

Research Article

A Novel Key Influencing Factors Selection Approach of P2P Lending Investment Risk

Pingfan Xia ^{1,2}, Zhiwei Ni ^{1,2}, Xuhui Zhu ^{1,2} and Liping Ni ^{1,2}

¹School of Management, Hefei University of Technology, Hefei 230009, China

²Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei 230009, China

Correspondence should be addressed to Zhiwei Ni; zhiwein@163.com

Received 3 September 2019; Accepted 15 October 2019; Published 28 November 2019

Academic Editor: Georgios Dounias

Copyright © 2019 Pingfan Xia et al. 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.

Recent frequent “thunderstorm incidents” of P2P lending industry have caused the panic of industry investors. To predict the investment risk of P2P lending, we should scientifically and rationally analyze the key influencing factors of P2P lending investment risk. Existing key influencing factors selection methods mainly involve traditional statistical approaches and artificial intelligence methods. The traditional statistical approaches cannot deal with the high-dimensional nonlinear problems, and it cannot find the exact key influencing factors of the P2P lending investment risk. The artificial intelligence methods cannot recognize and learn the application background, and the selected attributes without active thinking and personal perception may not be the key influencing factors of P2P lending investment risk. To address the above issues, a novel key influencing factors selection approach of P2P lending investment risk is proposed by combining the proposed fireworks coevolution binary glowworm swarm optimization (FCBGSO), multifractal dimension (MFD), probit regression, and artificial prior knowledge. First, multifractal dimension combined with the proposed FCBGSO is used to select the preliminary influencing factors of the investment risk; second, the nonsignificant relevant attributes in the preliminary influencing factors are removed using the probit regression, and we add the influencing factors extracted from the original dataset of P2P lending using the artificial prior knowledge into the retaining influencing factors after removing one by one. A small and reasonable number of influencing factor subsets are achieved. Finally, we evaluate each influencing factors subset using extreme learning machine (ELM), and the subset with the best classification accuracy is efficiently achieved, i.e., it is the key influencing factors of P2P lending investment risk. Experimental results on the real P2P lending dataset from the Renrendai platform demonstrate that the proposed approach performs better than other state-of-the-art methods and that it has validity and effectiveness. It provides a new research idea for the key influencing factors selection of P2P lending investment risk.

1. Introduction

Peer-to-peer (P2P) lending is a new financial model that integrates Internet platforms and private lending. Both lenders and borrowers can directly complete the transactions through P2P lending platforms without going through financial intermediaries [1, 2]. P2P lending is one of the most important modes of Internet finance. On one hand, it can serve the real economy; on the other hand, the recent frequent occurrences of P2P lending “thunderstorm incidents” have damaged the earnings of investors and hindered the healthy development of P2P lending industry. According to preliminary statistics, as of January 15, 2019,

there are 2,746 transferred or closed platforms in China’s P2P lending industry and 2,663 problematic platforms in total. Since June 2018, the risk incidents of P2P lending platforms have been continuously exposed. The large-scale “thunderstorm incidents” in the P2P lending industry have caused a strong impact on the healthy development of this industry. Meanwhile, it has attracted great attention of the Chinese government. In the 2018 report on the work of the Chinese government, it was definitely pointed out that “strengthen the overall coordination of financial supervision and improve the supervision of Internet finance.” Hence, Internet finance has been written into the report on the work of the government five times in a row. From the initial

“promoting development—standardizing development—being vigilant of risk” to the “improving Internet financial supervision” in 2018 indicates that the standardization control of Internet financial investment risk is imperative. Therefore, it is urgent to explore the investment risk of P2P lending. To research the investment risk, we should analyze the key influencing factors of P2P lending investment risk, which can provide high-quality data for the prediction of P2P lending investment risk.

A key influencing factors selection approach for P2P lending investment risk is essential to reduce irrelevant attributes with the investment risk in an original dataset of P2P lending and retain key influencing factors. In fact, there is a large amount of noisy or irrelevant features with investment risk in the real datasets of P2P lending. Existing key influencing factors selection approaches usually use traditional statistical methods and attribute selection algorithms based on artificial intelligence [3, 4]. The statistical methods are commonly used in the selection of key influencing factors of P2P lending investment risk, while there is no application of artificial intelligence methods, at least to our knowledge. The traditional statistical methods are only limited to the discussion of the impact of a single factor on the borrower for the order default risk, but it ignores the fusion and the crossover of multiple information. For instance, the credit rating of each loan has an important impact on investment risk, which is presented by Guo et al. [5]. Larrimore et al. examined the relationship between language use and investor decision-making [6]. The soft information in loan titles has a significant influence on whether the loan is successful. The results also suggest that investors do not invest blindly based on returns [7]. Xiao et al. proposed a visual analysis method which analyzes and detects risk in P2P lending deals [8]. Perceived age of P2P lending orders shows a strong signal of ability and experience, and more mature perceived age is more attractive to investors [9]. Chen et al. investigated the amount of punctuation used in loan descriptions can influence the investment default risk using data from Renrendai (one of the largest P2P lending platforms in China) [10]. To sum up, the traditional statistical methods such as regression analysis have small calculations and simple operations when they analyze the influencing factors of P2P lending investment risk.

For feature selection based on artificial intelligence, there are two main points: one is the evaluation criterion selection and the other one is the search strategy. With respect to evaluation criterion, various evaluation methods are used to evaluate feature subsets. Different evaluation methods have great relationship with the optimal subset. For example, information theory [11–13], distance analysis [4], rough sets [14–17], and fractal dimension [18–20]. Fractal dimension is treated as an evaluation criterion, which attracts many scholars’ attentions. It has two advantages [21]: on one hand, the number of an optimal feature subset can be determined by calculating its fractal dimension, which can dramatically reduce computational amount; on the other hand, the fractal dimension performs well when it comes to solving high-dimensional datasets and nonlinear problems. Most existing feature selection approaches based on fractal dimension use only a single fractal dimension, which may not precisely describe the original datasets [20] because of their complicated distribution. In contrast,

multifractal dimension (MFD) can describe the distribution of dataset in different aspects [19], which is regarded as the evaluation criterion of feature subsets in this work. In regard to searching strategy, finding an optimal feature subset of an original dataset is a combinatorial optimization problem [17]. Therefore, heuristic algorithms provide good searching strategies for the feature selection methods, for example, genetic algorithm (GA) [22, 23], ant colony optimization (ACO) [24–26], particle swarm optimization (PSO) [27, 28], and artificial fish swarm algorithm (AFSA) [29]. However, the complex coding process of GA is hard to be implemented [30, 31]; ACO has the disadvantages of the blindness search in the early stage, slow convergence speed, and huge computing resource consumption; PSO easily traps into local optimal solution [30, 31]; AFSA has the weaknesses of lack of population diversity and slow convergence rate in the later stage. In contrast, glowworm swarm optimization (GSO) has the advantages of simplicity of implementation, strong robustness, and good and fast global convergence [32], which can be used as a searching strategy for solving a feature selection problem [33]. We attempt to propose a fireworks coevolution binary glowworm swarm optimization (FCBGSO) as the searching strategy in this work.

