Data Security Risk Prediction of Labor Relationship Rights Protection Network Platform Based on Machine Learning

Min Yu(✉)

China University of Labor Relations, Beijing 100048, China
jhfg5413@163.com

Abstract. In order to improve the effect of data security risk prediction of labor relations rights protection network platform, this paper introduces machine learning algorithm into this field, and designs a new network platform data security risk prediction method. Determine the data risk index of the network platform of labor relationship rights protection, calculate the weight of the risk data in the website data, and find the risk data characteristics in the website. Build the decision tree, calculate the data entropy involved in the decision tree, summarize the characteristics of the risk data, create the nodes of the decision tree, and get the status of the risk data of the labor relationship rights protection network platform. The obtained risk data status is brought into the Bayesian network probability definition to analyze the risk degree of the risk data. The experimental results show that the design method can effectively shorten the evaluation time and improve the risk prediction accuracy.

Keywords: Machine learning · Labor relations rights and interests · Rights protection · Network platform · Data security · Risk forecast

1 Introduction

Employees are an important component of the production of social wealth, and the life object growing in the social environment. Any employee engaged in production and operation activities has their own rights and interests, and relies on these rights and interests to maintain their own operation and development. If employees lack the pursuit of rights and interests, they will lose the internal motivation to engage in production and operation [1].

To protect the rights and interests of workers is a responsibility entrusted by law, to safeguard the legitimate rights and interests of workers, is conducive to the establishment and maintenance of a labor system adapted to the socialist market economy, and to promote economic development and social progress. In order to discover the data security risks of the labor relations rights protection network platform in time and improve the data security of the labor relations rights protection network platform, it is necessary to predict the data security risks of the labor relations rights protection network platform.

Y. Xu et al. (Eds.): ML4CS 2022, LNCS 13657, pp. 227–241, 2023.
https://doi.org/10.1007/978-3-031-20102-8_18

At present, some scholars have studied this, but there are problems of poor prediction effect and low accuracy. Machine learning is an algorithm and statistical model used by computer systems. It relies on patterns and reasoning to effectively perform specific tasks. It is regarded as a subset of artificial intelligence. Machine learning algorithm can establish a mathematical model based on sample data, called "training data", and make predictions or decisions to perform tasks without explicit programming. Because it can obviously overcome the weakness of data analysis and prediction, in order to improve the effect of data security risk prediction of network platform, this paper introduces it into this field, a new data security risk prediction method is designed by using machine learning algorithm. Determine the data risk indicators of the labor relations rights protection network platform, calculate the weight of risk data in the website data, and find out the characteristics of risk data in the website. The decision tree is constructed, the data entropy in the decision tree is calculated, the characteristics of risk data are summarized, the decision tree nodes are generated, and the risk data status of the labor relations rights protection network platform is obtained, so as to complete the design of the data security risk prediction method of the labor relations rights protection network platform. Finally, the effectiveness of the design method is proved by experiments.

2 Risk Data Extraction of the Network Platform Based on Machine Learning

The risk data analysis process of the network platform based on machine learning is shown in Fig. 1 below:

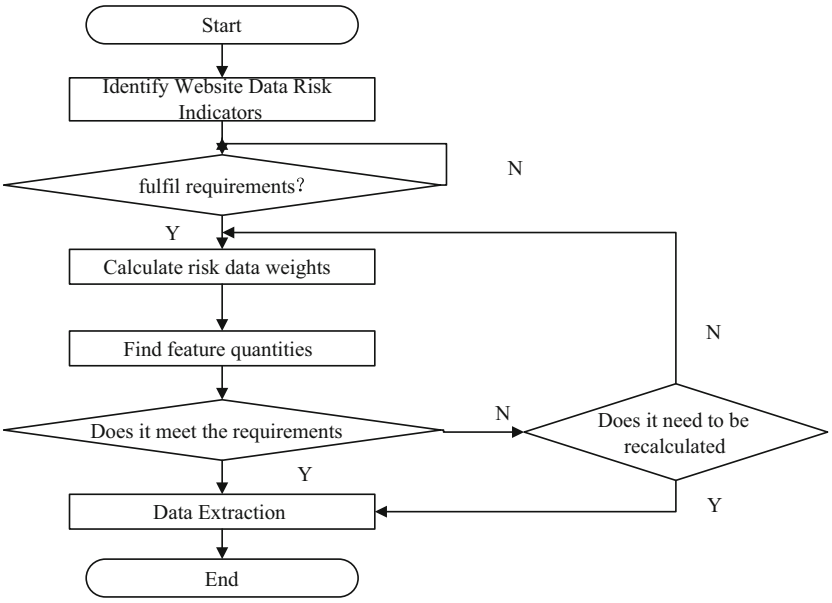

Fig. 1. Analysis process of risk data extraction method based on machine learning

