Research on the Control System and Risk Management Based on Internet Big Data and Cloud Computing

Xiujun Wang\(^1\), Dengchen Yu\(^2\)*, Fei Zhang\(^2\) and Xin Li\(^2\)

\(^1\)School of Business Administration, JiLin Engineering Normal University, Changchun 130052, China.
\(^2\)School of Economics and Trade, JiLin Engineering Normal University, Changchun 130052, China

*Corresponding author: whw_676893@jlenu.edu.cn

Abstract. The advancement of Internet information technology and the strong demand for private finance have promoted the vigorous development of Internet financial network lending. However, due to the complexity of the network environment and the volatility of the financial market, Internet finance has a higher risk. Aiming at the particularity of Chinese Internet financial big data platform, this article uses Internet lending as a case to screen out 15 financial risk indicators. The article uses the principle of game theory to set the principal component factor as a variable, and the operation of the platform in 2018 as the dependent variable to establish a financial risk early warning system for the online lending platform. At the same time, the paper cited cloud computing to optimize its performance, and designed a financial risk management and control system based on cloud computing. Construct system test links to test system performance. With the same amount of data and the same test platform, the computing performance of this system is better than that of the original system.

1. Introduction
In the environment of big data, the speed at which companies obtain information and wider channels have caused major changes in financial risk management. However, not all of these types and huge amounts of information are all valuable, and there are bound to be many useless or lost for corporate decision-making [1]. For time-sensitive information, companies should use data mining technology to filter out valuable information from a large amount of data, accurately identify and reduce various financial risks faced by the company, innovate financial management models, and improve corporate economic benefits.

Internet lending is a type of Internet finance. This mode of using Internet platforms to meet the needs of microfinance is a relatively form of Internet finance in China, which has promoted the process of HP Finance to a certain extent [2]. The risk early warning mechanism cannot only manage financial risks, but also diagnose the operating conditions of the platform. Therefore, the construction of financial risk warning for online lending platforms is a practical problem that China's online lending industry urgently needs to solve.
2. Construction of financial risk early warning indicator system for Internet financial companies under the big data environment

Because the operating model and profit model of online lending platforms are different from those of mature listed companies, the widely used financial risk warning indicators should not be directly applied. Therefore, the selection of indicators should comprehensively consider the information disclosure situation and the reality of online lending platforms. The CAMEL evaluation system is a system for assessing the overall level of financial institutions. It evaluates the business operations and credit levels of financial institutions through five aspects: capital adequacy, asset quality, management level, profitability and liquidity [3]. Online lending platforms involve lending behaviour and are similar to commercial banks, and there are many overlaps in their capital quality risks, asset adequacy risks, profitability risks, management level risks, and liquidity risks. Therefore, the use of the CAMEL evaluation system to screen financial risk indicators conforms to the principle of comprehensiveness in selecting indicators.

However, there are differences between online lending platforms and commercial banks, and their risk control capabilities and capital restraint capabilities are quite different. Therefore, the selection of financial risk indicators should show the characteristics of the industry. The financial risk early warning system of this article takes the five aspects of the CAMEL evaluation system as the first-level indicators, and then selects 15 indicators from the specific indicators as shown in Table 1.

| Indicator type          | Indicator code | Indicator name                        |
|-------------------------|----------------|---------------------------------------|
| Capital adequacy risk   | X1             | Registered capital                    |
|                         | X2             | Guarantee agency                      |
|                         | X3             | Cumulative pending balance            |
|                         | X4             | Number of targets                     |
| Asset quality risk      | X5             | Project quality                       |
|                         | X6             | Bank depository                       |
| Management level risk   | X7             | Executive education                   |
|                         | X8             | Executive experience                  |
|                         | X9             | 7-day average interest rate           |
| Profitability risk      | X10            | 7-day volume                          |
|                         | X11            | Number of investors in 7 days         |
|                         | X12            | Number of borrowers in 7 days         |
|                         | X13            | Balance to be repaid in the next 60 days |
| Liquidity risk          | X14            | 30-day net inflow of funds            |
|                         | X15            | Average loan period                   |

3. Analysis of the game behaviour between the platform and the borrower

3.1. Hypothetical model

Hypothesis 1: Both micro-enterprise borrowers and online lending platforms are randomly matched by large groups, and both meet the bounded rationality condition.

Hypothesis 2: Both parties of the game are playing the game under complete market conditions, and there is no government intervention.

Hypothesis 3: The benefit between the two parties of the game is complete information, and the game is an asymmetric evolutionary game.

Hypothesis 4: The online loan platform has two behaviours: agreeing to the loan and rejecting the loan. The probability of agreeing to the loan is $p_1$, and the probability of rejecting the loan is $1-p_1$. Micro-enterprises also have two behaviours: repaying loans and repaying debts. The probability of micro-enterprises repaying loans is $p_2$, and the probability of repaying debts is $1-p_2$. 
Hypothesis 5: The principal of the loan is $L$, the loan interest rate is $r$, and $S$ is some costs, including screening, payment, etc., that occur during the entire lending process of the online lending platform.

