Hindawi
Complexity
Volume 2018, Article ID 7616537, 16 pages
https://doi.org/10.1155/2018/7616537

Research Article

Research on Application of Big Data in Internet Financial Credit Investigation Based on Improved GA-BP Neural Network

Fei-Peng Wang

School of Business Administration, Shandong Institute of Business and Technology, Yantai, Shandong 264005, China

Correspondence should be addressed to Fei-Peng Wang; wfp2009@sdibt.edu.cn

Received 20 June 2018; Revised 11 September 2018; Accepted 18 September 2018; Published 2 December 2018

Guest Editor: Zhihan Lv

Copyright © 2018 Fei-Peng Wang. 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.

The arrival of the era of big data has provided a new direction of development for internet financial credit collection. First of all, the article introduced the situation of internet finance and traditional credit industry. Based on that, the mathematical model was used to demonstrate the necessity of developing big data financial credit information. Then, the Internet financial credit data are preprocessed, the variables suitable for modeling are selected, and the dynamic credit tracking model of BP neural network based on adaptive genetic algorithm is constructed. It is found that both LM training algorithm and Bayesian algorithm can converge the error to 10e-6 quickly in the model training, and the overall training effect is ideal. Finally, the rule extraction algorithm is used to simulate the test samples. The accuracy rate of each sample method is over 90%, and some accuracy rate is even more than 90%, which indicates that the model is applicable to the credit data of big data in internet finance.

1. Introduction

The credit system is the cornerstone to the development of the market economy and the financial industry. A sound credit system helps create a good consumer environment, effectively prevents the spread of credit risks, and promotes the healthy development of the economy [1]. Internet finance has been regarded by more and more scholars as the new engine of economic growth, and relying on big data to establish the credit system has become an inevitable choice for the development of internet finance.

Many researchers have put forward some creative ideas for the credit reporting system of big data [2]. However, due to the fact that large data credit is a brand-new concept, there are relatively few literatures directly related to it, and the related research lacks systematisms and depth. Only the whole idea of building a big data credit system was explored [3], and there was no reasonable evidence for feasibility analysis. In view of this, under the background of big data of Internet finance and domestic traditional credit investigation industry, the construction of personal credit investigation system is discussed. The development of the research is mainly reflected in three aspects. First, based on economic theory, concepts of credit information costs and potential risk costs are raised and the use of quantitative models to demonstrate the necessity of developing big data credits. Second, ingeniously introduce conducts and contacts of Internet data into BP neural network algorithms based on adaptive genetic algorithms. A credit-tracking model based on big data was constructed. Third, the application of the model was briefly described.

2. Internet Finance and Credit Industry

2.1. Internet Finance. Finance is at the core of modern economic operations, and financial viability determines the quality and potential of the overall economy. Financial dynamism depends on the ability of the financial system to accept new concepts and apply new technologies. However, with the application and popularization of Internet technology in the financial field, the traditional finance has been rapidly transformed from the economic field into a new industry—Internet finance [4]. Internet finance refers to the traditional financial institutions and Internet enterprises using Internet technology and information and communication technology to achieve financing, payment, investment,

المنارة للاستشارات
www.manaraa.com

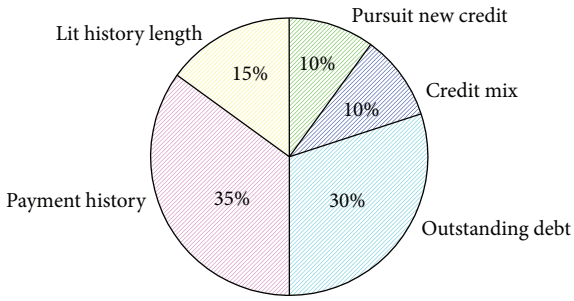

Figure 1: Reference elements of the traditional credit scoring model.

and information intermediary services. It is not only different from the indirect financing of commercial banks but also different from the direct financing of the capital market information finance business model. Internet finance is a new type of business generated by the cross-border integration of the traditional financial industry with the Internet industry with new technologies such as cloud computing, big data, and mobile payment [5].

2.2. Traditional Credit Investigation Industry. The essence of internet finance is finance, and the core of finance is credit. Therefore, the credit information system is an effective constraint mechanism to ensure the integrity of the financial market.

The traditional personal credit system in the financial industry is inefficient, narrow coverage, and slow update of data, which cannot satisfy the control of personal credit risk of Internet finance. At present, China's credit reporting system is mainly represented by the People's Bank of China Credit Reporting Center; there are also local government and its functional departments led by the government credit reporting system, as well as the market-oriented credit reporting system represented by the eight credit reporting companies that have obtained the licenses of individual credit institutions [6], as shown in Figure 1. At the same time, China also issued some more authoritative regulations and regulations on the establishment of a credit information system.

3. Mathematical Models

Based on the knowledge of economics, the concepts of credit costs and potential risk costs are proposed in this paper. Starting from the three characteristics of big data credit, using mathematical models to demonstrate the necessity of developing big data personal credit is put forward.

3.1. Analysis of Idea. Compared with the superiority of traditional credit information, big data credit can be measured from two dimensions of efficiency and cost [7]. Because of the three characteristics of timeliness, accuracy, and economies of scale in big data credits, the study believes that big data credits are superior to traditional credit information in terms of efficiency [8]. This paper will focus on the field of personal credit reporting from the cost dimension and then demonstrate the necessity of developing large data credit. The reasoning argument is based on the three characteristics of big data credits. By setting assumptions and constructing a mathematical model, the total social costs incurred by the two are compared.

The cost of traditional credit information mainly includes credit cost, potential risk cost, and other costs [9]. Firstly, each additional coverage of the credit system will result in the corresponding operating costs, which are derived from the data acquisition, data analysis, personnel employment, and other necessary aspects of credit, and defined as the cost of credit; secondly, even if the credit bureaus do not continue to expand their credit coverage, they still need to provide services to the creditworthy groups and pay for necessary equipment maintenance costs. This part of the cost is defined as other costs. Finally, based on economic rational people assumptions, any information subject is a potential defaulter. The purpose of credit investigation is to prevent the credit risk brought about by breach of contract. Conversely, without credit checking, potential risks can be generated, which is called potential risk cost. Credit evaluation of information subjects through credit information can effectively reduce potential risk costs but cannot eliminate them.

As a brand-new information network technology, large data credit reporting technology research and development, supporting infrastructure is needed. Therefore, large data credit has three costs of traditional credit, but at the same time, it also needs technical input, that is, technical cost [10].

3.2. Assumed Conditions and Variable Settings. For the sake of discussion, set the following six assumptions [11].

First, the information subject is homogenous, and there is no difference in the potential risk cost caused to the credit system. Assuming that the potential risk cost per unit time of the nonaccepting information subject is p, and the potential risk cost per unit time after accepting the traditional credit and large data credit is p_1 andp_2, respectively.

Second, the current population N is a fairly large value. Both the big data credit and the traditional credit information cannot include the entire population within the scope of the credit investigation within the time frame studied by the institute and do not consider the population growth.

Thirdly, the speed of credit information is defined as the number of people newly covered by the credit system in a given period of time. It is assumed that once an information subject has been included in the credit system, it will not need to reconduct it, and the traditional credit and big data credits are m_1 andm_1, respectively.

Fourth, the current credit system coverage is 0, and it is assumed that big data credits and traditional credits have the same credit cost β when the coverage is zero.

Fifth, big data credits need to invest a fixed technical cost A per unit of time, regardless of the diminishing cost of technology brought about by the maturity of big data credits.

Sixth, big data credit and traditional credit have the same other cost $R(T)$.

المنارة للاستشارات
www.manaraa.com

Based on the three characteristics of big data credit, this paper analyzes the difference in cost between big data credit and traditional credit information [12].

First, due to the timeliness of big data credits, the study considers that big data credits have higher credit reporting speeds than traditional credit information, that is, $m_1 < m_2$.

