IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 13, 2020, accepted January 23, 2020, date of publication January 30, 2020, date of current version February 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2970482

A Novel Prediction Method Based on Improved Binary Glowworm Swarm Optimization and Multi-Fractal Dimension for P2P Lending Investment Risk

PINGFAN XIA, ZHIWEI NI, XUHUI ZHU, JIAPENG ZHANG, AND YUANYUAN JIN

School of Management, Hefei University of Technology, Hefei 230009, China
Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei 230009, China

Corresponding author: Zhiwei Ni (zhiwein@163.com)

This work was supported in part by the National Nature Science Foundation of China under Grant 91546108 and Grant 71490725, in part by the Anhui Provincial Natural Science Foundation under Grant 1908085QG298 and Grant 1908085MG232, in part by the National Key Research and Development Plan under Grant 2016YFF0202604, in part by the Fundamental Research Funds for the Central Universities under Grant JZ2019HGTA0053 and Grant JZ2019HGBZ0128, and in part by the Open Research Fund Program of Key Laboratory of Process Optimization and Intelligent Decision-making(Hefei University of Technology), Ministry of Education.

ABSTRACT The frequent bankruptcy incidents of peer-to-peer (P2P) lending industry have damaged the benefits of investors in China, and how to accurately and efficiently predict the investment risks of P2P lending becomes an important problem. For this very reason, a novel prediction method based on improved binary glowworm swarm optimization and multi-fractal dimension (IBGSOMFD) for P2P lending investment risk is proposed. Firstly, we propose an improved binary glowworm swarm optimization, abbreviated IBGSO, by uniformly designing an initial population using the good-point set theory, improving the moving way of glowworms, and introducing the mechanism of population diffusion and variation. Secondly, IBGSO combined with multi-fractal dimension (MFD) is applied to feature selection, and the optimal subset extracted from the original dataset can be efficiently achieved utilizing IBGSO, which can reduce its redundant attributes, and retain its pivotal attributes of P2P lending investment risk. Finally, an investment risk prediction model of P2P lending based on support vector machine (SVM) is established, which can accurately and efficiently predict the investment risk of P2P lending. Experimental results on 6 University of California Irvine (UCI) benchmark datasets show that IBGSOMFD outperforms other state-of-the-art approaches in predictive ability and calculative efficiency, and its effectiveness and significance. After the performance verification of IBGSOMFD, this work looks at how it can be applied in the risk prediction of P2P lending investment in China to maintain a stable market order.

INDEX TERMS Improved binary glowworm swarm optimization, multi-fractal dimension, P2P lending investment risk, prediction.

I. INTRODUCTION

P2P lending is an innovative financial channel that can break the constraint of transaction costs and theoretically become one of the most efficient markets for allocating capital [1]. Both parties complete the transaction directly by the new financing model, which is a useful complement to the existing banking system [1]. P2P lending is an individual lending behavior independent of the formal financial system. To some extent, it has solved the problems of capital shortage of middle and low income people and financing difficulty of small and medium-sized enterprises. However, the frequent occurrences of P2P lending "bankruptcy incidents" have damaged the steady development of P2P lending market. As of January 10, 2020, there are 3,345 transferred or closed platforms in China's P2P lending industry, and 2,924 problematic platforms in total. Since summer 2018, a large number of risk incidents have exposed on P2P lending platforms.

The continually bankruptcy and fraud incidents of P2P lending industry have aroused the great attention of Chinese

The associate editor coordinating the review of this manuscript and approving it for publication was Marta Cimitile.

government. At the two sessions of March 2019 in China, Premier Li Keqiang stressed the need to ''strengthen financial risk monitoring, warning and resolution mechanisms''. To date, internet finance has been written into the Chinese government work report for six consecutive times. From the initial ''promoting development - standardizing development - being vigilant of risk monitoring and warning'' to the ''monitoring and warning'' in 2019, which denotes that it is extremely urgent to regulate and supervise the internet financial risk. As a result, it is necessary to accurately predict P2P lending investment risk, which can safeguard the improvement of internet financial supervision system and investment decision making of investors.

