International Journal of Security and Its Applications
Vol. 10, No. 12 (2016), pp.377-388
http://dx.doi.org/10.14257/ijsia.2016.10.12.31

GABP Neural Network Algorithm Applied in Evaluation of Computer Network Security

Xiaochun He*, JunJun Feng, Ruchun Jia

School of computer Information, Sichuan Information Technology College,
Guangyuan, Sichuan 628017, China
E-mail: hxc108@126.com

Abstract

In this paper, in order to assess the risk of network, network security assessment process being involved in the content in detail. The above-mentioned research-based support system platform security test and evaluate research of the safety situation assessment. Prediction subsystem detailed design and carry out the implementation. In this paper, network security issues, as a detailed study of neural networks knowledge. Focus on the evaluation methods and calculation rules of nerve network technology, it has been studied by specific examples. Calculation demonstrated the feasibility of neural network evaluation model through actual case, which pointed out the traditional methods. This paper focuses on the network security assessment based on neural network technology, extensive analysis of the proposed major modeling tool indicator system for network security analysis. The application of neural networks was a network security assessment and to optimize the network by genetic algorithm. The key parameter combination operated efficiency of neural networks to get better play.

Keywords: *Neural networks; improved genetic algorithm; network security assessment; complex network*

1. Introduction

Network made the size of Internet continued to expand its applications which are constantly expanding. The network has gradually penetrated into people's daily life and economic related fields, as well as military education and science and technology [1, 2]. It is not only the role gradually increased, and global basis and its position has also been strengthened [2]. Due to the continuous development of network technology, network security problems have severely affected the socio-economic development and national development strategies [4]. Information systems currently face serious security risks and threats, as many invaders spotted network structure complex and large-scale defects, using a variety of systems security shortcomings of the new means of carrying out continuous attacks [5].

Key finance, defense and government departments as well as e-commerce and other business organizations are important large-scale integration into the Internet. They are also increasingly become the target of attack. According to the latest statistics, a variety of network security incidents significant growth trend were tested annually. With the large increase in the presence of a large number of network users and network resources, as well as a variety of loopholes in the system and continue to find, so that network security issues become more complex, and showed a new trend [6-9]: (1) uncover vulnerabilities increasingly the faster, the coverage more widely; (2) more sophisticated attack tools; (2) the degree of automation and attack speed, gradually increase the lethality; (4) increasing asymmetric threats; (5) more firewall to higher permeability; (6) infrastructure will form a growing threat. Coupled with the increasing complexity of network attacks,

ISSN: 1738-9976 IJSIA

various methods are blended so that network security defense more difficult [10]. Therefore, the study has important practical significance based computer network security systems [11].

Constantly improve the intrusion detection system will be important to ensure network security. In this paper, in order to assess the risk of the network will be the starting point of view, network security assessment process being involved in the content in detail, and as a basis for research of network security situational awareness. And the above-mentioned research-based support system platform security testing and evaluation of research on the safety situation assessment and prediction subsystem detailed design and carry out the implementation. In this paper, network security issues, a detailed study of neural networks knowledge. Focus on the evaluation methods and calculation rules, nerve network technology has been studied by specific examples of calculation demonstrate. The feasibility of neural network evaluation model through actual case pointed out the traditional methods' limitations.

2. Basic Theoretical of Information Security Assessments

2.1 Characteristics of Complex Network Metrics

Clustering coefficient is called clustering coefficient, it measure extent network group and crucial parameter [12, 13]. In terms of social networks, the group is a very important form of clustering coefficient characteristics, the group is also known as the cluster phenomenon, for which scientists are clustering coefficient of concepts presented [14]. Node i is its clustering coefficient Ci describes the network into the node and inter-connected nodes directly connected relationship. Therefore, node i between its clustering coefficient Ci also refers to this node and its neighboring nodes stored in the number of edges which may exist with more than the maximum number of edges, Ci its expressions is:

$$C_i = 2e_i / k_i \left(k_i - 1\right) \tag{1}$$

In the formula ki is usually expressed as the i-node, ei indicates the number of edges in fact exist between the i-node its neighbor. Network clustering coefficient C is the average of the entire node clustering coefficient prescribing, its expression formula C can be expressed as:

$$C = \frac{1}{N} \sum\nolimits_{i=1}^{N} C_i \tag{2}$$

Among this, N refers to the network of its order.

