Hindawi
Scientific Programming
Volume 2018, Article ID 2613739, 14 pages
https://doi.org/10.1155/2018/2613739

Research Article

Optimizing the Borrowing Limit and Interest Rate in P2P System: From Borrowers' Perspective

Zhihong Li, Lanteng Wu, and Hongting Tang

College of Business Administration, South China University of Technology, Guangzhou 510641, China

Correspondence should be addressed to Hongting Tang; ht_tang@foxmail.com

Received 6 April 2018; Revised 2 July 2018; Accepted 19 July 2018; Published 23 August 2018

Academic Editor: Juan A. Gomez-Pulido

P2P (peer-to-peer) lending is an emerging online service that allows individuals to borrow money from unrelated person without the intervention of traditional financial intermediaries. In these platforms, borrowing limit and interest rate are two of the most notable elements for borrowers, which directly influence their borrowing benefits and costs, respectively. To that end, this paper introduces a BP neural network interval estimation (BPIE) algorithm to predict the borrowers' borrowing limit and interest rate based on their characteristics and simultaneously develops a new parameter optimization algorithm (GBPO) based on the genetic algorithm and our BP neural network predictive model to optimize them. Using real-world data from http://ppdai.com, the experimental results show that our proposed model achieves a good performance. This research provides a new perspective from borrowers in exploring the P2P lending. The case base and proposed knowledge are the two contributions for FinTech research.

1. Introduction

Recent research about P2P lending is mainly focused on two aspects. The first one is the empirical research of investor's behavior in the online loan platform to clarify the impact factors of investor's risk preference [1] and investor's choice [2–4]. The second aspect is borrowers' credit scoring and their probability of default [5–8], which is important to control the risk for P2P lending. For the borrowers, however, there are few efforts to investigate their borrowing benefits and costs or discuss the important principles of utility enhancing. This question is valuable since borrowers' experiences have the same significant influence as lenders' in ensuring the proper functioning of the P2P lending system. Meanwhile, a good guidance for borrowers will improve their benefits and further increase the platform's cash flow.

However, information asymmetry is also a universal phenomenon in P2P lending. It not only brings low efficiency to transaction but also brings more cost to borrowers. Specifically, the existence of information asymmetry makes borrowers overwhelmed by behavior choices since the corresponding evaluation mechanism in P2P lending is opaque for them. In that case, the key challenge is to examine profitable behaviors to optimize borrowers' borrowing parameters under limited information. Therefore, the purpose of this paper is to propose a comprehensive method to help borrowers predict and optimize their borrowing limit and interest rate simultaneously.

To build a target model, there are many methods to choose, for example, traditional regression model based on statistics and popular machine learning technology (e.g., association analysis, decision tree, and neural network model) derived from artificial intelligence. It is well known that the regression model is a widely applied mathematical model in data science. But there are also some limitations (e.g., high sensitivity to outliners and poor predictive accuracy) which are particularly remarkable in the case of complex and nonlinear problems [9, 10]. Therefore, we will give priority to the machine learning method in our context for two reasons. First, P2P lending is a data-intensive application which contains many nonnumeric contents. The second reason is the ability and robust of machine learning to process unstructured data. Given the situation of information explosion, we propose to adopt the three-layer BP neural network since it performs well in processing complex nonlinear functions.

Before predicting the borrowers' borrowing limit and interest rate, it is worth mentioned that obtaining a confidence interval of borrowing parameters is more credible and practical than an exact number for both researchers and borrowers. For this reason, we employ the three-layer BP neural network model to fit the complex relationship among different features in P2P lending and to further deduce the required confidence intervals. Specifically, we develop a BP neural network interval estimation (BPIE) method inspired by Duan and Xie's work [11], which obtains two-sided confidence interval for the real estate blessed price based on the BP neural network, to generate one-sided and two-sided confidence intervals for both maximum borrowing limit and lowest interest rate.

To optimize the borrowers' borrowing limit and interest rate, we construct a new parameter optimization algorithm (GBPO) based on the genetic algorithm (GA) and BP neutral network. Significantly, the most popular research topic of the BP neural network mainly involves sensitivity analysis, which aims to measure the effect of input parameters on output and acquire the most powerful input parameters [12, 13]. However, there is hardly any research attempted to identify appropriate combination of input to acquire better output, especially for P2P lending in perspective of borrowers. So, this paper reports the first work to obtain best behavior characteristics for borrowers to help them optimize their borrowing limit and interest rate. That is, the target of optimization analysis is obtaining improvement directions of behavior parameters within their adjustable range and also examining the most important parameters to maximize the borrowing limit or minimize the interest rate.

As mentioned previously, our proposed method (GBPO) is a combination of GA and BP algorithms, which is a common combination form in recent researches. A classic example of such a combination is the GA-BP algorithm, which is designed to overcome the limitations of the BP neutral network, that is, time-consuming in data training and undesired local optimization. In the GA-BP algorithm, the genetic algorithm performs a global search on weight ranges and finds out more reliable initial weights for the BP neural network [14]. In contrast to GA-BP algorithm, our method is designed to find fluent parameter combination of input from large number of available parameters for better experimental result. In specific, the BP neutral network is the basic construction of our optimized model; the optimal output of the BP neutral network was taken as the fitness function in the GA.

By taking the advantage of our proposed method, we conduct a systematic analysis in terms of both prediction and optimization for borrowers' borrowing parameters based on real-world data from http://ppdai.com. Experimental results show that our proposed method performs well and discovers more meaningful characteristics. Also, this paper provides a different perspective for research effort of the P2P lending system.

2. Data Preparation

In this section, we first give a brief look at basic knowledge of our adopted data. We then describe the selected parameters from target dataset. For a more convincing preference, we further conduct a relationship analysis for selected parameters.

2.1. PPdai Lending Platform. Our primary dataset is the publicly available data in the PPdai lending platform. The PPdai lending platform is the first online lending platform in China, and it was built in August 2007. As an unsecured credit lending platform for individuals, users in this system can be divided into two groups, one is borrowers and the other is investors. Borrowers post their loan information in the platform to borrow money at lower interest rates, while investors choose which borrowers to lend to by considering the borrower's loan materials and credit rating to earn higher returns. In general, the PPdai platform aims at small loans, and the loan limit is usually spanned from 100 to 30,000 yuan [15, 16].

To reduce the risk of the borrower defaulting, the PPdai platform also develops a credit rating system which is called the *Magic-Mirror System*. The *Magic-Mirror System* is a big data analysis system that simultaneously collects and processes more than 2000 kinds of characteristics of borrowers in order to precisely evaluate their credit score and risk of default. The main characteristics concerned by the *Magic-Mirror System* include personal information, third-party data, repayment history, personal debt, and credit history. Based on this credit rating system, each borrower will get a different level of credit rating ranging from AAA to F, whose risk is rising in turn.

2.2. Selected Parameters. To prepare our dataset, we carefully observe the information structure in the PPdai lending platform and obtain a total of five dimensions of 17 parameters (as shown in Table 1) as our research materials. The selected data are extracted by the following principles: (1) collecting transactional pages by replacing the id number of URL (i.e., replacing 37215007 with a specific number in "http://invest.ppdai.com/loan/info?id=37215007"), (2) obtaining corresponding user pages through collected transactional pages, and (3) crawling and saving each user's basic information and borrowing information from user pages. In addition, considering that user information is constantly updating with the completion of transactions, we filter out the outdated information and only select the latest transaction data of each users as our target dataset.

Through this process, we finally obtain our valid dataset containing 10192 documents. To get a general idea about our dataset, we give descriptive statistics of the numeric parameters, binary parameters, and categorical parameter of MMR, and the details can be seen in Table 2, Figure 1, and Table 3, respectively.

From Table 2, we can get the knowledge of statistics characteristics of each numeric parameter in our sample, which will help us understand the approximate distribution of each parameter. Figure 1 denotes the ratio of each category for each binary parameter. And in Table 3, we calculate the proportion of each category for the categorical parameter of MMR to clarify the user credit distribution in the PPdai lending platform.

Table 1: Parameters of borrowers in the PPdai lending platform.

