Network Security Situation Prediction Based on Grey Relational Analysis and Support Vector Machine Algorithm

Xiaoyi Hong

(Corresponding author: Xiaoyi Hong)

Xinxiang Vocational and Technical College

No. 6, Jingsan Road, Economic and technological development zone, Xinxiang, Henan 453006, China

(Email: hongxiaoy@yeah.net)

(Received Feb. 2, 2019; revised and accepted Dec. 2, 2019)

Abstract

At present, the Internet tends to be omni-directional and multi-angle, and the era of big data dominance has come. However, due to the complex network users and the huge amount of network data, the current situation of network security is worrying. Therefore, the prediction of network security situation is a key link. In this study, the network evaluation index was weighed using grey relational analysis (GRA) theory, the prediction process was simulated based on support vector machine (SVM) algorithm, GRA-SVM based network security situation prediction model was constructed, actual data were substituted into the model, and the results of GRA-SVM and SVM algorithms were compared. The results showed that, compared with single SVM algorithm, the model built by GRA-SVM algorithm had higher prediction precision, which was a reliable algorithm for predicting network security situation. The application of GRA-SVM algorithm can predict the various risks of network development, which can provide a reference for the early preparation of the protection system and minimize the damage to network security.

Keywords: Grey Relational Analysis Theory; Network Security; Support Vector Machine Algorithm

1 Introduction

With the gradual construction and improvement of the Internet, the number of users has a blowout growth. However, the openness of network information and the randomness of the use of network data make the network security problems behind the prosperity frequent. People gradually realize the urgency of improving network security, and network security situation prediction is the most important.

Wei *et al.* [8] proposed a weighted hidden Markov model (HMM), applied multi-scale entropy information to solve the problem of training data, and optimized the transfer matrix of HMM. In addition, they proved that the autocorrelation coefficient could reasonably use the correlation between the characteristics of historical data to predict future security conditions.

Jiang *et al.* [7] trained radial basis function (RBF) neural network to find the mapping relationship between the first N data and the subsequent Mdata and then adjusted its value. The results showed that the method had fast convergence and good prediction effect.

Huang *et al.* [6] proposed a new approach based on artificial immune system and phase space reconstruction, analyzed the time series of network security, reconstructed the appropriate time series phase space, and constructed the prediction model using immune evolution mechanism.

Zhang *et al.* [16] constructed a prediction model based on wavelet neural network (WNN) using the improved niche genetic algorithm (INGA), optimized the parameters of WNN by adaptive genetic algorithm (GA) to make it search more effectively, and solved the premature convergence problem of genetic algorithm using dynamic fuzzy clustering and elimination mechanism.

Hu *et al.* [5] proposed a new prediction model called cloud belief rule base (CBRB) model and represented belief parameter rules using cloud model, which made it more accurate to express expert knowledge. The experimental results proved the practicability of the proposed CBRB model. In this study, correlation analysis based on grey relational analysis (GRA) principle and support vector machine (SVM) algorithm were used for parameter integration. On the basis of them, a GRA-SVM network security situation prediction model was constructed, and it was applied to the specific network situation prediction and compared with the SVM model without correlation analysis.

2 Network Security and Situation Prediction

Network security [10] generally refers to various security problems occurring on the network. However, unlike traditional security issues, it is a special concept based on the development of network and the new challenges of information security in the process of network development. Issues such as personal information fraud, frequent network vulnerabilities and network system shocks are the consequences of network security damage.

Network security situation prediction [14] is the recognition of network security status, including the fusion processing of the old data obtained, extracting the background state and activity semantics of the network system, identifying the possible abnormal activities in the network, and finally outputting the network security situation prediction results based on the above characterization. It is an early warning of network security problems and an important defensive measure to protect the stability of network system, so it is widely used in the maintenance of network system.