Set risk source, hidden risk and joint risk characteristics as risk data characteristics of network platform, respectively with A, B, C, risk source index sample is risk level, risk probability, risk attribute; hidden risk index sample of network platform is change data information risk and hidden data risk sample; website joint risk sample is risk effect [4–7]. According to the above description of the risk indicators of the network platform, the evaluation index matrix is constructed, as follows:

$$e_j = -k \sum_{i=1}^{s} y_{ij} \ln y_{ij} \tag{1}$$

When it represents the entropy weight of the index data of the network platform, the website is the most, and the measure of the risk sample is a constant. In formula (1), k represents the entropy weight of the index data of the labor relations rights and interests protection network platform. When it is taken as 1, it represents that the risk confusion degree of the website is the largest and the risk degree is serious; y_{ij} represents the measurement value of the risk sample, which is a constant.

The index system is shown in Fig. 2 below:

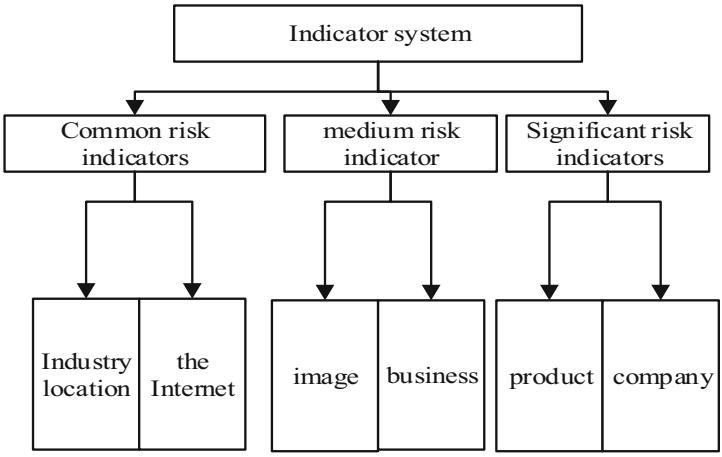

Fig. 2. Index system

There is a large amount of information data on the labor relations rights and interests protection network platform. In order to search quickly and not omit to retrieve the risk data containing risk factors, it is necessary to compress the site data first. In this paper, the vector data compression method is used to filter the risk data of e-commerce websites. First, the multidimensional data is used, and the calculation formula is as follows:

$$c = \varphi * P_1 \tag{2}$$

In formula (2), P_1 represents the power eigenvector of the actual website data information; φ represents the n-dimensional column vector whose data is converted to vector format. The middle component of the actual power feature vector of each type of the

network platform is replaced by the corresponding constant, and the safe data in the risk data set of the network platform is excluded, and the risk data set is obtained, which is shown as follows:

$$\mathrm{x} = \sum_{i=1}^{n} hp_i j_i \tag{3}$$

In formula (3), h represents the random vector of data; j_i represents the error of risk data eigenvector compression; p_i represents the actual data compression balance coefficient [8, 9].

Of labor relations rights protection network platform of risk data compression, can reduce the risk data feature extraction process and workload, then the labor relations rights protection network platform risk feature vector extraction, the compression of successful labor relations rights protection network platform risk data feature classification of all data collection, and then weighted the data, extract the characteristics of different risk data, the calculation formula is as follows:

$$0 = x\{\|\Delta x(m)\|^2\} = \sum_{i=m+1}^{n} p_i - b_{ij} \tag{4}$$

In formula (4), m represents the number of iterations for calculating the risk data vector of the labor relations rights and interests protection network platform; b_{ij} represents the initial center position of risk characteristic data vector calculation; $\Delta x(m)$ represents the weighted value of risk data characteristics [10].