Based on the above analysis, first and foremost, the traditional statistical methods cannot solve the high-dimensional nonlinear problem, and the analysis is one-sided, so it is difficult to exactly analyze the key influencing factors of the P2P lending investment risk. In addition, the artificial intelligence methods perform well when it comes to coping with high-dimensional and nonlinear datasets, but it cannot recognize and learn the application background, lack of active thinking and personal perception, and the selected attributes may not be the key influencing factors of P2P lending investment risk. Therefore, we proposed a novel approach to find the key influencing factors of P2P lending investment risk, which combines MFD, FCBGSO, the probit regression, and the artificial prior knowledge. The mission is attained in four steps: in the first step, we take the proposed FCBGSO as a search strategy and treat MFD as an evaluation criterion for feature subsets. Then, the preliminary attribute subset extracted from the original dataset of P2P lending is attained using the combination of FCBGSO and MFD. In the second step, the nonsignificant relevant attributes with the default risk are removed from the preliminary subset using the probit regression. In the third step, a small and reasonable number of attribute subsets are achieved by combining the retaining attributes after removing and the attributes obtained by the artificial prior knowledge. In the final step, considering the advantages of extreme learning machine (ELM) such as good generalization ability and the extremely fast learning speed, ELM is used to assess the classification accuracies of these subsets, and the attribute subset with the best accuracy is the key influencing factors of P2P lending investment risk.

The contributions of the proposed approach are presented as follows:

- (1) A novel approach for key influencing factors selection of P2P lending investment risk is proposed using the combination of FCBGSO, MFD, the probit regression, and the artificial prior knowledge

- (2) The proposed FCBGSO works well with respect to searching for the optimal solution in a binary space
- (3) Experiments on the real dataset of P2P lending from Renrendai platform demonstrate that the proposed method significantly performs better than traditional statistical approaches and artificial intelligence methods and that it has validity and effectiveness
- (4) It provides a novel research idea for the key influencing factors selection of P2P lending investment risk

The rest of this paper is organized as follows. In the next section, we briefly review the basic concept of a GSO, and then FCBGSO is proposed. The key influencing factors selection method of P2P lending investment risk and how to use it are presented in Section 3. Experimental results are shown in Section 4. In Section 5, the conclusions and the future work are presented.

2. Fireworks Coevolution Binary Glowworm Swarm Optimization (FCBGSO)

Swarm intelligence algorithms combined with MFD can be applied in attribute selection. Swarm intelligence algorithms are used as searching strategies. GSO has some advantages such as simplicity of implementation, strong robustness, and good global convergence. So, it can be used as a searching strategy, but there are still drawbacks, e.g., insufficient diversity, low convergence precision, and searching efficiency. To address the above drawbacks, FCBGSO is proposed, which significantly improves its convergence speed and precision. The preliminary influencing factors can be efficiently achieved. The outline of FCBGSO is presented as follows.

2.1. Glowworm Swarm Optimization (GSO). GSO is a relatively novel swarm intelligence algorithm proposed by Krishnanand and Ghose [34–36], which is a bionic swarm intelligent algorithm by imitating the luminous behavior in the process of foraging and courtship of glowworms in nature [37]. In GSO, each glowworm represents a solution, and it is randomly distributed in a solution space. The higher brightness the glowworm individual has, the more attraction it gains [38]. The glowworms move forward to their neighbors with higher luciferin, and these individuals can be updated. Thus, the global optimal solution is attained. The basic steps of GSO are listed as follows:

- (1) Updating luciferin of the glowworm $X_i(t)$ at the t th iteration is given by equation (1). The luciferin renewal depends on the objective function value $J(X_i(t))$ of the glowworm:

$$l_i(t) = (1 - \rho)l_i(t - 1) + \gamma J(X_i(t)), \quad (1)$$

where $l_i(t)$ is the luciferin level of $X_i(t)$ at the t th iteration, ρ represents the luciferin decay constant

($0 < \rho < 1$), and γ indicates the luciferin enhancement constant.

- (2) The glowworms in the dynamic decision domain of $X_i(t)$ whose luciferin is greater than $X_i(t)$ can be used to make up its set of neighbors $N_i(t)$, and it is expressed as equation (2). The probability $P_{ij}(t)$ of $X_i(t)$ moving to neighbor $X_j(t)$ in a set of neighbors is described as equation (3):

$$N_i(t) = \left\{ j : \|X_j(t) - X_i(t)\| < r_d^i(t); l_j(t) > l_i(t) \right\}, \quad (2)$$

$$P_{ij}(t) = \frac{l_j(t) - l_i(t)}{\sum_{k \in N_i(t)} l_k(t) - l_i(t)}, \quad (3)$$

where $r_d^i(t)$ is the dynamic radial range $0 < r_d^i < r_s$ and r_s is the radial range of the luciferin sensor.

- (3) Each glowworm selects a objective glowworm $X_j(t)$ with a higher luciferin at a probability $P_{ij}(t)$. Then, the position of $X_i(t)$ can be updated as the following equation:

$$X_i(t + 1) = X_i(t) + s \times \left(\frac{X_j(t) - X_i(t)}{\|X_j(t) - X_i(t)\|} \right), \quad (4)$$

where s is a moving step, set by the user.

- (4) After updating the positions of all the glowworms, the dynamic radial range of local-decision domain is noticed using the rule given as the following equation:

$$r_d^i(t + 1) = \min\{r_s, \max\{0, r_d^i(t) + \beta(n_t - |N_i(t)|)\}\}, \quad (5)$$

where β is a constant parameter and n_t is a parameter to control the number of neighbors.

2.2. Position Updating Modification Based on Dynamic Inertia Weight. Dynamic inertia weight strategies are categorized into four classes: linear decreasing inertia weight, nonlinear decreasing inertia weight, adaptive inertia weight, and stochastic inertia weight [39–41]. Consider that the stochastic inertia weight (SIW) in the position updating equation can balance the relationship between the local and the global search. It can obtain stable optimization results and quickly jump out of the local optima. Therefore, we use SIW to solve the drawback of slow convergence speed of basic GSO. The SIW is defined as follows:

$$w = r_{\min} + (r_{\max} - r_{\min}) \times \text{normrnd}() + \sigma \times \text{randn}(), \quad (6)$$

where r_{\min} denotes the lower limit value of SIW, r_{\max} indicates the upper limit value of SIW, $\text{randn}()$ shows a

random number which follows the normal distribution, $\text{normrnd}()$ expresses a random number of uniform distribution, and σ represents the deviation between inertia weights and their mean value.

The SIW is mainly used to update the positions of glowworms, and it is updated as follows:

$$X_i(t+1) = w \times X_i(t) + s \times \left(\frac{X_j(t) - X_i(t)}{\|X_j(t) - X_i(t)\|} \right). \quad (7)$$

To solve a binary combinational optimization problem, the positions of glowworms are mapped into 0 or 1 using a sigmoid function. The mapping process is presented as equations (8) and (9):

$$x_{ik} = \begin{cases} 1, & \text{if rand} \leq S(x_{ik}), \\ 0, & \text{else,} \end{cases} \quad (8)$$

$$S(x_i) = \frac{1}{1 + \exp(-x_i)}, \quad (9)$$

where $X_i(t) = (x_{i1}, x_{i2}, \dots, x_{ik}, \dots, x_{is})$, ($1 \leq k \leq s$), s is the dimension of the solution space of the problem, and $S(x_i)$ is a sigmoid function.

2.3. Coevolution Mechanism. To overcome the weakness of slow convergence speed in GSO, a coevolution mechanism is introduced into GSO, which can promote the process of evolution. To avoid invalid crossover caused by the excessive similarities between glowworms, the initial population is divided into three equal subpopulations by the proportion 1 : 1 : 1 according to their fitness values. They are elite subpopulation P_E , excellent subpopulation P_A , and common subpopulation P_B , respectively. Each subpopulation evolves independently and synchronously and keeps dynamic updating during the search process. The most excellent glowworm individual is selected from the elite subpopulation, and it performs a crossover with the optimal individual of P_A and P_B , respectively. Then, four new offspring are generated, which keeps the diversity of the population.

