A Comparative Study of Several Different Neural Network Approaches for Information Security Modeling

Jiahui Xie* 3458463068@QQ.COM
School of Electronic and Information Engineering, University of Science and Technology Liaoning, Lishan District, Anshan, Liaoning, China

Junpeng Li 59991625@QQ.COM
School of Science, Dalian Maritime University, Ganjingzi District, Dalian, Liaoning, China

Ping Qiu QIU13842619917@ICLOUD.COM
School of Electronic and Information Engineering, University of Science and Technology Liaoning, Lishan District, Anshan, Liaoning, China
Corresponding author

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

In the rapid development of the Internet, people's lives have been deeply bound to the Internet, and the network information data is explosive growth. However, along with it, there is an increasingly serious problem of network information security. In order to achieve more accurate network information security classification judgment, we use BP neural network, RBF neural network, based on genetic algorithm optimization of RBF neural network three models to compare the information security model respectively, used to assess their ability to assess the information security risk (threatening, vulnerability, asset identification). The experimental results show that the RBF neural network optimized based on genetic algorithm has higher accuracy and lower error in information security risk assessment, which has significant advantages over the traditional neural network and provides a strong basis for improving the level of information security protection and selecting the best neural network model.

Keywords: Genetic Algorithm, RBF Neural Network, BP Neural Network, Information Security.

1. Introduction

In the wave of digitalization, information security is crucial, and it is imperative to build an efficient security model. Traditional security protection often reveals problems such as low detection accuracy, high false alarm rate, and weak generalization ability when facing complex and changing network attacks (Wang and Lu, 2019).

BP neural network has some applications in the field of information security by virtue of its powerful nonlinear mapping ability, but there are shortcomings that are easy to fall into local optimization and slow convergence speed.RBF neural network with its unique radial basis function, fast learning speed and can effectively avoid the local optimization, but lack of adaptivity in the selection of parameters. Genetic algorithm has global optimization characteristics, based on the genetic algorithm optimization of RBF neural network, is expected to improve the RBF parameter determination difficulties, to improve the performance of the information security model (Gao and Shan, 2024; Li, 2018). The purpose of this paper is to study the performance of these three types of neural networks in the information security model in depth comparison, to provide theoretical support and practical reference for the construction of better information security protection system (Johan et al., 2024; Yao and Zhang, 2024).

2. Introduction to Several Different Neural Networks

2.1. RBF Neural Network

RBF neural network is a feed-forward three-layer neural network based on radial basis function, which is significant in the fields of time prediction, feature extraction and regression analysis. It contains an input layer, a hidden layer and an output layer. The input layer is responsible for receiving external signals, the hidden layer applies a suitable radial basis function to nonlinearly transform the input, and the output layer weights and sums the data coming from the hidden layer, and completes the linear transformation from the space of the hidden layer to the space of the output layer through the activation function. Its learning process covers initialization, forward propagation, error backpropagation and parameter updating, and usually adopts the radial basis function as the activation function.RBF neural networks rely on multiple radial basis function curves to approximate the target curve, and by adjusting the number and width of the function, the target curve can be fitted with high accuracy, and its local characteristics can automatically capture the key features of the data, enhancing the model generalization ability.

2.2. BP Neural Network

BP neural network is a widely used artificial neural network. It is a multi-layer feed forward neural network where the signal is passed forward from the input layer, processed in the hidden layer and finally the result is obtained in the output layer. The neurons are connected to each other by weights. In the training process, forward propagation is performed first and the input data is passed through the network to calculate the output value. If the output deviates from the expectation, backward propagation is started. The error is passed back from the output layer in reverse, and the connection weights between the neurons are adjusted according to the error to reduce the error.

2.3. Optimization of RBF Neural Network Model Based on Genetic Algorithm

Coding Phase. The chromosomes are encoded using binary coding technique which divides each chromosome into four key parts. Weights from the input layer to the hidden layer, thresholds for the hidden layer, weights from the hidden layer to the output layer, and thresholds for the output layer. This encoding ensures that the chromosomes represent the parameters of the neural network completely.