Dimensions' definition	Parameters' definition	Parameter type	Abbreviation	Remark
Identity information (IDI)	Head portrait	Binary	HPT	
	Gender	Binary	GED	Male = true, female = false
	Identity	Binary	IDT	True if identified as a lender, online seller, student, and so on
	Age	Numeric	AGE	
	Registration time	Numeric	RET	Number of days since registration
	Profile score	Numeric	PRS	A score based on additional information (e.g., age, education, work, and income)
Borrowing information (BRI)	Amount	Numeric	AMT	Yuan
	Rate	Numeric	RAT	
	Period	Numeric	PED	Months
Certification information (CFI)	Identity certification	Binary	IDC	
	Video certification	Binary	VDC	
	Education certification	Binary	EDC	
	Mobile phone certification	Binary	MPC	
	Online banking recharge certification	Binary	ORC	
Historical transaction information (HTI)	Transaction score	Numeric	TAS	Times of paying off loans (1 point each time)
	Transaction lost score	Numeric	TLS	Times of paying back loans delaying over 15 days (−2 points each time)
Risk rating information (RRI)	*Magic-Mirror* rating	Categorical	MMR	From AAA to F (8 levels)

Table 2: Descriptive statistics of numeric parameters.

Statistics	AMT	RAT	PED	AGE	RET	TAS	TLS	PRS
Sample size	10192	10192	10192	10192	10192	10192	10192	10192
Mean	2832.20	17.10	7.89	28.17	275.17	4.64	−0.08	0.63
Minimum	100	0	3	18	4	0	−16	0
Median	2100	18	6	27	232	3	0	0
Maximum	49000	36	24	55	3115	68	0	30

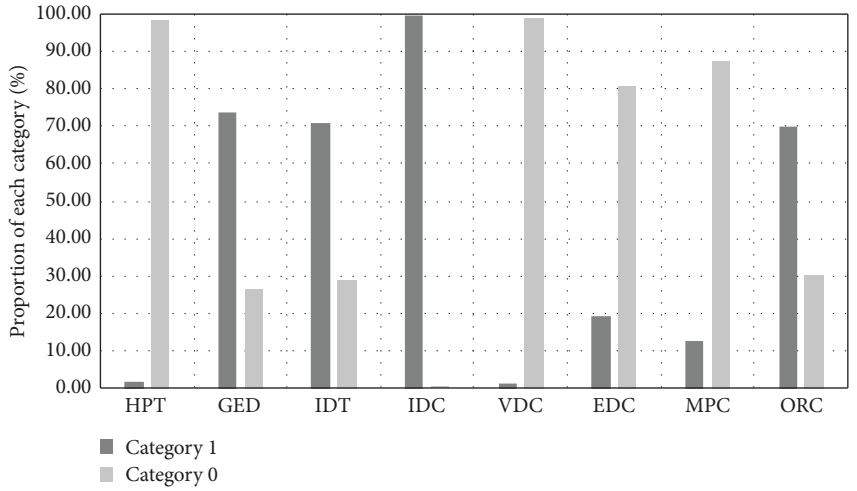

Figure 1: Descriptive statistics of binary parameters.

Table 3: Descriptive statistics of categorical parameter.

MMR	AAA	AA	A	B	C	D	E	F
Proportion (%)	0.18	47.5	0.83	3.35	14.83	19.38	12.16	0.7

Besides, in order to facilitate the parameter optimization, here we clarify the available adjustable direction for parameters and summarize the applicable adjustment of each parameter in Table 4.

2.3. Relationship Analysis. To establish an accurate predictive and optimized model, it is necessary to clarify the relationships of various parameters in the PPdai platform. There are several official statements about their determined relationships: (1) the *Magic-Mirror* rating mainly depends on repayment history, credit history, personal debt, personal information, third-party data, and others. (2) The effects of *Magic-Mirror* rating on users' borrowing can be seen in the following aspects: (a) whether your borrowing can be approved; (b) interest rates (the higher the rating, the lower the interest rate); and (c) borrowing limit (the higher the rating, the higher the borrowing limit).

Given the official statements, we further explore the relationship of parameters we selected in the PPdai platform as shown in Figure 2. From this figure, we can find that RRI (i.e., *Magic-Mirror* rating) of borrowers is mainly determined by parameters in dimensions of IDI, CFI, and HTI and in turn affects the borrowers' BRI (AMT and RAT included). It is worth mentioning that parameter dimension of BRI is not completely determined by RRI. RRI has a real impact on borrowers' borrowing limit and their interest rate, while IDI, CFI, and HTI also have an effect on them to some extent. Furthermore, borrowers freely choose their AMT (i.e., borrowing amount) within their borrowing limit. Significantly, since the borrowing limit of each borrower is not publicly accessible in the PPdai platform, we assume that each borrower uses the maximum value when borrowing and thus we take the borrowing amount as their borrowing limit in the following analysis.

Even though borrowing limit and interest rate of borrowers are available for oneself, borrowers in the PPdai platform have no access to the specific evaluation approach. It is therefore not possible for borrowers to take more targeted actions to optimize their borrowing parameters (i.e., borrowing limit and interest rate). To this end, we propose to develop a reasonable model to obtain and optimize borrowing parameters for borrowers using the available characteristics.

Table 4: Applicable adjustment for each parameter.

Abbreviation	Applicable adjustment
HPT	If the user adds a head portrait, the value will turn to 1.
GED	Not adjustable.
IDT	If the user translates his/her identity to a lender, online seller, or student, the value will turn to 1.
AGE	It will increase with user's age.
RET	It will increase with user's registration time.
PRS	It will increase with user's income, education level, and others.
IDC	If the user performs the authentication, the value will turn to 1.
VDC	If the user performs the authentication, the value will turn to 1.
EDC	If the user performs the authentication, the value will turn to 1.
MPC	If the user performs the authentication, the value will turn to 1.
ORC	If the user performs the authentication, the value will turn to 1.
TAS	If the user pays back on time, the value will be added 1.
TLS	If the user is 15 days behind on payment, the value will reduce by −2.
MMR	It will change with the parameter values above.

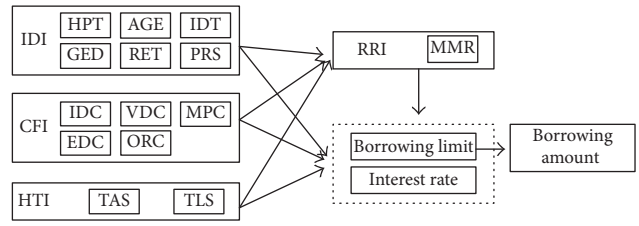

Figure 2: Relationship of selected parameters.

3. Proposed Prediction Model

3.1. BP Neural Network Model. Here, we build a BP neural network model as our basic model to predict borrowers' borrowing limit and interest rate. Based on the analysis above, we use 14 parameters that belong to four different dimensions as our independent materials and make the LMT (i.e., borrowing limit) and RAT (i.e., interest rate) as our target variables. The constructed architecture of our model is shown in Figure 3.

As shown in Figure 3, we adopt a three-layer neural network model which contains the input layer, hidden layer, and output layer. The input layer includes 14 neurons and a neuron bias, the hidden layer includes N_{hid} neurons and a neuron bias, while the output layer includes 2 neurons. This is a flexible method, which can smoothly process arbitrary nonlinear functions [17]. To achieve the desired result, the activation function of the input layer and hidden layer is tansig (i.e., $f_1(x) = 2/(1 + e^{-2x}) - 1$), and the activation function of the output layer is pure linear function (i.e., $f_2(x) = x$). Moreover, the error function is defined as follows:

$$E = \frac{1}{2S}\sum_{i=1}^{S}\left(Y_i - \widehat{Y}_i\right), \tag{1}$$

where S is the sample size of the dataset, Y_i is the confirmatory dependent variable values, and $\widehat{Y}_i$ is the predicted dependent variable values. Then, we use gradient descent training method as the learning algorithm to train the neutral network.