3 Data Security Risk Prediction of the Network Platform Based on Machine Learning

3.1 Network Platform for Protecting Labor Relations Rights and Interests

The IOT monitoring service platform is mainly divided into on-site information level and centralized control level. The on-site information level is mainly to set up multiple fixed monitoring points on the labor relations rights and interests protection network platform. Through the sensor time, the data information during real-time operation can be decoded, edited and transmitted anywhere. By setting the monitoring time and monitoring the collection of data records, through machine learning data sharing and real-time reporting, the collected data can be compressed, encoded and transmitted to the central control level.

The network platform of labor relations rights protection is shown in Fig. 3 below:

The information-level data acquisition formula is described in (5):

$$U_n = E\delta^{-2}[\tau_{an}\partial_n(T_n)] \tag{5}$$

In formula (5), U_n represents the value of different monitoring points at different times and converted into numerical value; $E\delta^{-2}$ is a constant value; τ_{an} represents real-time data section; ∂_n represents real-time data; T_n refers to the distance between

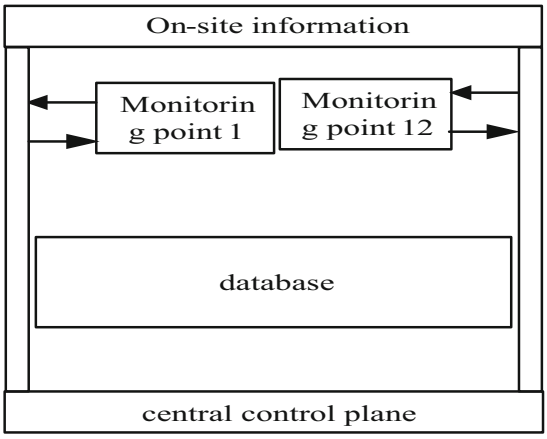

Fig. 3. Network platform for labor relations rights and interests protection

monitoring points. Through the information of different labor relations rights and interests protection network platforms monitored in different time periods, the converted values in different time periods can be calculated, and the values can be sorted out for subsequent statistical monitoring [5].

The central control level, as the central monitoring module of the whole system, integrates and stores the compressed and coded data transmitted by the sensors, analyzes and judges the data, and determines whether the labor relationship rights and interests at different time points and different monitoring points are stable and safe. The labor relations rights protection network platform based on machine learning will display the information and data transmitted from the field information layer in real time, so as to realize the regional positioning of risk positions.

Based on machine learning the rights and interests of labor relations protection network platform data security risk prediction architecture can realize hydropower station monitoring data collection, online monitoring data collection, provide data for digital cases, built a comprehensive integration of big data link platform, for the security and stability of information risk early warning system. Central control level can show different time running data of different monitoring points, at the same time when the abnormal value generated automatically alarm, operators through alarm warning and screen abnormal values and risk points, to track to preliminary judge risk situation and risk area, risk judgment data instructions to record and storage, so that in the future work to past risk query and determine whether will affect the future work process, can be timely prediction and regular maintenance.

3.2 Protection of Labor Relationship Rights and Interests Data Security Risk Prediction of the Network Platform

Machine learning is through the algorithm protocol for data depth analysis, in order to achieve some demand, this paper adopts machine learning technology decisions and Bayesian network algorithm to the labor relations rights protection network platform

risk data risk assessment, using Bayesian network algorithm to improve the accuracy of decision tree data analysis. The decision tree is shown in Fig. 4 below:

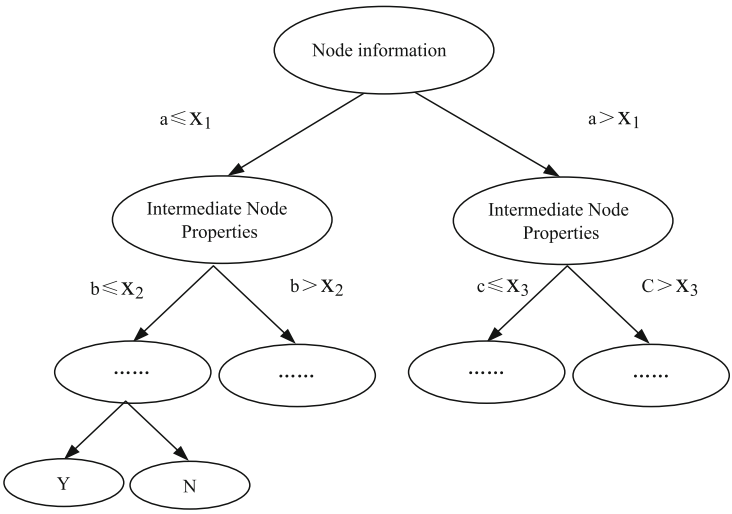

Fig. 4. Decision tree

Decision tree algorithm is one of the important methods of data risk analysis. As the name suggests, decision tree is to reasonably divide the overall data into similar state charts according to the hierarchy structure, state and data to complete the in-depth analysis of data. Each fulcrum in the decision tree structure is the key point connecting each data. In the data analysis, the data tree needs a data entropy for data judgment guidance. The calculation formula of data entropy is shown as follows:

$$E(S) = \sum_{i=1}^{n} (m_i + n_i)/(m + n) \tag{6}$$

In formula (6), S represents the root of the decision tree and the set of data to be analyzed m, n represents the number of data sets n_i, m_i represents possible nodes in the decision tree structure [14]. The directed acyclic diagram of Bayesian networks is shown in Fig. 5 below:

When the decision tree algorithm analyzes the risk data, its nodes will divide the binary nodes according to the actual situation, but the decision tree has a chance to the data analysis of the binary nodes, which reduces the analysis effect of the decision tree, so the accuracy of the decision tree analysis data is improved through the Bayesian network algorithm.

The essence of Bayesian network algorithm is to complete the forward analysis and reverse analysis based on conditional probability. On the one hand, it is to check the data analysis, and on the other hand, to ensure the depth and accuracy of the data analysis.

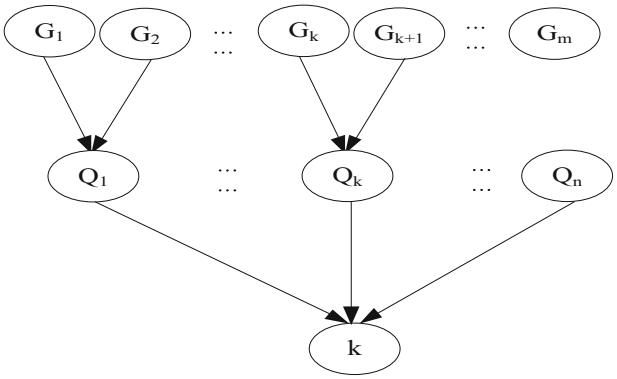

Fig. 5. Directed acyclic graph of Bayesian networks

That is, the formula of the Bayesian network algorithm, which is defined as:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \tag{7}$$

In formula (7), $p(B)$ represents the prior probability of data analysis; $p(B|A)$ represents the posterior probability of data analysis. The Bayesian network algorithm decision tree is shown in Fig. 5 below (Fig. 6):

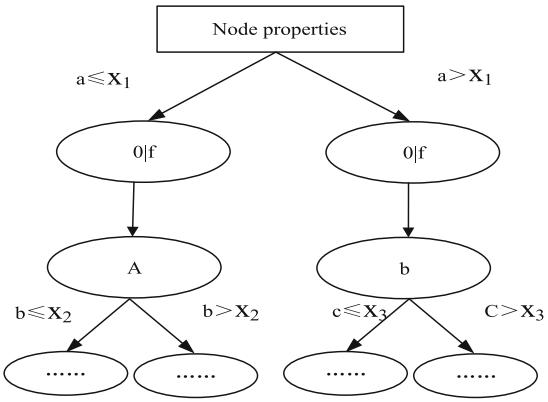

Fig. 6. A Bayesian network algorithm decision tree

According to the multiple states of the data analyzed by the decision tree, it can be brought into the Bayesian network probability algorithm, and the full probability of each state is obtained. The formula is as follows:

$$P(Y) = \sum_{i=1}^{n} p(B|A = a_i) \tag{8}$$

To sum up, summarized based on the decision tree and Bayesian network probability algorithm based on the rights and interests of labor relations network platform data security risk prediction analysis process, the research of machine learning risk data extraction method as the basis, this paper build based on data mining technology website risk assessment model, specific steps are as follows (Fig. 7):

