Risk Data Analysis of Cross Border E-commerce Transactions Based on Data Mining

Haohua Qing¹*, Guokai Zheng¹, Die Fu²

¹Department of Economic And Trade, Guangzhou College of Technology and Business, Guangzhou, Guangdong, 510850, China
²Sichuan University of Arts and Science, Dazhou, Sichuan, 635002, China

*Corresponding author e-mail: 13986536@asau.edu.cn

Abstract: At present, China's foreign trade volume of cross-border e-commerce is still relatively small, the targeted legal system has not been established, and the industry access threshold is relatively low. The development of cross-border e-commerce faces many risks. Relying on data mining technology, data docking and risk control. At the data level, the risk of cross-border e-commerce transaction is processed and analyzed, and the potential risk is predicted, and more targeted risk control methods are proposed to reduce the risk of cross-border e-commerce transaction business. Aiming at this problem, this paper establishes the risk data analysis of cross-border e-commerce transactions based on data mining. In order to further verify the data accuracy of data mining technology, this paper analyzes the test values of neural network algorithm, seizure rate and inspection rate. When the misclassification loss parameter value is increased from 1 to 4, the detection rate of the model increases and the detection rate decreases. When the misclassification loss value is higher than 4, the detection rate increases, but the seizure rate decreases significantly. Therefore, the target inspection rate of the model can be changed by adjusting the misclassification loss parameter value of the model. If we set the parameter value of misclassification loss reasonably, we can achieve the goal of preventing and controlling the maximum risk with the lowest data mining cost. Through the analysis, the research in this paper has achieved ideal results, and made a contribution to data mining in cross-border e-commerce transaction risk data analysis.

Keywords: Data Mining, Cross-border E-commerce, Risk Prevention and Control, Online Finance

1. Introduction
Data mining, also known as knowledge discovery in database, is a research hotspot in the field of artificial intelligence and database. The so-called data mining refers to the non-trivial process of
revealing hidden and potential valuable information from a large number of data in the database [1-2]. Data mining is a decision support process. It is mainly based on artificial intelligence, pattern recognition, visualization and other technologies, highly automated analysis of enterprise data, inductive reasoning, and mining potential patterns, help decision makers adjust market strategies, reduce risk, and make correct decisions [3-4]. The commonly used data analysis methods of data mining include classification, neural network, association rules, change and deviation analysis, etc. They mine data from different angles.

Cross border e-commerce has broken the space-time restrictions of traditional trade, and building a trading platform with Internet technology has become a new situation of China's import and export trade, and has effectively promoted the development of China's foreign trade and improved the transaction efficiency [5-6]. Before export, the downward pressure of international trade increases, and cross-border e-commerce is expected to become a new growth engine for China's trade and even the global economy [7-8]. Under the development trend of cross-border e-commerce in China, various risks are gradually emerging. Risk prevention should be done in advance to standardize the development of cross-border e-commerce [9-10].

This paper analyzes the actual situation of the application of data mining in cross-border e-commerce transaction risk data, and finds that compared with developed countries; there are still deficiencies in technology promotion and technical support. Therefore, this paper establishes the risk data analysis of cross-border e-commerce transactions based on data mining. Through the research of data mining technology, this paper analyzes the risk data of cross-border e-commerce transactions from the aspects of transaction form, risk assessment, regulatory module and so on. Through the analysis of the test results, this paper believes that the use of data mining technology can reduce the risk of cross-border e-commerce transactions, using data mining technology cross-border business transaction risk prevention and control has achieved good results.

2. Overview of Data Mining and Neural Network Algorithm

2.1 Overview of Data Mining

Data mining is a part of knowledge discovery (KDD), which is closely related to artificial intelligence, machine learning and deep learning. The problem to be solved is to extract potential, effective and understandable information and knowledge from a large number of messy data. Data mining can not only learn from known data, but also from unpredictable information. The difference between data analysis based on data mining and traditional data processing is that data mining technology can extract effective information without certain assumptions. At present, data mining technology has achieved milestone practice in credit risk assessment, manufacturing industry, quality prediction, genetic medicine data, judicial expertise and financial industry. Because the data with category label is used in e-commerce transaction risk assessment, this paper briefly introduces the classification algorithm of data mining. The commonly used classification algorithms are: random forest algorithm, support vector machine algorithm and neural network algorithm.