Over the last decade, prediction theory has been widely used in financial safety [1], food safety [2], fault diagnosis [3], traffic [4]–[6], medical diagnosis [7] and so forth. However, there are few researches in the field of P2P lending investment risk prediction. Common prediction methods include Logistic Regression [8], Multiple Discriminate Analysis [9], Fuzzy Comprehensive Evaluation method [10], Neural Network [2], Extreme Learning Machine (ELM) [11], *et al*. Nevertheless, the aforementioned approaches have some defects. The independent variable in Logistic Regression is sensitive to multicollinearity, which affects its predictive accuracy. Its capability of dealing with high-dimensional nonlinear problems is quite limited [12]. The hypothesis of Multiple Discriminate Analysis is strict, and it needs to satisfy the conditions of equal covariance matrix, linear independence and normal distribution [13]. The calculation of Fuzzy Comprehensive Evaluation is complex. The determination of membership degree and index weight is subjective [14]. The Neural Network is sensitive to the initial network weight setting, slow in convergence speed, and easy to fall into a local optimal solution [15]. The random selection of ELM hidden layer weight and deviation affects the optimization effect of the system. The number of hidden layer nodes is difficult to confirm. Too many nodes are easy to over-train, whereas too few nodes cause insufficient training and affect the predictive accuracy [16].

Support Vector Machine (SVM) is a machine learning algorithm based on the statistical theory [17]. SVM cannot only overcome the defects of traditional prediction methods, but also has strong generalization ability and robustness [18], [19]. SVM has accurate classification ability, but it cannot reduce the dimension of original datasets by itself. The irrelevant and redundant attributes contained in the original dataset will not only affect the calculative efficiency, but also affect the predictive results. Obviously, feature selection of the original dataset before prediction can improve the computational efficiency.

Previous work on feature selection roughly categorizes existing models into two points: evaluation criterion and search strategy [20]. In regard to the evaluation criterion, different evaluation criterions employed in feature selection can have a significant influence on the selection results. The optimal goal of feature selection is that the better predictive ability on the attribute subset after reduction is achieved employing less number of attributes. However, different evaluation criterions have different ways of assessing the attribute subset, and different results are obtained.

Diverse evaluation approaches are utilized to evaluate attribute subsets of great significance for big data analysis [21], [22]. For instance, information theory [23], rough sets [24], fractal dimension [20], [25] *et al*. Nevertheless, information theory is mainly used to find condition attributes that have an important influence on decision-making, which is only suitable for dealing with feature selection problems with small size [23]. Rough set theory is sensitive to noise. Discretization should be carried out before feature selection for dealing with continuous datasets, which may lead to information distortion [24]. In recent years, fractal dimension has been used as an evaluation criterion and achieved good results. Its advantages lie in the following two aspects [26]: Firstly, by calculating the fractal dimension of an optimal feature subset, the number of attributes can be directly determined. Simultaneously, the redundant and irrelevant attributes can be eliminated; Secondly, fractal dimension performs well in dealing with high dimensional problems. Most of the existing fractal dimension-based feature selection methods only employ a single fractal dimension. Considering the complex distribution of initial dataset, it is difficult to accurately describe all the characteristics of the dataset [25]. On the contrary, multi-fractal dimension (MFD) has the capability to denote the overall distribution of the dataset comprehensively [20]. Therefore, we regard MFD as the evaluation criterion for attribute reduction.

With respect to searching strategy, heuristic algorithm is a good choice [23]. For instance, Genetic Algorithm (GA) [26], [27], Ant Colony Optimization (ACO) [28], Particle Swarm Optimization (PSO) [29], Artificial Fish Swarm Algorithm (AFSA) [30], [31]. However, all these methods have some defects. GA has complex coding and slow searching speed; ACO has a large amount of computation and slow convergence speed, which is not capable to solve large scale problems; PSO is prone to premature convergence; The late convergence speed of AFSA is relatively slow, and it lacks population diversity. Inversely, Glowworm Swarm Optimization (GSO) has strong robustness and good global convergence [32], which is suitable as the searching strategy for attribute selection in this work. On the basis of GSO, we introduce some novel strategies to improve the algorithm performance, and propose an improved binary glowworm swarm optimization (IBGSO) as the searching strategy.