We introduce a competitive factor μ_1 into this work. The coevolution mechanism can be denoted as follows:

If $\text{rand} < \mu_1$, then

$$\begin{cases} X'_{EA}(t) = \text{round}\left(\frac{1}{2}((1+r) \times X_A(t) + (1-r) \times X_E(t))\right), \\ X''_{EA}(t) = \text{round}\left(\frac{1}{2}((1-r) \times X_A(t) + (1+r) \times X_E(t))\right), \\ X'_{EB}(t) = \text{round}\left(\frac{1}{2}((1+r) \times X_B(t) + (1-r) \times X_E(t))\right), \\ X''_{EB}(t) = \text{round}\left(\frac{1}{2}((1-r) \times X_B(t) + (1+r) \times X_E(t))\right), \end{cases} \quad (10)$$

where rand, r are randomly generated variables bounded between 0 and 1, $X_A(t)$, $X_B(t)$, and X_E are different glowworms in P_A , P_B , and P_E , respectively. $X'_{EA}(t)$, $X''_{EA}(t)$, $X'_{EB}(t)$, and $X''_{EB}(t)$ are the four new offspring. X_E will be replaced by the best glowworm selected from $X'_{EA}(t)$, $X''_{EA}(t)$, $X'_{EB}(t)$, and $X''_{EB}(t)$ if the best individual performs better than X_E . The architecture of the coevolution mechanism is presented in Figure 1.

2.4. Fireworks Evolution Strategy. To effectively avoid the defects of the premature convergence and the insufficient diversity of population in GSO, a fireworks explosion operation [42] is introduced. The current glowworm X_i produces multiple offspring by explosion with a certain probability. The best individual extracted from the multiple offspring can be retained to the next generation. We introduce a probability factor μ_2 , and the scale of the individual glowworms produced around X_i is formulized as follows:

If $\text{rand} < \mu_2$, then

$$S_i = H \times \frac{y_{\max} - f(x_i) + \varepsilon}{\sum_{i=1}^N (y_{\max} - f(x_i)) + \varepsilon}, \quad (11)$$

where S_i is the number of newly generated glowworms, y_{\max} shows the maximal fitness value of glowworms at the current iteration, H denotes a constant to adjust the amount of glowworm offspring, and ε is a small constant which can avoid zero division error.

The r th dimension in X_i is randomly selected to perform Gaussian mutation operation, namely, it is changed from 0 to 1 or 1 to 0:

$$r = \left\lfloor \frac{r_d^i \times e}{2} \right\rfloor, \quad (12)$$

where $e \sim N(1, 1)$ and $N(1, 1)$ indicates the Gaussian distribution with a mean value of 1 and a variance value of 1.

The glowworm offspring are produced by the fireworks evolution strategy, and their fitness values can be achieved. If the optimal glowworm in the generated offspring performs better than X_i , then X_i is replaced by it.

3. Key Influencing Factors Selection Method

3.1. Multifractal Dimension (MFD). Mandelbrot first proposed the concept of fractal in 1983 [43], which is used to describe the irregular geometry of the nature. There are two properties with respect to the fractal object: one is the self-similarity and the other one is the scale invariability, namely, there is a similar appearance when the fractal object is viewed in indifferent scales. Fractal theory is used in a wide variety of fields.

There are often two kinds of dimensions on datasets, i.e., the embedding dimension and the intrinsic dimension. The embedding dimension indicates the number of the original dataset's features; the intrinsic dimension represents the number of irrelevant features. Generally speaking, the intrinsic dimension is less than the embedding dimension. If all features are irrelevant with each other, the intrinsic

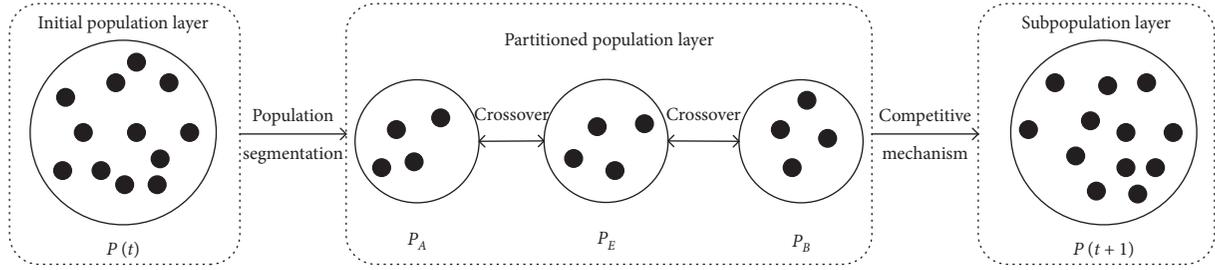


FIGURE 1: The architecture of the coevolution mechanism.

dimension is equal to the embedding dimension. The fractal dimension can represent the intrinsic dimension, and the upper bound of the fractal dimension is the number of key features required to characterize the original dataset. Traina et al. [44] showed that most of the datasets have fractal characteristic, and the fractal dimension can be regarded as an evaluation criterion for feature selection.

Fractal feature selection approaches were first proposed by Traina et al. [44]. The fractal dimension is taken as an evaluation criterion, which can measure the importance of features. The advantage of fractal feature selection algorithms is that the number of the selected features can be determined, but the fractal dimension needs to be recalculated after removing some features. To improve computational efficiency, GA [22, 23], ACO [24–26], PSO [27, 28], AFSA [29], and so on are employed as searching strategies to enhance efficiency of the fractal feature selection methods.

However, most existing fractal feature selection methods only take a single fractal dimension such as information dimension or correlation dimension. A single fractal dimension may not precisely describe a dataset [45]. In contrast, MFD can describe the dataset's distribution in different aspects, which can be calculated as the following equation:

$$D_q = \begin{cases} \lim_{r \rightarrow 0} \frac{1}{q-1} \times \frac{\log \sum p_i^q}{\log r}, & q \neq 1, \\ \lim_{r \rightarrow 0} \frac{\log p_i \sum p_i}{\log r}, & q = 1, \end{cases} \quad (13)$$

$r \in [r_1, r_2],$

where p_i stands for the probability of a data point dropped into the i th grid, r indicates the grid size, $[r_1, r_2]$ denotes the scale-free interval of a dataset, and q is an integer.

When $q < 0$, D_q shows the void distribution of a fractal dataset; when $q > 0$, D_q indicates the aggregation degree of a fractal dataset. Fractal dimension (FD) can just describe the distribution of a dataset in a single aspect. In contrast, the MFD can describe the distribution in many aspects. Hence, MFD is regarded as an evaluation criterion of feature subsets in this work.

3.2. Construct the Objective Function. By comparison with a single fractal dimension, MFD can accurately describe

datasets. So, the objective function can be expressed as the following equation:

$$f = \sqrt{\sum_q (\text{frac}_q - D_q)^2}, \quad (14)$$

where frac_q represents the q th-order fractal dimension of a feature subset, and D_q illustrates the q th-order fractal dimension of the original dataset.

We regard the difference between the MFD of a feature subset and the original dataset as the objective function. According to the definition of the objective function, we can see that the smaller the value of the objective function is, the better the solution is. D_q is specified with five fractal dimensions $D_2, D_3, D_4, D_5,$ and $D_6,$ respectively [19].

3.3. Extreme Learning Machine (ELM). ELM was first proposed by Huang et al. [45], which was developed for single hidden layer feedforward networks (SLFNs). By comparing with traditional neural networks, it requires great efforts in the adjustment of hyperparameter [46], ELM can provide good generalization ability and extremely fast learning speed. ELM contains input, hidden layers, and output nodes, and only hidden layer nodes required to be set in ELM. For given M different samples (x_i, y_i) , the model of ELM can be expressed as follows:

$$\sum_{i=1}^L b_i g_i(x_j) = \sum_{i=1}^L b_i g(\omega_i j_i + b_i) = y_i, \quad (15)$$

where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$, $y_i = [y_{i1}, y_{i2}, \dots, y_{im}]^T \in R^m$, L denotes hidden nodes, $g(x)$ indicates a hidden layer activation function, ω_i illustrates the weight vector connecting the i th hidden node and input nodes, and b_i is the threshold of i th hidden nodes. For all M samples, equation (15) can be written as

$$H\beta = Y, \quad (16)$$

where $H = \begin{bmatrix} g(\omega_1 x_1 + b_1) & \cdots & g(\omega_L x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\omega_1 x_N + b_1) & \cdots & g(\omega_L x_N + b_L) \end{bmatrix}_{N \times L}$,

$Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m}$, and $\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}$, H shows the hidden layer output matrix. The ELM theory states that the hidden node

learning parameters ω and b can be randomly assigned regardless of input data.