2.2 Neural Network Algorithm

Neural network is an active interdisciplinary subject developed on the basis of many subjects, which is used to simulate the structure and function of human brain nervous system. It is called a neural network because the structure used in this mathematical model is similar to that of the synapses in the brain. The mathematical model of neural network is a kind of structure formed by interconnection of "neural" nodes of big data, and it is also a kind of operation model. BP neural network is a multi-layer feed forward network with unidirectional propagation, which adopts back propagation algorithm. The number of neurons in the input layer is determined by the dimension of the input data, and the number of neurons in the output layer is determined by the classification tree of the sample set:

Output of the hidden layer node is:
\[ O_j = f \left( \sum X_i W_{ij} + \theta_j \right) \]  
(1)

Output of the output layer node is:

\[ Y_k = f \left( \sum O_j V_{jk} + \phi_k \right) \]  
(2)

Error is:

\[ E = \frac{1}{2} \sum_{m} \sum_{n} \left( o_{mn} - t_{mn} \right)^2 \]  
(3)

Where, \( X \) is the input, \( Y \) is the output, \( O \) is the output of the hidden layer, \( W \) is the weight between the input layer and the hidden layer, \( V \) is the weight between the hidden layer and the output layer, \( \theta \) is the offset of the hidden layer, \( \phi \) is the offset of the output layer, and the function \( f \) is the function \text{Sigmoid}.

3. Empirical Study on Risk Classification Test Model of Cross Border E-commerce Transactions Based on Data Mining

Accurate supervision of cross-border e-commerce requires the collection and classification of risk information of cross-border e-commerce supervision. However, the classification of risk level is more qualitative, that is, it is difficult to classify the risk level more accurately and quantitatively according to whether the risk information has a significant impact on cross-border e-commerce supervision. Through the mining and analysis of massive cross-border e-commerce big data, the customs can find the risk rules and characteristics of cross-border e-commerce supervision in big data, and classify and predict the risk of cross-border e-commerce supervision, so as to quantitatively divide the risk level of cross-border e-commerce supervision, and realize the quantitative division of cross-border e-commerce supervision risk level, so as to supervise more accurately cross-border e-commerce.

This paper uses predictive data mining method, as shown in Table 1, to analyze the customs declaration form data of Tianjin Customs in 2019. According to the important feature of customs seizure, the customs declaration data are classified and labeled according to whether they are seized or not, and a decision tree model for risk classification and prediction of customs declaration supervision is established to conduct empirical analysis on risk classification and prediction of customs declaration supervision, and extract risks. The relevant rules of classification can provide reference for risk assessment and prediction of existing customs supervision. According to the analysis of test results in Figure 1, among the actual number of seized customs declaration forms, the predicted number of seized customs declaration forms is 231, and the number of undetected seizures is 1335. The accuracy rate of prediction is 15.2%, while the inspection rate is 2.64%, and the seizure rate is 90.6%. Therefore, although the accuracy rate of prediction is low, it achieves a higher seizure rate with a lower detection rate.

| No seized "0" predicted | "1" is predicted to be seized |
|------------------------|-----------------------------|
| No "0" was found        | 7401                        | 24 |
| "1" actually found      | 1335                        | 231|

4. Discussion

4.1 Analysis on the Background of Cross Border E-commerce Risks in China
(1) Export scale of cross-border e-commerce is expanding year by year
As the goods sold by cross-border e-commerce enterprises in China have good performance price ratio and simple and convenient transaction process, many traditional cross-border foreign trade enterprises gradually transfer their sales activities from offline to online and become cross-border e-commerce, which accelerates the optimization and upgrading of China's traditional trade structure.