According to the analysis above, a novel prediction method based on improved binary glowworm swarm optimization and multi-fractal dimension (IBGSOMFD) for P2P lending investment risk is proposed. The tasks are completed in following steps. Firstly, an improved binary glowworm swarm optimization is proposed as a search strategy. Secondly, MFD is treated as an evaluation criterion, which is combined with IBGSO used for feature selection. Then the redundant and irrelevant attributes in P2P lending datasets can be eliminated.

Thirdly, the investment risk prediction model of P2P lending is constructed by SVM, which is applied to the practical problem of P2P lending. Finally, the IBGSOMFD is tested on the benchmark datasets to verify its effectiveness and significance. Then it is used for prediction of P2P lending investment risk based on Renrendai, Paipaidai and Yilongdai platforms in China. The contributions of this work can be demonstrated as below:

1) A novel prediction method based on improved binary glowworm swarm optimization and multi-fractal dimension for P2P lending investment risk is proposed. IBGSOMFD is a novel hybrid approach, which can make up for the shortages of traditional methods, and attain accurate and effective prediction.

2) IBGSO is proposed by uniformly designing an initial population using the good-point set theory, improving the moving way of glowworms, and introducing the mechanism of population diffusion and variation. It performs well in searching for the optimal solution.

3) MFD combined with IBGSO is used for feature selection. Then the redundant and irrelevant attributes in P2P lending datasets can be removed. It can provide high quality data for subsequent prediction. The combination of IBGSO and MFD provides a new research approach for P2P lending investment risk prediction.

4) The proposed method is compared with the state-of-the-art approaches on 6 benchmark datasets. The experimental results indicate that IBGSOMFD outperforms other methods in predictive ability. In the practical prediction of P2P lending investment risk on Renrendai, Paipaidai and Yilongdai platforms in China, IBGSOMFD can achieve better results.

The rest of this paper is organized as follows. In section II, we briefly review the basic concept of GSO, and then IBGSO is proposed. The prediction method of P2P lending investment risk and how to use it are illustrated in section III. Experimental results are denoted in section IV. In section V, the conclusions and the future work are demonstrated.

II. INPROVED BINARY GLOWWORM SWARM OPTIMIZATION (IBGSO)

Glowworm Swarm Optimization (GSO) has the advantages of easy implementation, strong robustness and outstanding search ability. It can be used as a search strategy, but it still has some defects such as uneven initial population distribution and lack of population diversity. To overcome the above weaknesses and improve the searching efficiency, IBGSO is proposed.

A. GLOWWORM SWARM OPTIMIZATION (GSO)

GSO is a bionic swarm intelligence algorithm that imitates the luminescence behavior of glowworms during their foraging and courtship in nature [33]. For an optimization problem, GSO can be employed to find its optimal solution. GSO starts to search for the optimal solution utilizing a population composed of randomly generated glowworms in a solution space [34]. In the population, each glowworm represents a feasible solution of the optimization problem. At the very start, the same luciferin value is assigned to each glowworm (the higher luciferin value of the glowworm has, the more attraction it gains), and their luciferin values are updated according to their fitness values (the fitness value is used to evaluate the merits of the glowworm). Each glowworm selects the individuals with great luciferin values to form its set of neighbors, and the probabilities of each glowworm moving to its neighbors can be achieved. Each glowworm can decides to move towards the objective glowworm gained using the roulette method. Its position can be updated, and its luciferin value can also be updated in terms of its fitness value. After several iterations, the optimal glowworm in the population can be attained, which is the optimal solution of the optimization problem. The basic steps of GSO are listed as follows:

The formula for updating the luciferin of glowworm $X_i(t)$ at the t-th iteration is as follows.

$$l_i(t) = (1-\rho)l_i(t-1) + \gamma J(X_i(t)) \tag{1}$$

The luciferin renewal depends on the objective function value $J(X_i(t))$ of the glowworm. Where $l_i(t)$ indicates the luciferin level of $X_i(t)$, ρ is the luciferin decay constant ($0 < \rho < 1$), γ denotes the luciferin enhancement constant.

