Construction of intelligent network anti-risk index evaluation system based on machine learning

Shunkai Wang[a,b,c,1], Bingjie Liu[a,b,2], Fan Zhang[a,b,c,*]
[a] China Automotive Technology And Research Center Co., Ltd.; [b] China Auto Information Technology (Tianjin) Co., Ltd.; [c] China Automotive Media (Tianjin) Co., Ltd, China, Tianjin, 300300.

ABSTRACT

With the growth of Internet, network attack means emerge in endlessly, and internet safety issues become more and more important. Anti-virus software and firewalls generate a large amount of alarm information. Faced with such a huge amount of data, network managers can hardly grasp the security status of the network system and cannot take corresponding preventive measures in time. In order to improve the data security transmission performance of the network, it is needy to evaluate the anti-risk index of intelligent networking. At present, most of the internet safety assessment methods are from the perspective of security vulnerabilities. By scanning whether there are some known vulnerabilities in the network, the corresponding assessment results and solutions are given. In this paper, an intelligent network anti-risk index assessment system based on improved machine learning algorithm is proposed. Through this system, the internet safety situation is predicted, and then the algorithm model of internet safety assessment decision system is constructed by using assessment rules. The simulation results show that this method has high accuracy in evaluating the anti-risk index of intelligent networking, strong ability in detecting intrusion information, and improved internet safety.

Keywords: Machine learning, internet safety, risk profile

1. INTRODUCTION

Nowadays, people enjoy the great convenience brought by the network in their work, study and life, but at the same time, they are puzzled by the frequent internet safety problems. Especially, in the rapid evolution of the Internet environment, the internet safety threats and the diversification of network attack means have surpassed the launching speed of preventive measures [1]. Society is increasingly dependent on computer networks. Moreover, the security problem of computer network has become increasingly prominent, and many network systems have suffered a lot of economic losses due to security problems [2]. With the growth of the Internet, network attack means emerge in endlessly, and internet safety issues are becoming more and more important. Antivirus software and firewalls generate a large amount of alarm information, so network managers can't grasp the security status of network system in the face of such massive data, and can't take corresponding preventive measures in time [3]. Evaluating the security of computer network system in advance is an effective means to prevent various internet safety problems [4]. In recent years, some computer internet safety assessment products have appeared on the market, which have played a certain role in strengthening the security of the network system. However, most of these internet safety assessment products are limited to the detection and analysis of security vulnerabilities in the network system, lacking in-depth analysis of internet safety assessment data, making it difficult to form an overall understanding of the security of computer network systems [5].

*m13502037247_3@163.com
[1] Wangshunkai@catarc.ac.cn; [2] liubingjie@catarc.ac.cn;

Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023), edited by Srikanta Patnaik, Tao Shen, Proc. of SPIE Vol. 12779, 1277922 · © 2023 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2688713

Internet safety situation prediction can provide network managers with the past and current internet safety status, at the same time, it can predict the future internet safety situation and reduce the data pressure of managers, so it has become a hot topic in current internet safety research [6]. At present, in essence, most of the internet safety assessment methods are based on security vulnerabilities, and then the corresponding assessment results and solutions are given by scanning the network for some known vulnerabilities [7]. The disadvantage of this method is that it takes a long time, occupies a large amount of bandwidth, and interferes with the normal operation of the network. In the design of network networking, a large number of data bit sequence streams are transmitted and controlled by the network. Because of the openness of network networking and the randomness of node distribution, network communication is easily invaded, which threatens the security of the network [8]. Accurately mastering the security level of the computer network system is of great significance to ensure the normal operation of the network system. In this paper, an intelligent network anti-risk index assessment system based on improved machine learning algorithm is proposed. The convergence of internet safety risk prediction is controlled by machine learning algorithm, and the internet safety situation is predicted by this system.