Therefore, the system equation (15) becomes a linear model. By finding the least squares solution of the linear system (15), the output weights can be analytically determined as follows:

$$\beta = H^\dagger T, \quad (17)$$

where H^\dagger indicates the Moore–Penrose generalized inverse of the hidden layer output matrix H [47].

3.4. Key Influencing Factors Selection Model Construction. The effective integration of FCBGSO, MFD, probit regression, and artificial prior knowledge is applied to the key influencing factors selection of P2P lending investment risk. Firstly, the MFD is treated as an evaluation criterion for a feature subset, and FCBGSO is used as a search strategy. The combination of FCBGSO and MFD (FCBGSO + MFD) is used for reducing the redundancy attributes in the original dataset, and the preliminary subset is attained. Secondly, we analyze the correlation between the selected attributes and the default risk of P2P lending investment using the probit regression, and those attributes that are nonsignificantly correlated with the investment risk will be removed. Finally, the attributes that have a significant impact on the investment risk are selected from the original dataset using the artificial prior knowledge, which are added into the retaining attributes one by one. Then, a small and reasonable number of attribute subsets are achieved, and we assess their classification accuracies using ELM. The attribute subset with the highest classification accuracy is the key influencing factors of P2P lending investment risk.

The pseudocode of Algorithm 1 is presented as follows. The main steps of the model construction are as follows:

Step 1: calculate the MFD of the original dataset of P2P lending and obtain the number of attributes in the preliminary subset m' ($m' = D$, $D = \max(D_q)$); the objective function $f = \sqrt{\sum_q (\text{frac}_q - D_q)^2}$, $q = 2, 3, 4, 5, 6$

Step 2: search the preliminary attribute subset B_1 of P2P lending orders with the minimal objective function value using FCBGSO

Step 3: eliminate attributes that are nonsignificantly related to default risk in B_1 using the probit regression and get the attribute subset B_2

Step 4: select the attributes extracted from the original dataset that have a significant influence on the investment risk and do not belong to B_2 using the artificial prior knowledge and form the attribute subset A

Step 5: add the attributes in A into B_2 one by one and get a small and reasonable number of attribute subsets B'_1, B'_2, \dots, B'_n

Step 6: calculate the classification accuracy of each attribute subset in B'_1, B'_2, \dots, B'_n using ELM, and then obtain their classification accuracies p_1, p_2, \dots, p_n

Step 7: assume p_i ($i = 1, 2, \dots, n$) is the highest classification accuracy in p_1, p_2, \dots, p_n , and then the attribute subset B'_i is the key influencing factors of P2P lending investment risk

4. Experimental Results

In this section, to assess the performance of the proposed approach, the experiments are implemented in MATLAB 2017a. The algorithm is tested on a computer running 64-bit Windows 10 with 2.81 GHz processor and 8 GB memory. Experimental parameters are set as follows: the population size $N = 30$, the maximum number of iterations $t_{\max} = 20$, luciferin volatile factor $\rho = 0.4$, luciferin renewal rate $\gamma = 0.6$, dynamic decision domain update rate $\beta = 0.08$, neighborhood threshold $n_t = 5$, and the remaining parameters are analyzed in Section 4.4.

4.1. Data Preprocessing and Indicator System Construction. Renrendai platform is one of the earliest P2P lending information intermediary service platforms in China, which has been steadily operating since its establishment. It has been ranked in the top 100 Internet companies in China twice. Hence, we used the P2P lending datasets of Renrendai as the empirical data in this work. We obtained more than 400,000 P2P lending transaction orders from the Renrendai platform, and 396,993 of them are valid. Then, the outlier orders and 295,589 orders of unsuccessful fundraising are removed. Finally, 99,469 orders are available for the key influencing factors selection of P2P lending investment risk. After the above procedure, the retaining dataset is an imbalanced dataset, and then the balanced dataset of P2P lending investment risk is achieved using the undersampling and the stratified sampling methods. On the basis of the relevant knowledge of the Internet finance and the research results on the key influencing factors of P2P lending investment risk [5, 6], its index system is shown in Figure 2. We take the default risk of the borrowers as the decision attributes in this work.

4.2. Experimental Results. The proposed key influencing factors selection method, using the combination of FCBGSO and MFD (FCBGSO + MFD), selects the preliminary attribute subset from the original dataset of P2P lending orders. The four attributes are retained after selection, i.e., they are H_1 (interest rate), H_4 (number of investors), H_7 (age), and H_{15} (occupation). The FCBGSO + MFD greatly reduces the redundant attributes in the original dataset. While, there is a question to discuss, that is, whether the retained four attributes are significantly related to the default risk. We use the probit regression model to assess the significance between the four attributes and the default risk.

We take the default state as the explained variable and regard interest rate, number of investors, age, and occupation as the explanatory variables. The probit regression model is established as follows:

$$P(\text{default} = 1) = f(\lambda S_i + \rho L_i), \quad (18)$$

Inputs: the initial parameters, the initial data of P2P lending, and MFD computing system.
Outputs: the key influencing factors of P2P lending B'_i .

- (1) Initialize the parameters.
- (2) N glowworms are generated randomly, and compute their MFD f using equation (14).
- (3) $X_{opt} \leftarrow \text{maxfitness}(X_1, X_2, \dots, X_N), f_{opt} \leftarrow \max\{f_1, f_2, \dots, f_N\}$.
- (4) $t \leftarrow 1$.
- (5) **while** $t \leq t_{max}$ **do**
- (6) **for** $i \leftarrow 1$ to N **do**
- (7) Select the objective glowworm X_j in the radial range local-decision domain r_d^i of the glowworm X_i .
- (8) Move a step to X_j using equations (6)–(9).
- (9) Update the luciferin l_i and the radial range local-decision domain r_d^i .
- (10) **if** $\text{rand} < r_1$ **do**
- (11) N glowworms are divided into three subpopulations according to their MFD.
- (12) Perform the coevolution mechanism to create offspring glowworms and update their parent glowworms.
- (13) **end if**
- (14) **if** $\text{rand} < r_2$ **do**
- (15) Perform the fireworks evolution strategy to create new glowworms and update the current glowworm.
- (16) **end if**
- (17) **end for**
- (18) $X_{opt} \leftarrow \text{maxfitness}(X_1, X_2, \dots, X_N), f_{opt} \leftarrow \max\{f_1, f_2, \dots, f_N\}$.
- (19) **end while**
- (20) Obtain the preliminary attribute subset B_1 which corresponds to X_{opt} .
- (21) Get the attribute subset B_2 by eliminating those attributes that are not significantly related to the default risk in B_1 using the probit regression.
- (22) Form an attribute subset A extracted from the original dataset of P2P lending using the artificial prior knowledge.
- (23) Generate a small and reasonable number of attribute subsets B'_1, B'_2, \dots, B'_n by adding the attributes in A into B_2 .
- (24) Get the classification accuracies by evaluating each subset in B'_1, B'_2, \dots, B'_n using ELM.
- (25) Achieve the key influencing factors of P2P lending B'_i with the highest classification accuracy.
- (26) **return** B'_i

ALGORITHM 1: The key factors selection approach.