3.2. Predictive Quality. To examine the predictive quality of our algorithm, we randomly assign the collected data into three parts which are called TRD (train data), TSD1 (test data-1), and TSD2 (test data-2), with the overall proportion of their size being 80 percent, 10 percent, and 10 percent, respectively. To process the model, we first set the value of

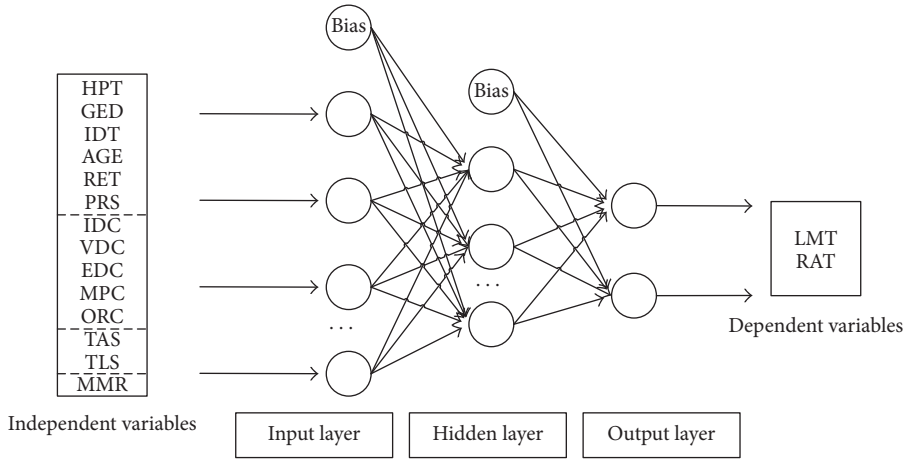

Figure 3: Architecture of the BP neural network model.

N_{hid} to 12, and other parameters in the process of training can be set as shown in Table 5 [18, 19]. On the basis of these preparations, we run the BP neutral network and get the performance of our predictive model, which is shown in Table 6. Here, we use the correlation coefficient to measure the accuracy of the prediction.

From Table 6, we can find that the predictive accuracy of LMT is not very high, while the predictive accuracy of RAT is especially high (up to 98 percent). Through analysis, we find that the RAT is determined, to a great extent, by the parameters shown in Figure 3 and borrowers are not authorized to change the value. LMT (i.e., AMT) is not just determined by the parameters but also can be freely chosen by borrowers within a specific limit. So, it is not difficult to understand why the predictive accuracy of the borrowing limit is not high.

3.2.1. Parameter Setting of N_{hid}. Here, we want to find a reasonable value for the parameter N_{hid}. There is an empirical formula for determining the number of neurons in the hidden layer, which is shown below:

$$N_{\text{hid}} = \sqrt{N_{\text{in}} + N_{\text{out}}} + a, \quad a = 1, 2, \ldots, 10, \tag{2}$$

where N_{in} denotes the number of neurons in the input layer, N_{out} denotes the number of neurons in the output layer, and a is a constant bounded between 1 and 10 [20–22]. Therefore, we can set N_{hid} in the range of 5~14 in our predictive model. To obtain the efficient value, we also process the model and acquire the predictive performance of our model using various N_{hid}. The results are shown in Figure 4.

As shown in Figure 4, the predictive accuracy of RAT has no significant difference, but when the value of N_{hid} is closer to 12, the predictive accuracy of RAT is slightly higher than others for three datasets. Therefore, we will use N_{hid} value of 12 in the following analysis.

3.3. Robust Check. We note that different forms of selection biases may affect our result. Thus, we further conduct an analysis using samples from some specific credit ratings as a robust check. In this case, we choose three group of samples, which come from credit rating of AA, C and D, respectively, to validate our proposed model. After randomly assigning the collected data into three parts (i.e., train data, test data-1, and test data-2) for each sample group, we utilize the train data to train the neural network and test the result using test data-1 and test data-2. The corresponding result can be seen in Table 7, and we employ the correlation coefficient to measure the accuracy of the prediction.

As shown in Table 7, the prediction accuracy of our predictive model slightly declines in predicting LMT for most of data samples. One possible explanation is that the selected sample sizes are smaller that original one (Table 6). For samples from credit rating D, the prediction accuracy of our predictive model approaches or briefly exceeds the original setting, which means that our model has better performance in predicting LMT for borrowers in the D level. So, basic information plays an important role in predicting LMT. From Table 7, we can also find that the prediction accuracy of our predictive model declines dramatically in predicting RAT for all selected samples compared with the original one. It is not just because of reduced sample size but

Table 5: Parameter setting of the predictive model and optimization model.

Algorithm	Parameter	Definition	Levels
GA	N_{hid}	Number of neurons in the hidden layer	5, 6, . . . , 14
	η	Learning rate	0.05
	ep	Maximum number of iterations	50000
BP	M	Population size	20, 40, . . . , 100
	P_{c}	Crossover probability	0.8
	P_{m}	Mutation probability	0.09
	G_{max}	Maximum number of iterations	100, 200

Table 6: Performance of the predictive model.

Target parameters	TRD	TSD1	TSD2
LMT	0.49	0.48	0.45
RAT	0.99	0.99	0.98

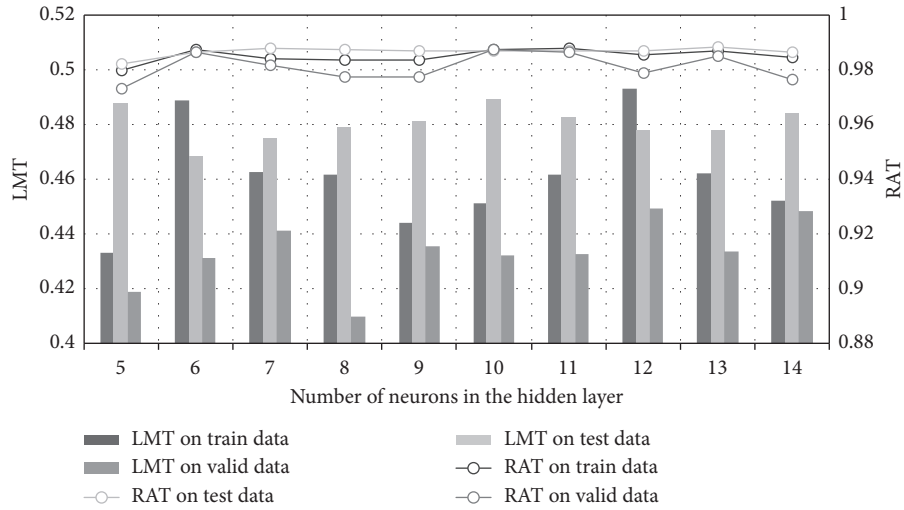

Figure 4: Predictive performance of the proposed model using various N_{hid}.

Table 7: Performance of the predictive model for different data samples.

Target	Predictive model of AA			Predictive model of C			Predictive model of D		
	TRD_{AA}	TSD1_{AA}	TSD2_{AA}	TRD_{C}	TSD1_{C}	TSD2_{C}	TRD_{D}	TSD1_{D}	TSD2_{D}
LMT	0.45	0.33	0.43	0.44	0.33	0.15	0.65	0.53	0.47
RAT	0.26	0.16	0.18	0.13	0.15	0.03	0.12	0.04	0.06

mainly results from the significance of credit rating to RAT. Based on the analysis above, it is necessary to consider both credit rating and basic information rather than only one of them in predicting LMT and RAT.

3.4. Prediction of Borrowing Limit. Here, we want to predict the borrowing limit (LMT) for borrowers using our proposed predictive model. As mentioned above, borrowing limit is the amount of the money an individual could borrow from others, and we will take borrowing amount as our borrowing limit here because of data deficiency. Meanwhile, it is worth noted that the expected result of our prediction is a confidence interval for borrowing limit but not an exact value since it is more credible and practical for us.

