Research on Rapid Identification of Infringement Risk in Financial Technology Data Transactions

Sibei Liu

Northeastern University, Liaoning, Shenyang 110167, China

Abstract: Aiming at the problems of low accuracy rate of transaction information mining, high error rate of identification of infringement risk of financial technology data transaction and long identification time in current data transaction infringement risk identification methods, a new rapid identification method of infringement risk in financial technology data transactions is proposed. The entropy of clustering is determined by using coverage density and weighted coverage density to mine the transaction information of financial technology data. The BP algorithm is used to train the T-S fuzzy neural network, and the financial technology data transaction information is input into the trained T-S fuzzy neural network to obtain the quick identification result of the infringement risk of financial technology data transaction. The experimental results show that this method has a high accuracy rate of fintech data transaction information mining, a low error rate of fintech data transaction infringement risk identification, and a short recognition time.

key word: Financial technology data; Risk of transaction infringement; Risk identification; Coverage density; t-s fnn

1. Introduction

Up to now, the continuous improvement of informatization and digital technology has laid a solid foundation for the interconnection and interworking of high-quality data, but has put forward higher requirements for the data governance mechanism [1-2]. As the technology finance industry with the closest connection with data, it is easy to have economic disputes caused by infringement of financial technology data transactions during its transformation, which has seriously disrupted the order of the technology finance market and caused huge potential risks for financial technology information security [3]. Therefore, analyzing the challenges faced in the governance process of financial technology data, and quickly identifying the infringement risk of financial technology data transactions are of great significance for optimizing the market of financial technology data elements, comprehensively improving the data governance capability in this field, and timely avoiding the infringement risk of financial technology data transactions, such as stabilizing growth, adjusting structure, promoting transformation and benefiting the people's livelihood.

For the important research topic of identification of infringement risk of financial technology data transaction, reference [4] proposed a WBS-RBS based identification method of infringement risk of data transaction. Based on the WBS-RBS framework, this method analyzes the influencing factors of data transaction infringement risk, so as to analyze the whole process of data transaction and build a related WBS model. The RBS model is built in combination with the data transaction process, the two models are combined, and the data transaction information is input into the model to obtain the data transaction infringement risk identification. But in practical application, it is found that this method has the problem of high error rate of data transaction infringement risk identification. Reference [5] proposed a method to identify infringement risk of data transaction based on ANP structural model. Based on the transmission and feedback path of data transactions, this method obtains the stakeholders and risk sources that exist in this process, so as to establish an index system for identifying data transaction infringement risk. The super decision-making software is used to calculate the index weight, and the ANP structural model is used to obtain the identification results of data transaction infringement risk. However, this method has the problems of low accuracy of data transaction information mining and long time consuming of data transaction infringement risk identification, which still lags behind the ideal application effect.

However, there are some problems in the process of applying the above methods to the identification of infringement risk of financial technology data transactions, such as low accuracy of transaction information mining, high identification error rate and long time consuming. This paper aims to solve the problems existing in these methods and designs a new rapid identification method of infringement risk in financial technology data transactions.

2. Design of rapid identification method for infringement risk of financial technology data transaction

2.1 Financial technology data transaction information mining

Suppose there is a financial technology data transaction information set D, and the amount of data contained in the set is represented by N, and one data item set is represented by $I=\{I_1,I_2,...,I_m\}$. The number of categories of financial technology data transaction information in D is C_K [6-7]. Assuming that the number of data items is M_K, the number of financial technology data transaction information is N_K, and the total number of occurrences of all data items is S_K, the coverage density of C_K is calculated by the following formula:;

$$C_D(C_K)=\frac{S_K}{N_K\times M_K}=\frac{\sum_{j=1}^{M_K} occur(I_{kj})}{N_K\times M_K} \quad (1)$$

When the contribution of financial technology data transaction information and data item is uniform, $T_i=T=\frac{1}{N_K}$ and $W_j=W=\frac{1}{M_K}$ exist. At this time, $C_D(C_K)$ can be calculated by the following formula:

$$C_D(C_K)=T_i\times\sum_{j=1}^{M_K} occur(I_{kj})\times W_j=T\times W\times\sum_{j=1}^{M_K} occur(I_{kj}) \quad (2)$$