2. METHODOLOGY

2.1 Internet safety situation prediction

The potential danger of the system can be found by using the technology of internet safety situation prediction, which deals with the original events of the network. However, the information that has certain correlation and can reflect the characteristics of some internet safety events can be extracted, and the occurrence and growth of security events can be evaluated and predicted through mathematical models and empirical knowledge, so as to provide reference for network management. Although there is no standard definition of network situational awareness at present, the role of network situational awareness is very clear, that is, the security situation of the whole network is observed through the awareness system, and then the internet safety events are judged in time according to the observed data, and provided to managers for decision-making in a visual way. Because there will be a large amount of data in the process of internet safety situation assessment, and to ensure the responsibility of calculation and assessment methods, especially the problems of redundant information and false alarm information need to be solved. In this sense, to improve the accuracy and scientificity of the security situation assessment, it is needy to have a high ability to use mathematical methods and network modeling [9]. Internet safety situation assessment refers to the overall situation of internet safety, which has an extremely important influence on the prediction of internet safety situation. The specific assessment steps are as follows: firstly, complete the collection of internet safety, and then normalize the security events and state information; Secondly, dynamically correlate the normalized security events to generate corresponding alarm information, Third update the reliability of security events in time; Finally, use the internet safety situation calculated by the situation index model to obtain the internet safety situation value.

The concept of data fusion was first applied to sensor information processing. It refers to some sensor observation information obtained in time series, and automatically analyzes and synthesizes them by computer technology based on certain criteria, so as to realize the decision-making and assessment of a certain need. Compared with single-source data, data fusion combines homologous data to show statistical advantages, and multi-sensors can also improve the accuracy of the system [10]. Usually, the process of data fusion embodies the characteristics of multi-level and multi-level data processing, which can automatically monitor, correlate, evaluate and combine data from multiple information sources. Therefore, data fusion is actually a process of integrating multiple sensors and multi-source information to process and improve the accuracy and reliability of the conclusion.

2.2 Internet safety situation prediction model

Internet safety situation prediction is a new technology to realize internet safety monitoring, which can find potential and malicious attacks and effectively reduce the harm caused by attacks. When a certain research object has a relatively large scope, a relatively complex structure and is influenced by many factors, the comprehensive performance of its state can be explained by the situation, and the battlefield situation in the military field is one of them. The main purpose of introducing the concept of situation into internet safety management is to establish a set of internet safety situation system, which requires high feasibility and accuracy, so that network managers can fully and timely understand the overall security situation of the network. On the basis of statistical analysis of network data, feature extraction analysis is carried out, and fusion clustering processing is carried out according to the feature extraction results of risk statistical distribution. The data processing platform deployment is shown in Figure 1.

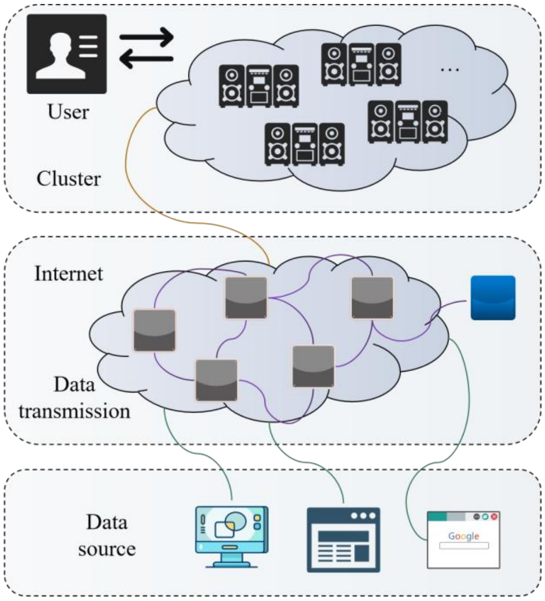

Figure 1. Deployment of data processing platform

Internet safety situation is a kind of data collected in chronological order, so it can be treated as a time series. As the input variable of the prediction model, the situation value of the previous time series is the situation value of the next time of the internet safety situation. In order to evaluate the anti-risk index of intelligent networking, it is needy to first analyze the statistical characteristics of big data transmitted by the network and the bit sequence data transmitted by the network. A multivariate statistical characteristic equation is used to describe the high-dimensional spatial storage state model of network transmission bit sequence data as follows:

$$\begin{pmatrix} X \\ P(X) \end{pmatrix} = \left\{ \begin{matrix} a_1, a_2, \ldots, a_m \\ p(a_1), p(a_2), \ldots, p(a_m) \end{matrix} \right\} \tag{1}$$

Among them:

$$\sum_{i=1}^{m} p(a_i) = 1 \tag{2}$$

Through the discrete analysis of data, the information entropy of distribution characteristics of network transmission bit sequence data is obtained:

$$H(X) = E(I(a_i)) = -\sum_{i=1}^{m} p(a_i) \log_2 p(a_i) \tag{3}$$

Network situation prediction management is weighted according to the relevant parameters of internet safety incidents, such as frequency, times, degree of threat to the network, etc., which integrates a large amount of internet safety information to obtain a situation value that can accurately reflect the running state of the network, and then combines the historical situation value and real-time situation value of the network to predict the future security development trend of the network [11]. The internet safety situation is collected according to the time sequence, so it can be regarded as a time series in the process of processing. The input variable of the prediction model is the situation value of the previous time series, and the situation value of the next time of the internet safety situation is taken as the output.