3 SVM Algorithm

SVM algorithm as a data mining technology is used to solve classification problems [9]. Based on the basic construction of statistical principle, a kernel function is added to the calculation process to map the low-dimensional problem to the high-dimensional space, and finally the optimal solution in the high-dimensional solution space is obtained [11]. It means that using SVM algorithm can unlock hidden patterns in a large amount of data, so as to discover the information behind the data. After uploading information to the system, the system can identify the time series or development trend of the data and make accurate judgments. The basic structure of the SVM algorithm is shown in Figure 1.

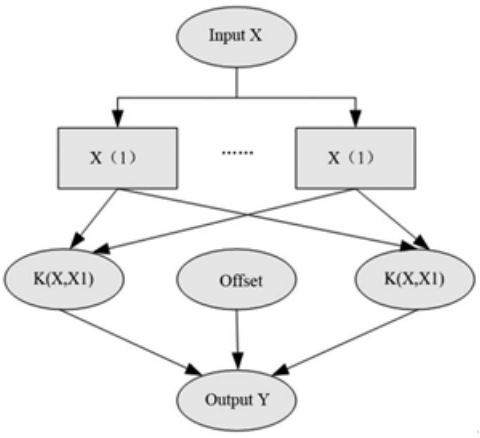

Figure 1: Basic structure of SVM algorithm

4 GRA-SVM Based Model Construction

4.1 Construction Method

4.1.1 SVM Algorithm

The regression equation [1] is an expression that reflects the regression relationship between one variable and another. The regression equation of SVM algorithm is [15]:

$$\begin{aligned} f(x) &= w \cdot \varphi(x) + b \\ \varphi &: R^n \to G, w \to G, \end{aligned}$$

where n represents n situation value training samples $\{x, y\}$, $i = 1, 2, \cdots .n$, x represents the training input value, y_i stands for the output results, w stands for weight vector, and b stands for offset vector.

The optimization function is used for optimization:

$$\min J = \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i^* + \xi_i),$$

where ξ_i^*, ξ_i are the relaxation factors and C is the punishment factor. The constraints in the formula are as follows.

In order to get the final regression result of SVM, it is necessary to substitute Lagrange multiplier. The Lagrange multiplier method [4] is an algorithm for network parameter error identification, which is often used to process support vectors and solve conditional extremum problems. Therefore a_i and a_i^* are substituted into the formula, and then the SVM regression expression is obtained:

$$f(x) = \sum_{i=1}^{R}(a_i - a_i^*)(\varphi(x_i), \varphi(x)) + b$$

Finally, $(\varphi(x_i), \varphi(x))$ is replaced by kernel function $k(x_i, x)$ to solve the problem of curse of dimensionality in the process of non-linear regression prediction. Finally, the following formula is obtained:

$$f(x) = \sum_{i=1}^{n}(a_i - a_i^*)k(x_i, x) + b.$$

4.1.2 GRA Principle

Because the parameters of GRA and SVM algorithm are repetitive for the same case, the same part will not be repeated. The situation value is set as (χ_0), and the evaluation index is set as χ_i; then the corresponding observation value can be obtained. In order to facilitate the calculation, the data need to be dimensionless, so as to simplify the calculation process [3]. The expressions are:

$$\begin{aligned} \chi_0 &= \{x_0(1), x_0(2), \cdots, x_0(n)\}, \\ \chi_i &= \{x_i(1), x_i(2), \cdots, x_i(n)\}, \end{aligned}$$

where

$$x_0(k) = \frac{y_0(k)}{\sum_{t=1}^{n} y_0(t)}$$
$$x_i(k) = \frac{y_i(k)}{\sum_{t=1}^{n} y_i(t)}$$

The correlation coefficient is the bridge and link to obtain the correlation degree. It can be calculated according to the following formula:

$$\xi(k) = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \rho \cdot \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \rho \cdot \max_i \max_k |x_0(k) - x_i(k)|}$$

where ρ represents the resolution coefficient.