Given this knowledge, we extend Duan and Xie's work [11] and introduce the BP neural network interval estimation (BPIE) algorithm to deduce a one-sided confidence interval for LMT, which is the predictive value of borrowing limit. Since there is no specific provision on the lower limit for borrowing amount, we only examine the one-sided confidence interval for borrowing limit. The derivation process is described as below:

Firstly, we define the promotion error of LMT as follows: given a test set, the confirmatory value and predicted value of LMT can be represented as LMT and $\widehat{\text{LMT}}$, respectively. Then, the promotion error of LMT can be represented as follows:

$$\sigma_{\text{LMT}} = \sqrt{\frac{\sum_{i=1}^{k}\left(\text{LMT}_i - \widehat{\text{LMT}_i}\right)}{k-1}}, \tag{3}$$

where k denotes the sample size of the dataset and LMT_i and $\widehat{\text{LMT}_i}$ denote the confirmatory value and predicted value of LMT for the ith individual, respectively.

Based on the promotion theory of machine learning introduced by Yan and Zhang [23], we can infer that LMT_i approximately obeys the normal distribution $\text{LMT}_i \sim (\widehat{\text{LMT}_i}, \sigma_{\text{LMT}})$. So, the probability distribution of LMT_i can be defined as follows:

$$P\left\{\frac{\text{LMT}_i - \widehat{\text{LMT}_i}}{\sigma_{\text{LMT}}} < z_\alpha\right\} = 1 - \alpha. \tag{4}$$

Then, we can get the following formula:

$$P\left\{-\infty < \text{LMT}_i < \widehat{\text{LMT}_i} + \sigma_{\text{LMT}} z_\alpha\right\} = 1 - \alpha. \tag{5}$$

Finally, we can get the one-sided confidence interval with the confidence $(1-\alpha)$ for LMT_i, which is as follows:

$$\left(-\infty,\ \widehat{\text{LMT}_i} + \sigma_{\text{LMT}} z_\alpha\right). \tag{6}$$

3.5. Prediction of Interest Rate. In this section, we conduct a predictive analysis for borrowers' interest rate. We have already got the predicted value of RAT in the previous analysis. Now, we want to deduce a confidence interval for RAT in order to improve the robustness and reference value of our prediction.

To obtain the two-sided confidence interval of RAT, we follow the inference process of LMT. Firstly, since RAT_i approximately obeys the normal distribution $\text{RAT}_i \sim (\widehat{\text{RAT}_i}, \sigma_{\text{RAT}})$ like LMT_i, the probability distribution of RAT_i can be defined as follows:

$$P\left\{\left|\frac{\mathrm{RAT}_i - \widehat{\mathrm{RAT}_i}}{\sigma_{\mathrm{RAT}}}\right| < z_{\alpha/2}\right\} = 1 - \alpha. \quad (7)$$

Then, we can get the following formula:

$$P\left\{\widehat{\mathrm{RAT}_i} - \sigma_{\mathrm{RAT}} z_{\alpha/2}, \widehat{\mathrm{RAT}_i} + \sigma_{\mathrm{RAT}} z_{\alpha/2}\right\} = 1 - \alpha. \quad (8)$$

Finally, we can get the two-sided confidence interval with the confidence $(1 - \alpha)$ for RAT_i, which is as follows:

$$\left(\widehat{\mathrm{RAT}_i} - \sigma_{\mathrm{RAT}} z_{\alpha/2}, \widehat{\mathrm{RAT}_i} + \sigma_{\mathrm{RAT}} z_{\alpha/2}\right). \quad (9)$$

3.6. Prediction Results. Based on the analysis above, we use the train dataset TRD to train the neutral network and employ the well-trained neutral network to get the predictive value $\widehat{\mathrm{LMT}}$ and $\widehat{\mathrm{RAT}}$ on dataset TSD1. Then, (6) and (9) derived from the BPIE method are utilized to calculate confidence interval for LMT_i and RAT_i. Before that, the significant level α needs to be determined so that we can obtain confidence intervals meeting the credibility. Normally, we set significant level α to 0.05 [24, 25], and we get the confidence interval with a confidence of 0.95 for both LMT_i and RAT_i. So, the corresponding intervals are $(-\infty, \widehat{\mathrm{LMT}_i} + 3511.62)$ and $(\widehat{\mathrm{RAT}_i} - 2.41, \widehat{\mathrm{RAT}_i} + 2.41)$, respectively.

In order to clearly justify the predictive accuracy of the BPIE method, we define the predictive accuracy rate of the BPIE method as follows:

$$\eta_{\mathrm{LMT}} = \frac{N_{\mathrm{LMT}^*}}{N_s} \times 100\%,$$
$$\eta_{\mathrm{RAT}} = \frac{N_{\mathrm{RAT}^*}}{N_s} \times 100\%, \quad (10)$$

where N_{LMT^*} and N_{RAT^*} denote the number of samples that fall into the confidence interval of LMT and RAT, respectively, and N_s denotes the sample size of a valid dataset.

By examining the predictive accuracy of the proposed BPIE method using the dataset TSD2, we get the predictive accuracy for LMT and RAT with $\eta_{\mathrm{LMT}} = 94.80\%$ and $\eta_{\mathrm{RAT}} = 97.84\%$, respectively, which indicates that our BPIE method has a high predictive accuracy.

3.7. Comparable Analysis. It is well known that the BP neural network has better performance than linear regression in fitting nonlinear relationship. In order to explore the performance of both methods in our context, here we conduct a comparable analysis.

In keeping with our proposed predictive model, we also build a linear regression model to predict LMT and RAT and corresponding confidence intervals (see Appendix A). To compare two methods, we use the CCDR (correlation coefficient difference rate) metric [26] and LDR (length difference rate) metric [27]. As shown in (11), CCDR is used to compare the prediction accuracy of both methods in predicting individual values:

$$\mathrm{CCDR} = \frac{r_{\mathrm{NN}} - r_{\mathrm{LR}}}{r_{\mathrm{LR}}} \times 100\%, \quad (11)$$

where r_{NN} and r_{LR} denote the correlation coefficient (between confirmatory value and predicted value) of the BP neutral network model and linear regression model, respectively. $\mathrm{CCDR} > 0$ means that prediction accuracy of the BP neural network model is higher than that of the linear regression model in individual value prediction.

LDR is used to compare the prediction precision of both methods in predicting confidence intervals, which can be noted as follows:

$$\mathrm{LDR} = \frac{L_{\mathrm{LR}} - L_{\mathrm{NN}}}{L_{\mathrm{NN}}} \times 100\%, \quad (12)$$

where L_{LR} and L_{NN} denote the length of the confidence intervals with a confidence of 0.95 derived from the linear regression model and BPIE method, respectively. $\mathrm{LDR} > 0$ means that confidence interval derived from the BPIE method is more efficient than that from linear regression model.

Based on the works above, we calculate the related indicators and present them in Tables 8 and 9. From the results, since the values of CCDR and LDR are greater than 0, we can conclude that our predictive model outperforms the linear regression model in predicting both individual values prediction and confidence intervals.

4. Proposed Optimization Model

In this section, we propose to use the GBPO algorithm for optimizing the borrowers' borrowing limit and interest rate based on the previous analysis. We firstly introduce the GBPO algorithm we developed and then describe the single-target optimization method and double-target optimization method to get improvement directions of behavior parameters for borrowers to maximize the borrowing limit and minimize the interest rate, respectively.

4.1. Preliminaries

4.1.1. Optimization Target. Before introducing the optimization model, the optimization target needs to be clarified so that we could determine what our optimization measures are. Referring back to the prediction stage in Section 3, we predict the confidence interval of maximum borrowing limit (i.e., $\widehat{\mathrm{LMT}}$) and lowest interest rate (i.e., $\widehat{\mathrm{RAT}}$) for borrowers using the prediction system (i.e., the well-trained BP neutral network). Based on these works, we propose to explore the promising striving direction for borrowers to acquire higher borrowing limit or lower interest rate given a set of behavior characteristics. That is, we process prediction for each generated sets of explanatory parameters to obtain corresponding $\widehat{\mathrm{LMT}}$ and $\widehat{\mathrm{RAT}}$ and find the best explanatory parameters collection. Therefore, the desired target in our optimization stage can be denoted as follows:

$$\max\ y_1 = \widehat{\mathrm{LMT}},$$
$$\min\ y_2 = \widehat{\mathrm{RAT}}. \quad (13)$$

Table 8: Result of CCDR.