(2) B2B is the main mode of export
Under the B2B mode of cross-border e-commerce, cross-border enterprises, as the main body of transactions, carry out product publicity, consultation and transaction through the network and cross-border trading platform. This kind of transaction is usually used for commodities, and the order amount is huge. Therefore, B2B mode has always been the most important transaction mode of cross-border e-commerce industry. This mode realizes the docking between the buyer and the seller. Buyers can understand the goods of different countries, eliminate the tedious intermediate links, reduce the operating costs, promote the rise of cross-border e-commerce industry, change the shopping habits of buyers, and improve the user experience. Through some e-commerce platforms, transactions between individuals can be realized directly. This mode needs to be carried out under the premise of ensuring the credit of both sides, but it can better meet the personalized consumption needs of consumers. The diversification of cross-border e-commerce trade mode makes the transaction process of cross-border e-commerce more complex, which is prone to a series of risks such as payment, logistics and credit.

(3) Low industrial concentration
Statistics show that in 2018, 86% of the sellers sold less than US $2.45 million, only 2.1% of the sellers sold more than US $11 million, and more than 75% of the total sellers sold less than 100000 units annually. In terms of sales volume and sales volume, small sellers still account for the mainstream of the market, indicating that China's cross-border e-commerce export industry concentration is low. As competition intensifies and capital investment increases, this situation may be improved.

(4) Trade structure is mainly export-oriented and unevenly distributed
In recent years, the scale of cross-border e-commerce transactions in China has been expanding. The living environment of China's traditional export and foreign trade enterprises is worrying. The rise of cross-border e-commerce industry has greatly promoted the transformation of foreign trade enterprises. Driven by the fierce competition of e-commerce and the advantages of government policies, China's cross-border e-commerce has achieved economic recovery through export trade. However, in recent years, the proportion of cross-border e-commerce exports has decreased, while the scale of cross-border imports has increased.

According to the analysis of Figure 1, China's cross-border trade structure is still dominated by exports, accounting for more than 76% of the transaction scale. However, in recent years, the proportion of cross-border e-commerce exports has decreased, from 81% in 2016 to 78.1% in 2019, while the scale of cross-border import has increased, accounting for 21.9% in 2019 from 19% in 2016. This also reflects the relatively broad development space of cross-border e-commerce import in China from another perspective.
Figure 1: China's cross border e-commerce import and export trade share in 2016-2019

In addition, this paper further tests the performance of the system. Figure 2 shows the graph of the model's capture rate and detection rate change due to different misclassification loss parameter values. The default misclassification loss parameter value of the model is 1. We can see that when the misclassification loss parameter value is increased from 1 to 4, the detection rate of the model increases, while the detection rate decreases, when the misclassification loss value is higher than 4 The detection rate increased significantly, while the seizure rate decreased significantly. We find that the target inspection rate of the model can be changed by adjusting the misclassification loss parameter value of the model. If we set the parameter value of misclassification loss reasonably, we can achieve the goal of preventing and controlling the maximum risk with the lowest cost of regulatory resources.

Figure 2: comparison of inspection rate and seizure rate under different misclassification cost
4.2 Coping Strategies for Cross Border E-commerce Risks

(1) Raise the entry threshold of cross-border e-commerce industry and strengthen market supervision
From the causes and specific manifestations of cross-border e-commerce risks in China, the number of cross-border e-commerce enterprises in China has increased significantly, but the product quality is uneven, the market order is chaotic, resulting in fraud and other credit risks, which is not conducive to the development of the industry.

(2) We will improve the customs, taxation and quarantine mechanisms and improve customs clearance services
Goods of cross-border e-commerce, especially those of cross-border retail e-commerce, are subject to customs inspection, tax collection and goods quarantine. Therefore, the government should establish a higher-level customs supervision mode, simplify the customs clearance process, reasonably plan the window, implement the new policy of cross-border e-commerce tax collection and management, establish a perfect e-commerce export inspection mode, strictly abide by the safety standards and the requirements of national laws and regulations, conduct quality inspection and safety inspection on products, and centrally handle the corresponding procedures.