Internet safety assessment refers to evaluating the possible design and implementation vulnerability of the system through an assessment procedure, which is used to ensure that the network system is protected from accidental or intentional damage. Security assessment is a very valuable method for establishing internet safety information system.

Internet safety assessment technology can effectively find the security defects in the system, and it is an effective and direct method to ensure the security of host system and network system to the maximum extent.

Let the probability distribution of random variable set $X=\{X_1,X_2,\ldots,X_n\}$ be $P(X_1,X_2,\ldots,X_n)$. If all variables are $\{0,1\}$, 2^n-1 parameters are needed to determine the joint distribution. And through Bayesian formula, the joint distribution can be written as:

$$\begin{aligned}P(X_1,X_2,\ldots,X_n)&=P(X_1)P(X_2|X_1)\ldots P(X_n|X_1,X_2,\ldots,X_{n-1}0)\\&=\prod_{i=1}^{n}P(X_i|X_1,X_2,\ldots,X_{i-1})\end{aligned} \tag{4}$$

For $\forall X_i\in X$, if $\pi(X_i)\subseteq\{X_1,X_2,\ldots,X_{i-1}\}$ exists, the conditions of X_i assimilation $\{X_1,X_2,\ldots,X_{i-1}\}/\pi(X_i)$ are independent when $\pi(X_i)$ is given, and the above formula can be changed to:

$$P(X_1,X_2,\ldots,X_n)=\prod_{i=1}^{n}P(X_i|\pi(X_i)) \tag{5}$$

3. RESULT ANALYSIS AND DISCUSSION

Network situation analysis system is a comprehensive system. The system mainly obtains security events and status information from some common security devices or assets, such as firewalls, intrusion detection, anti-virus, routers, servers, etc. Through data merging, cleaning, and correlation analysis, the internet safety situation values are calculated according to different security situation indicators, and the overall internet safety is evaluated. To analyze the security situation of network information system, an important part is to analyze the known weaknesses and establish the utilization rules of the weaknesses. Vulnerability rule is a description of the attacker's attack behavior by exploiting the vulnerability from the attacker's point of view, and it is an important basis for analyzing the possible attack path of a specific attacker. There may be many weaknesses in a network-based information system, which are distributed on different hosts. In network attacks, attackers usually make use of the weakness of the target host to gain more privileges on the attacked host, and after the attacker gains enough privileges on the target host, he can continue to attack other hosts from the captured host. The deviation curve of the main risk characteristics of networking is shown in Figure 2.

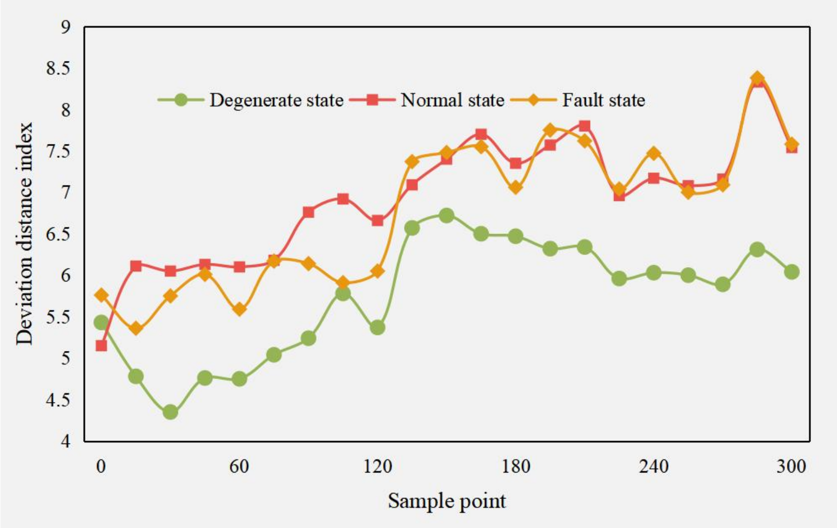

Figure 2. Selection of risk characteristic deviation curve

Security situation assessment is mainly to analyze and calculate the related security event information and asset status information cycle to assess the overall internet safety status. The security situation system needs to perform data format conversion, data filling, information merging and other processes on the security information elements reported by the two sources to form a unified format security incident message, which is convenient for correlation analysis and security. It is the process of assessment and other processes. The model in this paper is compared with random forest (RF) model and support vector machine (SVM) model. The recall rate comparison results are shown in Figure 3. The accuracy comparison results are shown in Figure 4.

