Big Data Credit Report in Credit Risk Management of Consumer Finance

Lu Gao¹ and Jian Xiao²

¹School of Business, Shandong University, Weihai 264209, China
²School of Business, Beijing International Studies University, Beijing 100024, China

Correspondence should be addressed to Jian Xiao; xiaojian@bisu.edu.cn

Received 14 April 2021; Revised 9 May 2021; Accepted 22 May 2021; Published 16 June 2021

Academic Editor: Wenqing Wu

Copyright © 2021 Lu Gao and Jian Xiao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traditional consumer finance is a modern financial service method that provides consumer loans to consumers of all classes. With the gradual improvement of China’s credit reporting system, big data credit reporting has effectively made up for the lack of traditional credit reporting and has been widely used in the consumer finance industry. In this context, the in-depth analysis of the specific application of big data credit reporting in the credit risk management of consumer finance and the strengthening of the research on the application of big data credit reporting in the credit risk management of consumer finance are urgently needed to be resolved in the economic and financial theoretical and practical circles’ problem. This article mainly studies the research on credit risk management of consumer finance by big data. The experimental results of this paper show that the model has a good forecasting ability, can distinguish between normal loan customers and default loan customers, and is suitable for practical personal credit risk control business. The prediction accuracy of the default model of the fusion model is 97.14%, and the default rate corresponding to the actual business is 2.86%. By combining the risk items such as the blacklist and gray list in the Internet finance industry, the bad debt rate and illegal usury can be well controlled to meet industry supervision.

1. Introduction

Personal credit evaluation is an important evaluation standard in personal credit reporting, and it has gradually entered the stage of market-oriented development. However, for a long period of time, China has relied on the personal credit report of the personal credit center of the central bank as the standard, with the rise of big data.

There are many studies of personal credit risk assessment nationally and internationally, including studies of multi-source data, Internet data, and Internet behavior data, but few studies exist to assess credit risk for personal big data. In theory, this research can enrich and perfect the theoretical system of personal credit risk assessment [1, 2]. It will help implement China’s inclusive financial strategy, lower the threshold for financial services, benefit more people, eradicate poverty, and achieve social equity. At the same time, it has a certain reference significance for the application of big data risk control in the industry [3, 4]. In terms of interpretability, new big data algorithms can evaluate the importance of statistical indicators, and statistical indicators with the highest comprehensive ranking of statistical indicators have better interpretability. On the whole, some big data algorithms are excellent in accuracy and stability and can be used as a strategic reserve for my country’s new generation of credit risk assessment models.

In theory, Morris and Shin decompose bank credit risk into bankruptcy and liquidity risk and define liquidity risk as the possibility of reverse bankruptcy caused by the bank’s operation for free [5]. The “liquidity ratio” (that is, the ratio of cash to current liabilities) on the balance sheet has been shown to reduce liquidity risk, reduce excessive debt yields, and increase solvency uncertainty (a measure of portfolio volatility index). For the method in [6], Petrone and Latora believe that the interconnectedness of financial institutions will affect instability and credit crises. In order to quantify systemic risks, research shows that this mechanism is highly contagious; that is, the lower the correlation between...
bankrupt banks, the greater the loss. This is in sharp contrast with the different advantages of standardized credit risk models adopted by banks and regulatory agencies. Therefore, this may depreciate the capital needed to overcome the crisis and lead to instability of the financial system [6]. And this mechanism has a negative impact on consumer finance. Aolin believes that joint loan guarantee agreements and mutual guarantee agreements between SMEs form the basis of the SME guarantee network. Therefore, a risk control plan is formulated according to the situation and importance of the company in the network. Use real-time mortgage data to determine the company’s node location on the mortgage network (including Coriolis and near the company) to understand the protection mechanism and prevent systemic credit risks before the crisis [7].

The innovations of this paper are as follows: (1) Internet financial risk prevention strategies. From the unique perspective of big data, the actual business big data is applied to personal credit risk assessment, and method theory is derived to better provide services for Internet financial risk control. (2) Use the machine learning algorithm model to conveniently construct an individual credit risk assessment index system. XGBoost machine learning algorithm method application innovation and XGBoost in the processing of large amounts of data; its distributed, parallel computing; and GPU graphics card computing advantages significantly improve model training efficiency and can output feature importance scores, for noncorrelation. The index is filtered to facilitate the rapid establishment of a personal credit risk assessment index system. (3) Using the big data technology platform, the financial industry can establish a comprehensive risk view, and management and risk managers can understand the risk view of the financial industry from different dimensions and grasp the control status of its main risk points.