Second, because the assessment results of big data credits are more accurate, the study considers that big data credits can reduce the potential risk cost more effectively than traditional credit information, that is, $p_1 > p_2$.

Thirdly, due to the scale effect of big data credits, the study considers that the credit cost of big data credits does not increase with the expansion of credit coverage, that is, the credit cost β of the unit population remains unchanged. The credit information cost of the traditional credit reporting unit population will increase with the increase in the number of people covered. For ease of discussion, the article sets the growth of its credit cost as a linear model $(n-1)\alpha+\beta$ [13], where n is the number of people covered by the current credit system.

3.3. Model Construction and Solution. The study will use time T as an independent variable and set the current time point as $T=0$ to analyze the total social cost accumulated on $[0, T]$ for traditional credit and big data credits. When the time is T, the populations covered by traditional credit and big data credits are m_1T and m_2T, respectively. And the uncovered populations are $N-m_1T$ and $N-m_2T$, respectively.

The total social cost of traditional credit information on $[0, T]$ is $C_1(T)$, which is the sum of credit cost, potential risk cost, and other costs:

$$C_1(T)=\sum_{n=1}^{m1t}[(n-1)\alpha+\beta]+\int_0^T p(N-m_1t)dt+\int_0^T p_1m_1tdt+R(T). \tag{1}$$

The total social cost $C_2(T)$ of $[0, T]$ for big data credit is the sum of credit cost, potential risk cost, other cost, and technical cost:

$$C_2=m_2T\beta+\int_0^T p(N-m_2t)dt+\int_0^T p_2m_2tdt+R(T)+AT. \tag{2}$$

For all T, there are

$$\begin{aligned}C(T)&=C_1(T)-C_2(T)\\&=\sum_{n=1}^{m_1t}[(n-1)\alpha+(m_1T-m_2T)\beta+\int_0^T p(m_2t-m_1t)dt\\&\quad+\int_0^T(p_1m_1-p_2m_2)tdt-AT.\end{aligned} \tag{3}$$

Finding the first and second derivatives of $C(T)$ yields

$$\begin{aligned}C'(T)&=m_1^2T\alpha+[m_2(p-p_2)-m_1(p-p_1)]T\\&\quad+(m_1-m_2)\beta-\frac{m_1\alpha}{2}-A,\end{aligned} \tag{4}$$

$$C''(T)=m_1^2\alpha+[m_2(p-p_2)-m_1(p-p_1)]. \tag{5}$$

Since $m_1<m_2$ and $p_1>p_2>p_3$, $C''(T)>0$ is obtained, and the derivative $C''(T)$ monotonously increases in the interval $[0, T]$, and $C''(T)<0$. Judging that $C(T)$ shows a trend of decreasing first and then increasing, there is a minimum value and no maximum value. Based on the assumption of N infinity, there must be a bit of TE in the interval of $[0, T]$, making$C(T)=0$.

Let $C_1(T)=C_2(T)$, get

$$T_E=\frac{2A+(m_1-m_2)\beta}{m_1^2+[m_2(p-p_2)-m_1(p-p_1)]}>0. \tag{6}$$

When $T<T_E$, $C(T)<0$, that is, the total social cost of crediting using big data is greater than the total social cost of traditional credit information; and when $T>T_E$, the total social cost of credit using big data is less than the total social cost of traditional credit information.

Defining $T<T_E$ and $T>T_E$ as short term and long term, the following conclusions are drawn: in the short term, due to the high technical cost of investing in big data credits, traditional credit has more advantages in terms of cost; but in the long run, large data credit has three characteristics: timeliness, accuracy, and economies of scale. Whether from the perspective of efficiency or cost, large data credit is superior to traditional credit. Therefore, in the theoretical level, it is necessary to develop large data credit reporting.

4. Management of Big Data Based on Internet Financial Credits

After demonstrating the necessity of big data credit, the credit tracking model based on adaptive GA-BP neural network will be studied. Before this, first explain big data and do preprocessing.

4.1. Credit Market Big Data Overview. The unbalanced opening of the central bank's credit system still remains unsolved. At least the central bank's attitude is clear and supports the development of internet finance, and it is believed that internet finance is a useful complement to traditional finance [14]. Most of them leave credit data in the databases of other institutions outside the bank credit system and Internet companies. Internet credit companies have a strong demand for credit ratings for lenders' credit ratings. The market spontaneously formed a unique risk-control ecological field. Large companies use self-built credit rating systems through data mining; small companies obtain credit rating consulting services through third-party information sharing.

المنارة للاستشارات
www.manaraa.com

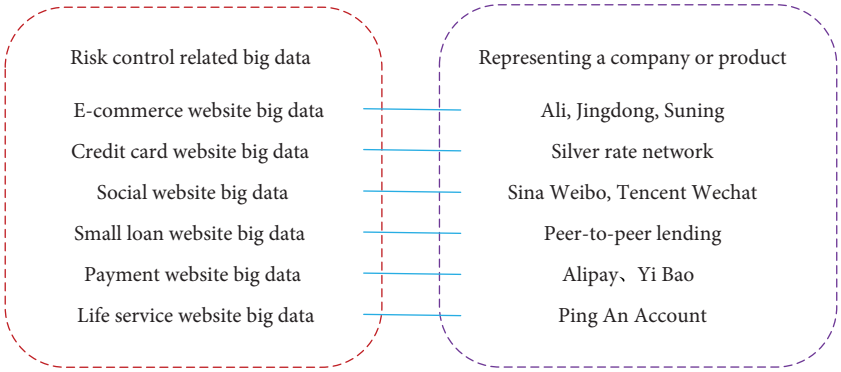

Figure 2: Data sources.

Table 1: Summary of data smoothing methods.

Method name	Specific steps
Smooth by average	Average the data in the box and replace all values in the box with the average
Smoothed by boundary values	Replacing each value in the box with the smallest boundary value
By median smoothing	Take the median of the box to replace all the data in the box

4.2. The Source of Big Data. E-commerce big data are used for risk control. After all the information is aggregated, the values are entered into a network behavior scoring model for credit rating. Big data on credit card websites is also very valuable for risk control of Internet finance. The year of credit card application, whether it passes, credit line, card type, credit card repayment amount, attention to preferential information, etc. can be used as reference data for credit rating: use social network relationship data and mutual trust between friends to aggregate popularity [15].

The borrower is divided into several credit ratings, but it is not necessary to publish their own credit history. In addition, water and electricity bill payment information in Taobao, credit card repayment information, payment, and transaction information have become data all-round players: the credit big data includes credit limits and default records. The direction of third-party payment platform payment, the amount of monthly payments, and the purchase of product brands can all be used as important reference data for credit rating: big data such as water for life service websites. Electricity, gas, cable television, telephone, network fees, and property fee payment platforms objectively and truly reflect the basic information of individuals and are an important type of data in credit rating, as shown in Figure 2.

4.3. Internet Credit Data Preprocessing. Through the construction of the user's credit image, it is possible to rationally organize and store the dimensions of the complex, diverse, widely distributed, and heterogeneous platforms on the Internet. However, big data are sparsely populated and online and offline behaviors of users are widely distributed and extremely difficult. Full collection and coverage and user behavior preferences are also different; there are significant differences in the behavior of different categories, resulting in the possibility of user behavior information missing rate of more than 50%, plus the data source instability caused by the lack of data and inconsistencies. The problem has caused us to preprocess large data before using it to model financial information, ensuring that the data meet the modeling requirements. Pretreatment mainly starts from the following aspects.

4.3.1. Data Cleaning. The purpose of data cleaning is mainly to deal with the data problems found in the process of data verification. Its purpose is to solve the problems of incomplete data, inconsistent data, and data noise. During the data cleaning process, the data are properly processed and adjusted for problematic data so that it can meet the requirements of modeling as much as possible because the quality of the model largely depends on the amount of modeling data; if the adjusted data are still unavailable, they need to be deleted [16]. In the specific cleaning process, the uniqueness, completeness, validity, relevance, timeliness, and consistency must be ensured.