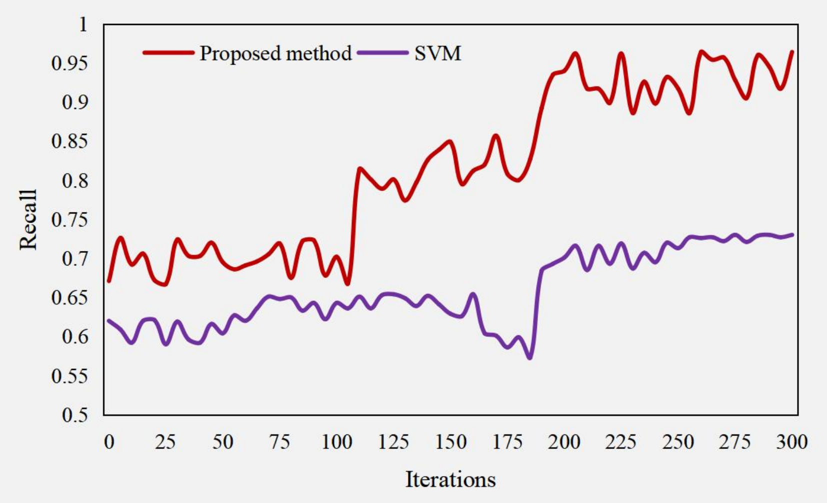

Figure 3. Comparison of recall rates of different models

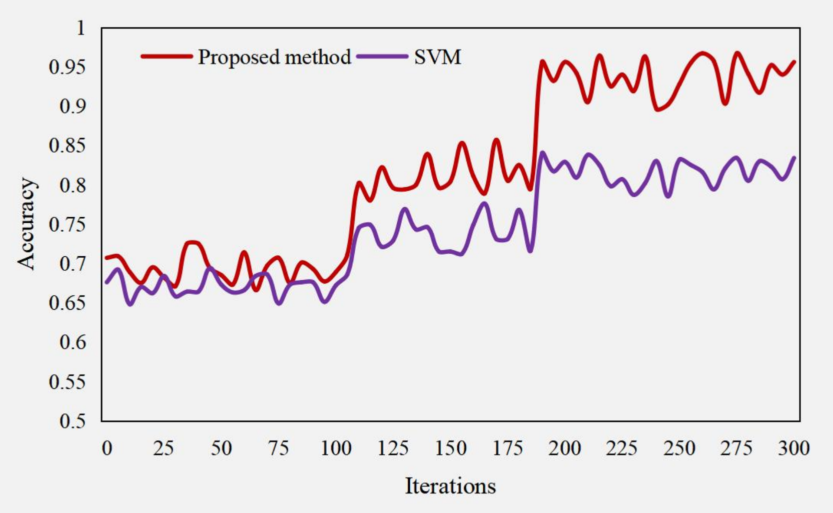

Figure 4. Comparison of accuracy of different models

The results show that this model can get 93.7% recall and 94.5% accuracy. The results are all better than the compared models. The internet safety situation factor is the weight of the security situation index, which can be determined by the analysis of internet safety experts, or by data mining based on historical data according to the influence of the three factors on the internet safety situation. Due to the numerous products and different types of internet safety equipment, the collected raw data is rather messy. After sorting and analyzing, the information which is important to internet safety and exists in most events is screened out and formatted into a unified form, which not only ensures the normalized data expression ability, but also integrates the efficiency of data message transmission in the system. The security comparison results of different network risk prediction schemes are shown in Figure 5.