When the data is large, there will be many correlation coefficients, so in order to facilitate comparison, the average correlation coefficient is taken as the correlation degree in GRA principle [13]. Correlation degree r_i can be expressed by:

$$r_i = \frac{1}{n} \sum_{k=1}^{n} \varphi_i(k).$$

Finally, by weighting the sequence, the final weighted correlation degree is obtained:

$$r_i = \frac{1}{n} \sum_{k=1}^{n} W(k)\varphi_i(k).$$

4.2 Construction Process

The preliminary construction process of GRA-SVM based network security situation prediction model can be represented by Figure 2.

After initializing the historical data of samples, the weight of correlation degree can be calculated by GRA correlation formula. According to these data, training set and prediction set are generated. Then SVM algorithm is used to model the training samples. The evaluation values of each class are stored in training set List1 in descending order for training [2]. The data in List1 is transmitted into the training module of the prediction model. 24 h is taken as a round, and the attack situation values of different indicators per hour are predicted and output. The actual values of historical data are input and compared with the predicted value. GRA-SVM based prediction model can be established when certain accuracy is satisfied.

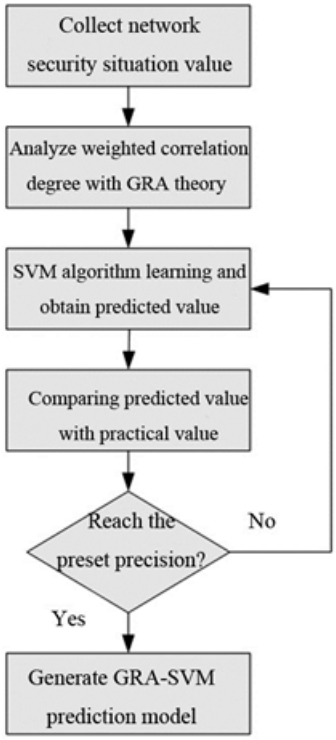

Figure 2: Overall flow of GRA-SVM based model construction

5 Experimental Verification

5.1 Prediction Results of GRA-SVM Based Model

A GRA-SVM prediction model can be obtained by inputting the weighted correlation degree of the index after GRA analysis into the training program of SVM prediction. After the model is established, the original stored historical data are cleared, and only the running program of the algorithm itself is left. Then the new prediction results can be obtained through the analysis of the GRA-SVM based prediction model.

1) Index Collection:
In order to ensure the accuracy of the prediction system and the scientificity of the experiment, a lot of information needs to be collected. The historical data of 200 alarm messages from July 1 to July 10, 2019 was collected and classified according to different network security issues. After analyzing the corresponding historical data, the GRA-SVM based prediction model can get the final result. Similarly, the alarm information from July 20 to July 30 was selected as the actual situation value in the future.

   As the selected data set is very large, the data are normalized [12] to simplify the process. The following formula is used:

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

   where x stands for the original situation value, $\hat{x}$ stands for the normalization number, $x_{\max}$ stands for the maximum situation value, and $x_{\min}$ stands for the minimum situation value.

2) Model Analysis Data
The average correlation coefficient in the model is cal-

culated after substituting the formula, and the historical data are selected. The category item with few alarms which are selected from the 200 alarm information was abandoned. The remaining categories are sorted, and the top six evaluation indicators of the correlation degree are selected and ranked in descending order. The results of GRA analysis for each index are shown in Table 1.

According to the correlation degree, the weighted correlation degrees, and the weights of different evaluation indexes, can be obtained. The weights are input into the SVM algorithm program, and the final prediction results of different indexes are numbered in descending order. The results are shown in Table 2.

5.2 Result Comparison

After calculation, the predicted value of the GRA-SVM based network security situation prediction model was compared with the actual value of the future security situation, as shown in Figure 3. The indexes are ranked again and numbered according to the weight of correlation degree, corresponding to the x axis.