The binning method is a commonly used method in the data cleaning process. Its core idea is to smooth the value of the current data by the value of the surrounding data. There are mainly three methods for data smoothing, as shown in Table 1.

Three data smoothing methods and specific steps are given in Table 1, so we can choose the appropriate data smoothing method according to the characteristics of the data.

4.3.2. Univariate Analysis. The purpose of univariate analysis is to determine the variables that satisfy the following two conditions [17]:

(1) Conforming to actual business significance

(2) High discrimination ability for the analysis object

المنارة للاستشارات
www.manaraa.com

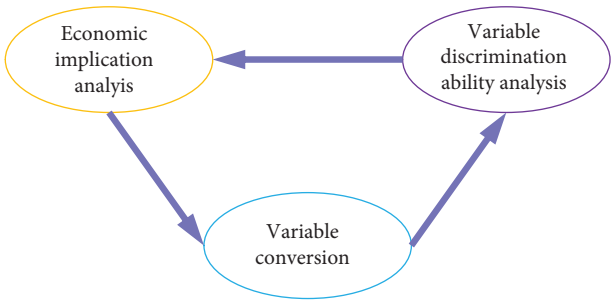

Figure 3: Univariate analysis.

The results obtained through univariate analysis are a set of variables that are basically suitable for modeling, while reducing the complexity of later multivariate analysis.

At this stage, usually from the analysis of the ability to distinguish the variables, the following method is used to analyze the variables several times, until the variables meet the above requirements, as shown in Figure 3:

4.3.3. Multivariate Analysis. Through multivariate analysis, the combination of model variables that meet the following three conditions can be determined:

(1) Low correlation between variables

(2) The model has stable and high discrimination ability

(3) The model contains as many different types of information as possible

4.3.4. Big Data Processing Flow. The first stage of the process is shown in Figure 4. Through the steps of data collection, data cleaning, and variable analysis, the user's portrait data are converted to data available for risk assessment modeling of large data; then, the effectiveness of univariate analysis and cross-variables is performed through binning, and then, multivariate analysis is performed. According to the principles of comprehensiveness, scientific, feasibility, and measurability, the variables suitable for modeling are selected and the BP neural network-based credit tracking model is introduced, as shown in Figure 4.

So, a tracking indicator system was constructed, as shown in Table 2.

According to expert experience, knowledge, and intuition, initial values of various indicators are obtained in Table 2.

5. A Dynamic Credit Tracking Model of BP Neural Network Based on Adaptive Genetic Algorithm

Firstly, a dynamic credit tracking model was built. Then, the model was simulated using the MATLAB neural network toolbox. Finally, the model was used to train 20 samples collected through the network using LM training algorithm, Bias algorithm, and momentum gradient algorithm, and the remaining 8 samples were simulated. There are three reasons for the number of samples selected:

(1) There are many indicators, and the horizontal data are more

(2) The data preprocessing is reduced by a part

(3) The amount of data is large

However, only 20 samples were selected for analysis due to objective reasons. Very good simulation results were achieved, as shown in the flowchart in Figure 5.

As a powerful tool for studying complexity, BP (Back Propagation) neural network technology has demonstrated extraordinary advantages in pattern recognition, classification, prediction, and rating in recent years. It has a powerful parallel processing mechanism and is highly self-learning and self-developed. Adaptability, there are a large number of adjustable parameters inside, thus making the system more flexible and able to handle any type of data, which is unmatched by many traditional methods. Through continuous learning, neural networks can discover its laws from a large amount of complex data of unknown patterns. It overcomes the complexity of the traditional analysis process and the difficulty of selecting the appropriate model function form. It is a natural nonlinear modeling process. It is necessary to distinguish what kind of nonlinear relationship existed and brought great results to modeling and analysis. When the method is applied to credit risk analysis, the input of a series of popular credit indexes can be processed, and the corresponding credit rating output can be produced, and the experience, knowledge, and intuitive thinking of experts can be reproduced, thus ensuring the objectivity of the evaluation and prediction results. Because the initial weights of the BP neural network are randomly generated and the training speed is slow, there are problems such as local minimum values [18]. Therefore, these defects can be improved to some extent through the combination of genetic optimization and BP neural networks.

5.1. Dynamic Tracking Model Index System Construction. Based on the principles of comprehensiveness, scientific, feasibility, and testability, after preprocessing the data, this paper finally selected two first-level indicators and a total of 11 second-level indicators to construct a dynamic credit tracking evaluation index system. The design steps are as follows.

Firstly, the initial indicators were determined. Through drawing lessons from FICO's credit scoring index system in the United States, the personal credit evaluation index system of commercial banks, and the actual experience and expert opinions of China's public credit management, this paper obtained a number of preliminary indicators that reflect the public credit rating and then classify the goals by connotation. It is a set of three key indicator sets, that is, a first-level indicator. Each indicator set contains multiple secondary indicators that reflect its connotation and are operational.

Secondly, according to the three principles of the same indicators with the same connotation to be merged, the cause and effect of the indicators of cause and effect, and poor operability to find alternative indicators, the indicators were screened by the Delphi expert assessment method and related analysis methods; for example, we removed the obstacles

المنارة للاستشارات
www.manaraa.com

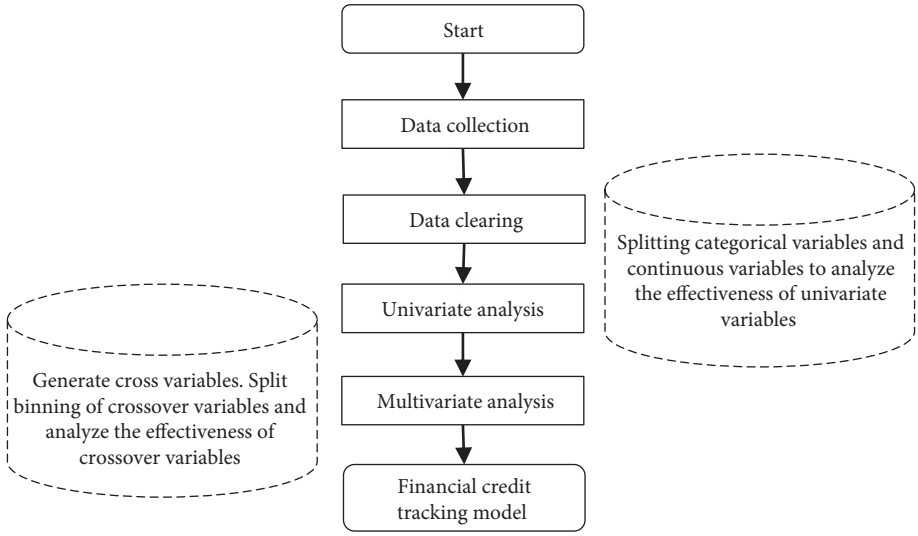

Figure 4: Big data processing flow.

Table 2: Sample.

Serial number index	U11	U12	……	U15	U16	U21	U22	……	U25	U26	U31	……	U34	Credit rating score
1	5	5	……	5	5	1	1	……	3	1	4	……	3	2
2	3	3	……	3	3	5	3	……	3	3	4	……	3	2
3	5	1	……	3	1	5	1	……	1	3	5	……	3	2
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……
8	5	4	……	5	1	3	3	……	4	3	3	……	5	2
9	3	4	……	5	5	3	1	……	5	3	4	……	3	2
10	3	3	……	1	5	5	1	……	1	1	5	……	5	1
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……
15	5	5	……	3	5	3	1	……	1	4	4	……	5	2
16	5	3	……	1	1	1	3	……	3	3	5	……	1	1
17	3	1	……	3	3	5	1	……	4	1	4	……	5	2
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……
22	3	5	……	5	3	3	1	……	5	5	4	……	5	2
23	5	1	……	3	5	1	1	……	4	4	3	……	3	2
24	3	4	……	5	5	1	1	……	3	1	4	……	3	2
……	……	……	……	……	……	……	……	……	……	……	……	……	……	……

such as "asset-liabilities ratio" and "ineffectiveness" and then removed the indicators "vehicles" and "blacklist of personal credit information." The presurvey shows that the indicator has strong homogeneity. Finally, it simplifies the indicator "personal annual expenditure," and the correlation analysis shows that it is strongly related to "U21 Contact the bank."