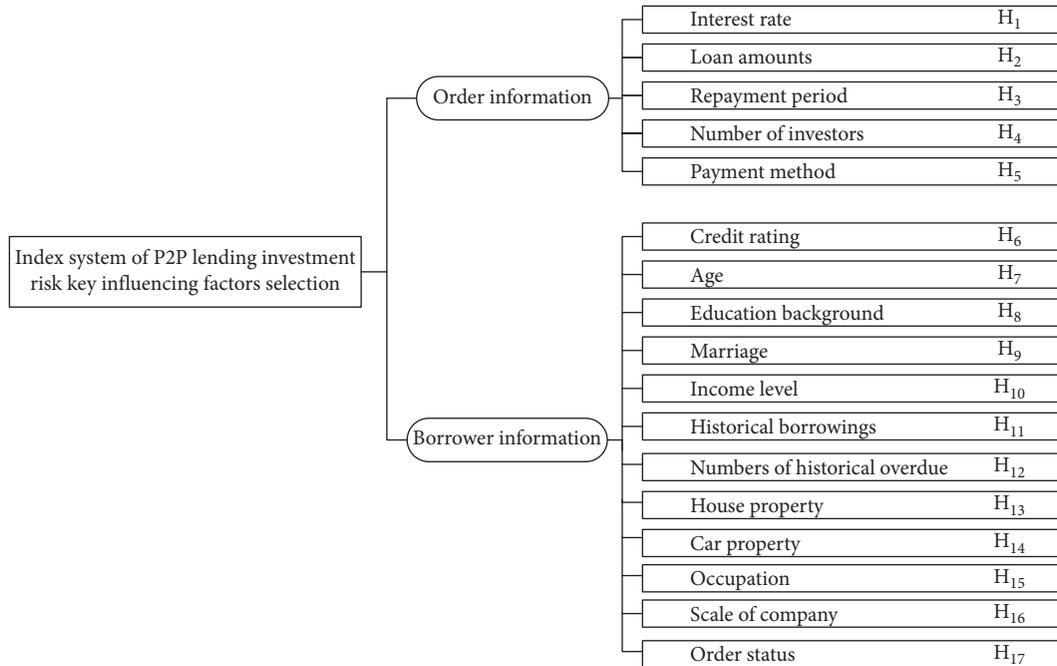


FIGURE 2: Index system of P2P lending investment risk key influencing factors selection.

where default denotes default risk, S indicates explained variable, and L demonstrates control variable.

As reported in Table 1, the regression coefficient of interest rate is 0.0573 and the marginal utility is 0.0221, which reveal that there is a positive significance between the interest rate and the default risk at 1% significance levels. Age and occupation are also significantly positive at the 1% level. But, the number of investors has no significant impact on the default risk in comparison with other three factors. Therefore, when analyzing the key influencing factors selection of P2P lending investment risk, H_4 should be removed and H_1 , H_7 , and H_{15} are retained.

Considering that FCBGSO + MFD cannot recognize and learn the application background, lack of active thinking and personal perception, we extract the attributes with a significant impact on default risk using the artificial prior knowledge in this work. Credit rating plays an important role in the process of investors making investment decisions, as illustrated in Table 2. In the P2P lending industry, investors need to consider on whom the funds are invested in and the specific amount allocated for each order, so as to maximize the expected investment income and reduce the return risk. Credit rating is an important input to solve such combinatorial optimization problem, so it has important reference value for the key influencing factors selection of P2P lending investment risk [5, 51]. In addition, the borrower's historical information is a nice complement to the credit rating. The higher the repayment rate of historical borrowings on time, the lower the ratio between historical overdue times and historical borrowing times, which indicates the borrowers convey a message to investors that the borrowers are trusted and welcomed by the market. The lower the default risk perceived by investors, the smaller the risk compensation. Therefore, H_{10} (historical borrowings) and H_{11} (historical overdue times) of borrowers are of great significance in the analysis of key influencing factors selection of P2P lending investment risk [48, 49].

In summary, the results achieved by the key influencing factors selection method of P2P lending investment risk are shown in Table 3. The attributes selected by the artificial prior knowledge are H_6 , H_{10} , and H_{11} , which are added into the attribute subset (H_1 , H_7 , and H_{15}) one by one. Then, a small and reasonable number of attribute subsets are achieved, which are shown in Table 4. We use ELM to calculate the classification accuracy of each attribute subset, and the subset with the highest accuracy is the key influencing factors of P2P lending investment risk. Because the higher the classification accuracy of the subset is, the more relevant between the subset's attributes and the default risk.

The maximal and average classification accuracies of combinations 1–10 are displayed in Table 4. In Table 4, combination 1 is the original dataset, combination 2 is the preliminary attribute subset attained by FCBGSO + MFD, combination 3 is the retaining attributes after removing the nonsignificant correlation variable in combination 2 using the probit regression method, and combinations 4–10 are the attribute subsets by adding H_6 , H_{10} , and H_{11} into combination 3 one by one.

TABLE 1: Regression analysis between different influencing factors and default risk.

Variable names	Probit regression equation coefficient	Explained variable (default)	
		$P > z $	dy/dx
Interest rate	0.0573***	≤ 0.001	0.0221
Number of investors	0.0007	0.402	0.0003
Age	0.0182***	0.008	0.0070
Occupation	0.0495***	0.004	0.0191
Persudo $R^2 = 0.304$			
LR $\chi^2(4) = 33.67$			
Prob $> \chi^2 = 0.0000$			

***, **, and * indicate statistical significance at 10%, 5%, and 1% significance levels, respectively.

The maximal and average classification accuracies of the attribute subsets (combinations 4–10) are markedly higher than that of combination 2, which indicates the proposed approach can achieve a better result than the FCBGSO + MFD, namely, the combination of the artificial intelligence method, the traditional statistical method, and the artificial prior knowledge performs better than every single one of them. After removing H_4 in combination 2 by the probit regression, the accuracy of combination 3 is slightly lower than that of combination 2, but the decrease is within the acceptable range. It implies that H_4 is not a key influencing factor of P2P lending investment risk. The maximal and average accuracies of combination 9 are higher than the other combinations. Therefore, H_1 , H_7 , H_{10} , H_{11} , and H_{15} in combination 9 are the key influencing factors of P2P lending investment risk. It indicates that the proposed approach dramatically reduces the redundant attributes. The key influencing factors of P2P lending investment risk are exactly achieved, which provides high-quality data for the prediction of P2P lending investment risk.

4.3. Comparison Analysis. To verify the effectiveness and credibility of the proposed approach, we compare it with the following methods in literatures [19, 29, 50, 52]. Literatures [19, 50] adopt swarm intelligence algorithms combined with MFD for the key influencing factors selection. The literature [29] uses a rough set theory combined with artificial fish swarm algorithm for attribute selection. The literature [52] employs the statistical method and the artificial prior knowledge to extract the key influencing factors. In Table 5, the maximal and average classification accuracies of the proposed approach are superior to that of other algorithms, which denotes its validity and effectiveness. Moreover, in comparison with the literatures [19, 29, 50, 52], the maximal classification accuracies achieved by the proposed approach are increased by 19 percentage points, 18 percentage points, 23 percentage points, and 4 percentage points, respectively. The average accuracies are raised by 19 percentage points, 18 percentage points, 21 percentage points, and 2 percentage points, respectively. Given the above, the key influencing factors selected by the proposed method perform the best, followed by literature [19, 29, 52] and literature [50] is the

TABLE 2: Key influencing factors analysis of P2P lending investment risk achieved by artificial prior knowledge.