In the above formula, W_j represents the contribution of the data item [8-9]. The calculation formula is as follows:

$$W_j=\frac{occur(I_{kj})}{S_K} \quad (3)$$

Substitute formula (3) into formula (2) to get the calculation result of weighted coverage density. The calculation formula is as follows:

$$\begin{aligned} W_{C_D}(C_K)&=T\times\sum_{j=1}^{M_K} occur(I_{kj})\times W_j \\ &=\frac{1}{N_K}\times\sum_{j=1}^{M_K} occur(I_{kj})\times\frac{occur(I_{kj})}{S_K} \\ &=\frac{\sum_{j=1}^{M_K} occur(I_{kj})^2}{S_K\times N_K} \end{aligned} \quad (4)$$

Suppose that the frequency of data items in a financial technology data transaction information set is represented by X and the density of information set is represented by $f(X)$, then $C_D(C_K)$ and $W_{C_D}(C_K)$ can be calculated by the following formula:

$$C_D(C_K)=T\times E[X] \quad (5)$$

$$W_{C_D}(C_K)=T\times E[X^2] \quad (6)$$

In the above formula, $E[X]$ and $E[X^2]$ represent the first and second moments of $C_D(C_K)$ and $W_{C_D}(C_K)$ respectively.

Based on the above calculation results, the entropy [10] of clustering in the process of financial technology data transaction information mining is calculated. The specific calculation formula is as follows:

$$R=\sum_{j=1}^{M_K} p(X=I_{kj})\log p(X=I_{kj}) \quad (7)$$

When the occurrence frequency of data items in the financial technology data transaction information set is the same, $Var(X)=0$ is true, that is, the variance is the minimum. The entropy value of the financial technology data transaction information cluster is the highest. At this time, a higher quality clustering result is generated to realize the financial technology data transaction information mining.

2.2 Rapid identification of transaction infringement risk based on t-s fuzzy neural network

T-S fuzzy neural network structure [11-12] is shown in Figure 1.

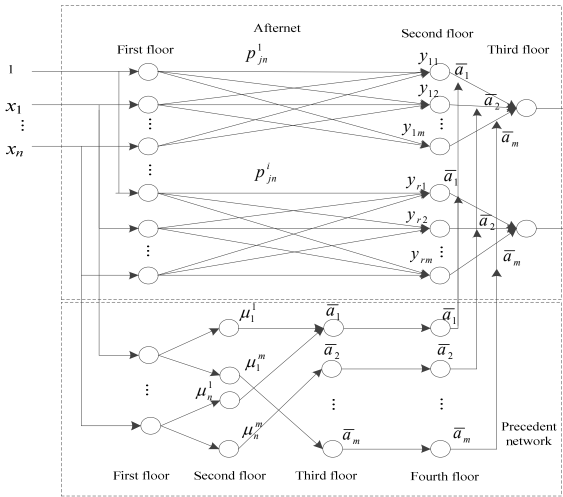

Figure 1 T-S fuzzy neural network structure

T-S fuzzy neural network structure is composed of antecedent network and consequent network. For the input rule part, the antecedent network is used to match, while the consequent network does weighted average processing for the output of each rule, so as to minimize the model error [13].

(1) Precedent network

The first layer is the input layer, and the input vector is represented by $x=\left[x_1,x_2,...,x_n\right]$.