Thirdly, expert circular appraisal and revision are performed on the selected credit index to form the final credit index system. In order to ensure the quality of the index system, we must further find relevant experts to demonstrate. Experts who are invited here should not repeat the experts who collect and screen indicators. After the experts have demonstrated and revised the credit index system, they have conducted presurveys and solicited opinions from the surveyed people. Such a cycle can then be determined as the final credit index system.

Finally, determine the scoring criteria. Through the Delphi expert evaluation method, the opinions and opinions of internal and external experts were investigated and repeatedly integrated. Finally, they obtained the unanimously agreed opinions and opinions as the basis of the index scoring standards. For example, with the "Contact with Banks" as an example, the more contacts there are, the stronger the repayment willingness and the better the credit status, so the "contact with the bank" is assigned 5 points in descending order, 3 points, 0 point is similar to other indicators are assigned in order. So, a tracking indicator system was constructed, as shown in Table 3.

المنارة للاستشارات
www.manaraa.com

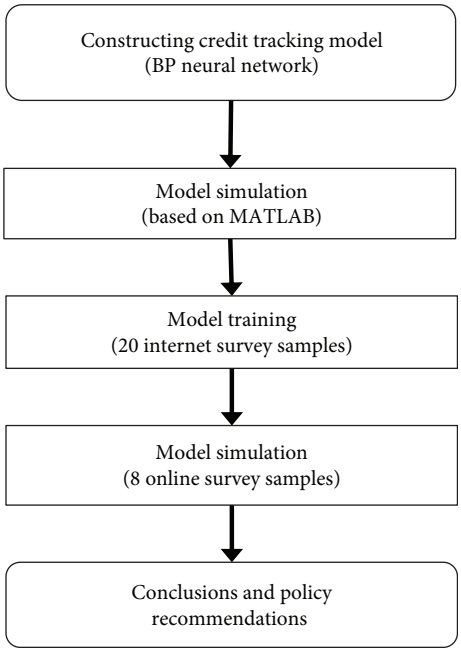

Figure 5: Construction and application flowchart of dynamic credit tracking rating model.

5.2. Construction of Dynamic Credit Tracking Model Based on BP Neural Network Based on Adaptive Genetic Algorithm

5.2.1. Genetic Algorithm. The genetic algorithm is an algorithm that simulates the rules of the survival of the fittest in the natural world. It selects, crosses, and mutates the population to obtain the optimal individual population method. The optimization process of the traditional genetic algorithm is as follows [19]:

(1) According to the characteristics of the problem to be dealt with, select the code corresponding to the problem solution, and give an initial population, which includes N chromosomes

(2) Calculate the fitness function value for each chromosome in the initial population

(3) When the result of an iteration of the genetic algorithm meets the condition of stopping the iteration, the algorithm stops iterating. Otherwise, a random probability value is used to randomly select N chromosomes from the old population, and the new population composed of these chromosomes is next iteration

(4) Crossing to obtain a cross set of N chromosomes, the new generation of individuals will inherit the previous generation of information

(5) Set a small mutation probability to allow certain genes in the chromosome to mutate, obtain new populations to enhance individual fitness, and repeat (2) the calculation process

The setting of the crossover probability P_c and the mutation probability P_m in the genetic algorithm will largely affect the convergence of the genetic algorithm and increase the error of the optimal solution and the real solution.

In general, the greater the value of the crossover probability P_c, the faster the new individual will be produced, and the value of the crossover probability P_c is too large, so that the individual structure with a high fitness value is destroyed. For the mutation probability P_m, if the value of P_m is too small, it is not easy to generate a new individual; if the value of P_m is too large, the algorithm is similar to the random search algorithm.

Therefore, the crossover probability P_c and the mutation probability P_m have a profound effect on the performance of the algorithm. Therefore, the crossover probability and mutation probability that can be adaptively adjusted are used to ensure the diversity of the group:

$$P_c = \frac{1}{e^{(-k_1 \Delta f)} + 1}, \tag{7}$$

$$P_c = 1 - \frac{1}{e^{(-k_2 \Delta f)} + 1}. \tag{8}$$

Among them, Δf means the difference between the maximum fitness value and the average fitness value of each chromosome individual; k_1 and k_2 represent the adjustment rate, and they all take a value of 1. Through the adaptive genetic algorithm, the ability of the genetic algorithm to search the global optimal solution can be effectively improved, and the problem of falling into local optimization can be avoided.

5.2.2. Neural Network. Neural networks have been widely used in the field of bank risk management. A certain structure of BP neural network has the ability to predict and classify similar data after training with given sample data. Applying this feature of BP neural network, according to the established personal credit risk tracking index system, a three-layer BP neural network model was established. The number of input nodes is equal to the number of feature variables in the index system; the number of output nodes is one, and the level of personal credit can be determined according to the output value; the method of determining the hidden layer nodes is to determine the excessive hidden layer nodes first, after training, and then according to the training results are pruned [20].

The connection weight between nodes selects the random number in the interval [−1, 1], and the optimal initial weight of the BP neural network is determined by the adaptive genetic algorithm. The transfer function of each layer is tansig, and the transfer function of the second layer is purelin, as shown in Figure 6.

5.3. Dynamic Credit Tracking Model Training

5.3.1. Model Training Steps. According to the previously established tracking index system, a certain amount of personal data is selected in the personal credit database, each of which contains a corresponding personal credit level L, represented by the vector $A_j = (a_1, a_2, \ldots, a_n, d)$, where a_i is

المنارة للاستشارات
www.manaraa.com

Table 3: Tracking indicator system.

First-level indicators	Secondary indicators	Index score							
U1 personal basic information	U11 sex	Male	5	Female	3				
	U12 age	Ages 23 to 27	3	Ages 28 to 32	4	32 years of age or older	5	Age 22 and below	1
	U13 education level	Junior college and below	1	Undergraduate course	3	Master or above	5		
	U14 marriage	Married	5	Unmarried	3	Divorced	1		
	U15 registered permanent residence	Rural	1	In this city	5	Other city	3		
	U16 administrative level/position	Division/general manager or above	5	Section/part manager	3	Other	1		
	U17 personal annual income	Over 100,000 yuan	5	30,000 to 100,000 yuan	4	10,000 yuan to 30,000 yuan	3	Less than 10,000 yuan	1
U2 personal credit index	U21 contact the bank	Keep in touch	5	Frequent contact	4	Occasionally contact	3	Never contact	1
	U22 loan bank account status	Have a loan account	3	Savings account	5	Have a credit card account	4	There is no account	1
	U23 credit card, mortgage delinquency record	0	5	1–2 times	3	3 times and above	1		
	U24 bad records on taxes, utilities, etc.	0	5	1–2 times	3	3 times and above	1		