In the dynamic decision domain of $X_i(t)$, the glowworm with a luciferin value greater than $X_i(t)$ can be used to form its set of neighbors $N_i(t)$. The probability $P_{ij}(t)$ of $X_i(t)$ moving towards its neighbor $X_j(t)$ in $N_i(t)$ is as follows:

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

$$P_{ij}(t) = \frac{l_j(t) - l_i(t)}{\sum\limits_{k \in N_i(t)} l_k(t) - l_i(t)} \tag{3}$$

where $r_d^i(t)$ denotes the dynamic radical range $0 < r_d^i < r_s$, r_s express the radical range of the luciferin sensor.

Glowworms move towards their objective glowworms that are brighter than themselves with a certain probability $P_{ij}(t)$. Individuals are updated in the following way.

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

where s is the moving step size.

The dynamic radial range of local-decision domain is defined as below.

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

where β represents a constant parameter, n_t denotes a parameter used to limit the number of neighbors.

B. IMPROVED BINARY GLOWWORM SWARM OPTIMIZATION (IBGSO)

In order to solve the combinatorial optimization problem in binary space and improve the search efficiency of GSO, IBGSO is proposed. Firstly, the population is initialized by using the good-point set theory, so that the population is

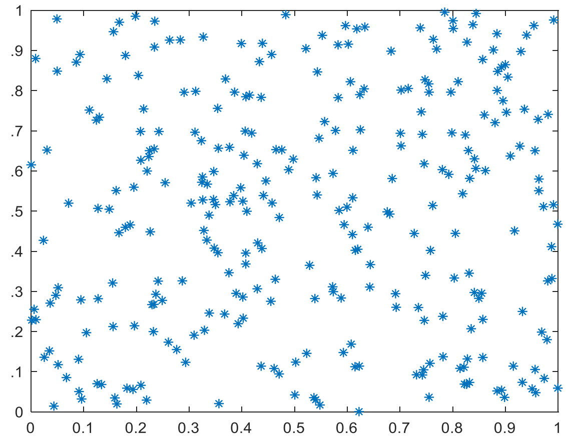

FIGURE 1. Initial population distribution generated by random method.

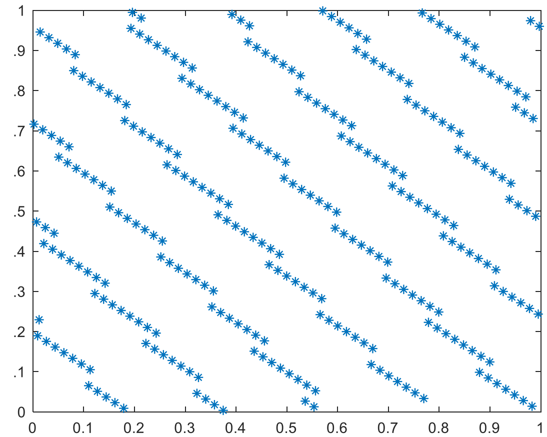

FIGURE 2. Initial population distribution generated by square root sequence.

evenly distributed in a solution space. Secondly, the location update strategy of glowworm is used for searching in the binary solution space. Finally, the population diffusion and variation mechanism are introduced to improve the search efficiency of GSO.

1) POPOLATION INITIALIZATION STRATEGY BASED ON GOOD-POINT SET (GPS)

The population initialization of GSO is essentially an optimization problem to represent the overall characteristics of solution space comprehensively and accurately using limited number of glowworms. Nowadays, random population initialization method is commonly used. Although it can simplify the computational complexity of GSO, the population initialization cannot perform uniformly, which affects the convergence speed and accuracy [35]. To make up for the deficiency, good-point set (GPS) is used for population initialization.