The second layer is the fuzzification layer, which uses the Gaussian membership function as the fuzzification function [14-15] to ensure the fuzzification processing architecture. The specific calculation formula is as follows:

$$\mu_{ij}=e-\frac{\left(x_i-c_{ij}\right)^2}{\sigma_{ij}^2} \quad (8)$$

The third layer is the fuzzy rule layer. The membership function of the second layer is processed and calculated to obtain the incentive intensity of the fuzzy rule. The specific calculation formula is as follows:

$$\alpha_j=\left[\mu_1^{i_1},\mu_1^{i_2},...,\mu_n^{i_n}\right] \quad (9)$$

The fourth layer is the normalization layer, whose main task is to normalize the incentive intensity of fuzzy rules. The specific calculation formula is as follows:

$$\overline{\alpha}_j=\frac{\alpha_j}{\sum_{i=1}^{m}\alpha_i} \quad (10)$$

In the above formula, $\overline{\alpha}_j$ represents the ratio of the excitation intensity of the j-th fuzzy rule to the total excitation intensity. There are m nodes in this floor.

(2) Afternet

For r output variables, there are r subsequent subnets with the same structure, and each subsequent subnetwork corresponds to one network output.

The first layer is the input layer of rules, and the input variables are directly output to the next layer.

The second layer is the output layer of rules, and the output is fuzzy rules. The output result is expressed by the following formula:

$$y_{ij}=p_{j0}^i+p_{j1}^ix_1+...+p_{jn}^ix_n=\sum_{k=0}^{n}p_{jk}^ix_k \quad (11)$$

The third layer is the output layer of the subsequent subnetwork, which stacks all the valid results to produce a clear output. The calculation formula is as follows:

$$y_i=\sum_{j=1}^{m}\overline{\alpha}_jy_{ij} \quad (12)$$

BP algorithm is used to train T-S fuzzy neural network, and the specific implementation steps are as follows:

(1) Calculate the difference between T-S fuzzy neural network output and expected output by the following formula:

$$e=\frac{1}{2}\left(yd-ye\right)^2 \quad (13)$$

(2) The correction rule output coefficient is calculated by the following formula:

$$p_j^i\left(k\right)=p_j^i\left(k-1\right)-a\frac{\partial e}{\partial p_j^i} \quad (14)$$

Among them,

$$\frac{\partial e}{\partial p_j^i}=\frac{\left(yd-yc\right)\omega_i}{\sum_{i=1}^{n}\omega_i\times x_j} \quad (15)$$

In the above formula, $p_j^i\left(k\right)$ represents the output coefficient of the network rule, x_j represents the network input vector value, ω_i represents the excitation intensity of the input rule, and a represents the learning rate of the network.

(3) The modification results of membership function parameters are expressed by the following formula:

$$c_j^i\left(k\right)=c_j^i\left(k-1\right)-\beta\frac{\partial e}{\partial c_j^i} \quad (16)$$

$$b_j^i\left(k\right)=b_j^i\left(k-1\right)-\beta\frac{\partial e}{\partial b_j^i} \quad (17)$$

In the above formula, b_j^i and c_j^i represent the central position of the membership function and its correction vector respectively, and β represent the correction coefficient.

Input the trading information of financial technology data into the trained T-S fuzzy neural network, and get the quick identification result of infringement risk of financial technology data trading.

3. Experimental design

In order to verify the effectiveness of the rapid identification method for infringement risk of financial technology data transactions designed in this paper, relevant experimental tests were carried out. During the experiment, in order to ensure that the experimental results are representative, the experimental parameters need to be set, as shown in Table 1.

Table 1 Experimental parameter setting

Environmental parameters	Configuration requirements
Hardware environment	CPU:Inter(R)Core(TM)i5-8400 CPU@2.80Ghz Memory: 8G Hard disk 500G
The operating system	Microsoft Windows 10
The software platform	Pycharm, jupyter
Development of language	Python 3.6
Simulation software	Matlab 7.2

Take a large financial technology data trading platform as the research object, use the web crawler to obtain the background data of the financial technology data trading platform, and preprocess the obtained data. Test the sample set and experimental sample set respectively after processing the data, input the data in the test sample set into the simulation software for trial operation, and test the software operation. When the software operation status is normal. Input the data from the experimental sample set to the simulation software for experiment, and get the relevant experimental results.

The accuracy rate of financial technology data transaction information mining, the error rate of financial technology data transaction infringement risk identification and the identification time consumption are taken as experimental indicators to test the application effect of the reference [4] method, the reference [5] method and method of this paper.