المنارة للاستشارات
www.manaraa.com

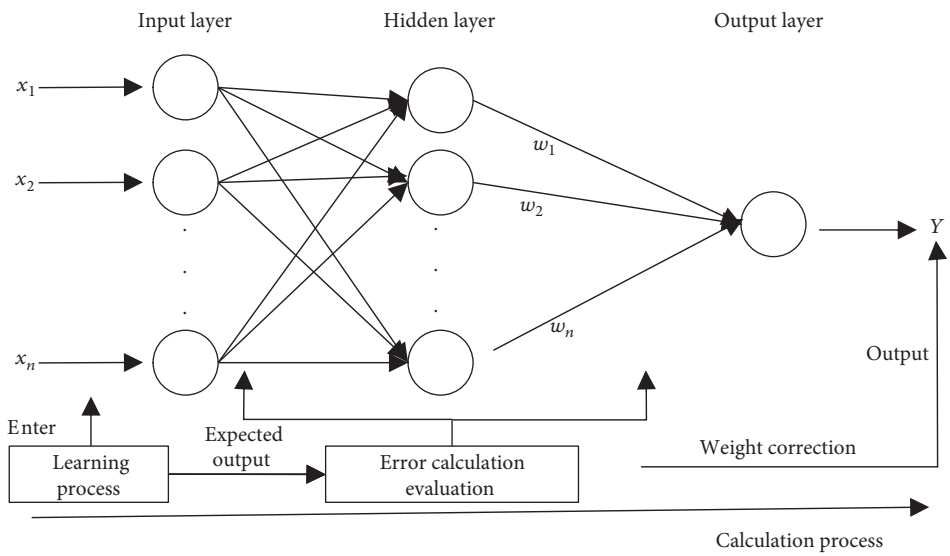

Figure 6: Structure of credit tracking model based on BP neural network.

the corresponding feature variable in F',a_i=5, 3 or 0, n is the number of elements in F', $n = 11$, $L \in D$, $D = \{A, B, C, D\}$, and $n(L)$ is the position number of L in set D, $1 \le n(L) \le 4$, $d = (d_1, d_2, \ldots, d_4)$, and $d_n(l) = 1$, $d_i = 0$, $1 \le i \le 4$, $i \ne n(L)$ is personal credit rating vector. In the data set, part of the data is randomly selected as a test data set for verification of model training results; the remaining data are used as a training data set for model learning.

$a_1, a_2, \ldots, a_n$ of each vector A_j of the training data set is taken as the input of the model, and d is the target output of the model. Let w_{lm} be the connection weight of input layer node l to hidden layer node m, v_{mp} be the connection weight value of hidden layer node l to output layer node p, and θ and w are the determined threshold values. For a given input $a_1, a_2, \ldots, a_n$, the activation function of hidden node m is tansig, and the activation function of output node p is purelin.

When the genetic algorithm optimizes the initial weights of the BP neural network, a fitness function needs to be set to determine the probability that the individual is selected. Since the fitness value of the genetic algorithm is continuously increased during the search process of the optimal value, the fitness function can be set as the following expression:

$$\text{fitness} = \frac{1}{\sum_{i=1}^{N} |y_i - x_i| + \zeta}, \tag{9}$$

where x_i represents the actual credit score data; y_i represents the credit score data predicted by the BP neural network; N represents the number of samples; ζ is a smaller value, avoiding 0 as the denominator.

Then, according to the genetic algorithm optimization step introduced, the initial weight of the BP neural network is determined.

For each sample data in the training data set, if

$$\max_p \left|\beta_p - d_p\right| \le \eta_1 \max_p \left|\beta_p - d_p\right| \le \eta_1, \tag{10}$$

then the training data set can be correctly classified, where $0 < \eta_1 < 1/2$. The test data set is used to test the trained model. If (10) is satisfied, the model training is completed; otherwise, the model is trained again. The purpose of model training is to obtain a set of weights (w, v) that can correctly classify personal credit rating data.

After the model training is completed, a fully connected three-layer BP network is obtained. The model pruning is to delete some of the connections in the model according to certain rules without affecting the accuracy of the model classification. In the weighted set (w, v) of the trained model, under the condition that (10) is satisfied, for any w_{lm}, if it satisfies

$$\max_p \left|v_{mp} - v_{lp}\right| \le 4\eta_2, \tag{11}$$

then delete w_{lm} in the model. For any v_{mp}, if satisfied

$$\max_p \left|v_{mp}\right| \le 4\eta_2, \tag{12}$$

then delete v_{mp} in the model, where $\eta_2 > 0$ and $\eta_1 + \eta_2 < 1/2$. If there is no weight that satisfies (11) and (12) in the weight set (w, v), ω_{lm} associated with the vector minimum product $\max_p |v_{mp} - v_{lp}|$ is deleted in the weight set (w, v).

After the model training is completed, MATLAB will automatically extract the model rules. The rule extraction is to extract the classification rules in the pruned model, and

المنارة للاستشارات
www.manaraa.com

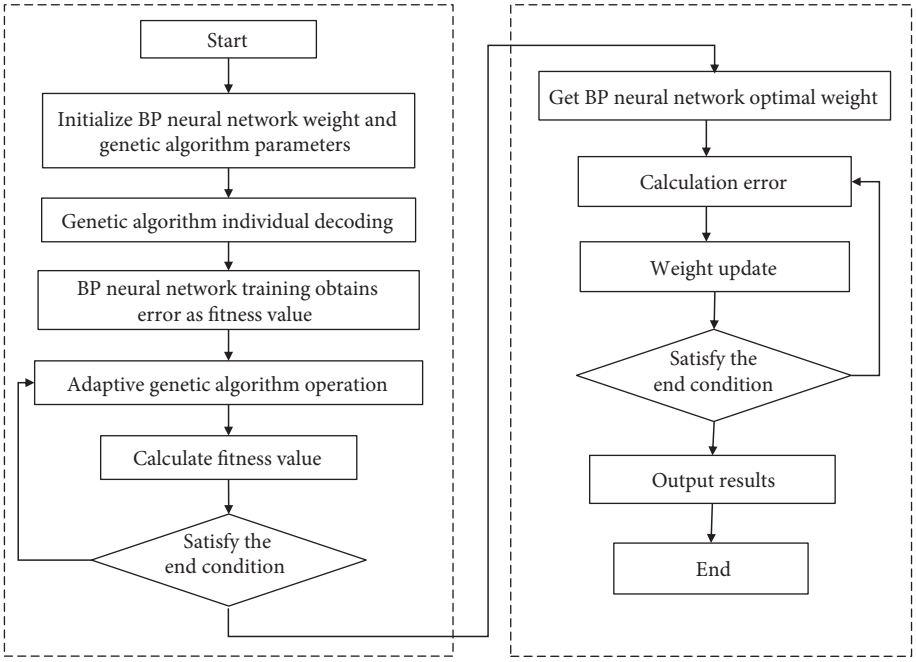

Figure 7: Flow diagram of BP neural network algorithm based on optimization of adaptive genetic algorithm.

the relationship between the input and output of the model performance after the pruning is still more complicated. The rule extraction algorithm firstly uses a clustering method to discretely process the hidden layer activation value. When certain accuracy of the model is ensured, the input value and the hidden layer activation value are discretized, and the number of discrete values can be conveniently managed. Second, enumerate the activation values of the discretization, calculate the output of the model, and generate complete rules from the hidden layer activation value to the output. Third, for each hidden layer activation value that appears in the above rules, the enumeration can generate these hidden layer activations. Enter the value and generate a complete rule from input to output, as shown in Figure 7.

Using the above adaptive GA-BP neural network model, it is possible to dynamically track the implementation of credit rating changes and obtain the following credit rating transition table, as shown in Table 4.

The credit rating is divided into four levels: A, B, C, and D. When the probability of default does not exceed 5%, we determine the risk classification of "Normal." When the probability of repayment does not exceed 5%, we determine the risk classification of "Track."

5.3.2. Result Analysis. After completing the design of the model, the model was simulated by the MATLAB program. 20 samples were used for training. The results are shown in Figures 8, 9, and 10. Both the LM training algorithm and the Bayesian algorithm can converge to 10e-6, but momentum gradient algorithm convergence is slow, but overall, the model has achieved a good training effect and has a certain application significance.

(1) LM Training Results. It can be seen from Figure 8 that the target error 10e-6 is reached after five network training times; the training speed is fast, and the training effect is good.