Assuming that Gm is a unit cube in m-dimensional Euclidean space, the initial population size is n, where $r \in G_m$. Then the GPS is denoted as $p_n(j) = \{((r_1 \times j), (r_2 \times j), \cdots, (r_m \times j)), j = 1, 2, \cdots n\}$, which represents the good-points of uniform design from the m-dimension space. There are three common methods to generate the glowworm population:

1) Square root sequence: $r_i = \sqrt{d_i}, 1 \leq i \leq m$, Where d_i represents different prime numbers;

2) Cyclotomic field method: $r_i = 2\cos 2\pi i/d, 1 \leq i \leq m$, Where d is the minimum prime number satisfying $(d-3)/2 \geq m$;

3) Exponential sequence: $r_i = \{e^i, 1 \leq i \leq m\}$.

In Figure 1to 4, each point represents a glowworm. The initial population distribution pictures with the size of 300 are shown in Figure 1 to 4. The initial population distribution generated by random method is denoted in Figure 1. Additionally, Figures 2 to 4 indicate that the initial population distribution pictures are generated by three common GPS methods. By comparing random sequence and three common GPS methods using the same population size, the three

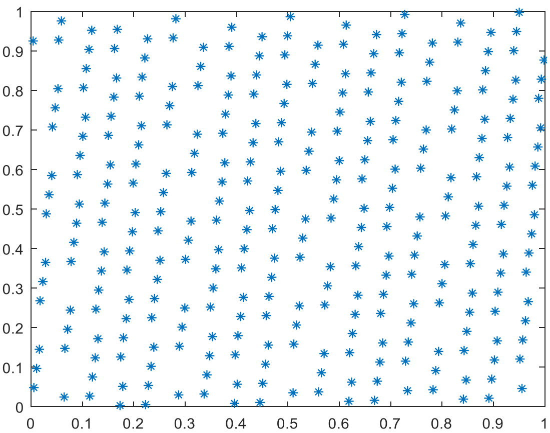

FIGURE 3. Initial population distribution generated by cyclotomic field.

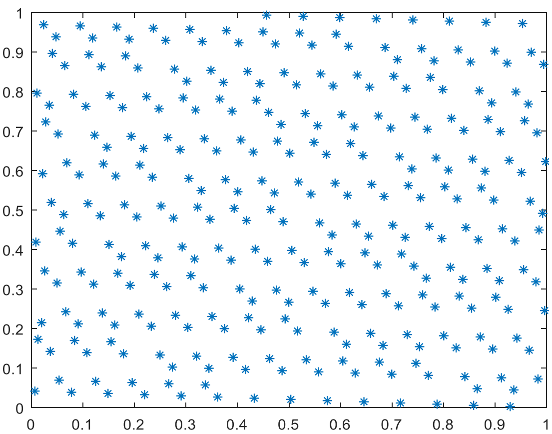

FIGURE 4. Initial population distribution generated by exponential sequence.

commonly used GPS methods are more evenly distributed. Then, compared with square root sequence, exponential sequence and cyclotomic field methods have more uniform population distribution, and the former is better than the latter. Therefore, the exponential sequence is employed to initialize the population of glowworms, which can achieve a relatively initial population and maintain specific diversity.

The construction of GPS is independent of the spatial dimension, which can adapt to a high - dimensional problem. Furthermore, the GPS has a good stability, and the same distribution can be achieved using the same population size. In general, a good initial glowworm population can be obtained by mapping the generated good-points to the solution space of GSO.

2) IMPROVE THE MOVING WAY OF GLOWWORMS

To solve a binary combinational optimization problem, it is necessary to improve the location updating way of glowworms. Let $X_i = (x_{i1}, x_{i2}, \mathrm{L}, x_{ik})$ be the current glowworm, $X_j = \left(x_{j1}, x_{j2}, \mathrm{L}, x_{jk}\right)$ is the objective glowworm, k denotes the dimension of the solution vector, and the hamming distance between X_i and X_j is used as the moving step. Change the $s = \lceil rand \times step \rceil$ bits of the current glowworm randomly. Namely, select the current glowworm's s bits and change them from 1 to 0 or vice versa, to realize the location updating of individuals.