### 3.2. Analysis of Stable Strategy

For the online lending platform, the income of the agreed loan is:

$$\pi_{a1} = p_2 (Lr - s) + (1 - p_2) (-L - S)$$

The income of rejecting the loan is: $\pi_{a2} = 0$.

Therefore, the expected income of the online loan platform:

$$\pi_{a} = p_1 \pi_{a1} + (1 - p_1) \pi_{a2} = p_1 p_2 (Lr - s) - p_1 (1 - p_2) (L + S)$$

Applying the replication dynamic equation to the online lending platform, the percentage change rate at which the online lending platform chooses to agree to the loan is obtained:

$$F(p_1) = \frac{dp_1}{dt} = p_1 (\pi_{a} - \pi_{a1}) = p_1 (1 - p_1) [p_2 (Lr - s) + (1 - p_2) (-L - S)]$$

Let $F(p_1) = 0$, get two stable states $p_1^* = 0$, $p_1^* = 1$. For all $p_1$s, it is a stable state, which means that the initial proportion of whether the online lending platform agrees to the loan is stable, and the online lending platform adopts a mixed strategy [4]. To sum up, the absence of a sound credit system and legal system, online lending platforms appearing on the market to issue loans, and honest repayment of loans by micro-enterprises is just an unstable strategy. In the end, the entire market will tend to be a complete failure where companies cheat on accounts and platforms no longer lends.

### 4. Financial risk management and control system design

#### 4.1. Hardware design

In this hardware design, the design is aimed at cloud computing network construction and a single cloud computing host. In the design of the cloud computing host, a host with high computing power is selected as the basis of the design. Install the corresponding high-precision computing chip in this host, and improve the computing power of the system through the chip.

#### 4.2. Software design

In view of the poor computing power of the original financial risk management and control system, the cloud computing method is used to improve the data processing capacity of the system, and the system software modules after applying the cloud computing method are shown in Figure 1. According to the above software module framework, the software design is completed [5]. In the software design, the cloud computing method is used to optimize the original calculation part, and the optimized module is connected with other modules to realize the computing performance of the system.
Figure 1. System software module diagram.

The paper chooses SaaS service as the service mode of this cloud computing method. This method can effectively reduce the cost of the enterprise and lower the threshold and risk of enterprise informatization. The implementation of the cloud computing method is divided into three parts: cloud collection of financial data, cloud processing, and financial abnormal data. In the cloud collection part, the big data network is used to complete the acquisition of financial information, and the financial information is used to complete the processing process using a distributed computing method to obtain the abnormal data part of the financial data. At this point, the cloud computing part of the design is completed.

4.3. Realization of financial risk assessment process
The paper adopts the risk level assessment method to carry out risk assessment on the abnormal financial data processed by cloud computing. The information is initially processed in the form of an indicator system. In this risk assessment, five scoring levels will be set, namely, no risk, low risk, medium risk, increased risk, and major risk [6]. The standard coefficients for setting 5 levels are: 1, 0.8, 0.6, 0.4, 0.2. Assuming that the first-tier basis of the evaluation level is divided into A, the second-tier basis is divided into B, and the actual value of the score is O, the evaluation efficiency coefficient H can be obtained as:

\[ H = \frac{(O - B)}{(A - B)} \]  \hspace{1cm} (4)

From formula (4), the adjustment score T of the level can be obtained as:

\[ T = H \times (A - B) \]  \hspace{1cm} (5)

The single index score of a known level is the sum of the basic score and the adjustment score, and the comprehensive score is the sum of the individual scores, then the grade score is:

\[ w = \frac{H \times \sum T}{\sum T \times J} \]  \hspace{1cm} (6)
In the formula, J is set as the index weight. The risk level of financial information is calculated by the above formula, and the content of the level is set in the form of a table, as shown in Table 2.

| Evaluation index interval | Risk status   | Alarm         |
|---------------------------|---------------|---------------|
| 0.85-1.0                  | No risk       | No police     |
| 0.70-0.85                 | Focus on risk | Light police  |
| 0.5-0.7                   | Less risk     | Police        |
| 0.3-0.5                   | Greater risk  | Heavy warning |
| 0-0.3                     | Significant risk | Giant police |

5. Simulation of risk early warning system

The key independent variables are selected from the loan information by regression analysis, and the obtained coefficients pass the significance test. There is no multicollinearity between the variables, so the positive and negative coefficients can be assumed to be consistent with the actual situation, that is, the direction of the independent variable's influence on the dependent variable Consistent with the actual situation [7]. On this basis, the obtained coefficients can be used for regression prediction analysis. This article uses a decision tree algorithm to replace the traditional regression prediction method in the prediction. Based on the LCIS_2.csv data set containing 291336 samples and 39 fields, in order to apply the machine learning CART algorithm, the data set is divided into train, training set and test set of csv and test. Both the training set and the test set contain 10 fields including loan amount, loan period, initial rating, loan type, first bid, historical successful loan amount, my investment amount, principal repaid, principal to be repaid, and bad debt status. The sample size of the training set is 250,000, and the sample size of the test set is 41336.