Objective parameters	TRD	TSD1	TSD2
LMT (%)	36.11	14.29	18.42
RAT (%)	2.09	1.02	1.03

Table 9: Result of LDR.

LMT	RAT
6.08%	130.29%

4.1.2. Explanatory Parameters. Given the optimization target, finding a reasonable way to optimize them is a natural demand. To reach this, we firstly determine the adjustable range of all explanatory variables. By analyzing the feature of each independent parameter, we obtain the adjustable ranges of them as shown in Table 10.

As shown in Table 10, these ranges are divided into two categories. The first category is applied to all borrowers, including all numeric variables and MMR. This category can be divided into two subtypes. One is used to add to initial values and another is used to replace the initial values. The second category is applied to individual borrowers, mainly including binary variables. Since the value of these independent parameters (e.g., certification information) only goes from 0 to 1, the value adjustment will just be suitable for borrowers whose corresponding parameters are zero. Besides, it is worth mentioning that we do not consider all the binary variables in our optimization algorithm for two reasons. First, some variables, such as gender, cannot be changed in reality for a specific borrower. Second, the sample size of some independent variables varies greatly between categories, while the predictive results are extremely sensitive to such disparities.

4.2. GBPO Algorithm. It is worth mentioning that our optimization system based on the BP neutral network is a mix-integer nonlinear programming (MINLP) problem for two reasons. First, the predictive system (i.e., BP neutral network) we got in prediction stage is a complex nonlinear function. Second, several integer variables are included in target explanatory parameters [28]. Common solution for this problem is the branch-and-bound approach proposed by Land and Doig [29]. However, it always takes an amount of time that increases exponentially with problem size. To alleviate this, we proposed to use the genetic algorithm, a typical evolution algorithm to solve our discrete and combinatorial optimization problem. Actually, the genetic algorithm is proved to be an effective approach for MINLP problems by many works [30–32], and it is extensible and easy to combine with other algorithms [33]. Therefore, we propose a parameter optimization algorithm (GBPO) based on the BP neutral network we constructed above and the genetic algorithm to obtain optimal results. The idea of GBPO algorithm can be listed as follows:

(a) For a specific borrower, the set of his/her characteristics (i.e., independent variables) can be denoted as $x = \{x_1, x_2, \ldots, x_{14}\}$ according to the order of independent variables in Figure 3. While the set of adjustable independent variables are represented by $I = \{i_1, i_2, \ldots, i_N\}$ according to the order in Table 10, N is the number of adjustable variables.

Table 10: Adjustable range of explanatory variables.

Type	Variables	Range
Adjustable ranges for all borrowers	RET	0–120*
	TAS	0–4*
	TLS	0, −2, −4*
	PRS	0, 1, 2*
	MMR	1–8
Adjustable ranges for individual borrowers	IDT	0, 1
	EDC	0, 1
	MPC	0, 1
	ORC	0, 1

Note: the range with *symbol represents that it should add to the initial values; otherwise, it should replace the initial values.

(b) Encode adjustable variables: we here adopt a binary code in optimization algorithm while most of our parameters are numeric or categorical. So, recoding is necessary for further analysis. We use the following formulas to get the required binary code length:

$$L_i = \left\lceil \log_2\left(R_i^b - R_i^a\right) \right\rceil, \quad i \in I,$$
$$L = \sum_{i \in I} L_i, \tag{14}$$

where L_i denotes the required length of each variable in the set I, L denotes the total length of each chromosome, R_i^b denotes the upper limit of the range of variable i, and R_i^a denotes the lower limit of the range of variable i.

(c) Initialize the population: randomly generate M chromosomes, which can be denoted as $S = \{S_1, S_2, \ldots, S_j\}$, $j = 1, 2, \ldots, M$. The notation S_{ji} represents a part of chromosome that corresponds to the variable i in jth chromosome.

(d) Decode the chromosomes: for each chromosome, we encode it using the following formula:

$$V_{ji} = \sum_{k=1}^{L_i} S_{jik} 2^{k-1} \frac{R_i^b - R_i^a}{2^{L_i} - 1}, \quad i \in I, j = 1, 2, \ldots, M, \tag{15}$$

where V_{ji} denotes the random number which we generated for jth chromosome of variable i. After that, we add a specific V_{ji} to the initial variable with symbol * in Table 10 and reset the other initial variables by corresponding V_{ji}. Then, we can get a new set of independent values corresponding to jth chromosome, which is denoted as $x'_j = \{x'_{j1}, x'_{j2}, \ldots, x'_{j14}\}$,

(e) Evaluating the fitness of each chromosome in the population: for each new set of independent values, we can calculate the corresponding target values which are introduced in Section 4.1.1 using our

predictive model, noted as ($\widehat{\text{LMT}}$, $\widehat{\text{RAT}}$). To evaluate them, our fitness function can be defined as follows:

$$F_1 = \widehat{\text{LMT}},$$
$$F_2 = \frac{1}{\widehat{\text{RAT}}}, \tag{16}$$

where F_1 denotes the fitness function for the parameter of LMT, and F_2 denotes the fitness for the parameter RAT.

(f) Perform the selection operation using the roulette method.
(g) Perform the crossover operation using the 1-point crossover method, where crossover probability is set as P_c.
(h) Perform the mutation operation using a simple mutation method, where mutation probability is set as P_m.
(i) Repeat steps (c) to (h) until the iteration times meet the given maximum iteration threshold G_{max}.

4.3. Single-Target Programming GBPO Algorithm. In order to observe the optimization effect, we randomly choose four groups of independent variables from the dataset TSD2, which are No. 107, No. 377, No. 422, and No. 455 individuals, respectively. In the first stage, we optimize the borrowing limit and interest rate separately; that is, we conduct a single-target programming analysis.

We use F_1 and F_2 defined before as our fitness function of LMT and RAT. To perform fair comparisons, we use the same parameter settings for all combinations as shown in Table 5. For the remaining parameters, M is set to 60 and G_{max} is set to 100 (discussed below). By processing the selected data using our model, we obtain the evaluation results as shown in Table 11.

From Table 11, we can find that both LMT and RAT have a better optimized result than the initial value. Taking the individual of No. 107 as an example, his/her borrowing limit has increased by 12793 yuan and his/her interest rate has decreased by 4 percent. These results show that our proposed optimization algorithm is effective.

To get more detail about the result, we represent the optimized input variables of No. 107 as shown in Table 12. Line 2 denotes the original value of each variable, and line 3 and line 4 denote the optimized value of each variable where aiming at LMT and RAT, respectively. From this table, it is easy to obtain improvement directions for No. 107 to optimize his/her LMT. First, No. 107 should conduct the certification of EDC and MPC. Second, No. 107 needs to pay the bills on time for four times and also have two clear chances to delay loans and to adjust the value of TAS and TLS to 4 and −4, respectively. Third, the *Magic-Mirror rating* of No. 107 should be adjusted to B. However, this credit rating is not altered for users but determined by P2P lending based on the user's behavior. So, No. 107 should commit to improve his/her credit rating by adjusting his/her most crucial adjustable behavior (see Appendix B). In this case, readers may arise another question that our proposed model is meaningless, since users could provide as much as personal information in order to optimize their borrowing parameters. Actually, there is a trade-off between privacy protection and better borrowing parameters. Our study is proposed to obtain better LMT and RAT by providing as little personal information as possible given existing stage.

4.4. Parameter Setting of M and G_{max}. Even though the increased value of M and G_{max} might get better optimization results, but at the same time, it spends more time to convergence. In order to make a trade-off between better optimization result and time-consuming question, we conduct a comparative analysis using different threshold of M and G_{max} as shown in Table 5. We also use the materials of No. 107 and the experiment result is shown in Figure 5.

As shown in Figure 5, when $M \geq 60$, there is barely any major fluctuation for optimal values of the LMT and RAT, which means that degrees of optimization are similar when $M \geq 60$. Therefore, we choose the parameters combination with minimum time consumption; that is, M is set to 60 and G_{max} is set to 100.