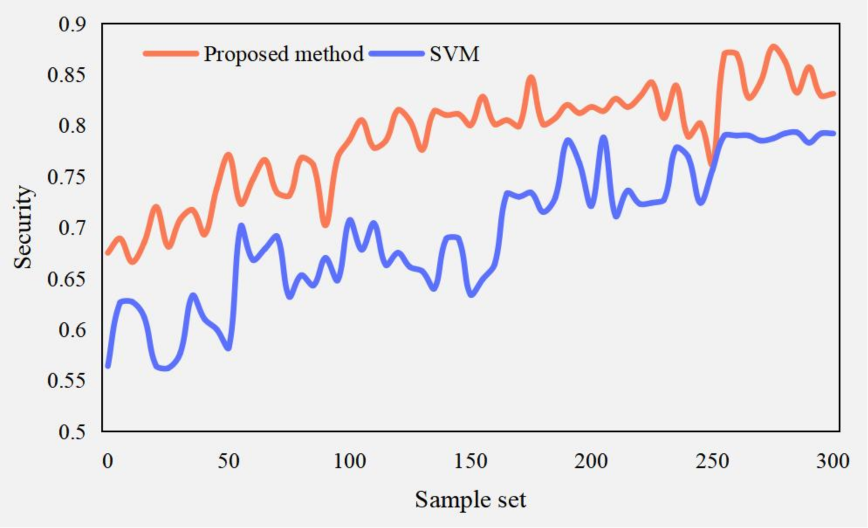

Figure 5. Security comparison of different schemes

According to the data in Figure 5, the network risk prediction method in this paper is safe and has practical application value for the safe operation of intelligent networking. Compared with the traditional forecasting methods, this method can accurately predict the situation and accurately reflect the overall change of internet safety situation. The forecasting results are helpful to guide network administrators to take appropriate countermeasures for possible future internet safety incidents and internet safety change trends.

4. CONCLUSIONS

Society is increasingly dependent on computer networks. Moreover, the security problem of computer network has become increasingly prominent. Internet safety situation prediction can provide network managers with the past and current internet safety status, and predict the internet safety situation in the future. In this paper, an intelligent network anti-risk index assessment system based on improved machine learning algorithm is proposed. The convergence of internet safety risk prediction is controlled by machine learning algorithm, and the internet safety situation is predicted by this system. The results show that the model can get 93.7% recall and 94.5% accuracy. The results are all better than the compared models. The proposed network risk prediction method is safe and has practical application value for the safe operation of intelligent networking. Due to the small number of internet safety samples used, the obtained decision rules can't fully reflect the decision-making process of computer internet safety, but basically reflect the law that the larger the assessment index value, the better the security.

REFERENCES

[1] Zhao, H., Ding, Y., Li, A., et al., Research on Safety Early Warning of Vehicle-Bridge Vibration for Long-Span Multi-Track Steel-Truss Arch Bridge of High Speed Railway[J]. Zhongguo Tiedao Kexue/China Railway Science, 39(2):2 (2018)

[2] Artigiani, E. E., Wish, E. D., Introducing the National Drug Early Warning System[J]. Current Opinion in Psychiatry, 33(4):1 (2020).

[3] Zhang, S., Shang, C., Fang, X., et al., Wireless Monitoring-Based Real-Time Analysis and Early-Warning Safety System for Deep and Large Underground Caverns[J]. Journal of Performance of Constructed Facilities, (2):35 (2021).

[4] Suo, C., Sun, H., Zhang, W., et al., Adaptive Safety Early Warning Device for Non-contact Measurement of HVDC Electric Field[J]. Electronics, 9(2):329 (2020).

[5] Gomez, F., Masmono, M., Nicolau, V., et al., De-RISC - Dependable Real-Time Infrastructure for Safety-Critical Computer Systems[J]. Ada user journal, (2):41 (2020).

[6] Li, X., Li, Y., Lu, X., et al., An online anomaly recognition and early warning model for dam safety monitoring data:[J]. Structural Health Monitoring, 19(3):796-809 (2020).

[7] Niu, H., Smart safety early warning model of landslide geological hazard based on BP neural network[J]. Safety Science, 123:104572 (2020).

[8] Chen, K., Wang, C., Chen, L., et al., Smart safety early warning system of coal mine production based on WSNs[J]. Safety Science, 124:104609 (2020).

[9] Patriarca, R., Falegnami, A., Nicola, A, D., et al., Serious games for industrial safety: An approach for developing resilience early warning indicators[J]. Safety Science, 118:316-331 (2019).

[10] Zhang, H., Li, Y., Zhang, H., Risk early warning safety model for sports events based on back propagation neural network machine learning[J]. Safety Science, 2019, 118:332-336 (2019).

[11] Su, H., Wen, Z., Yan, X., Liu, H., Yang, M., Early-warning model of deformation safety for roller compacted concrete arch dam considering time-varying characteristics[J]. Composite Structures, 203(11):373-381 (2018).