Decision tree algorithm classification and regression tree (CART) is a relatively effective non-parametric classification and regression method. It achieves the purpose of prediction by constructing a binary tree model [8]. It uses a completely different method from traditional statistics to construct prediction criteria, and is given in the form of a binary tree, which is easy to understand, use and explain. In many cases, the prediction tree constructed by the CART model is more accurate than the algebraic model prediction criteria constructed by traditional statistical methods, and the more complex the data and the more variables, the better the algorithm. The key to the model lies in the construction of the prediction criteria. Here we use the decision tree CART algorithm in the python machine learning library sklearn to analyze the samples. Import train.csv and test locally as training samples and test samples, and extract the independent variables and dependent variables X_train, X_test, y_train and y_test (all in vector form) from the training samples and test samples, to get the training Confusion matrix on the set and test set (see Table 3).

| Use case number | The amount of data | Data format |
|-----------------|--------------------|-------------|
| Test 1          | 1000               | CHAR        |
| Test 2          | 3000               | CHAR        |
| Test 3          | 5000               | CHAR        |
| Test 4          | 7000               | CHAR        |
| Test 5          | 10000              | CHAR        |
| Test 6          | 12000              | CHAR        |
| Test 7          | 15000              | CHAR        |
| Test 8          | 17000              | CHAR        |
| Test 9          | 20000              | CHAR        |
| Test 10         | 30000              | CHAR        |
According to the accuracy rate of the bad debt status, it is predicted that 86% of all samples marked as bad debts by the borrower are indeed bad debts; from the recall rate, it can be seen that only 13% of all samples of the borrowers marked as bad debts are actually bad debts. The results obtained from this sample data show that the accuracy of predicting the borrower’s bad debts can reach 86%, but the prediction is conservative (see Table 4).

Table 4. Precision rate and recall rate.

| Accuracy/Total | Accuracy  | Recall rate | Comprehensive evaluation index | Number of samples |
|----------------|-----------|-------------|--------------------------------|-------------------|
| 0.86           | 0.99      | 1.00        | 0.99                          | 40824             |
| 0.99           | 0.99      | 0.98        | 0.98                          | 41336             |

6. Conclusion
Constructing a scientific and effective risk warning model for Internet financial platforms is of great practical significance for identifying and warning platform risks as soon as possible, and taking effective measures in a timely manner to prevent and defuse risks. Some traditional risk early warning models mostly use mathematical statistical analysis, logistic regression and other methods to predict. However, due to some difficult nonlinear relationships and excessive reliance on historical data and other constraints, these methods are not suitable for Internet finance, a new financial on the issue of risk early warning of the model. Therefore, the introduction of game theory into the risk warning research of Internet financial platforms in this article provides a brand-new idea for the risk management of the platform.

References
[1] Nurfalah, I., Rusydiana, A. S., Laila, N., & Cahyono, E. F. Early warning to banking crises in the dual financial system in Indonesia: The markov switching approach. Journal of King Abdulaziz University: Islamic Econ., 31(2) (2018) 133-156.
[2] Park, E., Kim, D. H., & Kim, K. The Study on the Development of the Financial Sector Early Warning System. Journal of Economic Theory and Econometrics, 28(3) (2017) 36-67.
[3] Jemović, M., & Marinković, S. Determinants of financial crises—An early warning system based on panel logit regression. International Journal of Finance & Economics, 26(1) (2021) 103-117.
[4] Gholizadeh, A., Fallah Shams, M. F., & Afshar Kazemi, M. A. The Designing Early Warning System of Financial Crisis Outbreak in Tehran Stock Exchange by Logit & Probit Model. Financial Knowledge of Securities Analysis, 13(48) (2021) 135-148.
[5] Zhang, R., Xian, X., & Fang, H. The early-warning system of stock market crises with investor sentiment: Evidence from China. International Journal of Finance & Economics, 24(1) (2019) 361-369.
[6] Wang, Q., Hui, F., Wang, X., & Ding, Q. Research on early warning and monitoring algorithm of financial crisis based on fuzzy cognitive map. Cluster Computing, 22(2) (2019) 3689-3697.
[7] Tang, Y. Financial Risk and Early Warning Based on Qingdao Marine Economic Forecast. Journal of Coastal Research, 112(1) (2020) 195-198.
[8] Naeem, M. Z., Spulbar, C., Ejaz, A., Birau, R., Gorun, T. H., & Rebegea, C. Measuring the effectiveness of signals approach in an early warning system for crises and its impact on textile industry: a case study for South-East Asia. Industria Textila, 72(2) (2021) 175-190.