4.5. Double-Target Programming GBPO Algorithm. In this section, we conduct an optimization analysis for both LMT and RAT, that is, double-target programming problem. Different from the single-target programming method, the double-target programming method is dedicated to optimizing two borrowing parameters simultaneously if their directions of optimization are different.

To execute this procedure, we firstly examine the changing trend of one target parameter as another parameter is getting better. We also take No. 107 as an example here and the experiment results are shown in Figure 6.

As shown in Figure 6, as one target parameter is getting better, another parameter does not show a clear changing trend. So, it is hard for us to intuitively acquire the best results for both parameters. Besides, Figure 6(b) shows that when RAT is getting better and better, the value of LMT always stays at a low level. Thus, if we want to optimize one of the parameters solely, another parameter will be uncontrollable and undesirable. A comprehensive analysis about two target parameters is imperative. Besides, this figure also demonstrates that we may just achieve relatively satisfactory results but optimal results for both parameters.

To optimize two target parameters simultaneously, there are many methods that can be chosen. This paper introduces a simple and understandable method named the linear weighting method to solve our double-target programming problem [34]. The linear weighting method is implemented by giving different weights to both goals based on the different importance of them. And our goal is to obtain the maximum weighted sum of two target parameters. In our context, since the importance of the two target parameters varies from borrower to borrower, we will examine a series of weight combinations of two parameters in our analysis. And stakeholders are free to choose the weight combination according to their actual requirement.

Table 11: Optimized values using the single-target programming method.

Parameters	LMT				RAT			
Individual	No. 107	No. 377	No. 422	No. 455	No. 107	No. 377	No. 422	No. 455
Initial value	500	3500	2500	500	11	11	20	11
Optimized value	13293	10811	9831	7873	7	8	6	8

Table 12: Optimized values of No. 107 using the single-target programming method.

Variables	HPT	GED	IDT	AGE	RET	PRS	IDC	VDC	EDC	MPC	ORC	TAS	TLS	MMR
x_0	0	1	0	21	13	0	1	0	0	0	0	0	0	2
x^*_{LMT}	0	1	0	21	13	0	1	0	1	1	0	4	−4	4
x^*_{RAT}	0	1	0	21	14	2	1	0	1	1	0	4	0	1

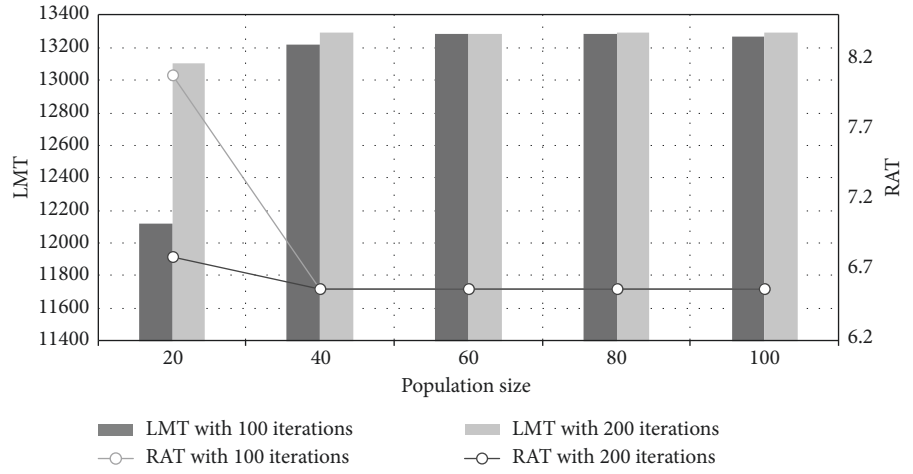

Figure 5: Optimization performance of various M and $G_{\max}$.

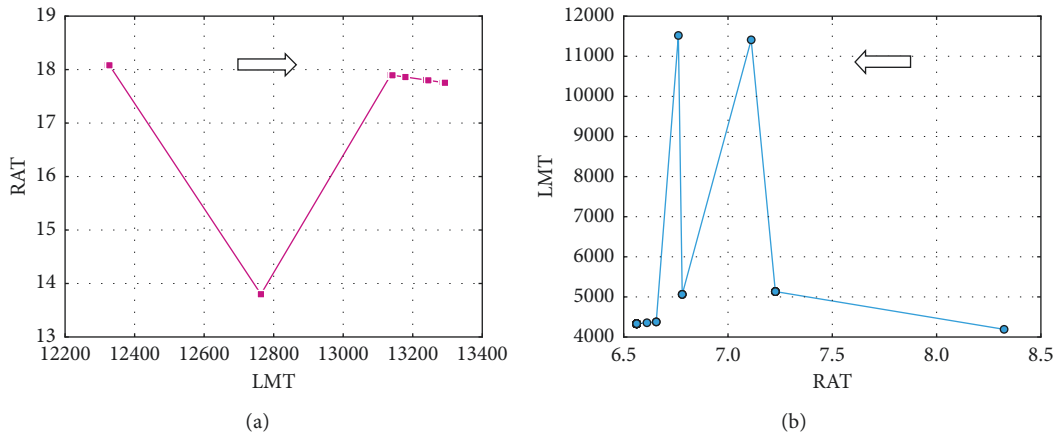

Figure 6: Changing trend of one target parameter as another parameter is getting better: (a) the optimization direction of LMT; (b) the optimization direction of RAT.

To match our proposed GBPO algorithm with the linear weighting method, we need to modify the fitness function and the details are shown below:

(a) Normalize the value of $\widehat{\mathrm{LMT}}$ and $\widehat{\mathrm{RAT}}$, which we obtain in GBPO Algorithm. $\widehat{\mathrm{LMT}}$ and $\widehat{\mathrm{RAT}}$ are predicted values generated by GA. The formulas are defined as follows:

$$\mathrm{stdLMT}_j = \frac{\widehat{\mathrm{LMT}}_j - \widehat{\mathrm{LMT}}_{\min}}{\widehat{\mathrm{LMT}}_{\max} - \widehat{\mathrm{LMT}}_{\min}},$$
$$\mathrm{stdRAT}_j = \frac{\widehat{\mathrm{RAT}}_j - \widehat{\mathrm{RAT}}_{\min}}{\widehat{\mathrm{RAT}}_{\max} - \widehat{\mathrm{RAT}}_{\min}}, \tag{17}$$

where stdLMT_j and stdRAT_j denote the jth standardized value of LMT and RAT, respectively. $\widehat{\text{LMT}}_{\max}$ and $\widehat{\text{RAT}}_{\max}$ denote the maximum value of LMT and RAT, respectively.$\widehat{\text{LMT}}_{\min}$ and $\widehat{\text{RAT}}_{\min}$ denote the minimum value of LMT and RAT, respectively.

(b) Modify the fitness function as follows:

$$F_3 = w_1 \text{stdLMT}_j + w_2 \text{stdRAT}_j, \tag{18}$$

where w_1 and w_2 denote the weights of stdLMT_j and stdRAT_j, which represent the importance of LMT and RAT, respectively.

Here, we examine the optimization results for different weight combinations of two parameters. We select nine different weight combinations in our case and the details are shown in Table 13.

We also take No. 107 as our analysis materials and use the proposed double-target programming GBIO algorithm to check the effects of different weight combinations. The experiment results are shown in Figure 7.

From Figure 7, we can find that with the declining weight of RAT and the rising weight of LMT, the optimal value of LMT and RAT are both showing a rising trend, which is consistent with our expectations. In the meantime, all the optimal values of LMT and RAT are meeting the satisfactory level, which shows that our double-target programming GBPO algorithm is effective. Moreover, it is noted that borrowers can select different weights for two parameters to obtain their desired optimization results based on their actual requirements.

5. Conclusion

In this paper, we firstly build a three-layer BP neural network to predict the borrowing parameters LMT (borrowing limit) and RAT (interest rate). Based on the constructed BP neural network, we develop a BPIE method to obtain the confidence intervals of the LMT and RAT, which represent the prediction ranges of borrowing limit and interest rate for borrowers. Using the real-word data from http://ppdai.com, we conclude that the predictive accuracy of the proposed BFIE method are 94.80% and 97.84% for borrowing limit and interest rate, respectively. After that, we propose to optimize the borrowing parameters, LMT and RAT. By considering the number of target parameters, we transform our problem into a single-target programming problem and a double-target programming problem. To solve the problems, we introduce a new GBPO algorithm based on the BP neural network predictive model and genetic algorithm. Using the randomly selected data from valid dataset, the experiment result shows that our proposed algorithm is effective in optimizing the target parameters.