Average path length (APL) is a network into another key feature of a comparison metric, which generally refers to the average shortest distance among all pairs of nodes of the network. In this process, the distance between nodes is from a node to another node experienced side with a minimum number, during which the largest among all the nodes on the network diameter of distance called. Transmission efficiency and performance metrics to measure the diameter and length of the network is the average path in the network. Wherein the average path length of the calculation formula is:

$$APL = \frac{1}{N\left(N-1\right)} \sum_{i \neq j \in V} D_{ij} \tag{2}$$

Where j is the shortest distance between nodes i and j. Degree of correlation which refers to the relationship among different network nodes interconnected described. If a greater degree of connection nodes tend to the earth node, said network is a positive correlation, whereas the network is said to be negative correlation. Only the vertices of their Pearson correlation coefficient r (-1<r<1) is calculated on the network will be able to describe their degree of correlation, r is defined as:

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}\left(j_i + k_i\right)\right]^2}{M^{-1}\sum_i \frac{1}{2}\left(j_i^2 + k_i^2\right) -- \left[M^{-1}\sum_i \frac{1}{2}\left(j_i + k_i\right)\right]^2} \tag{4}$$

2.2 Risk Factors

Risk assessment mainly around the basic elements of security, vulnerabilities, threats, assets and expanded in the process of assessment. To the business strategy, the value of assets, security requirements, residual risk, security events and various essential elements of its related to fully consider all types of property. Figure 1 is a risk assessment of the relationship among the various elements, in partial block diagram of a risk assessment is the basic element which is substantially elliptical partial elements of its associated attributes.

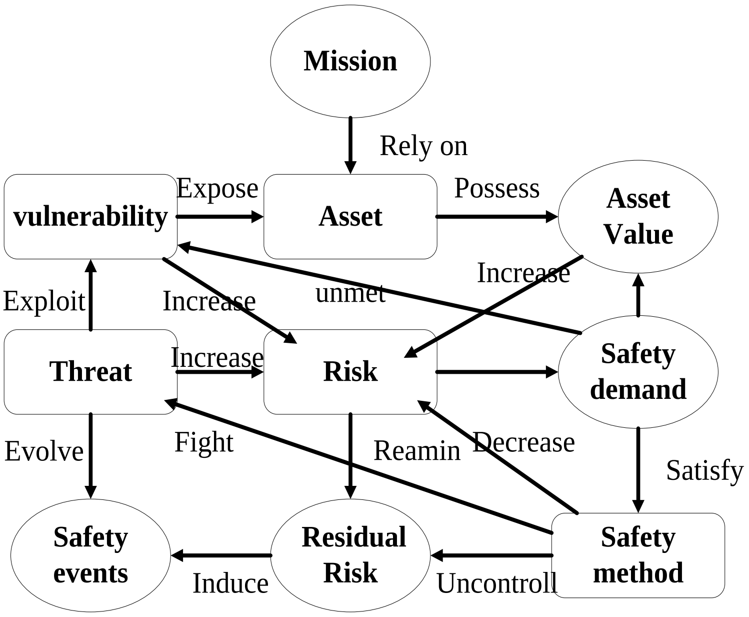


Figure 1. Elements of the Relationship between Risk Assessments

In the process, the risk of the preparation process is to assess the effectiveness of risk guarantee. Identifying asset identification, threat and vulnerability identification is the foundation of its risk assessment. They provide basic ground test data to calculate the risk, under the protection of the integrity. Accuracy of the test data can only be provided on the network information system to their real value at risk calculation to accurately assess the risks. The risk calculated its risk assessment of the critical content. It is a result of evaluation based on lies.