2. Proposed Method

2.1. Big Data Credit. Credit evaluation in the traditional sense is mainly based on the interbank financial lending relationship; that is, based on the historical economic data and behavior of the debtor, the overall credit level of the debtor is evaluated by a simple linear analysis. It can be seen that the advantage of this method of analyzing historical information based on the individual user is that the credibility of the data is often high, so the risk control is also relatively good [8]. In addition, due to the small dimensions of data collection, the time lag effect is obvious, and credit service products are relatively few [9, 10]. The emergence of big data has largely overturned this traditional credit reference model and concept. Big data credit can be regarded as “big data technology + traditional credit,” which uses computer and Internet technology to support the model analysis and processing of data analysis and processing involved in credit information activities, which can more fully reflect the economic activities and the credit situation involved [11–13].

From the perspective of China’s current credit information system, there are great differences in the sources, statistical calibers, and methods of processing credit information. From the perspective of foreign countries, it was originally only a financial product that provided online alternatives based on credit information. Later, it was the first to introduce big data technology in the industry, classify data sources, and use computers to perform simulation analysis to reduce credit risk. Control capacity has increased by about 40%, and credit information service efficiency has increased by nearly 90%.

2.2. Theory of Information Economics. Information economics is based on a society with a certain level of information technology development. Information economists believe that information is another important resource in the market in addition to labor and capital. Information is also an element of transaction costs. People who are in a strong position of information can obtain excess returns, while those who are weak in information pay excess costs, which will lead to a lack of fairness in market transactions, increase the cost of transactions, and ultimately reduce market efficiency [14, 15].

The personal credit information system can well solve the problem of adverse selection and moral hazard in information asymmetry, and it is an effective system that can transmit signals and punish unbelievers and encourage trustworthy people. The information sharing platform provided by the personal credit information system can allow the lender (such as a bank) to have a deep understanding of the credit status of the borrower before making a loan decision, effectively eliminating the problem of adverse selection, and after the loan, the penalty for personal credit information. The existence of mechanisms can prevent the occurrence of moral hazard. From the perspective of the borrower, the borrower continuously improves its information in the personal credit information system and then improves its credit status by keeping promises, so that it can obtain a more favorable interest rate in future loans [16, 17]. In the long run, the establishment and improvement of the personal credit information system are beneficial not only to both borrowers and lenders but also to the healthy development of the entire market economy on the macro level.

2.3. Credit and Credit. Credit information in English is “credit reporting” or “credit investigation,” and its specific meaning is to collect credit data from nature. From this perspective, credit information is a series of activities with purposeful and directional information collection, processing, and evaluation [18, 19]. Literally, credit also means honesty and trust. It can be seen that credit often appears in the moral and ethical dimensions of economic life and is a relatively broad concept [20]. On a narrower level, credit specifically refers to the behavior of the goods or currency holder (creditor) to provide the borrower (debtor) with the goods or currency under the condition that the other party promises to repay, which is essentially a debt-debt relationship. Therefore, credit can be understood as the debtor’s ability to repay, as well as credit and trust activities based on commodities or currencies [21, 22].

2.4. Construction and Optimization of Personal Credit Risk Assessment Index System Based on Big Data. With the continuous improvement of the personal credit system, personal
credit behavior has gradually been recognized and become a necessary means to reflect personal morality and maintain social and economic order. It can be seen that personal credit risk assessment is of great significance for both commercial banks and residents. The construction of indicators requires high matching rate, high saturation, good timeliness, and multidimensional data. In this way, indicators are comparative and easy to understand, and it is easy to understand the actual situation of personal credit. The data studied in this project come from the company's big data platform. The data of the big data platform come from multiple sources, such as company APP product data, central bank credit data, Internet credit company credit data, and e-commerce company's online shopping data. Data are from three-party cooperative enterprises, Internet public data captured by crawlers, and data released by public inspection. The data used in the research of this project are desensitized from multiple levels to ensure data security [23, 24]. The construction of a personal credit risk assessment indicator system requires engineering steps such as data exploration and preprocessing, feature engineering, preliminary screening of indicators, and optimization of indicators. The specific data flow diagram is shown in Figure 1.

2.5. Construction and Optimization of Personal Credit Risk Assessment Model Based on Big Data Index System. The main idea of logistic regression for classification is to establish a classification formula based on existing data to establish a regression formula for classification boundaries. The purpose of logistic regression is to find the best-fitting parameters of the nonlinear function sigmoid, which can be done by optimization algorithms [25, 26]. Among the optimization algorithms, the most commonly used is the gradient ascent algorithm. The gradient ascent algorithm can be simplified to a stochastic gradient ascent algorithm. The stochastic gradient ascent algorithm is equivalent to the gradient ascent algorithm, but it occupies less computing resources [27]. In addition, stochastic gradient ascent is an online algorithm that updates parameters when new data arrives without having to reread the entire data set of the batch operation. The advantage of logistic regression is that linear calculation is low in cost, easy to understand, and simple to implement. The disadvantage is that it is easy to underfit, and the classification accuracy may not be high. The applicable data types are numerical and nominal data.