Different from priori studies, this paper provides a new perspective from borrowers to predict and optimize the borrowing limit and interest rate given the limited information. The proposed method and findings of our experiment study have practical implications for researchers and borrowers in the P2P system.

Table 13: Weight combinations of two parameters.

Combinations	1	2	3	4	5	6	7	8	9
Weight of LMT	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Weight of RAT	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1

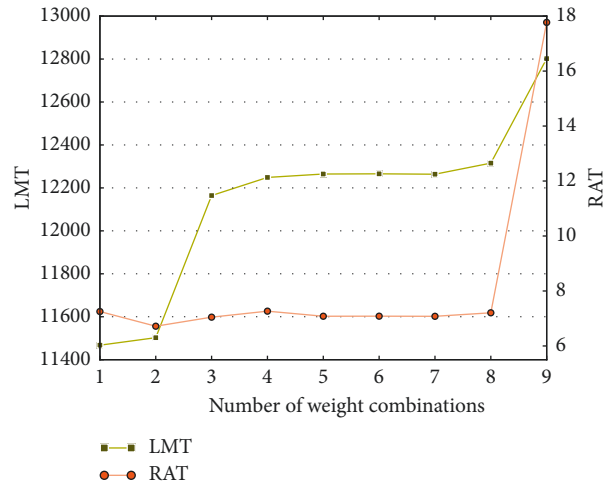

Figure 7: Performance of various weight combinations on borrowing parameters.

Table 14: Likelihood ratio test.

Parameters	−2LL	P value
HPT	22992.94	0.000
GED	23412.63	0.000
AGE	23050.42	0.000
RET	23024.71	0.000
PRS	26941.21	0.000
IDC	22965.82	0.000
VDC	22983.97	0.000
EDC	23639.35	0.000
MPC	22992.91	0.934
ORC	24595.18	0.004
TAS	37600.36	0.000
TLS	23093.18	0.000

Appendix

A. Building Linear Regression Model

We first build linear regression models for individual value prediction, which can be written as follows:

$$\begin{aligned} \text{LMT} &= C_{\text{LMT}} + \alpha_{\text{LMT}}\text{IDI} + \beta_{\text{LMT}}\text{CFI} + \gamma_{\text{LMT}}\text{HTI} \\ &\quad + \omega_{\text{LMT}}\text{RRI} + \varepsilon_{\text{LMT}}, \\ \text{RAT} &= C_{\text{RAT}} + \alpha_{\text{RAT}}\text{IDI} + \beta_{\text{RAT}}\text{CFI} + \gamma_{\text{RAT}}\text{HTI} \\ &\quad + \omega_{\text{RAT}}\text{RRI} + \varepsilon_{\text{RAT}}, \end{aligned} \tag{A.1}$$

where C_x denotes the intercept vector; α_x, β_x, γ_x, and ω_x are coefficient vectors; and ε_x denotes the residual.

Second, corresponding confidence intervals are derived from constructed linear regression models. The process is as follows.

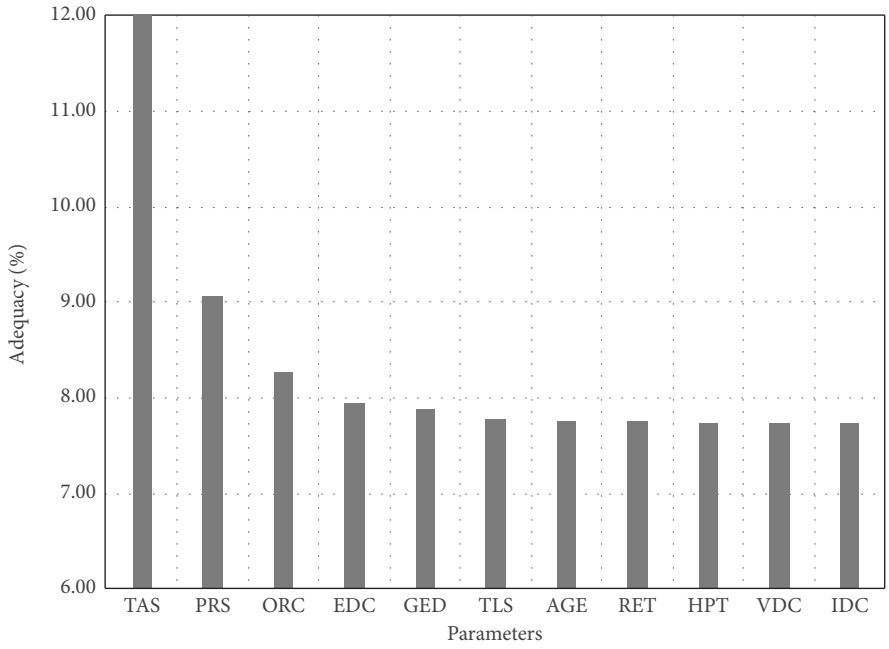

Figure 8: The adequacy value of each parameter.

(1) Taking LMT as an example, LMT_i obeys the following distribution according to its statistics [25]:

$$\frac{\widehat{\mathrm{LMT}}_i - \mathrm{LMT}_i}{\widehat{\sigma}_{e_0}^{(\mathrm{LMT})}} \sim t(n-k-1), \tag{A.2}$$

where LMT_i denotes the confirmatory value of the ith individual, $\widehat{\mathrm{LMT}}_i$ denotes the predicted value of the ith individual, n denotes the sample size, and k denotes the number of independent variables. $\widehat{\sigma}_{e_0}^{(\mathrm{LMT})}$ represents the estimation of standard deviation of prediction error, and its calculation formula is as follows:

$$\widehat{\sigma}_{e_0}^{(\mathrm{LMT})} = \widehat{\sigma}^{(\mathrm{LMT})}\sqrt{1 + X_i\left(X'X\right)^{-1}X_i'}, \tag{A.3}$$

where $\widehat{\sigma}^{(\mathrm{LMT})}$ represents the estimation of standard error, X_i represents the value vector of ith individual's dependent variables, and X represents the value matrix of the dependent variables.

(2) Then, it is easy to get the one-sided confidence interval for LMT_i, that is:

$$\left(-\infty,\ \widehat{\mathrm{LMT}}_i + t_\alpha \widehat{\sigma}_{e_0}^{(\mathrm{LMT})}\right), \tag{A.4}$$

where α denotes the significance level and t_α denotes the α quantile of t-distribution.

(3) Similarly, two-sided confidence interval for RAT_i is as follows:

$$\left(\widehat{\mathrm{RAT}}_i - t_{\alpha/2}\widehat{\sigma}_{e_0}^{(\mathrm{RAT})},\ \widehat{\mathrm{RAT}}_i + t_{\alpha/2}\widehat{\sigma}_{e_0}^{(\mathrm{RAT})}\right), \tag{A.5}$$

where α denotes the significance level and $t_{\alpha/2}$ denotes the $\alpha/2$ quantile of t-distribution.

B. Ranking the Importance of Parameters on MMR

In order to simply compare the importance degree of each parameter on MMR, we established the following multinomial logistic regression:

$$\mathrm{logit}\left(p_j\right) = \ln\left(\frac{p_j}{1-p_j}\right) = \alpha_j + \beta_1\mathrm{IDI} + \beta_2\mathrm{CFI} + \beta_3\mathrm{HTI}, \tag{B.1}$$

where $p_j = p(\mathrm{MMR} \le j|X)$ indicates the conditional probability of $\mathrm{MMR} \le j$ given X; β_1, β_2, and β_3 denote the corresponding coefficient vector; and α_j denotes the intercept of the jth model.

By using SPSS Statistics on our samples, we can get the following result as shown in Table 14.