2.2 Genetic Algorithm and BP Neural Networks

Neural network can be directly input data and learning. In the learning process, it can adaptively found embedded in the sample data. The regularity inherent characteristics of the object be processed in the distribution of sample space without having to make any assumptions, but to learn between the sample directly from the data relations. Thus it can solve the problem because they do not know who to identify the sample distribution which can not be resolved.

BP neural network is currently the most widely used network, usually by an input layer. A number of hidden layer and output layer, the network structure was shown in Figure 2. Theoretical proof, for any continuous function on a closed interval can be used only one hidden layer of BP network to approximate, so a three BP network can achieve any given n-dimensional to m-dimensional mapping.

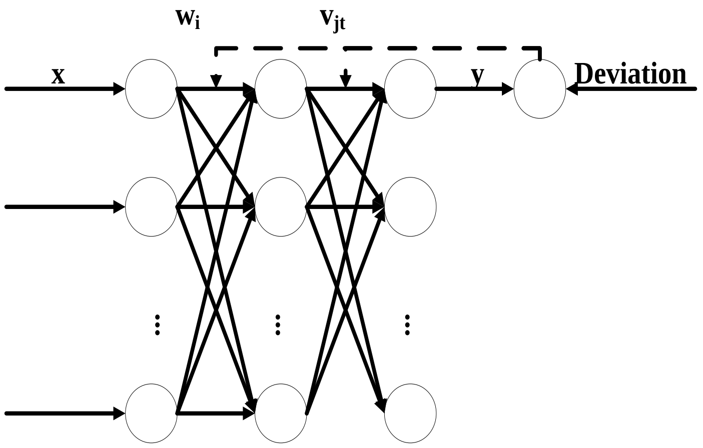


Figure 2. BP Neural Network Model

GA is not a simple comparison of random search algorithm, through fitness assessment of the role of chromosomes and chromosome gene. The effective use of existing information to guide the search for the most optimized state wants to improve the quality. The basic genetic algorithm follows the flow chart was shown in Figure 3.

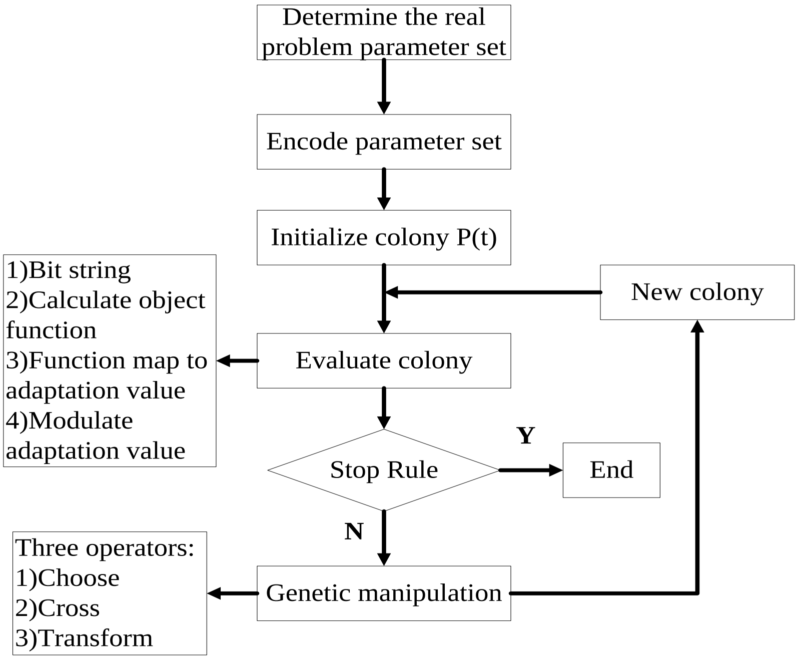


Figure 3. Flow Chart of Simple Genetic Algorithm

2. Complex Network Security Evaluation Modeling based on GABP Algorithm

2.1 Improved Modeling of GABP Algorithm

Genetic algorithm combined with BP neural network optimization to the basic idea: BP algorithm which depends on gradient information to the guidance into changes to adjust the network weights their method. Genetic algorithm its global searches and other characteristics of use, from optimal network structure and network connection weights were looking for. Because hidden layer and output layer to the input layer of the neural network of three layers, but the sample modeling determined the input layer and output layer node number, and therefore when BP network to optimize the structure, it should be hidden node the number of optimization.