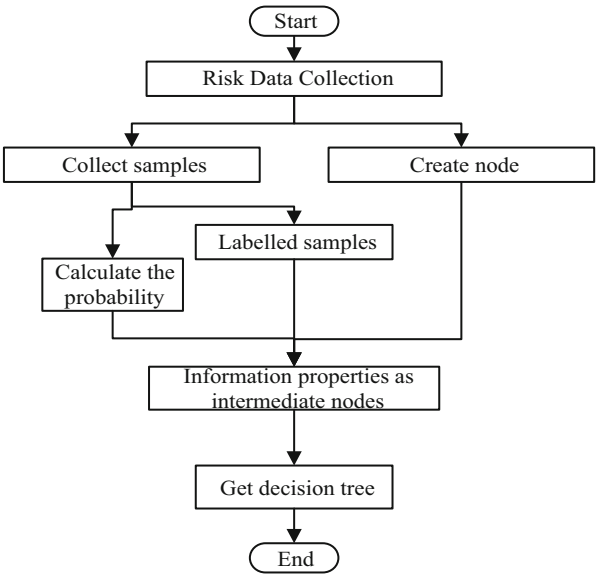

Fig. 7. Data security risk prediction process of the network platform based on machine learning

(1) First, define the collection of risks existing in the network platform of labor relationship rights protection, and the collection form is as follows:

$$Y = F_{vt} * loss \quad F \in \{F1, F2, F3, F4, F5\} \tag{9}$$

In formula (9), $F1$ means that there is a risk vulnerability in the confidentiality of the data of the labor relations rights and interests protection network platform; $F2$ indicates that there is a risk vulnerability in the data integrity of the labor relations rights and interests protection network platform; $F3$ indicates that there is a risk vulnerability in the reliability of the data of the labor relations rights and interests protection network platform; $F4$ indicates that there is a risk vulnerability in the principle of the data of the labor relations rights and interests protection network platform; $F5$ indicates that there is a risk vulnerability in the data defense of the labor relations rights and interests protection network platform; F_{vt} represents the probability of data risk caused by website attack; *loss* the loss in the process of the risk of the labor relations rights and interests protection network platform is affected by the amount of website data risk;

(2) Then complete the data in the labor relationship rights protection network platform for risk data extraction, compress the extracted risk data according to the decision tree algorithm, and simplify the workload of risk assessment on the website;
(3) Secondly, in the compressed website risk data set, the website data risk state is calculated according to the decision tree theory and the Bayesian network probability algorithm;
(4) Finally, the risk probability of the data risk status and the hidden risk probability of the data risk of the network platform is calculated, and the risk assessment model of the e-commerce website is shown as follows:

$$F(t) = 1 - \lim_{2} F_{vt} + e_j * x \tag{10}$$

In formula (10), the unknown significance is shown above.

The results of data security risk prediction of the network platform are presented in the form of 100%. The evaluation result is 0–30%, and the low-risk website; the evaluation result is 30–60%, 30% −60%; the evaluation result is 60% and high-risk website.

The key network information is first extracted. The number of network information searched by network users is huge, and the variety of information is various and complex, so it is easy to be mixed together during transmission. If the prediction is made at the same time in a certain period of time, the efficiency of the prediction results will be greatly reduced, and the real needs of network users cannot be accurately predicted. Therefore, when predicting the search target of network information, the first thing to do is to extract the key network information. This paper uses clustering technology to extract the target mask, identify the candidate box, filter and extract the user search data, update the extracted target mask, and update the results in the text information, which reduces the difficulty of extraction. Due to the high complexity of the network transmission layer, in order to meet its needs, it is necessary to constantly update the target mask and strengthen the refresh of the network. When extracting and analyzing the key network information, the time is recorded and analyzed in real time to complete the set extraction target.

Then you filter the target information. The network information search target is iteratively classified to obtain the phased target information classified for the first time. If the uncoordinated information appears in the target information, store the lower right and left corner of the target box, store the first information screening results in the upper left and right corners of the candidate box, and move the central target information to the edge of the candidate box during secondary screening to filter the last information at any time. In the final screening of target information, control the time of classification and screening, and try to get the final target information in the shortest time, while the accuracy of the results and the timeliness of screening must be guaranteed.