In Table 14, −2LL (i.e., −2*log-likelihood) is an index which is commonly used to measure the fitting degree of models; the smaller the value, the better the fit [35]. P value reflects the significance of each parameter. From Table 14, we can find that almost all of P values are less than 0.05 except for MPC, which means that all parameters except MPC have significant impact on MMR.

In order to rank the importance of each parameter on MMR, we calculate the statistical adequacy, which is related to the explanatory value of each predictor relative to the entire set [36]. The adequacy is the proportion of the total explained variation in the outcome that is explained by the individual predictor [37]. Then the values of adequacy can be calculated for each parameter which is significant. The results are shown in Figure 8.

From Figure 8, we can obtain the ranking of importance degree of each parameter on MMR. And it is easily found that TAS, PRS, and ORC have the greatest influence on MMR.

Data Availability

Data, as supplemental material, are available at https://pan.baidu.com/s/1BGrENUwFILVkMqDsyGroVQ.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the National Science Foundation of China (Project no. 71571073), Guangdong Natural Science Foundation (Project no. 2014A030313243), and the Twelfth Five-year Planning Project for Guangdong Philosophy and Social Science (GD14CGL09).

References

[1] D. Paravisini, V. Rappoport, and E. Ravina, "Risk aversion and wealth: evidence from person-to-person lending portfolios," *Management Science*, vol. 63, no. 2, pp. 279–297, 2017.

[2] Y. Li, Y. J. Gao, Z. N. Li et al., "The influence of borrower's description on investors' decision-analysis based on P2P online lending," *Economic Research Journal*, no. S1, pp. 143–155, 2014.

[3] M. Lin, R. Prabhala, and S. Viswanathan, "Judging borrowers by the company they keep: social networks and adverse selection in online peer-to-peer lending," *Management Science*, vol. 59, no. 1, pp. 17–35, 2013.

[4] M. Herzenstein, S. Sonenshein, and U. M. Dholakia, "Tell me a good story and i may lend you money: the role of narratives in peer-to-peer lending decisions," *Journal of Marketing Research*, vol. 48, pp. S138–S149, 2011.

[5] C. Serrano-Cinca and B. Gutiérrez-Nieto, "The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending," *Decision Support Systems*, vol. 89, pp. 113–122, 2016.

[6] L. C. Thomas, "A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers," *International Journal of Forecasting*, vol. 16, no. 2, pp. 149–172, 2000.

[7] C. Serrano-Cinca, B. Gutiérrez-Nieto, and L. López-Palacios, "Determinants of default in P2P lending," *PLoS One*, vol. 10, no. 10, Article ID e0139427, 2015.

[8] U. Rajan, A. Seru, and V. Vig, "The failure of models that predict failure: distance, incentives, and defaults," *Journal of Financial Economics*, vol. 115, no. 2, pp. 237–260, 2015.

[9] Y. Miao, G. Chen, W. Wang, and X. Gong, "Application of gray-tone difference matrix-based features of pavement macrotexture in skid resistance evaluation," *Journal of Southeast University*, vol. 31, no. 3, pp. 389–395, 2015.

[10] F. Xiao and S. N. Amirkhanian, "Artificial neural network approach to estimating stiffness behavior of rubberized asphalt concrete containing reclaimed asphalt pavement," *Journal of Transportation Engineering*, vol. 135, no. 8, pp. 580–589, 2009.

[11] Y. H. Duan and J. H. Xie, "Price estimation of real estate market-based on BP neural network interval estimation method," *Statistics and Decision*, no. 5, pp. 18-19, 2004.

[12] M. Gevrey, I. Dimopoulos, and S. Lek, "Review and comparison of methods to study the contribution of variables in artificial neural network models," *Ecological Modelling*, vol. 160, no. 3, pp. 249–264, 2003.

[13] X. Li, L. Shu, L. Liu, D. Yin, and J. Wen, "Sensitivity analysis of groundwater level in Jinci Spring Basin (China) based on artificial neural network modeling," *Hydrogeology Journal*, vol. 20, no. 4, pp. 727–738, 2012.

[14] D. J. Montana and L. Davis, "Training feedforward neural networks using genetic algorithms," in *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 762–767, Detroit, MI, USA, August 1989.

[15] D. Y. Chen, "Research on credit transaction trust of P2P network based on social cognition theory," *Nankai Business Review*, vol. 17, no. 3, pp. 40–48, 2014.

[16] D. Y. Chen, H. Zhu, and H. C. Zheng, "Risk, trust and willingness to lend–an empirical study based on the registered users on the ppdai platform," *Management Review*, vol. 26, no. 1, pp. 150–158, 2014.

[17] E. Ozgan, "Artificial neural network based modelling of the Marshall stability of asphalt concrete," *Expert Systems with Applications*, vol. 38, no. 5, pp. 6025–6030, 2011.

[18] D. Zheng, Z. D. Qian, Y. Liu, and C. Liu, "Prediction and sensitivity analysis of long-term skid resistance of epoxy asphalt mixture based on GA-BP neural network," *Construction and Building Materials*, vol. 158, pp. 614–623, 2018.

[19] F. Yu and X. Xu, "A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network," *Applied Energy*, vol. 134, pp. 102–113, 2014.

[20] J. N. Fan, Z. L. Wang, and F. Qian, "Research progress of hidden layer structure design for BP artificial neural network," *Control Engineering*, vol. 12, no. S0, pp. 109–113, 2005.

[21] Y. S. Ding, *Computational Intelligence: Theory, Technology and Application*, Science Press, Beijing, China, 2004.

[22] L. M. Zhang, *Artificial Neural Network Model and Its Application*, Fudan University Press, Shanghai, China, 1993.

[23] P. F. Yan and T. S. Zhang, *Artificial Neural Network and Simulated Evolutionary Computation*, Tsinghua University Press, Beijing, China, 2nd edition, 2005.

[24] D. Groebner, *Business Statistics: A Decision Making Approach*, China Machine Press, Beijing, China, 2007.

[25] X. Y. Li and C. Yuan, *Application of Economic Statistics*, Peking University Press, Beijing, China, 2015.

[26] S. Raschka, *Python Machine Learning*, Packt Publishing, Birmingham, UK, 2015.

[27] Z. Sheng, *Probability Theory and Mathematical Statistics*, Higher Education Press, Beijing, China, 3rd edition, 2001.

[28] F. Trespalacios and I. E. Grossmann, "Review of mixed-integer nonlinear and generalized disjunctive programming methods," *Chemie Ingenieur Technik*, vol. 86, no. 7, pp. 991–1012, 2014.

[29] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, no. 3, pp. 497–520, 2016.

[30] M. Sakawa, K. Kato, and S. Ushiro, "Operational planning of district heating and cooling plants through genetic algorithms for mixed 0–1 linear programming," *European Journal of Operational Research*, vol. 137, no. 3, pp. 677–687, 2002.

[31] M. Sakawa, K. Kato, S. Ushiro, and M. Inaoka, "Operational planning of district heating and cooling plants using genetic

algorithms for mixed integer programming," *Applied Soft Computing*, vol. 1, no. 2, pp. 139–150, 2001.

[32] M. Sakawa, K. Kato, and S. Ushiro, "Operation planning of district heating and cooling plants through genetic algorithms for nonlinear 0–1 programming," *Computers and Mathematics with Applications*, vol. 42, no. 10-11, pp. 1365–1378, 2001.

[33] A. S. Chuang and F. Wu, "An extensible genetic algorithm framework for problem solving in a common environment," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 269–275, 2002.

[34] F. F. Leimkuhler, *Introduction to Operations Research*, McGraw-Hill, New York, NY, USA, 2005.

[35] F. E. Harrell Jr., *Regression Modeling Strategies*, Springer, New York, NY, USA, 2001.

[36] D. Thompson, *Ranking Predictors in Logistic Regression*, SAS Institute, Article ID D10–2009, Cary, NC, USA, 2009.

[37] D. J. Mclernon, E. W. Steyerberg, E. R. T. Velde et al., "Predicting the chances of a live birth after one or more complete cycles of in vitro fertilisation: population based study of linked cycle data from 113 873 women," *BMJ*, vol. 355, article i5735, 2016.