(2) Bayesian Training Results. It can be seen from Figure 8 that the target error 10e-6 is reached after 77 network training times, and the training speed is fast and the training effect is good.

(3) Momentum Gradient Training Results. From Figure 10, it can be seen that after the number of network trainings of 4381 times, the target error is 10e-2 and the training speed is slow, but the training result is still within the allowable error range.

5.4. Dynamic Credit Tracking Model Simulation. To test the dynamic credit tracking model established above, the remaining 8 test samples were input into the trained network for testing. The network test output results are shown in Figures 11, 12, and 13. Comparing the output of the network with the expected output, it can be seen that the error is less than the allowable error. Therefore, it is considered that the network output is reasonable, and the trained network has better generalization performance.

After simulation, the simulation results obtained are shown in Table 5. The simulation accuracy rate of each sample method has reached more than 80%, and some even reached more than 90%. The simulation results are very good. It can be concluded that the evaluation model is feasible for comprehensive credit evaluation. This model can be directly used to rating unknown samples, thus reducing the evaluation workload, reducing the subjectivity of evaluation, and improving the rationality of the rating results. The

المنارة للاستشارات
www.manaraa.com

Table 4: Credit rating shift table.

Change in credit rating	Risk classification	Risk definition
A	Normal	Repayments may be stable and will not be breached. Probability of default will not exceed 5%.
B	Attention	Repayment may have certain doubts, default probability is generally between 5% and 20%.
C	Focus on	Repayment may have big doubts, the probability of default is roughly between 20%~95%.
D	Track	Repayment may be minimal, no more than 5%.

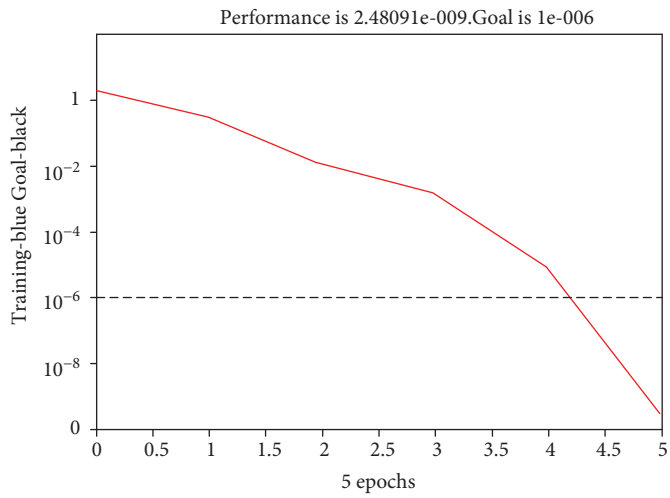

Figure 8: LM training results.

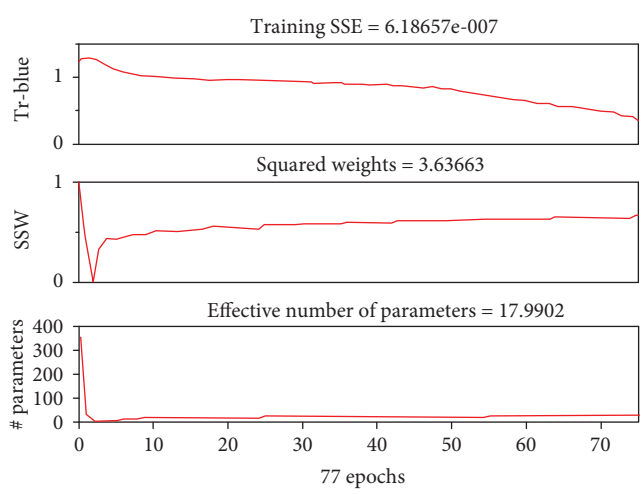

Figure 9: Bayesian training results.

dynamic concept in the dynamic credit tracking model is reflected in the time dimension. The index data of each person are different in different periods, so the results predicted by the model evaluation are different. The results are displayed in time series to reflect the dynamic concept.

As can be seen from Table 5, the simulation accuracy of each sample method is over 80%, and some even reach 100%. The simulation results are very good.

5.5. Application of GA-BP Neural Network Model. The micro-level dynamic credit tracking model corrects the information asymmetry of the loan, thereby regulating the behavior of the relevant market entities. At the macro level, it is the service supervision that promotes financial stability. Specifically, the credit tracking model mainly achieves the following five major functions: (1) automatically identify the credit risk of the loan; (2) dynamically monitor the credit status of the lender; (3) automatically alert the credit risk; (4) provide risk analysis report and risk operation suggestion; and (5) reports and risk operational advice.

The credit status of the lender has changed, and the rating system needs to be rerated to reflect the true credit level in a timely manner. A personal credit rating system based on the lender's personal basic information indicators including personal economic ability indicators and personal credit indicators is established. The creditor's credit file is established based on the results of the credit rating in the personal credit rating system. These files also contain subfiles in the modules that will be mentioned below. Since the credit rating weight and indicator scores in the personal credit rating system need to be dynamically adjusted according to the changes of the lender's actual situation, the credit status in the personal credit file is constantly changing.

In this paper, strict postloan risk monitoring was implemented through the credit tracking model, and its functional structure was divided into four modules according to different degrees of loan expected default probability and

المنارة للاستشارات
www.manaraa.com

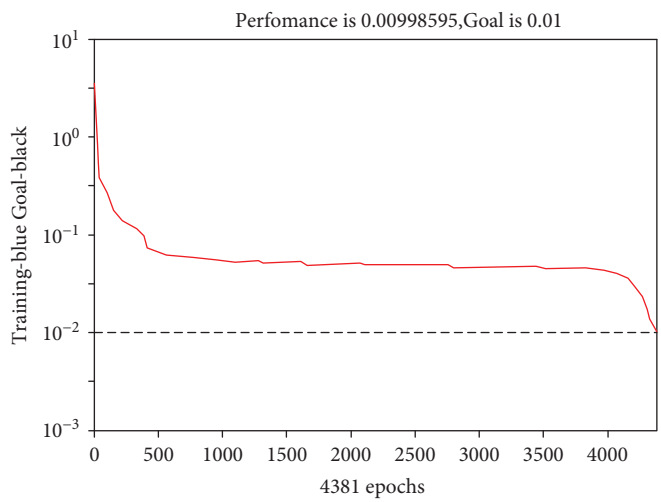

Figure 10: Momentum gradient training results.

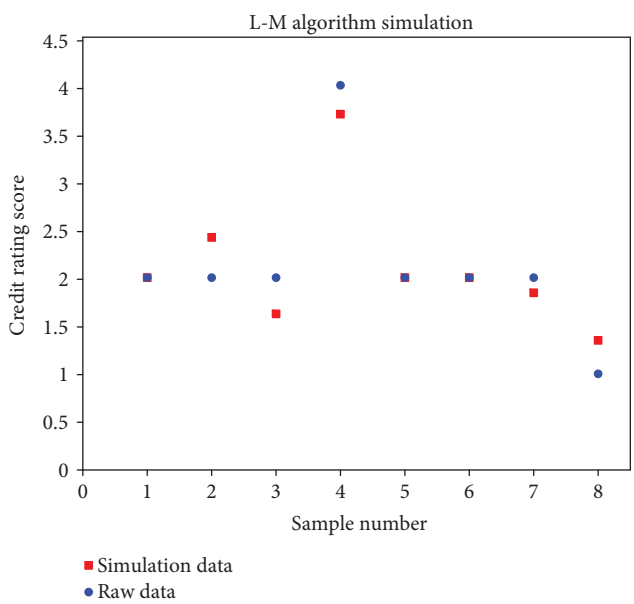

Figure 11: LM test results.