Finally, get the network information that users search. Find specific network information from coarse to fine. The information content searched by users can be divided into many aspects. From different positioning angles, analyze the edge information of

the target information according to the central content, move the central target information to the edge of the candidate box, overlap with the edge information, and accurately predict the location through the overlapping information. Traditional software can only use density based detection methods, Through iterative classification of different aspects of loose network information, to improve the accuracy of network information search targets.

So far, the design of data security risk prediction method based on network platform has been completed.

4 Experimental Research

In order to verify the effectiveness of the data security risk prediction method of the labor relations rights protection network platform based on machine learning designed in this paper, the flow change trend of network information will be detected. When the information of the network platform is in the normal state, the flow change trend is shown in Fig. 8 below:

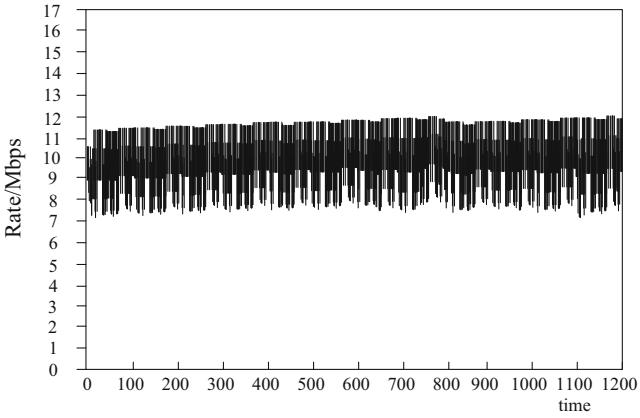

Fig. 8. Trend of flow rate under normal condition

According to the above figure above, in the case of no attack or the network receives normal requests, the fluctuation of network traffic is relatively stable, basically at 9 – 12 Mbps.When there is a network information leakage point, the change trend of the network throughput is detected, and the resulting network throughput change is shown in Fig. 9 below:

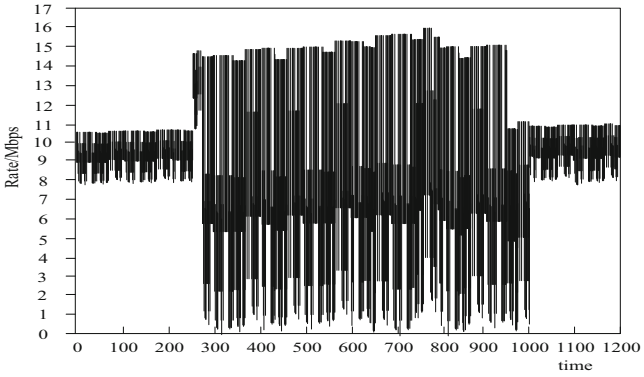

Fig. 9. Changes in network throughput during information leakage

According to the figure above, when information leakage occurs, the network throughput decreases significantly, and the information throughput rate decreases by about 72%. It can be seen that the network information node leakage has a great impact on the normal work of the network platform for the protection of labor relations rights and interests. After verifying the working ability of the detection method, choose the traditional hidden Markov model of labor relations protection network platform data security risk prediction method and the Fourier algorithm data security risk prediction method comparison experiment, after the network information attack efficiency experiment results as shown in Fig. 10:

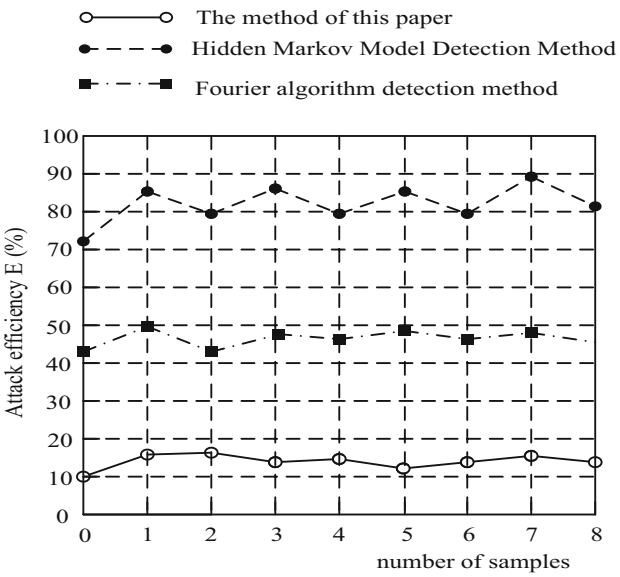

Fig. 10. Experimental results of the postdetected network information attack rate