Table 2 shows the comparison results of the accuracy of the three methods of financial technology data transaction information mining.

Table 2 Comparison of fintech data transaction information mining accuracy

Number of experiments	Reference [4] method/%	Reference [5] method/%	Method of this paper/%
10	84.7	85.6	98.7
20	85.6	87.5	96.8
30	86.3	86.4	97.5
40	81.4	87.2	97.6
50	79.6	84.6	98.6
60	75.6	85.3	97.6
70	78.5	87.2	98.1
80	84.1	84.3	96.8
90	88.7	85.2	97.6
average value	82.7	84.3	96.8

Analysis of the data in Table 2 shows that with the increase of the number of experiments, the accuracy rates of the three methods of financial technology data transaction information mining all show a fluctuating trend. Among them, the maximum accuracy rate of financial technology data transaction information mining by the reference [4] method is 88.7%, the minimum value is 75.6%, and the average value is 82.7%; the accuracy rate of financial technology data transaction information mining by the reference [5] method The maximum value is 87.5%, the minimum value is 84.3%, and the average value is 84.3%; the maximum accuracy rate of financial technology data transaction information mining of the method in this paper is 98.7%, the minimum value is 96.8%, and the average value is 96.8%. On the whole, the information mining accuracy of the method in this paper is higher, and the mining results are more accurate.

Figure 2 shows the comparison results of the error rate of infringement risk identification in financial technology data transactions of the three methods.

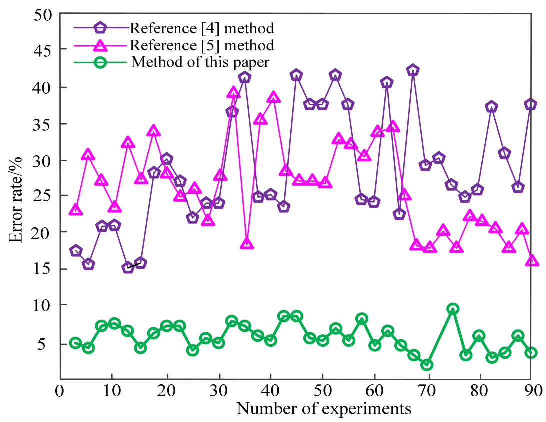

Figure 2 Comparison of error rate of infringement risk identification of financial technology data transactions

Analyzing the results in Figure 2, it can be seen that the error rate of infringement risk identification of financial technology data transactions in the reference [4] method varies between 42 and 45%, and the error rate of infringement risk identification of financial technology data transactions in the reference [5] method is variation between 16 and 39%. Compared with the experimental comparison methods, the error rate of the method in this paper varies between 2 and 10%, and the error rate is lower, indicating that the method has a higher identification accuracy of the infringement risk of financial technology data transactions.

The time-consuming identification of infringement risks in financial technology data transactions by three methods is compared, and the results are shown in Table 3.

Table 3 Time consuming comparison of infringement risk identification of financial technology data transactions

Number of experiments	Reference [4] method/ms	Reference [5] method/ms	Method of this paper/ms
10	356.4	158.9	69.8
20	263.9	178.3	68.7
30	389.1	146.1	83.2
40	369.7	156.3	75.6
50	278.4	155.7	74.3
60	269.8	149.3	72.9
70	243.6	169.8	69.8
80	239.8	158.1	67.1
90	257.4	163.9	68.4
average value	296.4	159.6	72.2

From the analysis of the data in Table 3, it can be seen that the average time-consuming time of fintech data transaction infringement risk identification by the reference [4] method is 296.4ms, and the average time-consuming of the reference [5] method is 159.6 ms. ms, the average time taken to identify the infringement risk of financial technology data transactions in this method is 72.2ms, which is the shortest identification time among the three methods, which can achieve the important goal of rapid identification of financial technology data transaction infringement risks.