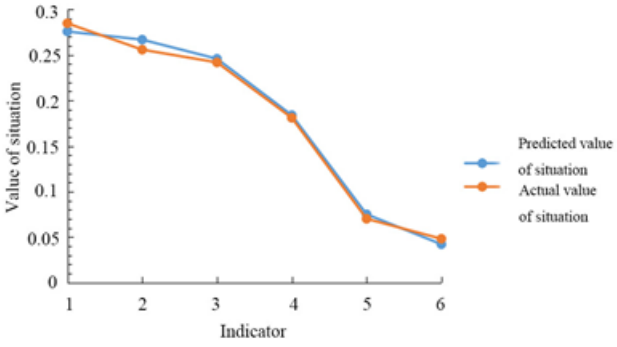

Figure 3: Comparison of predicted and actual situation values

It can be seen from Figure 3 that the two curves are approximately equal and the endpoints coincide approximately. The numerical difference of network bandwidth situation value numbered 4 is the smallest, which is only 0.003, while the network vulnerability number index numbered 2 is the largest one among all situation values, which is 0.011. However, on the whole, the predicted situation value obtained by the analysis of GRA-SVM based network security situation prediction model is close to the actual situation value obtained from statistical investigation. So it can be said that GRA-SVM based network security situation prediction model is an effective model to predict network security situation.

Since Figure 3 is only a simple comparison of absolute values, it is not enough to get the final result. In order to ensure the rigor of the experiment and further verify the accuracy of GRA-SVM based prediction model, a single SVM prediction model is selected as the control group, which directly inputs the data into the SVM algorithm system without GRA processing. The error between GRA-SVM based prediction model and SVM based prediction model is shown in Figure 4. Two decimals of the error rate are kept.

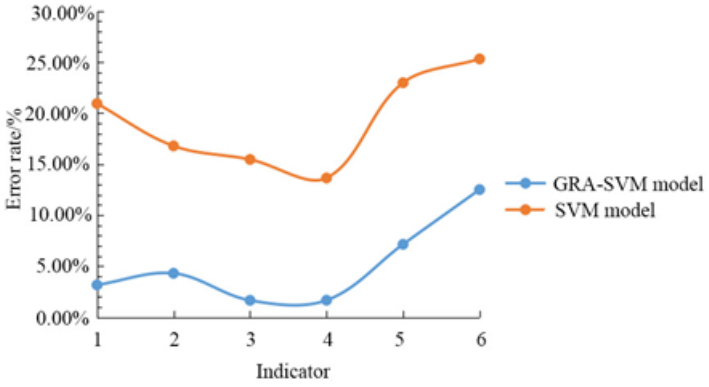

Figure 4: Comparison of error rates between GRA-SVM based prediction model and SVM based prediction model

Figure 4 shows that the error rate curve of the SVM based prediction model is always under the SVM based model. Both of them have small errors in network bandwidth, but in the aspect of the average fault-free time of subnetworks numbered 6, both the results of GRA-SVM and SVM based models have large errors. The greatest error appears when calculating the frequency of critical devices accessing secure websites. The error rate of GRA-SVM based model is 3.16%, while that of SVM based model is 20.93%; the difference between the two models is 17.77%. Thus the accuracy of GRA-SVM algorithm is 17.77% higher than that of the traditional SVM method. Therefore, it is concluded that the application of GRA-SVM based model can significantly improve the accuracy of network security situation prediction and ensure the reliability of the results.

In the practical application of network security situation prediction model, the prediction time and detection range are also important parts. On the one hand, it is because of the complexity of the network security situation value samples; on the other hand, fast prediction of the situation value results is also an indispensable means to solve the network security problems in the first time. Therefore, taking the GRA-SVM based prediction model and SVM based prediction model as the comparison objects, the detection coverage and average detection time of the two models are shown in Figures 5 and 6.