Calculate the Fitness. The weights and thresholds represented by the selected chromosomes are applied to the RBF neural network and the network is trained using the training set data to produce the predicted output. Subsequently, the sum of the squares of the errors between the predicted and actual values is calculated, the reciprocal of which is defined as the fitness score for that chromosome. This process ensures that the fitness assessment reflects the quality of the solution encoded by the chromosome. The formula for the fitness is as follows

Selection Phase. In the information security risk assessment model, we use the roulette wheel selection method, where each chromosome enters the selection phase with a probability P_i, P_i of the formula.

$$P_i = \frac{f_i}{\sum_{i=1}^{m} f_i} \tag{1}$$

Where, f_i is the fitness of the chromosome i and m is the number of chromosomes in the population.

Crossover Phase. In performing crossover operation, we used arithmetic crossover technique. This technique generates new offspring chromosomes by performing linear mixing of a pair of chromosomes. The process by which crossover occurs is as follows.

$$\begin{cases} p_{mi} = p_{mi}\left(1-a\right) + p_{ni}a \\ p_{ni} = p_{ni}\left(1-a\right) + p_{mi}a \end{cases} \tag{2}$$

Where, a is a random number in the interval [0,1].

Mutation Phase. At the time of chromosomal gene mutation, we mutate the jth gene of the *i*th chromosome and the mutation process is shown.

$$p_{ij} = \begin{cases} p_{ij} + \left(p_{ij} - p_{\max}\right) * F\left(x\right) a > 0.5 \\ p_{ij} + \left(p_{\min} - p_{ij}\right) * F\left(x\right) a \leq 0.5 \end{cases} \tag{3}$$

Where, $p_{\max}$, $p_{\min}$ are the upper and lower gene p_{ij} bounds, respectively, a is a random number in the interval [0,1], $F\left(x\right) = b \cdot \left(1\text{-g}/\text{G}_{\max}\right)^2$, b is a random number, is the number of iterations currently iterated, and $G_{\max}$ is the maximum number of genetic generations.

3. A Comparative Study of Several Different Neural Network Approaches for Information Security Modeling

3.1. Data Sources of Information Security Risk Assessment Samples

Information security risk factor analysis usually relies on a large number of samples, but because information security involves state secrets, it is extremely difficult to obtain samples. In this study, we selected the source data in the information security risk assessment based on RBF fuzzy neural network (Ruan and Dang, 2011), and used MATLAB to carry out simulation experiments based on the genetic algorithm of RBF neural network, RBF neural network, and BP neural network (Li et al., 2021). The experimental results show that the RBF network optimized with genetic algorithm can effectively assess the information security model, the first 15 groups of learning data, and the last 5 groups of test data, the specific data are shown in Table 1. the data have been normalized, and the data in the table, X1, X2, X3, indicate the three major risk factors of threat identification, vulnerability identification, and asset identification.

3.2. Selection of Input, Hidden and Output Layers

The RBF neural network of this experiment is a 3-layer neural network, the accuracy is set to 0.0001, the maximum number of nodes in the hidden layer is 15, the hidden layer function is a Gaussian function, the input layer is three neurons, and the output layer is one neuron.

The BP neural network in this experiment is a 3-layer neural network model. After calculating the hidden layer neurons are three, thus the model of BP network is determined. Our this simulation experiment is done through Matlab by calling the neural network toolbox. There are 3 risk factors in this model, so the input neurons are set to 3 and the output layer is 1. The neural network units are entered as per the data inputs of the risk factors given in Table 1 and after repeated training, we found that the hidden units are set to 5 for the best experimental results (Chen et al., 2018).

The genetic algorithm based optimization RBF neural network of this experiment, with the help of GA to optimize the weights of the hidden layer to the output layer, as well as the center width vector of the Gaussian basis function, as a way to improve the training effect of the information