Optimized genetic-neural network mathematical descript its main problems are as follows:

$$\begin{cases} \min E\left(w,v,\theta,r\right)=\dfrac{1}{2}\sum_{k=1}^{N_1}\sum_{t=1}^{n}\left[y_k\left(t\right)-y_k\left(t\right)\right]^2 \\ s.t\ \ w\in R^{m\times p},v\in R^{p\times n},\theta\in R^p,\mathrm{r}\in R^n \end{cases} \tag{5}$$

Because genetic algorithm optimization processes the objective function to its maximum value as a function of its fitness, so the fitness function is defined as:

$$F\left(w,v,\theta,r\right)=\frac{1}{\sqrt{\sum_{k=1}^{N_1}\sum_{t=1}^{n}\left[y_k\left(t\right)-y_k\left(t\right)\right]^2}} \tag{6}$$

$$\begin{cases} \max \quad F\left(w,v,\theta,r\right) \\ s.t\ \ w\in R^{m\times p},v\in R^{p\times n},\theta\in R^p,\mathrm{r}\in R^n \end{cases} \tag{7}$$

Control code which is mainly used to control the number of its hidden nodes, which is strung by 0-1 composed, in this being 0 means no connection, and a means is connected, its long string 11. It can be entered by the node 0.5 and 1.5 times the number to be determined. The weights of the digital system are mainly used to connect to the network right to control use of floating point encoding. The string length l_2 = m x l_1+l_1+l_1 x n (here m is the number of its input node, n, compared with its output node a number). Encoded according to a certain order level and became linked long strings, each string corresponds to a set of connection weights and network structure of the network. We have three cases input node as a word, then the hidden layer to only a maximum of six points. Figure 4 is a graphical diagram of the network structure.

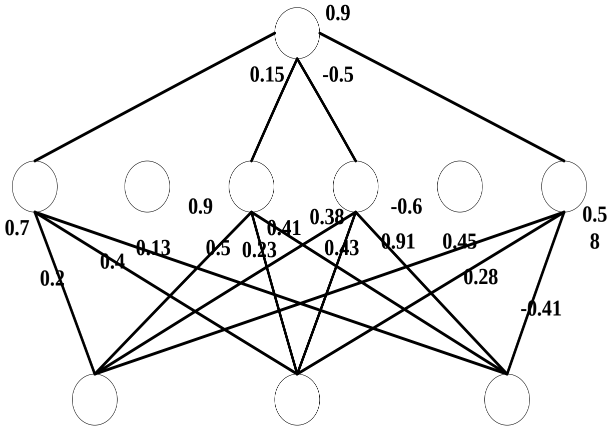


Figure 4. Factor Structure of the Network with Neighboring Rights

2.2 Lateral Distribution Structure of the System

"Security zone" should be the best method of dividing the physical and logical combination of approaches. In particular, many cases distinguished between physical, distinction is often closely related to logic. Due to the complexity of the system, sometimes inclusive relationship between domains, which is a big security domain may contain further subdivided into several smaller security domains. Sometimes there are relations between the security crossing that part of the system components may belong to two or more security domains. It can be given security domain defined below.

Figure 5 shows an abstract representation of information system security domains. Figure 5 region dotted line indicates the security domain, which contains an organization consisting of WAN. The security domain contains three sub-security domains D1, D2, D2, which are located in different geographical locations.