(3) Promoting the strategy of talent training and improving the quality of employees
In order to reduce the risk of cross-border e-commerce enterprises in China, the government should have a clear direction and positioning for cross-border e-commerce personnel training, understand the industry development law, and cultivate industry leading talents. During the implementation of the policy, colleges and universities can set up cross-border e-commerce personnel training mode, teaching materials, business practice teaching and other courses to cultivate high-end cross-border e-commerce talents integrating e-commerce knowledge, international trade and foreign language ability. The state should also strive to provide them with employment opportunities and development space.

5. Conclusions
Cross border e-commerce risk is a variety of risk events that lead to the increase of cost, the decrease of economic profits and the loss of interests of cross-border e-commerce export enterprises in cross-border trade activities. Based on the data mining of cross-border e-commerce transaction risk data analysis, this paper from the data mining algorithm, transaction risk classification and other aspects of cross-border e-commerce transaction risk analysis. Through the empirical study and analysis of the risk classification test model of data mining, it is concluded that data mining can play an ideal role in predicting and seizing the risk data of cross-border e-commerce transactions. Although the accuracy of the prediction needs to be improved, it has achieved a higher seizure rate of 90% with a lower inspection rate. Therefore, the value of data mining has been recognized by cross-border e-commerce. In the background analysis of cross-border e-commerce risks, we compare and analyze the inspection rate and seizure rate under different misclassification loss costs. Through the analysis, we find that the target inspection rate of the model can be changed by adjusting the misclassification loss parameter value of the model. If we set the parameter value of misclassification loss reasonably, we can completely achieve the goal of preventing and controlling the greatest risk with the lowest cost of regulatory resources, which has obvious advantages over the traditional data. This study has achieved ideal results and provided technical support for data mining of cross-border e-commerce transaction risk data analysis.

References
[1] Zheng Y. Trajectory Data Mining: An Overview[J]. ACM Transactions on Intelligent Systems
and Technology, 2015, 6(3):1-41.

[2] Wilson, A. M., Thabane, L., & Holbrook, A. (2015). Application of data mining techniques in pharmacovigilance. Br J Clin Pharmacol, 57(2), 127-134.

[3] Buczak, A., & Guven, E. (2017). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.

[4] Helma, C., Cramer, T., Kramer, S., & Deraedt, L. (2018). Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. J Chem Inf Comput, 35(4), 1402-1411.

[5] Valarezo, A., Perez-Amaral, T., Garin-Munoz, T., Herguera Garcia, L., & Lopez, R. (2018). Drivers and barriers to cross-border e-commerce: evidence from spanish individual behavior. Telecommunications Policy, 42(6), 464-473.

[6] Lin, A. J., Li, E. Y., & Lee, S. Y. (2018). Dysfunctional customer behavior in cross-border e-commerce: a justice-affect-behavior model. Journal of Electronic Commerce Research, 19(1), 36-54.

[7] Yang, N., Fu, J., & Wang, Y. (2017). Study of the impact of the cross-border e-commerce model based on the belt and road on china's international trade system. Revista De La Facultad De Ingenieria, 32(8), 490-496.

[8] Wang, Y. K. (2017). Model for evaluating the logistics service quality of cross-border e-commerce enterprises with intuitionistic fuzzy information. Journal of Computational and Theoretical Nanoence, 14(2), 1136-1139.

[9] Rui-Fen, L. I. (2015). Analysis on the cross-border e-commerce logistics distribution situation and development trend. Value Engineering, 52(3), 143-147.

[10] Choi, H. J., Jung, H. J., & Lee, D. H. (2019). A study on port improvement with the activation of cross-border e-commerce: a study of pyeongtaek port. Journal of Korea Trade, 23(7), 34-44.