Figure 5 shows that the detection coverage rate of GRA-SVM based prediction model is 26.42% higher than that of SVM based prediction model, i.e. 1.4 times. Figure 6 shows that the average time used by GRA-SVM based prediction model in detecting each index data is 0.22 s less than that of SVM based prediction model, and 38.6% of time can be saved. According to Figures 5 and 6, the GRA-SVM based prediction model is superior to

Table 1: GRA analysis results of correlation degree

Evaluation indicator	Original situation value	Correlation degree
Number of vulnerabilities in network	0.285	0.945
Average failure-free time of subnet	0.256	0.938
Network bandwidth	0.242	0.826
Alarm probability	0.181	0.746
Network bandwidth usage frequency	0.070	0.713
Frequency of critical devices accessing secure websites	0.048	0.695

Table 2: Predicted situation value of network attack

Number	Evaluation indicator	Predicted situation value
1	Frequency of critical devices accessing secure websites	0.276
2	Number of network vulnerabilities	0.267
3	Alarm probability	0.246
4	Network bandwidth	0.184
5	Network bandwidth usage frequency	0.075
6	Average failure-free time of subnet	0.042

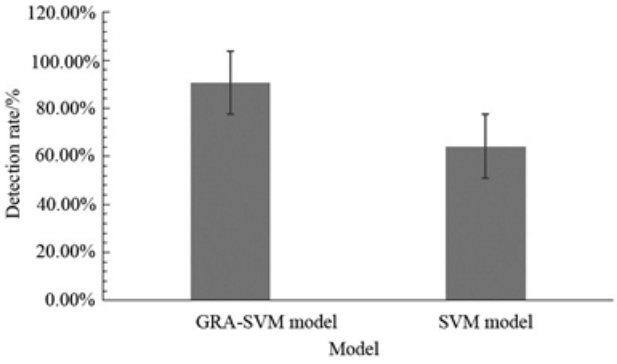

Figure 5: Comparison of detection rate between GRA-SVM based prediction model and SVM based prediction model

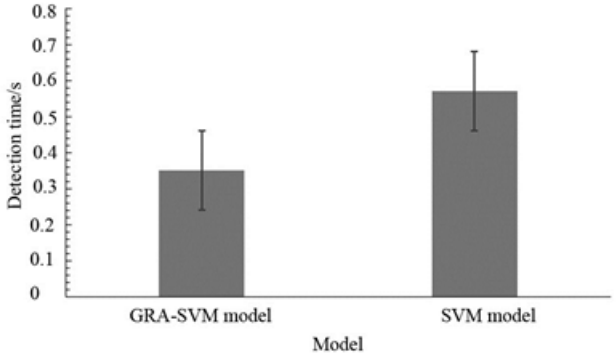

Figure 6: Comparison of detection time between GRA-SVM based prediction model and SVM based prediction model

the SVM based prediction model in terms of overall detection coverage and average detection time. So it can be said that GRA-SVM based prediction model is relatively efficient and more suitable for the practical application of network security issues.

6 Conclusion

Network security situation prediction is an indispensable measure and condition to maintain network security. Based on GRA principle and SVM algorithm, a GRA-SVM network security situation prediction model is constructed. Through the analysis of historical data, we can get accurate prediction of future situation value. In order to verify the relative reliability of the results, it is compared with the model constructed by the traditional SVM algorithm. The research process of this paper shows that:

1) The construction of GRA-SVM based network security situation prediction model can effectively improve the efficiency of situation value prediction, thus gaining more time for repairing network security problems. Moreover it is conducive to making full preparations for network security threats in the future and reducing the risk of accidents.

2) This experiment contributes to maintaining the stability of the network environment and is conducive to providing a more secure and unblocked network environment for Internet users, thus promoting the development of the Internet industry.

References

[1] A. Bhatnagar, A. Sinha, S. Chaudhary, N. Manuja, H. Kaur, T. R. Chaitra, "Accuracy and evaluation of a new regression equation in predicting the width of unerupted permanent canines and premolar teeth," *European Archives of Paediatric Dentistry Official Journal of the European Academy of Paediatric Denstry*, vol. 18, no. 1, pp. 31–37, 2017.