Attributes	Names	Explanation
H ₆	Credit rating	Literatures [5, 48] indicate that credit rating can reflect a borrower’s credit status, reveal his credit risk, and avoid adverse selection in investment. Credit rating is an important input for combinatorial optimization problem to balance investment earnings and return risk, which is of great significance to key influencing factors selection of P2P lending investment risk
H ₁₀	Historical borrowings	Literatures [49, 50] illustrate that the borrower’s historical information is an important factor affecting the investment risk. It embodies in the number of historical overdue and historical borrowings. The higher the repayment rate of historical borrowings on time, the lower the ratio between historical overdue times and historical borrowing times. It indicates that the borrowers convey a message to investors that the borrowers are trusted and welcomed by the market, and the risk compensation is smaller
H ₁₁	Numbers of historical overdue	

TABLE 3: Key influencing factors selection analysis of P2P lending investment risk.

	Original dataset	Preliminary influencing factors	Nonsignificant relevant attributes	Attributes selected by artificial prior knowledge
Number of attributes	17	4	1	3
Attribute subsets	H ₁ , H ₂ , . . . , H ₁₇	H ₁ , H ₄ , H ₇ , H ₁₅	H ₄	H ₆ , H ₁₀ , H ₁₁

TABLE 4: Classification accuracy analysis before and after key influencing factors selection of P2P lending investment risk.

Combinations	Attribute subsets	Classification accuracy (%)	
		Max	Mean
Combination 1	H ₁ , H ₂ , . . . , H ₁₇	85.6250	77.8227
Combination 2	H ₁ , H ₄ , H ₇ , H ₁₅	76.6234	66.1543
Combination 3	H ₁ , H ₇ , H ₁₅	74.5342	65.7832
Combination 4	H ₁ , H ₆ , H ₇ , H ₁₅	86.4198	78.3466
Combination 5	H ₁ , H ₇ , H ₁₀ , H ₁₅	77.6398	70.6136
Combination 6	H ₁ , H ₇ , H ₁₁ , H ₁₅	89.3750	79.5277
Combination 7	H ₁ , H ₆ , H ₇ , H ₁₀ , H ₁₅	86.9565	80.0076
Combination 8	H ₁ , H ₆ , H ₇ , H ₁₁ , H ₁₅	88.8199	82.4711
Combination 9	H ₁ , H ₇ , H ₁₀ , H ₁₁ , H ₁₅	93.1250	83.9844
Combination 10	H ₁ , H ₆ , H ₇ , H ₁₀ , H ₁₁ , H ₁₅	90.0621	82.6736

“Max” and “Mean,” respectively, indicate the maximal and the average classification accuracies of the P2P lending subsets.

worst. It also illustrates that the proposed key influencing factors selection approach by combining qualitative and quantitative analysis is more reasonable and scientific.

4.4. *Parameter Analysis.* In the proposed selection method of key influencing factors of P2P lending investment risk, FCBGSO is employed as a search strategy. To improve the performance of FCBGSO, its main parameters should be analyzed, including iterations, population size, initial local-decision range, and maximal local-decision range.

TABLE 5: Comparison analysis between the proposed approach and other method.

Methods	Selected key influencing factors	Reduction rate (%)	Classification accuracy (%)	
			Max	Mean
Literature [19]	H ₁ , H ₃ , H ₄ , H ₇	76.47	74.1290	64.9641
Literature [29]	H ₁₀ , H ₁₂ , H ₁₃	82.35	75.0000	65.9587
Literature [50]	H ₁ , H ₄ , H ₇ , H ₁₆	76.47	70.6250	62.6507
Literature [52]	H ₁ , H ₂ , H ₃ , H ₆	76.47	88.7500	81.9134
Proposed method	H ₁ , H ₇ , H ₁₀ , H ₁₁ , H ₁₅	70.59	93.1250	83.9844

To verify the performance of FCBGSO, it is compared with GSO [53], IGSO [54], DGSO [55], and BGSO [56] as shown in Figure 3(a). As the iterations increase, the MFD difference curves between attribute subset selected by the five algorithms and the original dataset of P2P lending go down first and level off (the smaller the MFD difference is, the better the algorithm performs). Additionally, the convergence speed and precision of FCBGSO are significantly better than GSO, IGSO, DGSO, and BGSO. We advise to set the maximum of iterations at 20.

In Figure 3(b), with the increasing of population size, the MFD difference decreases continuously. When the size of the population reaches 30, the performance of FCBGSO

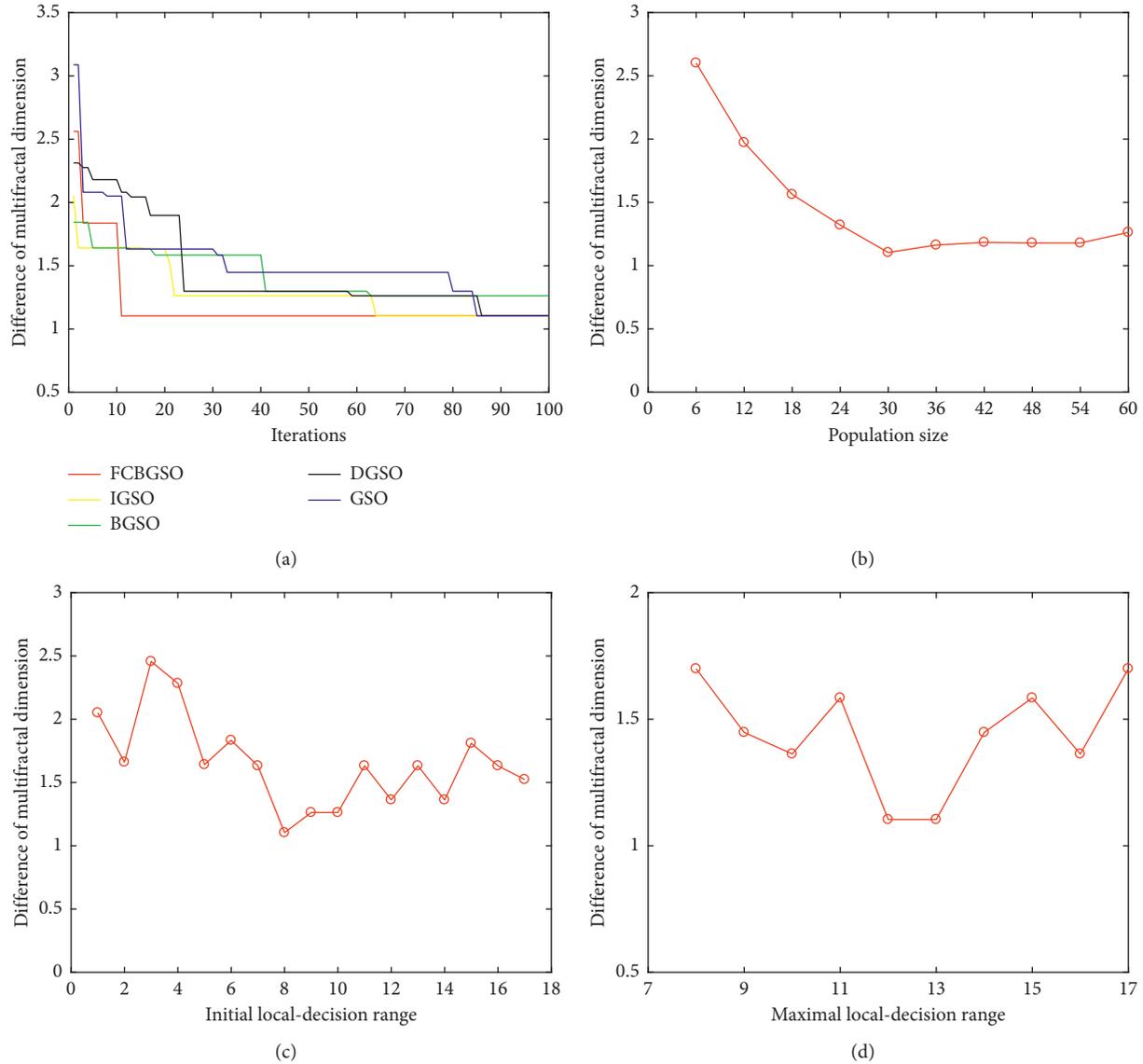


FIGURE 3: Performance impact analysis of FCBGSO with different parameters. (a) Iterations. (b) Population size. (c) Initial local-decision range. (d) Maximal local-decision range.

tends to be stable. So, the population size should be set at 30.