4. Conclusion

With the continuous improvement of social productivity, the volume of financial technology data is showing a rising trend, but the following is the increasing risk of infringement of financial technology data transactions, which brings severe challenges to data risk supervision. Aiming at the problems existing in the current methods for identifying the infringement risk of financial technology data transactions, such as the low accuracy rate of financial technology data transaction information mining, the high error rate of identifying the infringement risk of financial technology data transactions, and the long identification time, a new method for rapid identification the infringement risk of financial technology data transactions is proposed. The experimental results show that the accuracy rate of financial technology data transaction information mining is high, the error rate of financial technology data transaction infringement risk identification is low, and the identification time is short, which can achieve the important goal of rapid identification of financial technology data transaction infringement risk. This method effectively breaks through the bottleneck of the development of financial science and technology, promotes the rapid development of financial science and technology, and thus provides data support for providing users with financial science and technology products and services.

References

1. Feng H. Application of Big Data Blockchain Technology in Financial Support of Real Economy[J]. Journal of Physics: Conference Series, 2021, 1955(1):012018-012029.
2. Li Z. The Path Analysis of Data Mining Technology in Internet Financial Asset Management[J]. Francis Academic Press, 2021, 15(10):1-9.
3. Du G, Liu Z, Lu H. Application of innovative risk early warning mode under big data technology in Internet credit financial risk assessment[J]. Journal of Computational and Applied Mathematics, 2021, 386(12):113260-113272.
4. Wang W, Zhang M J Risk Identification of Data Transaction Infringement Based on WBS-RBS[J]. Information Studies:Theory & Application, 2021, 44(1): 83-88.
5. Li C, Tao L, Zhang W D. Construction of Evaluation Index System for Copyright Infringement Risks of Paid Contents of Video Knowledge[J]. Information Studies:Theory & Application, 2021, 44(3):84-90, 108.
6. He H, Hong Y, Liu W, et al. Data mining model for multimedia financial time series using information entropy[J]. Journal of Intelligent and Fuzzy Systems, 2020, 15(1):1-7.
7. Zhang Z. Mining method of massive data of mobile library under information asymmetry facing large-scare database[J]. Microprocessors and Microsystems, 2021, 81(7):103730-103741.
8. Bakiyalakshmi S, Priyadharshini R. A Random Forest Fraud Transaction Detection System using Data Mining Techniques[J]. JETIR(www.jetir.org), 2021, 8(11):1-6.
9. Xu B, Huang D, Mi B. Research on E-Commerce Transaction Payment System Basedf on C4.5 Decision Tree Data Mining Algorithm[J]. Computer Systems Science and Engineering, 2020, 35(2):113-121.
10. Mukherjee A, Chaki R, Chaki N. Data mining-based hierarchical transaction model for multi-level consistency management in large-scale replicated databases[J]. Computer Standards & Interfaces, 2021, 74(3):103485-103496.
11. Jiang X Q, Chen W F, Guo L J, et al. Application of T-S Fuzzy-Neural Network Model in Water Quality Comprehensive Evaluation[J]. Procedia Computer Science, 2020, 166(1):501-506.
12. Chen J, Gu S, Chen G, et al. Water Resources Carrying Capacity Assessment in Dali Prefecture Based on T-S Fuzzy Neural Network Model[J]. IOP Conference Series Earth and Environmental Science, 2020, 525(1):012104-012125.
13. Zhang S Z, Li P L, Zhong B Z, et al. Speed Grade Evaluation of Public-Transportation Lines Based on an Improved T-S Fuzzy Neural Network[J]. Hindawi Limited, 2020, 12(8):8838084.1-8838084.13.
14. Wang N , Pedrycz W , Yao W , et al. Disjunctive Fuzzy Neural Networks: A New Splitting-Based Approach to Designing T-S Fuzzy Model[J]. IEEE Transactions on Fuzzy Systems, 2020, 30(2):370-381.
15. Wei L, Jqa B, Xjz C, et al. Identification and simplification of T-S fuzzy neural networks based on incremental structure learning and similarity analysis[J]. Fuzzy Sets and Systems, 2020, 394(1):65-86.