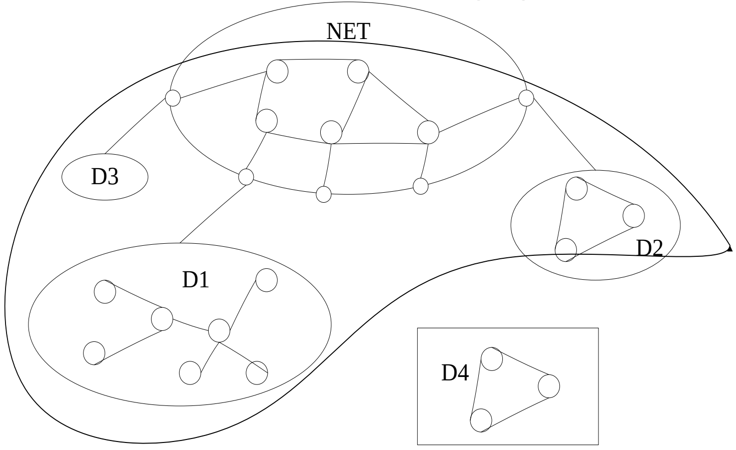


Figure 5. Security Domain Model

IT security research can be from the following four aspects: security domain; inter-domain security; intermediate domain security; extraterritorial protection.

4. Experiment and Results

4.1 Data Processing

Neural network weights and thresholds unify the threshold seen as input for the connection weights. Neural network weights and thresholds of binary encoding, 70 and 11 weight threshold value corresponding to 0/1 series together to give a long binary string, a chain gene (chromosome). As shown in Figure 6, it shows the value combination of the network, that is an individual. Population size is set to N, a randomly generated initial population.

With the corresponding decoding method, N to N individuals decoded set of network weights to give the N network having the same structure. The given input and output sample set is divided into training and testing samples, with improved BP algorithm. Since adapt the learning rate momentum gradient descent method to this N network set separately weights training, group obtained N network weights corresponding to the N network output, if after this set of N weights after training at least has a group to meet the accuracy requirements, then end algorithms.

This paper presents an improved genetic neural network-based intrusion detection model. The model of the main modules: network data acquisition module, data preprocessing module, feature extraction module and genetic neural network intrusion detection module of the main ideas and ways to achieve its described in detail. Each data set contains 41 data characteristic attributes (dimension), the end of the data type of attack is described. Typical data format is as follows:

0,tcp,discard,RSTO,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,226,9,0.00,0.00,1.00,1.00,0.04,0.07,0.00,255, 9,0.04,0.07,0.00,0.00,0.00,0.00,1.00,1.00,neptune.

0,tcp,finger,S0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,77,2,1.00,1.00,0.00,0.00,0.01,0.07, 1.00,1,6,1.00,0.00,1.00,0.50,1.00,0.17,0.00,0.00,land.

0,icmp,ecr_i,SF,520,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,472,472,0.00,0.00,0.00,0.00,1.00,0.00,0.00,255,255,1.00,0.00,1.00,0.00,0.00,0.00,0.00,0.00,smurf.

Since 2, 2, 4-dimensional data are non-numerical form each for subsequent experiments to identify needs, must be numerical. Each content dimension appears to statistics, in alphabetical order, and the serial number instead of the original content, numerical processing, the above data is converted to:

0,2,7,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,226,9,0.00,0.00,1.00,1.00,0.04,0.07,0.00,255,9,0.04,0.07,0.00,0.00,0.00,0.00,1.00,1.00,neptune.

0,2,16,6,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,77,2,1.00,1.00,0.00,0.00,0.01,0.07,1.00,1, 6,1.00,0.00,1.00,0.50,1.00,0.17,0.00,0.00,land.

0,1,12,10,520,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,472,472,0.00,0.00,0.00,0.00,1.00,0.00, 0.00,255,255,1.00,0.00,1.00,0.00,0.00,0.00,0.00,0.00,smurf.