[2] B. Biggio, B. Nelson, P. Laskov, "Poisoning attacks against support vector machines," in *Proceedings of the 29th International Coference on International Conference on Machine Learning*, pp. 1467–1474, 2012.

[3] M. Conesa, J. F. Sánchez, I. Alhama, F. Alhama, "On the nondimensional of coupled, nonlinear ordinary differential equations," *Nonlinear Dynamics*, vol. 84, no. 1, pp. 91–105, 2016.

[4] M. D. R. De Pinho, I. Shvartsman, "Lipschitz continuity of optimal control and Lagrange multipliers in a problem with mixed and pure state constraints," *Discrete & Continuous Dynamical Systems-Series A*, vol. 29, no. 2, pp. 505–522, 2017.

[5] G. Y. Hu, P. L. Qiao, "Cloud belief rule base model for network security situation prediction," *IEEE Communications Letters*, vol. 20, no. 5, pp. 914–917, 2016.

[6] T. Q. Huang, Y. Zhuang, "An approach to real-time network security situation prediction," *Journal of Chinese Computer Systems*, vol. 35, no. 2, pp. 303–306, 2014.

[7] Y. Jiang, C. H. Li, L. S. Yu, B. Bao, "On network security situation prediction based on RBF neural network," in *36th Chinese 0Control Conference (CCC'17)*, 2017.

[8] W. Liang, Z. Chen, X. L. Yan, X. D. Zheng, P. Zhuo, "Multiscale entropy-based weighted hidden markov network security situation prediction model," in *IEEE International Congress on Internet of Things*, 2017.

[9] M. D. C. Moura, E. Zio, I. D. Lins, E. L. Droguett, "Failure and reliability prediction by support vector machines regression of time series data," *Reliability Engineering & System Safety*, vol. 96, no. 11, pp. 1527–1534, 2017.

[10] Z. E. Mrabet, N. Kaabouch, H. E. Ghazi, "Cybersecurity in smart grid: Survey and challenges," *Computers & Electrical Engineering*, vol. 67, pp. 469–482, 2018.

[11] L. R. Quitadamo, F. Cavrini, L. Sbernini, F. Riillo, L. Bianchi, S. Seri, G. Saggio, "Support vector machines to detect physiological patterns for EEG and EMG-based human-computer interaction: A review," *Journal of Neural Engineering*, vol. 14, no. 1, ID: 011001, 2017.

[12] A. Suzuki, K. Yamanishi, "Exact calculation of normalized maximum likelihood code length using fourier analysis," Jan. 11, 2018. (`https://arxiv.org/pdf/1801.03705v1.pdf`)

[13] D. Wen, J. Li, P. F. He, "Grey correlation analysis of agro-meteorological disasters and soybean yield in Heilongjiang province," *Journal of Natural Disasters*, vol. 26, no. 4, pp. 56–62, 2017.

[14] R. F. Wu, G. L. Chen, "Research of network security situation prediction based on multidimensional cloud model," in *International Conference on Innovative Mobile & Internet Services in Ubiquitous Computing*, 2012.

[15] Z. Xue, R. Zhang, C. Qin, X. Zeng, "An adaptive twin support vector regression machine based on rough and fuzzy set theories," *Neural Computing and Applications*, 2018. (`https://link.springer.com/article/10.1007\%2Fs00521-018-3823-4`)

[16] H. Zhang, Q. Huang, F. Li, J. Zhu, "A network security situation prediction model based on wavelet neural network with optimized parameters," *Digital Communications and Networks*, vol. 2, no. 3, pp. 139–144, 2016.

Biography

Xiaoyi Hong, born in December 1982, female, master of engineering, is currently teaching in Xinxiang vocational and technical college, and her research direction is computer science and technology.