The logistic regression model is a binary classification model, which can be expressed by a conditional probability distribution \( p(y = 1 \mid x) \). The form is a parameterized logistic distribution. Assuming that the vector \( x = (x_1, x_2, x_3, \ldots, x_n) \) has \( n \) independent variables, the condition rate \( p(y = 1 \mid x) = p \) is the probability that the observed value occurs for \( x \). Therefore, the logistic regression model can be expressed as follows:

\[
p(y = 1|x) = \pi(x) = \frac{1}{1 + e^{-g(x)}}, \tag{1}
\]

\( f(x) = 1/(1 + e^{-g(x)}) \) is called logistic function.

\[
g(x) = w_0 + w_1x_1 + \cdots + w_nx_n. \tag{2}
\]

The probability that \( y \) does not occur under \( x \) conditions is

\[
p(y = 0|x) = 1 - p(y = 1|x) = 1 - \frac{1}{1 + e^{-g(x)}} = \frac{1}{1 + e^{g(x)}}. \tag{3}
\]

The ratio of the probability of occurrence of \( y \) to the occurrence of \( y \) is

\[
\frac{p(y = 1|x)}{p(y = 0|x)} = \frac{p}{1 - p} = e^{g(x)}. \tag{4}
\]

This ratio is called the occurrence ratio of events (odds), and taking the log of odds gives

\[
\ln \left( \frac{p}{1 - p} \right) = g(x) = w_0 + w_1x_1 + \cdots + w_nx_n. \tag{5}
\]

It can be seen that the dependent variable and the independent variable are nonlinear relationships, and the linear conversion can be performed by odds ratio of logarithmic occurrence. Usually, the maximum likelihood estimation is used to find the parameters of the logistic model. The likelihood function is
The basic idea of applying the logistic model to personal credit risk assessment is as follows: a sample of \((X_{i1}, X_{i2}, X_{i3}, \ldots, X_{ik})\) \((i = 1, 2, 3, \ldots, k)\) sample data of \(n\) groups of loan customers is given, where \(Y\) is a 0-1 variable and \(Y_i = 1\) indicates that the \(i\)th customer is a bad credit customer.

The logistic equation is

\[
P_i = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik})}
\]

The above equation can be linearly changed to obtain

\[
\ln \left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} \quad i = 1, 2, \ldots, n.
\]

For ease of use, take the natural logarithms on both sides of formula (9), and obtain the log-likelihood function (LLF) as

\[
\begin{align*}
\ln L(w) &= \prod_{i=1}^{n} \left(\pi(x_i)\right)^{Y_i} \left(1 - \pi(x_i)\right)^{1-Y_i}.
\end{align*}
\]

3. Experiments

3.1. Experimental Design

3.1.1. Index Selection. According to the foregoing theoretical analysis, we found from the basic information of individual users in a bank that the borrower’s basic personal information and loan-related information will affect the borrower’s credit limit. In view of the fact that all the borrowers’ default records disclosed by the Renrendai platform are 0 times and the credit ratings are all grade A, select the total amount of the subject as the explanatory variable. The specific situation of each indicator is as follows:

1. The total amount of the subject. The total amount of the target refers to the amount of the loan that the borrower wishes to post on the Renrendai platform. The minimum value is 5250 yuan, and the maximum value is 193,500 yuan, so the value range of the target total is 5250–193,500. This article uses the target total as a proxy for credit lines.

2. Age. Age refers to the actual age of the borrower. The youngest borrower is 22 years old, and the oldest borrower is 58 years old. The age range is 22–58.

3. Educational background. The borrower’s education includes four levels: graduate or above, undergraduate, college, and high school or below. This variable is a nominal variable, so it needs to be quantified and converted to a dummy variable. Setting “high school or below” as 1, “college” as 2, and “undergraduate” as 3, and the value of “graduate or above” is 4.

4. Marital status. Borrowers have three types of marital status: divorced, unmarried, and married. This variable is nominal, so it needs to be quantified and converted to a dummy variable. The value of “divorce” is 1, the value of “unmarried” is 2, and the value of “Married” is 3.

3.1.2. Research Hypotheses. Based on the relevant research results of the existing literature, this paper proposes the following research hypotheses:

Hypothesis 1. In terms of the personal characteristics of the borrower, the older the borrower, the higher the credit limit; the higher the borrower’s education, the higher the credit limit; the more stable the marriage status of the borrower, the higher the credit limit; the larger the credit limit, the higher credit limit; the longer the borrower’s working time, the higher the credit limit; the more developed the province of the borrower’s economy, the higher the credit limit.