Figure 3(c) analyzes the relationship between the initial local-decision range and the performance of FCBGSO. If the initial local-decision range is undersize, it may affect its convergence speed. If the initial local-decision range is oversize, the algorithm easily traps into local optima. As there are 17 attributes in the P2P lending dataset, the radius of the initial local-decision range varies from 1 to 17. When the initial local-decision range is 8, the algorithm performs at its best. We advise to set the initial local-decision range at 8.

Figure 3(d) investigates the relationship between the maximal local-decision range and the performance of FCBGSO. The maximal local-decision range should be greater than or equal to the initial local-decision range, so the range of maximal local-decision range varies from 8 to 17. The algorithm achieves the best result when the maximal

local-decision range is 12 or 13. Therefore, the maximal local-decision range should be set at 12 or 13.

5. Conclusion

To exactly predict the investment risk of P2P lending, we need to scientifically and rationally analyze its key influencing factors. But, existing traditional statistical approaches cannot find the exact key influencing factors of the P2P lending investment risk, and the attributes achieved by artificial intelligence methods may not be the key influencing factors of P2P lending investment risk. To tackle the above issues, a key influencing factors selection approach of P2P lending investment risk is proposed using the combination of FCBGSO, MFD, probit regression, and artificial prior knowledge. On one hand, the proposed FCBGSO with a high searching efficiency combined with MFD tends to perform

well when it comes to dealing with the high-dimensional original dataset of P2P lending, and the preliminary attribute subset is achieved. On the other hand, the nonsignificant relevant attributes with the default risk in the preliminary attribute subset are removed using the probit regression method. After that, a small and reasonable number of attribute subsets are attained by combining the retaining attributes and the attributes achieved by the artificial prior knowledge. The attribute subset with the best accuracy assessed using ELM is efficiently achieved from the attribute subsets, namely, it is the key influencing factors of P2P lending investment risk. Finally, the experimental results on the real P2P lending dataset of Renrendai demonstrate the validity and effectiveness of the proposed approach. In addition, the proposed FCBGSO performs better than other binary heuristic algorithms with respect to the convergence speed and precision.

In future work, we will attempt to use an ensemble classifier of ELMs with a high classification ability to predict the investment risk of P2P lending. We believe that promising results can be achieved, which can provide new research ideas for the investment risk prediction of P2P lending.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Anhui Provincial Natural Science Foundation under grant nos. 1908085QG298 and 1908085MG232, the National Nature Science Foundation of China under grant nos. 91546108 and 71490725, the National Key Research and Development Plan under grant no. 2016YFF0202604, the Fundamental Research Funds for the Central Universities nos. JZ2019HGTA0053 and JZ2019HGBZ0128, and the Open Research Fund Program of Key Laboratory of Process Optimization and Intelligent Decision-making (Hefei University of Technology), Ministry of Education.

References

- [1] H. Zhang, H. Zhao, Q. Liu, T. Xu, E. Chen, and X. Huang, "Finding potential lenders in P2P lending: a hybrid random walk approach," *Information Sciences*, vol. 432, pp. 376–391, 2018.
- [2] M. Herzenstein, U. M. Dholakia, and R. L. Andrews, "Strategic herding behavior in peer-to-peer loan auctions," *Journal of Interactive Marketing*, vol. 25, no. 1, pp. 27–36, 2011.
- [3] Q. Tuo, H. Zhao, and Q. Hu, "Hierarchical feature selection with subtree based graph regularization," *Knowledge-Based Systems*, vol. 163, pp. 996–1008, 2019.
- [4] M. Dash and H. Liu, "Feature selection for classification," *Intelligent Data Analysis*, vol. 1, no. 3, pp. 131–156, 1997.
- [5] Y. Guo, W. Zhou, C. Luo, C. Liu, and H. Xiong, "Instance-based credit risk assessment for investment decisions in P2P lending," *European Journal of Operational Research*, vol. 249, no. 2, pp. 417–426, 2016.
- [6] L. Larrimore, L. Jiang, J. Larrimore, D. Markowitz, and S. Gorski, "Peer to peer lending: the relationship between language features, trustworthiness, and persuasion success," *Journal of Applied Communication Research*, vol. 39, no. 1, pp. 19–37, 2011.
- [7] J. Yao, J. Chen, J. Wei et al., "The relationship between soft information in loan titles and online peer-to-peer lending: evidence from RenRenDai platform," *Electronic Commerce Research*, vol. 18, pp. 1–19, 2018.
- [8] Z. Xiao, Y. Li, and K. Zhang, "Visual analysis of risks in peer-to-peer lending market," *Personal and Ubiquitous Computing*, vol. 22, pp. 1–14, 2018.
- [9] L. Gonzalez and Y. K. Loureiro, "When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans," *Journal of Behavioral and Experimental Finance*, vol. 2, pp. 44–58, 2014.
- [10] X. Chen, B. Huang, and D. Ye, "The role of punctuation in P2P lending: evidence from China," *Economic Modelling*, vol. 68, pp. 634–643, 2018.
- [11] Z. Zhang, L. Bai, Y. Liang, and E. Hancock, "Joint hypergraph learning and sparse regression for feature selection," *Pattern Recognition*, vol. 63, pp. 291–309, 2017.
- [12] J. Liu, Y. Lin, M. Lin, S. Wu, and J. Zhang, "Feature selection based on quality of information," *Neurocomputing*, vol. 225, pp. 11–22, 2017.
- [13] D. Huang and T. W. S. Chow, "Effective feature selection scheme using mutual information," *Neurocomputing*, vol. 63, no. 1, pp. 325–343, 2005.
- [14] J. Liu, Y. Lin, Y. Li, W. Weng, and S. Wu, "Online multi-label streaming feature selection based on neighborhood rough set," *Pattern Recognition*, vol. 84, pp. 273–287, 2018.
- [15] A. Ferone, "Feature selection based on composition of rough sets induced by feature granulation," *International Journal of Approximate Reasoning*, vol. 101, pp. 276–292, 2018.
- [16] X.-Y. Luan, Z.-P. Li, and T.-Z. Liu, "A novel attribute reduction algorithm based on rough set and improved artificial fish swarm algorithm," *Neurocomputing*, vol. 174, pp. 522–529, 2016.
- [17] Y. Chen, Q. Zhu, and H. Xu, "Finding rough set reducts with fish swarm algorithm," *Knowledge-Based Systems*, vol. 81, pp. 22–29, 2015.
- [18] K. Mukherjee, J. K. Ghosh, and R. C. Mittal, "Variogram fractal dimension based features for hyperspectral data dimensionality reduction," *Journal of the Indian Society of Remote Sensing*, vol. 41, no. 2, pp. 249–258, 2013.
- [19] C. Zhang, Z. Ni, L. Ni, and N. Tang, "Feature selection method based on multi-fractal dimension and harmony search algorithm and its application," *International Journal of Systems Science*, vol. 47, no. 14, pp. 3476–3486, 2016.
- [20] Z. W. Ni, H. W. Xiao, Z. J. Wu et al., "Attribute selection method based on improved discrete glowworm swarm optimization and fractal dimension," *Pattern Recognition and Artificial Intelligence*, vol. 26, no. 12, pp. 1169–1178, 2013.
- [21] C. A. M. Lima, A. L. V. Coelho, R. C. B. Madeo, and S. M. Peres, "Classification of electromyography signals using relevance vector machines and fractal dimension," *Neural Computing and Applications*, vol. 27, no. 3, pp. 791–804, 2016.
- [22] H. Dong, T. Li, R. Ding, and J. Sun, "A novel hybrid genetic algorithm with granular information for feature selection and optimization," *Applied Soft Computing*, vol. 65, pp. 33–46, 2018.