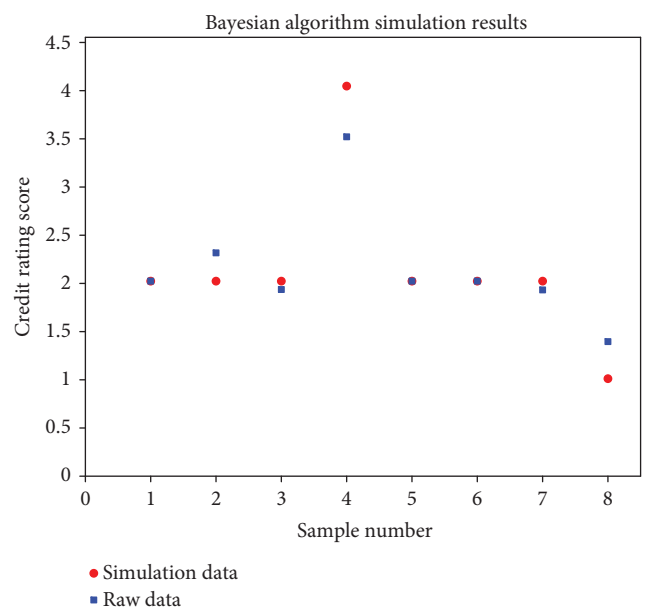

Figure 12: Bayes test results.

corresponding credit rating of the lender, namely contact module, tracking and supervision module, warning module, and disciplinary module. These four submodules are at different expected default probability and risk level, which may exist at the same time, and may exist in turn as the risk level changes with the extension of the loan period. The specific design is as follows.

5.5.1. Contact the Module. The credit rating is attributed to the contact module in the normal risk category, indicating that the expected default probability is low, the credit rating is increased, and the loan repayment may be stable. In the contact module, a contact interval based on the probability of default is set and issue different warning signals according to different levels of expected default probability. Based on this, the bank establishes a contact subfile in the credit file and maintains different degrees of contact with the lender accordingly, supervising its timely repayment during the credit period. At the same time, the discount on the repayment can be appropriately given.

5.5.2. Tracking and Urging Modules. Concerned about the risk level in the personal credit rating system, a tracking and supervision interval based on the corresponding expected default probability is set. In the credit tracking model, the expected default tracking index is determined, that is, the best default probability under the bank utility maximization (expected default loss and income equilibrium state), and the default probability is used as a standard to divide the three subintervals of high, medium, and low to

المنارة للاستشارات
www.manaraa.com

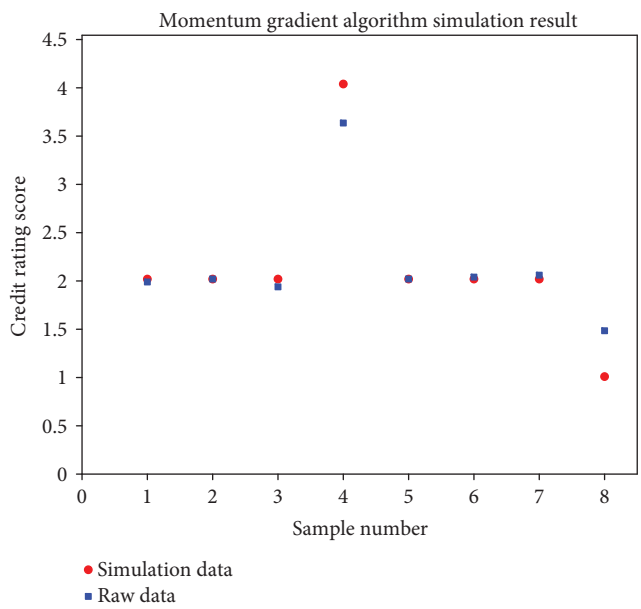

Figure 13: Momentum gradient test results.

reflect the change in credit status. Banks are advised to adopt different tracking strategies. At the same time, a tracking subfile is established in the credit file to record the credit status of such lenders. For lenders who do not contact the bank on time or do not contact the bank, they should reevaluate their credit level, pay close attention to and promptly send them a warning signal, or even urge them to repay in advance. Lenders who do have difficulties in repaying loans should be appropriately extended or reduced in order to promote their repayment enthusiasm and help to improve their credit level.

5.5.3. Early Warning Module. With the extension of the repayment period, the credit status of the lenders is constantly changing. The data analysis and evaluation in the above files are combined to classify the lenders whose risk level is on the suspicious level to establish an early warning subfile to record their credit during the warning period. Similarly, an early warning index that maximizes the utility of the bank is established in this paper, which is the best expected default probability, and correspondingly determines a warning expected default probability interval. For changes in the probability of default within the warning interval, i.e., changes in credit status, it is recommended that banks adopt different strategies to reduce losses.

5.5.4. Disciplinary Modules. For lenders at risk levels of loss, the probability of default is high and almost no repayment is possible. It is recommended that banks determine the probability of punishment and the amount of punishment for such lenders so that their default losses are much less than their default gains, but they cannot exceed the level of possible tolerance. At the same time, reasonable intervals for breach of contract and penalizing the classification of defaults by lenders in different situations are recommended, which will also be recorded in the disciplinary file to increase the cost of refinancing it to the financial system.

The specific operation and implementation of the above functions can help our dynamic credit tracking model be put in place by constructing a perfect and strict debt collection system, and the detailed steps are completed in turn: the establishment of a personal consumer credit collection department within the bank, the main function is to track and collect debts for individuals who do not repay their personal consumer credit in a timely manner. At the same time, it is necessary to strengthen the construction of the network, strengthen the internal information exchange between banks on the credit status of lenders through the interconnection of the Internet and information resources, and strengthen the relationship between the bank and the lender's unit. The steps of the debt collection system can be planned like this:

(1) For lenders who do not contact the bank on time or contact the bank, the bank's debt collection department should immediately contact and negotiate with the lender's unit. At this time, the bank and the unit can urge the repayment to assist the tracking system to play a tracking and urging role

(2) Once the unit is unknown or the lender does not know where to go, the bank debt collection department should immediately contact the lender's family, and the parents provide the lender's whereabouts and urge the lender to repay the loan, if necessary, by their parents to return it. The bank debt collection department should immediately contact the public security department through a loan lender whose family is still unable to contact and conduct inquiries nationwide through its unique identity card number. In short, it is necessary to achieve the supervision and contact functions through the dual channels of family and society and increase the intensity of tracking

(3) If the above situation still does not work, the bank can immediately freeze or stop its basic account, recover the loan, and resort to the law if necessary. In the event that a credit loan cannot be recovered, the banking institution will suffer huge losses. In order to minimize such losses, the borrower may be required to provide mortgage guarantees or use credit insurance and personal credit insurance to pass on some risks. In addition, personal credit risk can be passed on through marketization or corresponding insurance business. These are the risk warning functions of the debt collection system. Once the risk warning signal appears, not only should the above-mentioned functions such as tracking, supervision, and contact be linked, but it should also be transferred to the fourth step, that is, disciplinary action will be implemented immediately

(4) The realization of the disciplinary function should also establish a matching reward and punishment mechanism to assist the dynamic tracking model to implement debt collection

المنارة للاستشارات
www.manaraa.com

Table 5: Summary table of multiple algorithm test results.

Sample number	Credit rating score (original)	LM simulation Epoch = 5 Goal = 1e-6	Simulation effect MSE=2.409e009	Bayesian simulation Epoch = 5 Goal = 1e-6	Simulation effect MSE=3.0933e008	Dynamic gradient simulation Epoch = 43,381 Goal = e-2	Simulation effect MSE=0.0100
001	2	2.00	100.00%	2.00	100.00%	1.97	99.00%
002	2	3.02	90.00%	2.30	90.00%	2.01	99.67%
003	2	1.66	88.67%	1.91	97.00%	1.93	97.67%
004	4	3.88	96.00%	3.48	82.67%	3.61	87.00%
005	2	2.00	100.00%	2.00	100.00%	1.98	99.33%
006	2	2.00	100.00%	2.00	100.00%	2.03	99.00%
007	2	1.81	93.67%	1.91	97.00%	2.05	98.33%
008	1	1.19	93.67%	1.39	87.00%	1.47	84.33%

Annotation: (1) Simulation effect = 1 − |Simulation value − True value| ÷ (max(True value) − min(True value)) = 1 − |Simulation value − True value| ÷ 3. (2) Epochs: network training times; goal: network training target error; MSE: error vector.