Table 1: Normalized data for information security risk assessment samples

sample number	X_1	X_2	X_3	Y (Assessment results)
1	0.3	0.2	0.9	0.11
2	0.9	0.8	0.9	0.72
3	0.4	0.2	0.7	0.11
4	0.5	0.9	0.9	0.61
5	0.4	0.3	0.7	0.16
6	0.9	0.7	0.8	0.59
7	0.2	0.5	0.6	0.17
8	0.9	0.6	0.8	0.51
9	0.7	0.3	0.5	0.18
10	0.4	0.8	0.9	0.48
11	0.8	0.3	0.6	0.21
12	0.7	0.6	0.5	0.35
13	0.5	0.4	0.6	0.22
14	0.6	0.4	0.8	0.28
15	0.6	0.6	0.5	0.33
16	0.7	0.4	0.6	0.2591
17	0.3	0.6	0.9	0.3228
18	0.1	0.4	0.7	0.1085
19	0.8	0.4	0.3	0.1959
20	0.5	0.6	0.3	0.3794

security risk assessment prediction. When the population of the genetic algorithm tends to be stable and the fitness value meets the evolutionary standard of the overall genetic algorithm, the optimized parameters are mapped and converted into the new network structure parameters of the RBF, and ultimately the high-precision prediction of information security risk assessment is realized (Liao et al., 2023; Liu and Wang, 2024).

First of all, for the activation function based on the data samples for the neurons of the hidden layer of the RBF neural network is used is a Gaussian function, the formula is as follows:

$$\phi\left(\|x-t\|\right)=e^{\frac{-\|x-t\|^2}{2\sigma^2}} \tag{4}$$

Where, σ represents the center width vector of Gaussian basis function, x is the input sample, and t is the center point.

Secondly, the RBF neural network is a linear neuron output layer, and the weights from the hidden layer to the output layer are designed using the least squares method, which is calculated as.

$$w_{ij}=\exp(\frac{m}{C_{\max}}(||x-c_j||)^2) \tag{5}$$

Where, $C_{\max}$ is the maximum distance between the centers and w_{ij} is the weights from the hidden layer to the output layer.

3.3. Comparison of the Results of Optimizing RBF Neural Network with BP Neural Network and RBF Neural Network Based on Genetic Algorithm

In the GA-RBF neural network constructed in this paper, after fitting by MATLAB, it is calculated that the initial population in the genetic algorithm is 30, the maximum number of hereditary generations is 300, the optimal crossover probability is 0.65, and the variance probability is 0.24. The variance of the Gaussian basis function in the hidden layer is optimal.

The specific prediction result data for the comparison of the three types of algorithms are shown in Table 2, and the specific comparison graph is shown in Figure 1.

Table 2: Three types of modeling algorithms for information security risk factor assessment target values

sample number	real value	GA-RBF	MAE	RBF	MAE	BP	MAE
1	0.2591	0.2515	0.0076	0.2620	0.0021	0.2404	0.0187
2	0.3228	0.3346	0.0124	0.3114	0.0112	0.2998	0.0239
3	0.1085	0.1119	0.0038	0.1032	0.0051	0.1212	0.0121
4	0.1959	0.2021	0.0057	0.1677	0.0289	0.2185	0.0220
5	0.3794	0.3840	0.0049	0.3528	0.0258	0.3609	0.0184

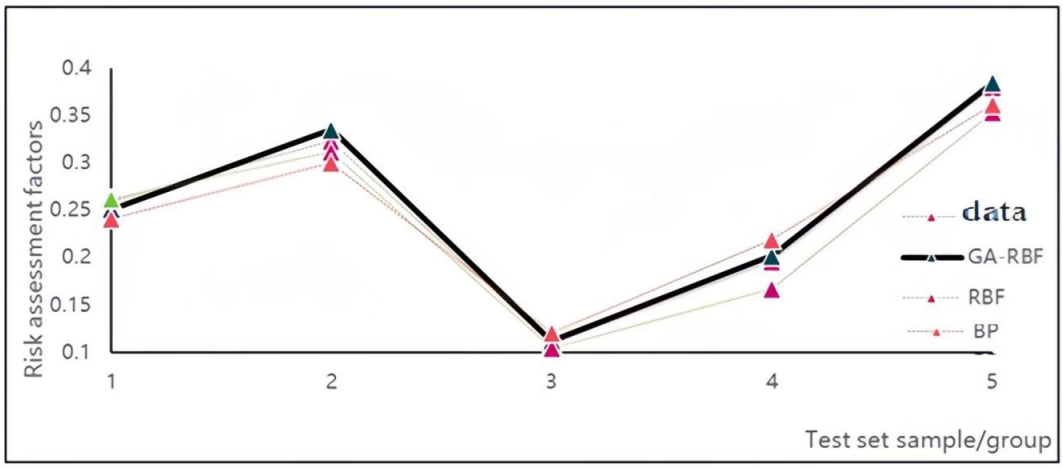

Figure 1: Comparison of the prediction results of the three types of algorithmic models