Hypothesis 2. In terms of the financial characteristics of the borrower, the higher the borrower’s income, the higher the credit limit; if the borrower owns real estate or car production, the higher the credit limit; if the borrower has the mortgage or car loan, the lower the credit limit.
Hypothesis 3. In terms of the creditworthiness of the borrower, the more times the borrower successfully applies for a loan, the higher the credit limit obtained; the more the borrower pays off, the higher the credit limit.

Hypothesis 4. In terms of borrowing characteristics of borrowers, the higher the annual interest rate, the lower the credit line; the longer the repayment period, the lower the credit line.

3.1.3. Credit Report. The information in the personal credit report mainly includes six aspects: the results of the identity information verification by the Ministry of Public Security, basic personal information, bank credit transaction information, nonbank credit information, personal declarations and objections, and query historical information.

4. Discussion

4.1. Regression Results of Factors Affecting Credit Lines

4.1.1. Personal Characteristics. As shown in Table 1 and Figure 2, among the several variables representing personal characteristics, the three variables of marital status, working hours, and working place have no significant effect on the borrower’s credit line. Age is statistically significant at a significant level of 1%, and the sign of the coefficient is positive, indicating that the age of the borrower has a positive impact on its credit limit. The older the borrower is, the larger the amount of borrowing that can be obtained. Older borrowers tend to have richer social experience and richer asset accumulation, so the larger the amount of borrowing they can get.

4.2. Financial Characteristics. As shown in Figure 3, the experimental object of the survey is my country’s multiple variables representing financial characteristics in 2014-2019; whether or not owning a real estate or car production has no significant effect on its credit line. When the borrower cannot repay on time, it is difficult to conduct mortgage auctions on the borrower’s real estate and car properties, so whether or not owning the real estate or car properties has no obvious effect on the borrower’s credit limit.

Housing loans and car loans are statistically significant at significant levels of 5% and 10%, respectively, and the signs of the coefficients are positive, indicating that housing loans and car loans have a positive impact on borrowers’ credit lines. On the contrary, borrowers who have a home loan or car loan can also obtain a higher amount of borrowing because investors have very limited information about the borrower. If the borrower can obtain a house loan or car loan through bank inspection, it means good solvency, so the amount of borrowing that can be obtained may also be higher.

4.3. Empirical Analysis of Personal Credit Risk Assessment Based on Big Data Platform

4.3.1. Analysis of the Company’s Personal Credit Risk Assessment Results. Using the above-built logistic regression
Table 2: Forecast results of personal credit risk model of personal credit business data.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Actual normal customers</th>
<th>Actual default customers</th>
<th>Predict the number of normal customers</th>
<th>Predicted default customers</th>
<th>Accuracy in predicting default customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic + XGBoost fusion model</td>
<td>22,470</td>
<td>6701</td>
<td>19,738</td>
<td>6509</td>
<td>97.14%</td>
</tr>
</tbody>
</table>

Figure 4: ROC curve of the fusion model.

+ XGBoost decision tree fusion model to make empirical predictions on the company’s personal credit data, there were 29,171 loan customers, 22,470 normal customers, and 6,701 default customers.

As shown in Table 2, compared with the single model, the personal credit risk assessment model based on the fusion model in the company’s big data business environment has higher classification accuracy, can well predict normal customers and default customers, and is suitable for practical individual’s credit risk control business. These characteristics are highly interpretable and can be cross-validated with the industry’s professional knowledge, which further promotes the improvement of personal credit business and facilitates the construction of a comprehensive personal credit risk assessment system in the Internet finance industry.

As shown in Figure 4, the AUC of this project fusion model is 0.93 and KS is 0.66; it shows that the predictive ability of the personal credit risk assessment model is good, and it can accurately distinguish between normal loan customers and default loan customers, which is suitable for practical personal credit risk control in business. The prediction accuracy of the default model of the fusion model is 97.14%, and the default rate corresponding to the actual business is 2.86%. The default rate of personal credit loans was well controlled, from the original default rate of 4% to 2.86%, and the default rate decreased by 1.14%.

5. Conclusions

Currently, most Chinese companies are facing many challenges in terms of data infrastructure system architecture and data analysis. In the context of big data and the Internet, big data credit reporting and consumer finance have developed rapidly. The prerequisite for the healthy development of the consumer finance industry is efficient and accurate credit risk management. In the context of the timeliness and comprehensiveness, this paper studies and analyzes the limitations of the classic credit risk rating system, avoids the limitations of traditional indicators when designing a personal big data risk rating system, and creates complete and logical personal credit rating data. In the context of system, after the index is created for the first time, the index may have relevance issues and unequal predictability. Therefore, the focus of research is to use effective feature selection methods to optimize indicators when creating indicators.

Data Availability

No data is available. The article does not touch on data.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