المنارة للاستشارات
www.manaraa.com

On the one hand, an incentive mechanism with incentives is established, which is contrary to the penalty mechanism of default, and the purpose is to contrast with punishment, that is, to form an incentive compatibility mechanism, rewarding those who default on the lender, rewarding those in order to maximize the effectiveness of bank credit tracking; the bank should determine the reward ratio and the amount of reward in the credit tracking model to achieve zero-sum between the bank and the lender. Game equilibrium state: the specific design can be as follows: the preferential treatment for the loan repayment of the loaner who pays in advance. Regularly publish online or media information on the credit rating of the lender. For lenders with high credit ratings, the bank will reduce the interest rate or reduce the principal discount to promote its repayment or early repayment, thereby reducing the bank collection cost. For those who have difficulties in repaying loans, they should be given a principal or interest reduction or a graceful loan period; for the support of the branch, the western volunteers, etc., in accordance with national policies, the principal and interest reduction or grace period should be granted.

On the other hand, a penalty mechanism with notice and coercion is enacted to force the arrears to repay. Specific measures can be taken as follows: First, lenders whose credit rating is lower than the warning line are announced in time in the financial system and warned that the cost of refinancing will increase or it will be difficult to finance. Second, for the untrustworthy lender, the breach of contract will be recorded in the personal credit file of the loan and further incorporated into the national personal credit information system, so that the individual transaction behavior will be affected in the future, and the social supervision effect of a life-long loss affecting life will be achieved. Third, if the circumstances are serious, the legal liability of the defaulting borrower will be investigated according to the law, and the name of the introducer or witness who fails to perform the duties will be announced. The fourth is to implement a personal bankruptcy system and a credit guarantee mechanism. With restrictions on consumption and harsh constraints on bankrupt individuals, people can't travel abroad, can't use credit cards, can't enjoy loan services, and can't buy high-end goods. At the same time, a social credit guarantee company is established to guarantee personal credit.

6. Conclusions

This paper analyzes the status quo of the credit industry and uses mathematical models to demonstrate the inevitability of financial information industry development of big data. On this basis, a dynamic credit tracking model of BP neural network based on adaptive genetic algorithm was constructed. Using MATLB software to simulate, it can be seen from the training results that the LM training algorithm and Bayesian algorithm can converge within 10e-6 quickly. While momentum gradient algorithm convergence is slow, the overall training effect is good, and the simulation results are all up to 90%, indicating that the model can be well applied to the big data in Internet finance credit.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there is no commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments

This research is supported by the Shandong Province Natural Science Fund Project "the Simulation Research on Industry, Population, Employment, Education and Social Security Policy under the new normal of Shandong province" (Project number ZR2015GL013), the Shandong Social Science Program Research Project "Study on the Value Orientation and the United Front Strategy of the Young Generation of Private Entrepreneurs of Shandong" (Project number 17CTZJ05), and the Shandong Institute of Business and Technology's teaching reform research project, "Collaborative Education mode of Applied Talents in Local Universities"—Taking Human Resources Management as an Example (Project number 11688G201720).

References

[1] R. H. Wang, "Research on personal credit consumer finance based on mobile terminal," *Chinese Commercial Theory*, vol. 34, pp. 29-30, 2017.

[2] X. G. Chen and Y. Q. Li, "Research on the transmission mechanism of internet credit risk based on the perspective of internet finance," *Innovation*, vol. 11, no. 1, pp. 80–90, 2017.

[3] C. H. Liu, J. Jiang, and J. Li, "Business strategy and perceived benefits of internet banking: their impact on banks' strategic responses to China's entry to WTO," *Journal of Dong Hua University*, vol. 20, no. 3, pp. 63–68, 2003.

[4] Y. C. Zhang and J. Z. Yang, "Research on internet financial credit issues in the background of big data," *E-commerce*, vol. 5, pp. 55-56, 2016.

[5] W. F. Feng and C. M. Li, "Research on big data credit construction under the background of internet finance," *International Finance*, vol. 10, pp. 61–66, 2015.

[6] Z. Chen, "Practice of comprehensive credit reporting and differentiated development at home and abroad," *Credit Reporting*, vol. 36, no. 06, pp. 50–53, 2018.

[7] G. S. Pan, "Building a well-developed Chinese credit market," *Credit Information*, vol. 32, no. 11, pp. 1–4, 2014.

[8] N. Ghatasheh, "Business analytics using random forest trees for credit risk prediction: a comparison study," *International Journal of Advanced Science and Technology*, vol. 72, pp. 19–30, 2014.

[9] R. A. Bartlett, *Practitioner's Guide to Business Analytics: Using Data Analysis Tools to Improve Your Organization's Decision Making and Strategy*, McGraw Hill Professional, 2013.

المنارة للاستشارات
www.manaraa.com

[10] W. Guo, Y. Z. Wang, and J. Y. Qin, "Research on the construction of personal credit reporting system in China under the background of big data," *Modern Management Science*, vol. 3, pp. 3–8, 2018.

[11] Z. F. Chang, *Research on the Construction and Risk Management of Internet Financial Credit System*, Nanjing University, 2018.

[12] L. Chen, T. Y. Zhou, J. Y. Ren, and Y. X. Gu, "Research on the application of large data credit reporting in the internet background: taking "ant golden clothes" as an example," *Market Weekly*, vol. 7, pp. 108–110, 2018.

[13] J. L. Wen and L. H. He, "Cooperation of credit investigation between mainland China and Taiwan under the new situation," *Asian Agricultural Research*, vol. 8, no. 10, pp. 9–12, 2016.

[14] B. A. Kusi, E. K. Agbloyor, K. Ansah-Adu, and A. Gyeke-Dako, "Bank credit risk and credit information sharing in Africa: does credit information sharing institutions and context matter?," *Research in International Business and Finance*, vol. 42, pp. 1123–1136, 2017.

[15] B. J. Yang and X. H. Liu, "Analysis of basic credit data mining," *Modern Management Science*, vol. 8, pp. 54–56, 2015.

[16] W. Wei, Z. Sun, H. Song, H. Wang, X. Fan, and X. Chen, "Energy balance-based steerable arguments coverage method in WSNs," *IEEE Access*, vol. 6, pp. 33766–33773, 2018.

[17] L. Ji, H. Chi, and J. M. Chen, *Application of Multivariable Analysis in Performance Evaluation of Bank and Savings Deposits*, China Management Science, 2001.

[18] J. Yi, Q. Wang, D. Zhao, and J. T. Wen, "BP neural network prediction-based variable-period sampling approach for networked control systems," *Applied Mathematics and Computation*, vol. 185, no. 2, pp. 976–988, 2007.

[19] Y. Wang, T. Q. Luan, and M. Z. Pu, "Optimization of acetic acid fermentation medium based on neural network and genetic algorithm," *Chinese Journal of Food Science*, vol. 12, no. 5, pp. 88–94, 2012.

[20] K. Cui and X. Qin, "Virtual reality research of the dynamic characteristics of soft soil under metro vibration loads based on BP neural networks," *Neural Computing and Applications*, vol. 29, no. 5, pp. 1233–1242, 2018.

المنارة للاستشارات
www.manaraa.com

Copyright © 2018 Fei-Peng Wang. This is an open access article distributed under the Creative Commons Attribution License (the “License”), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Notwithstanding the ProQuest Terms and Conditions, you may use this content in accordance with the terms of the License. http://creativecommons.org/licenses/by/4.0/

المنارة للاستشارات

www.manaraa.com