- [23] S. Jadhav, H. He, and K. Jenkins, "Information gain directed genetic algorithm wrapper feature selection for credit rating," *Applied Soft Computing*, vol. 69, pp. 541–553, 2018.
- [24] H. Ghimatgar, K. Kazemi, M. S. Helfroush, and A. Aarabi, "An improved feature selection algorithm based on graph clustering and ant colony optimization," *Knowledge-Based Systems*, vol. 159, pp. 270–285, 2018.
- [25] S. Tabakhi, P. Moradi, and F. Akhlaghian, "An unsupervised feature selection algorithm based on ant colony optimization," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 112–123, 2014.
- [26] Y. Wan, M. Wang, Z. Ye, and X. Lai, "A feature selection method based on modified binary coded ant colony optimization algorithm," *Applied Soft Computing*, vol. 49, pp. 248–258, 2016.
- [27] P. Moradi and M. Gholampour, "A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy," *Applied Soft Computing*, vol. 43, pp. 117–130, 2016.
- [28] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: novel initialisation and updating mechanisms," *Applied Soft Computing*, vol. 18, pp. 261–276, 2014.
- [29] Y. Chen, Z. Zeng, and J. Lu, "Neighborhood rough set reduction with fish swarm algorithm," *Soft Computing*, vol. 21, no. 23, pp. 6907–6918, 2017.
- [30] X. Zhu, Z. Ni, L. Ni, F. Jin, M. Cheng, and J. Li, "Improved discrete artificial fish swarm algorithm combined with margin distance minimization for ensemble pruning," *Computers & Industrial Engineering*, vol. 128, pp. 32–46, 2019.
- [31] X. Zhu, Z. Ni, M. Cheng, F. Jin, J. Li, and G. Weckman, "Selective ensemble based on extreme learning machine and improved discrete artificial fish swarm algorithm for haze forecast," *Applied Intelligence*, vol. 48, no. 7, pp. 1757–1775, 2018.
- [32] X. Chen, Y. Zhou, Z. Tang, and Q. Luo, "A hybrid algorithm combining glowworm swarm optimization and complete 2-opt algorithm for spherical travelling salesman problems," *Applied Soft Computing*, vol. 58, pp. 104–114, 2017.
- [33] R. Karthikeyan and P. Alli, "Feature selection and parameters optimization of support vector machines based on hybrid glowworm swarm optimization for classification of diabetic retinopathy," *Journal of Medical Systems*, vol. 42, no. 10, p. 195, 2018.
- [34] K. N. Krishnanand and D. Ghose, "Glowworm swarm optimisation: a new method for optimising multi-modal functions," *International Journal of Computational Intelligence Studies*, vol. 1, no. 1, pp. 93–119, 2009.
- [35] K. N. Krishnanand and D. Ghose, "Glowworm swarm optimization for simultaneous capture of multiple local optima of multimodal functions," *Swarm Intelligence*, vol. 3, no. 2, pp. 87–124, 2009.
- [36] K. N. Krishnanand and D. Ghose, "Glowworm swarm based optimization algorithm for multimodal functions with collective robotics applications," *Multiagent and Grid Systems*, vol. 2, no. 3, pp. 209–222, 2006.
- [37] H. Cui, J. Feng, J. Guo, and T. Wang, "A novel single multiplicative neuron model trained by an improved glowworm swarm optimization algorithm for time series prediction," *Knowledge-Based Systems*, vol. 88, pp. 195–209, 2015.
- [38] B. Wu, C. Qian, W. Ni, and S. Fan, "The improvement of glowworm swarm optimization for continuous optimization problems," *Expert Systems with Applications*, vol. 39, no. 7, pp. 6335–6342, 2012.
- [39] M. Taherkhani and R. Safabakhsh, "A novel stability-based adaptive inertia weight for particle swarm optimization," *Applied Soft Computing*, vol. 38, pp. 281–295, 2016.
- [40] C. Gan, W. Cao, M. Wu, and X. Chen, "A new bat algorithm based on iterative local search and stochastic inertia weight," *Expert Systems with Applications*, vol. 104, pp. 202–212, 2018.
- [41] H. T. Liang and F. H. Kang, "Adaptive mutation particle swarm algorithm with dynamic nonlinear changed inertia weight," *Optik*, vol. 127, no. 19, pp. 8036–8042, 2016.
- [42] R. Cheng, Y. Bai, Y. Zhao, X. Tan, and T. Xu, "Improved fireworks algorithm with information exchange for function optimization," *Knowledge-Based Systems*, vol. 163, no. 1, pp. 82–90, 2019.
- [43] B. B. Mandelbrot and J. A. Wheeler, "The fractal geometry of nature," *American Journal of Physics*, vol. 51, no. 3, pp. 286–287, 1983.
- [44] C. Traina Jr., A. Traina, L. Wu et al., "Fast feature selection using fractal dimension," *Journal of Information and Data Management*, vol. 1, no. 1, pp. 158–171, 2000.
- [45] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [46] Y. Cai, X. Liu, Y. Zhang, and Z. Cai, "Hierarchical ensemble of extreme learning machine," *Pattern Recognition Letters*, vol. 116, pp. 101–106, 2018.
- [47] X. Li, W. Mao, and W. Jiang, "Multiple-kernel-learning-based extreme learning machine for classification design," *Neural Computing and Applications*, vol. 27, no. 1, pp. 175–184, 2016.
- [48] H. Yum, B. Lee, and M. Chae, "From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms," *Electronic Commerce Research and Applications*, vol. 11, no. 5, pp. 469–483, 2012.
- [49] K. Xie, Z. Mao, and J. Wu, "Learning from peers: the effect of sales history disclosure on peer-to-peer short-term rental purchases," *International Journal of Hospitality Management*, vol. 76, pp. 173–183, 2019.
- [50] Y. J. Lu, Z. W. Ni, X. H. Zhu et al., "Attribute reduction method based on MapReduce-based improved discrete glowworm swarm algorithm and multi-fractal dimension," *Pattern Recognition and Artificial Intelligence*, vol. 31, no. 6, pp. 537–547, 2018.
- [51] 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.
- [52] J.-T. Han, Q. Chen, J.-G. Liu, X.-L. Luo, and W. Fan, "The persuasion of borrowers' voluntary information in peer to peer lending: an empirical study based on elaboration likelihood model," *Computers in Human Behavior*, vol. 78, pp. 200–214, 2018.
- [53] M. Marinaki and Y. Marinakis, "A glowworm swarm optimization algorithm for the vehicle routing problem with stochastic demands," *Expert Systems with Applications*, vol. 46, pp. 145–163, 2016.
- [54] Y. Chen, S. Wang, W. Han, Y. Xiong, W. Wang, and L. Tong, "A new air pollution source identification method based on remotely sensed aerosol and improved glowworm swarm optimization," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 8, pp. 3454–3464, 2017.
- [55] Y. Q. Zhou, Z. X. Huang, and H. X. Liu, "Discrete glowworm swarm optimization algorithm for TSP problem," *Acta Electronica Sinica*, vol. 40, no. 6, pp. 1164–1170, 2012.
- [56] M. Li, X. Wang, Y. Gong, Y. Liu, and C. Jiang, "Binary glowworm swarm optimization for unit commitment," *Journal of Modern Power Systems and Clean Energy*, vol. 2, no. 4, pp. 357–365, 2014.