The numerical data retention after 41 or numerically, a 4: 1 ratio, respectively, from the normal data and attack data were randomly selected as training data, press 1: 1 ratio of normal data and attack again from randomly selected training data 1/5 of the amount of data as the test data, and based on the training data, the number of data structure t.txt (by a number of 0 and 1, 0 corresponds to normal data, a data corresponding to the attack), calls for subsequent experiments.

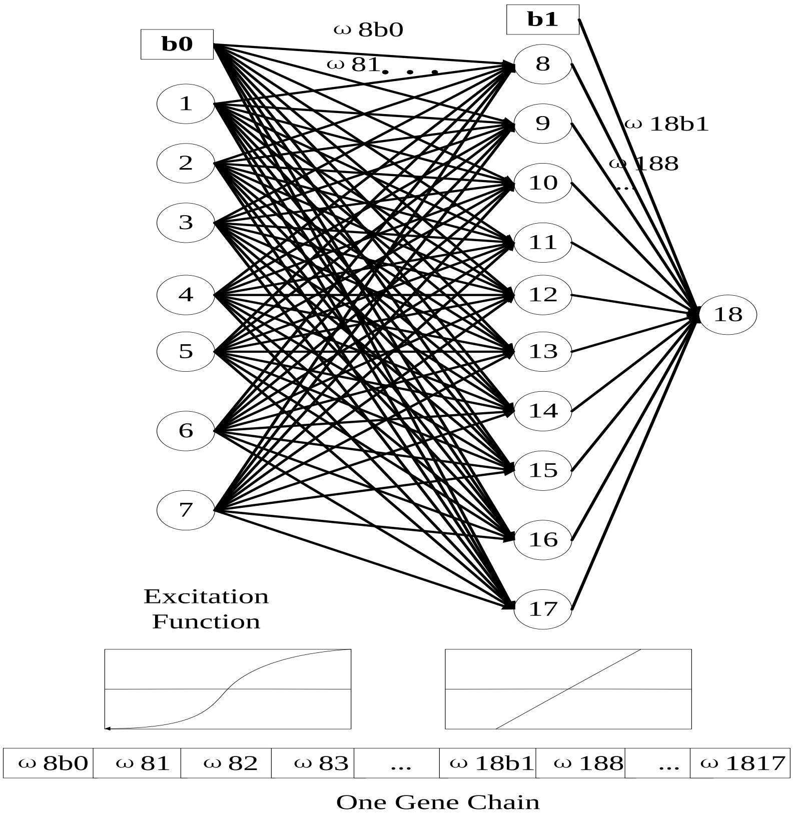
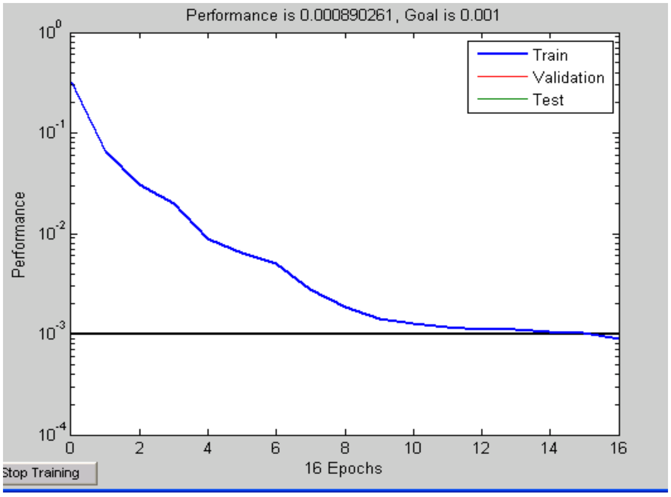