According to the figure above, after using the detection of the detection method proposed herein, the attack efficiency of external attacks is significantly reduced, always between 10% and 20%. After using the detection method based on Fourier algorithm, the attack efficiency is between 40% and 50%, while after using the Markov model detection method, the attack efficiency is between 70% and 80%.Thus, the detection method proposed has the highest attack detection capability, which can effectively reduce the external attack efficiency and improve the network security after detection.

The cause of this phenomenon is based on the hidden Markov model Ad hoc network information leakage point of labor relations rights protection network platform data security risk prediction method and Fourier algorithm Ad hoc network information leakage point of labor relations rights protection network platform data security risk prediction method, the two methods have certain limitations, only suitable for continuous network signal.

Based with wavelet reconstruction, greatly improves the security of network information transmission. Machine learning has its own advantages and fast transmission rate. Because machine learning is not connected, it is disconnected after each response, so the possibility of information leakage is low. This paper enables machine learning to optimize the server mode of traditional network information transmission and determine the appropriate detection parameters. Based on the classical Fourier algorithm, the paper improves and proposes the rapid detection method for wavelet reconstruction, so the overall detection ability is stronger.

The data security risk prediction model of labor relations security network platform studied in this paper has certain logic, but in order to verify whether the model has application significance and achieve the expected purpose, this paper conducts comparative testing, and verifies and analyzes according to the test conclusion. In order to avoid the chance of comparison results, this paper selects data security risk prediction method and data testing network security platform data security risk prediction method as the traditional control risk assessment model, collaboration to complete the test. Before the experiment, two network platforms of labor relationship rights protection were randomly selected as the trial subjects, and the evaluation efficiency of different risk assessment models was determined to combine the average accuracy of the two experimental results.

Before the start of the trial, the randomly selected labor relationship rights protection network platform is risk assessed according to the professional software, and the evaluation results are encrypted and stored, which are the important reference data for the results of the risk assessment model after the trial. In the process of the test, the data analyzer recorded the evaluation process and important data of the three models on the labor relationship rights protection network platform in real time. All the three evaluation models submitted two evaluation results respectively. After the test, the staff verifies the data, summarizes the data, draws the test conclusion, and arranges the test site and test equipment. Since the test operation eliminates possible external interference factors, the test conclusion has credibility and authenticity. The obtained experimental results are shown in Table 1 below:

Table 1. Evaluates the risk index

	Assess the risk index	
	Squirrel selling book net	Daily E-commerce website
Professional evaluation software	55%	30%
Data security risk prediction method based on data mining	54.5%	30%
Data security risk prediction method based on data analysis	50%	27%
Data security risk prediction method based on machine learning	53%	26%

According to the traditional prediction method and the actual prediction effect of the prediction method of this paper, record the accuracy of two methods, the traditional prediction method and the network information search target prediction, the target information and real target information comparison results, the accuracy results as shown in Fig. 11:

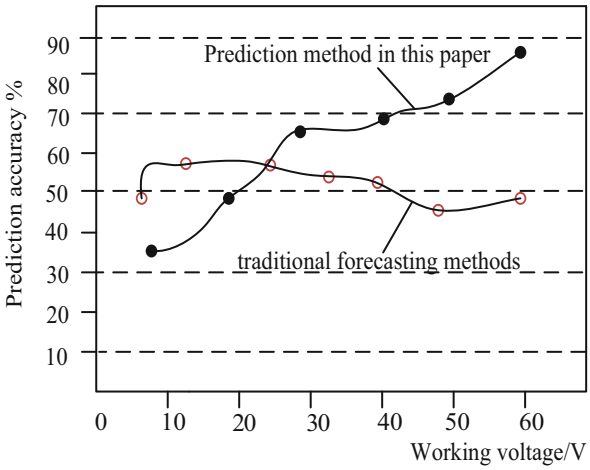

Fig. 11. Accuracy experiment results

According to Fig. 11, the two prediction methods classify the network information search target, but the prediction results are very different. The traditional prediction method, the accuracy of the target information is 70%, while the prediction method designed in this paper is 88%, 18% higher than the traditional method, indicating that the prediction method designed in this paper has a higher accuracy of the target information. The results of the prediction time experiment are shown in Table 2 below:

