorithm, composed of cross probability and cross position. Firstly, a random number between 0 to 1 for each pair of chromosomes is generated. If the number is less than the crossover probability, cross it. In this paper, the method of uniform crossover is used to select the cross position. The specific location is determined by a binary string equal to the encoded string. The "1" bits on the randomly generated binary string are the corss position. The crossover popability is usually between 0.5 and 0.9.

(5) Mutation

The mutation is a change in some of the information on the chromosome ofsmall probability. In the process of crossing, as some high quality individuals can be lost inevitably, the mutation operator can enhance the global search ability to some extent as an auxiliary method, and avoid premature of the population. Like the crossover, the probability is used to determine whether to mutate, and determine the location of the mutation. Mutation is a rare phenomenon, so the rate is usually between 0.001 and 0.05.

Then the genetic algorithm will repeat selection, crossover and mutation process to the maximum number of iterations or adapt to meet the corresponding requirements.

4.Elman Neural Network based on GA

As the optimal individual is decodedfrom the genetic algorithm, the initial neural network is more close to the optimal solution than the totally random initial one. On the basis of the network, we can improve the stability of the random initialization network.

Step1: initialize fixed parameters of neural network. Set input layer number, output layer number, hidden layer neurons number. Initial weight matrix randomly, and normalize the training data to the range from -1 to 1.

$$x' = \frac{x - min(x)}{max(x) - min(x)} \cdot 2 - 1 \tag{4.1}$$

Step 2: check whether the chromosome meets the requirements. If so, end the iteration, and go to Step 7; otherwise continue iteration.

Step 3: encode the chromosome with gray code.

Step 4: use the optimal preservation strategy to select the best individuals of the population to be preserved.

Step 5: check whether or not the iteration reachs the maximum. If so, then end the genetic, go to step 7; otherwise, cross and mutate the remaining individual, and generate the next generation.

Step 6: decode chromosomes. Go to Step 2.

Step7: train the network by back-propagation training algorithm with momentum and adaptive learning rate method.

① Additional momentum.

On the basis of back propagation, a value proportional to the change of the previous weight is added to the current weight change. A new weight change is generated according to the back propagation.

$$\Delta w_{ji}(k+1)=(1-mc)\eta\delta_j y_j+mc\cdot\Delta w_{ji}(k) \tag{4.2}$$

Parameter k is the time of iteration; mc is momentum factor, which is usually about 0.95; η is learning rate; δ_j is local gradient; y_j is the input of the j_{th} neuron.

Condition to momentum method:

$$mc=\left\{\begin{array}{c} 0, E(k)>1.04E(k-1) \\ 0.95, E(k)\leqslant E(k-1) \\ mc\text{ , }otherwise \end{array}\right\} \tag{4.3}$$

$E(k)$ is sum of the squared errors of k_{th} iteration.

② Adaptive learning rate.

According to the training process, the learning rate can be adjusted automatically.

$$\eta(k+1)=\left\{\begin{array}{c} 1.05\eta(k), E(k+1)<E(k) \\ 0.7\eta(k), E(k+1)\leqslant 1.04E(k) \\ \eta(k), otherwise \end{array}\right\} \tag{4.4}$$

$E(k)$ is sum of the squared errors of k_{th} iteration.

Step 8: output the trained network.

5.Experiment Analysis

The experiment data originated from HoneyNet [7], which collected 77 days of 2000 alarm messages.The number of daily equipment alarm is defined as the network security situation value. The first 70 data are divided as training data and the last 7 as the test data. This paper uses 5 days of historical data to predict the value of the sixth day. The input dimension is 5 and the output dimension is 1.This ENN has simple three layers and one delay layer. With the genetic algorithm to optimize the Elman network to predict the last 7 days, the result is shown as Fig. 3.

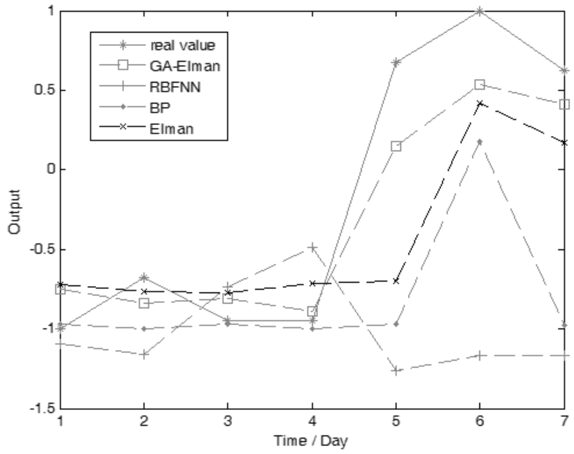

Figure 3: Prediction of Different Algorithms

Fig. 3 shows that the predicted value of GA-Elman algorithm is more close to the real value, indicating the method features a better effect on predicting network situation. In this paper, the mean square error (MSE) and the mean absolute error (MAE) are used to quantitatively evaluate the accuracy of the model.

MSE:

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left(y_{\exp,i}-y_{text,i}\right)^{2} \tag{5.1}$$

MAE:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|y_{\exp,i}-y_{text,i}\right| \tag{5.2}$$

$y_{\exp,i}$ is the expected value; $y_{test,i}$ is the predicted value; n is the sample number.

Algorithm	GA-Elman	RBFNN	BPNN	ELMAN
MSE	0.0928	1.7395	0.8678	0.3716
MAE	0.2582	1.0201	0.6426	0.4547

Table 1: Prediction Errors of Different Algorithms

As seen from Table 1, the MSE and the MAE of GA-Elman are less than RBFNN, BPNN, ELMAN, indicating that GA-Elman can output more precise results. With the improvement of the genetic algorithms, the effect of prediction has been promoted to a certain extent.

6.Conclusion

The GA-Elman model proposed in this paper can better predict the time series data, and it features the advantage of being closer to the final solution than the other models in the initial stage. Therefore, the results can be obtained more quickly in fulfillment of the requirements on the prediction of rapid changesin network situation. The number of network nodes can be dynamically adjusted according to the actual situation, but whether the model is suitable to large scale data flow still remains to be studiedf urther.

Reference

[1] Huiqiang Wang, Jibao Lai, Liang Zhu, Ying Liang. *Survery of Network Situation Awareness System*[J]. Computer Science, 2006,33(10):5-10.(InChinese)

[2] Jing Zhai, Jun Cao. *Combination forecasting model based on time series ARIMA and BP neural network*[J]. Statistics and decision making, 2016(4):29-32.(InChinese)

[3] Wanwan Lan, Limin Xue, Qinyu Zhao. *Preditction Method for Network Security Situation Based on Generalized RBF Neural Network*[J]. Command Information System and Technology, 2015,6(1):6-9.(InChinese)

[4] Wenming Huang, Shuangshuang Xu, Rongzhen Deng. *Shor-term traffic flow prediction of optimized RBF neural networks based on the modified ABC algorithm*[J]. Computer Engineering & Science, 2016, 38(4):713-319.(InChinese)

[5] Tao Chen, Zhenghu Gong, Ning Hu. *A prediction model of network situation based on proved BP algorithm*[J]. Proceedings of the 2009 National Conference on computer network and communication, 2009:93-99.(InChinese)

[6] Mayan You, Jie Ling, Yanjun Hao. *Prediction Method for Network Security Situation Based on Elman Neural Network*[J]. Computer Science, 2012, 39(6):61-63.(InChinese)

[7] Honeynet Project. *Know Your Enemy: Statics.* Http://old.honenet.org/papers/stats/