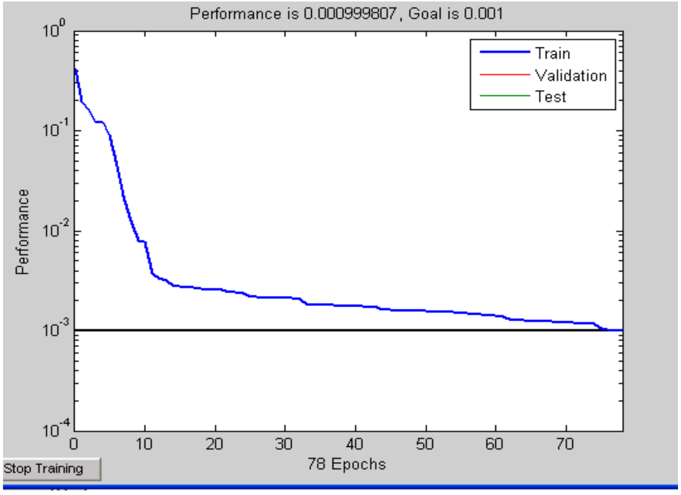
Figure 6. Neural Network Weight-codings

4.2 Analysis of Results

Improved algorithm train the neural network of GABP used matlab in simulation experiments. With smurf (DOS) attack test data, genetic algorithm squared error and curves, curves fitness training goals and BP curves are shown in Figure 7 (a) and 7 (b).



(a) Training goal curvilinear (smurf)



(b) Training goal curvilinear (neptune)

Figure 7. Curves Fitness Training Goals and BP Curves

When blended attacks against training data convergence maximum number of steps required, the training takes the longest time. This attack mixed with data containing many different types of attack data, which contain information about the complex. And with a single attack, we can also find training time and number of steps required for convergence is not the same. This reflects the different types of attacks in the amount of information related to the different experimental data.

4.2 Comparison GABP Algorithms and Several other Models

In addition to the neural network, we also examined the use of several other commonly used safety assessment models such as Logit model, nearest neighbor, LDA model and decision tree model. It is found the accuracy of these methods were lower than genetic neural network assessment models, the specific results were shown in Table 1:

Table 1. Comparison of GABP Model Prediction Structure with other Types of Models

Model	Training Samples			Test Samples		
	Accuracy	1st error	2nd error	Accuracy	1st error	2nd error
Logit	72.19%	9.26%	4%	72.22%	17.17%	9.01%
Nearest Neighbor	77.22%	4.02%	6.22%	70.52%	12.42%	10.22%
LDA	92.24%	2.7%	2.7%	75.26%	12.27%	12.25%
Decision Tree	97.2%	4.05%	0%	72.72%	15.90%	12.0%
GA-BP	100%	0	0	76.4%	4.5%	10%

Test samples from the overall prediction accuracy point of view. GABP prediction model is higher than the Logit model and decision tree model 9.26%, 4% higher than the LDA model, described GABP model prediction ability to better than the other four models. In addition, the proportion of two types of errors from the point of view, GABP model Type I error rate is the lowest in five models, respectively, compared with Logit model fell by 9.01%; down by 7.77% than the LDA model; ratio decision tree model decreased by 0.40%. The first reduce the error rate for the credit risk assessment is a very practical significance and value of the advantages, in terms of time in the field of view, GABP credit risk prediction model also has certain advantages.

5. Conclusions

Constantly improve the intrusion detection system will be important to ensure network security. In this paper, in order to assess the risk of the network will be the starting point of view, network security assessment process being involved in the content in detail, and as a basis for research of network security situational awareness. The above-mentioned research-based support system platform security testing and evaluation of research on the safety situation assessment and prediction subsystem detailed design and carry out the implementation. This paper focuses on the network security assessment based on neural network technology. The application of neural networks was a network security assessment and to optimize the network by genetic algorithm the key parameter combination, so that the operation efficiency of neural networks get better play.

References

[1] Luo, Y., & Jackson, D. O. CEO compensation, expropriation, and the balance of power among large shareholders. Advances in Financial Economics, 15, **(2012)**, pp. 195-227.

[2] Xiao, L., Wang, J., Yang, X., & Xiao, L. A hybrid model based on data preprocessing for electrical power forecasting. International Journal of Electrical Power & Energy Systems, 64, **(2015)**, pp. 211-227.