Comprehensive comparison of the above experimental results can be seen, as shown in Fig. 1, in which the GA-RBF fitting curve and the fitting curve of the true value almost show the phenomenon of overlap, while the fitting curve of the BP and RBF neural networks deviates from the true value compared with the GA-RBF.

We also made a corresponding comparison for the performance of the algorithms at the same time, due to the uncertainty of the neural network, all take the MSE and MAE mean values between the predicted and real values of 20 experiments, and the relevant data are shown in Table 3.

Table 3: Three types of modeling algorithms for information security risk factor assessment error

arithmetic	MSE	MAE
RBF	0.0332%	0.0145
BP	0.0374%	0.0189
GA-RBF	0.0056%	0.0069

By comparing the MSE and MAE of BP, RBF and GA-RBF algorithms, it is found that GA-RBF algorithm has significant advantages. Compared with traditional RBF neural network, GA-RBF has 0.0276% lower mean square error and 0.0076 lower mean absolute error.The error analysis shows that the GA-RBF algorithm predicts values closer to the true values. Compared with RBF and BP neural networks, the prediction effect of RBF neural network based on genetic algorithm is better.

Comprehensive comparison of the above experimental results shows that the MSE and MAE errors of the GA-RBF algorithm are smaller than those of the BP and RBF algorithms, and the image fitting of the predicted values and the real values is better, and this study illustrates that the prediction effect of the GA-RBF algorithm for the experimental data of the information security modeling study is superior to that of traditional BP and RBF algorithms on the whole.

4. Conclusion

In this paper, BP neural network, RBF neural network, and RBF neural network optimized based on genetic algorithm (GA - RBF) are used to evaluate the information security model. By choosing different genetic algorithms and optimizing relevant parameters, the optimized model is used to validate the prediction, and the results show that the prediction index of the test set is better than the training set. Compared with BP and RBF neural network models, GA - RBF neural network model predicts information security factors with significant advantages and smaller errors. Effective risk assessment can clarify the security needs of information systems, grasp the security situation, help build a protection system, and improve the effectiveness and efficiency of protection.

References

Xingyu Chen, Huanxie Chen, Can Wang, and et al. Research on information security risk assessment based on bp neural network. *Information Communication*, (04):165–166, 2018.

Yu Gao and Fangfang Shan. Research on the construction of information security risk assessment model and index system based on improved neural network. *Journal of Jiamusi University (Natural Science Edition)*, 42(02):28–31, 2024.

W.D. Johan, P. Wolter, and G.V. Pieter. Bias and noise in security risk assessments, an empirical study on the information position and confidence of security professionals. *Security Journal*, 37 (1):170–191, 2024.

Jiawei Li, Kehe Wu, and Bo Zhang. A power grid information security risk assessment model based on small habitat genetic algorithm. *Electric Power Construction*, 42(03):89–96, 2021.

Senyu Li. Research on information security risk assessment method based on improved neural network, 2018.

Ruiquan Liao, Longwei Li, Wei Wang, Bin Ma, and Yuan Pan. Optimization of ga-based rbf neural network gas-liquid two-phase flow holding rate prediction model. *Journal of Changjiang University*, (12):01–05, 2023.

Hailin Liu and Tingyou Wang. Improved ga-rbf neural network for water plant coagulation dosing prediction. *Water Supply Technology*, 18(01):40–45, 2024.

Hui Ruan and Depeng Dang. Information security risk assessment based on rbf fuzzy neural network. *Computer Engineering and Design*, 32(06):2113–2115, 2128, 2011.

Minjie Wang and Enqiong Lu. An overview of information security risk assessment. *Communication Management and Technology*, (06):54–55, 59, 2019.

J. Yao and T. Zhang. Side-channel attack security risk assessment model based on mutual information game. *Journal of Physics: Conference Series*, 2732(1), 2024.