Table 2. Prediction time for the experimental results

Experimental times / times	Prediction time, / min	
	Conventional method	The method of this paper
1	15.22	5.23
2	16.04	5.42
3	15.87	5.07
4	15.44	5.88
5	15.96	5.09

According to the above table, the text mining network information prediction method proposed in this paper takes much less time than the traditional methods and has better prediction ability.

Traditional prediction methods have low power to predict uncoordinated information data, The capability value is only 0.1, Much different from the standard capacity value of 1.5. However, the prediction method designed in this paper has a strong ability to predict uncoordinated information data. The capability value reaches 1.8, It is also 0.3 higher than the standard capability value. Although both the traditional prediction methods and the prediction methods designed in this paper can predict the network information search targets. However, the traditional prediction method is poor. The accuracy of the prediction results is much lower than the prediction method designed in this paper. And the prediction performance is poor. Therefore, the target prediction method based on network information search based on text mining is better than the traditional prediction method. The prediction effect is better, Higher effectiveness and feasibility.

5 Conclusion

This paper studies a machine-based network platform data security risk prediction method, first according to the characteristics of the risk data, and then according to the risk assessment characteristics of the decision tree algorithm and Bayesian network algorithm of labor relations security network platform risk assessment model, complete the study of this paper. Finally, through comparative test analysis, prove that the assessment of risk assessment method high efficiency, can achieve the expected effect of this paper, shorten the data risk in the labor relations rights and interests protection network platform data time, ensure the rights and interests of labor relations network platform data transaction security, has certain application value.

References

1. Li, Y., Zhang, Z.: Network security risk loss assessment method based on queuing model. Comput. Simul. **4**, 258–262 (2021)
2. Cui, S.Y., Li, C., Chen, Z., Wang, J., Yuan, J.: Research on risk prediction of dyslipidemia in steel workers based on recurrent neural network and lstm neural network. IEEE Access **3**(99), 1–1 (2020)
3. Wang, S., et al.: Human short-long term cognitive memory mechanism for visual monitoring in IoT-assisted smart cities. IEEE Internet of Things J. **9**, 7128–7139 (2021). https://doi.org/10.1109/JIOT.2021.3077600
4. Zhou, X., Li, W., Wen, Z.: Data analysis for risk prediction of cervical cancer metastasis and recurrence based on DCNN-RF. J. Phys. Conf. Ser. **1813**(1), 012033 (2021)
5. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. Mob. Netw. Appl. **26**(5), 1891–1903 (2021). https://doi.org/10.1007/s11036-020-01725-x
6. Silva, G.M., Leo, L.D.S., Eller, C.C., et al.: Similar gaps, different paths? Comparing racial inequalities among BA holders in Brazil and the United States. Int. J. Comp. Sociol. **62**(5), 359–384 (2021)
7. Gao, P., Li, J., Liu, S.: An introduction to key technology in artificial intelligence and big data driven e-learning and e-education. Mob. Netw. Appl. **26**(5), 2123–2126 (2021). https://doi.org/10.1007/s11036-021-01777-7
8. Lee, S., Kim, J.H., Park, J., Oh, C., Lee, G.: Deep-learning-based prediction of high-risk taxi drivers using wellness data. Int. J. Environ. Res. Public Health **17**(24), 9505 (2020)
9. Li, D., Xu, J., Li, L.: Research on network lending risk analysis based on platform efficiency. J. Finan. Risk Manag. **10**(4), 453–472 (2021)
10. Liang, H., et al.: Data mining-based model and risk prediction of colorectal cancer by using secondary health data: a systematic review. Chin. J. Cancer Res. **32**(02), 124–133 (2020)