[3] G. Szekely, I. B. Valtcheva, J. F. Kim and A. G. Livingston, React. Funct. Polym, 2014 DOI:10.1016/j.reactfunctpolym.**(2014)**.02.007.

[4] Kim, M. K. Short-term price forecasting of Nordic power market by combination Levenberg–Marquardt and Cuckoo search algorithms. IET Generation, Transmission & Distribution, 9(12), **(2015)**, pp. 1552-1562.

[5] Zhang J, Ackerman MS, Adamic L, Expertise Networks in Online Communities: Structure and Algorithms. In: Proceedings of the International World Wide Web Conference Committee (IW2C2) ACM, **(2007)**, pp. 221–220. doi:10.1145/1242572.1242602

[6] Lu R, Mucaki E J, Rogan P K. Discovery of Primary, Cofactor, and Novel Transcription Factor Binding Site Motifs by Recursive, Thresholded Entropy Minimization[J]. bioRxiv, **(2016)**: 042752.

[7] Wallace, M. L., & Rafols, I. Research Portfolio Analysis in Science Policy: Moving from Financial Returns to Societal Benefits. Minerva, 52(2), **(2015)**, pp. 79-115.

[8] Wallace, M. L., & Rafols, I. Research portfolios in science policy: moving from financial returns to societal benefits. Available at SSRN 2500296. **(2014)**.

[9] Pinto, F. S., da Cruz, N. F., & Marques, R. C. Contracting water services with public and private partners: a case study approach. Journal of Water Supply: Research and Technology-Aqua, 64(2), **(2015)**, pp. 194-210.

[10] Kim, M. K. Short-term price forecasting of Nordic power market by combination Levenberg–Marquardt and Cuckoo search algorithms. IET Generation, Transmission & Distribution, 9(12), **(2015).** pp. 1552-1562.

[11] Adamic LA, Zhang J, Bakshy E, Ackerman MS Knowledge sharing and Yahoo Answers: Everyone knows something. In: Proceedings of the 17th International Conference on World Wide Web (WWW'07), New York: AC, **(2007)**, pp. 665–674. doi:10.1145/ 1267497.1267577.

[12] Sayer, J. R. A determination of the key factors and characteristics that SME-scale commercial biomedical ventures require to succeed in the South African environment. **(2015).**

[13] Gopalan, R., & Jayaraman, S. Private control benefits and earnings management: evidence from insider controlled firms. Journal of Accounting Research, 50(1), **(2012)**, pp. 117-157.

[14] Chen, H. Y. The Research of Computer complex network reliability evaluation method based on GABP algorithm. In Applied Mechanics and Materials (Vol. 556, pp. 6207-6210). Trans Tech Publications. **(2014).**

[15] Zou, L. K., Liu, S. K., & Ma, G. F. Intrusion Detection Model Based on Improved Genetic Algorithm Neural Network in Computer Integrated Process System. In Applied Mechanics and Materials (Vol. 270, pp. 2707-2711). Trans Tech Publications. **(2012).**

Authors

Xiaochun He. Male, born in Feb. 1981. He received his Bachelor Degree in College of Computer Science and Technology, Shaanxi University of Science & Technology, Xi'an, China in 2005, He is a lecturer and Engineer working in the Computer Information Department at Sichuan Vocational College of Information Technology, He has published more than ten academic papers. He has six patents authorization. He current research interests include computer application technology, network technology.

JunJun Feng. Male, born in Jun. 1989. He received his master degree in College of Computer System Structure, Beifang University of Nationalities, Yinchuan, China in 2015. He is a lecturer and Engineer working in the Computer Information Department at Sichuan Vocational College of Information Technology. He current research interests include computer application technology, artificial intelligence, information safety.

Ruchun Jia. Male, born in Feb. 1989. He received his master's degree of engineering of Sichuan University, Chengdu, China in 2014.He is a lecturer and Engineer working in the Computer Information Department at Sichuan Vocational College of Information Technology. He current research interests include Information security technology and big data technology.