The hamming distance hm_{ij} between X_i and X_j can be calculated as follows:

$$hm_{ijl} = \begin{cases} 1, & if\ x_{il} \neq x_{jl} \\ 0, & if\ x_{il} = x_{jl} \end{cases} \tag{6}$$

$$hm_{ij} = \sum_{l=1}^{k} hm_{ijl} \tag{7}$$

where hm_{ijl} denotes the hamming distance of X_i and X_j on the l-th dimension.

3) POPULATION DIFFUSION MECHANISM

In the basic GSO, glowworms may gather together in the later searching stage, which may lead to the prematurity convergence. To expand the search range of glowworm individuals, population diffusion mechanism is introduced to GSO.

The calculation of the offspring quantities produced by the glowworms in the diffusion process [36] is shown as below.

$$N_c = \left\lfloor (D_j - D_{\min}) \times \frac{N_{\max} - N_{\min}}{D_{\max} - D_{\min}} + N_{\min} \right\rfloor \tag{8}$$

where N_c is the number of offspring glowworms, D_j is the value of individual fitness, D_{max} and D_{min} denote the best and worst fitness values of the offspring generation, N_{max} and N_{min} are the maximum and minimum of the offspring quantities, and $\lfloor\ \rfloor$ shows a rounded down function.

In the process of spatial diffusion, the parent glowworm is taken as the axis and the offspring diffuse in space in the form of normal distribution.

$$\lambda = \lambda t + (\lambda_0 + \lambda t) \times \frac{(\omega_{\max} - \omega)^{\alpha}}{{\omega_{\max}}^{\alpha}} \tag{9}$$

where λt and λ_0 are the maximum and minimum standard deviations respectively. ω is the number of iterations, and ω_{max} is the maximal number of iterations, and α is a nonlinear harmonic index.

When the value of α is large, the convergence accuracy of the algorithm cannot be guaranteed, but it performs a fast convergence speed; When the value of α is small, its convergence accuracy of the algorithm can be improved, but it has a slow convergence speed, and may fall into a local optimum. The dependent variable of function $\tan z$ decreases with the decrease of the independent variable, and the decline speed also shows a decreasing trend. By introducing tangent function, the standard deviation is as follows:

$$\lambda = \lambda_t + (\lambda_0 - \lambda_t) \times \tan(0.875 \times \frac{\omega_{max} - \omega}{\omega_{max}}) \tag{10}$$

Consistent with the evolution of glowworms, as the number of iterations increases, the attenuation rate λ decrease gradually. In the initial stage of algorithm, the larger λ is, the farther the offspring are from their parent. With the evolution of the algorithm, the smaller λ is, the offspring are relatively uniformly closer to their parents. Then we calculate the fitness function values of offspring glowworms. We select the best offspring individuals, and compare them with current glowworms. If the selected offspring individuals perform better than current glowworms, update current glowworms. Otherwise, it does not perform the operation.

4) VARIATION MECHANISM

In order to increase the population diversity, a variation mechanism is added. The variation factor γ is introduced to enable the glowworm to achieve variation at a certain probability, which can increase the randomness of glowworm individual search, and avoid it falling into a local optimum. If $rand < \gamma$, randomly select two bits of the current glowworm individual, and change them from 0 to 1 or vice versa.

Based on the above analysis, an improved binary glowworm swarm optimization is denoted as Algorithm 1, and the architecture of IBGSO is shown as Figure 5.

C. COMPLEXITY ANALYSIS OF IBGSO

1) TIME COMPLEXITY ANALYSIS

Assuming that the initial population size of IBGSO is n, the time complexity of IBGSO is analyzed as follows.

1) Initialization of population and parameters of IBGSO, and its cost is $O(n)$;
2) Searching process of glowworms, and its cost at each iteration is $O(n^2)$;
3) Iteration process of IBGSO, and its total computational cost after T_{max} iterations is $O(T_{max} \times n^2)$.

In general, the total time complexity of IBGSO is $O(T_{max} \times n^2)$.

2) SPACE COMPLEXITY ANALYSIS

Assuming that the length of glowworm in IBGSO is l, the space required to store the basic parameters of glowworm is n, the space required to store the location of n glowworms is $n \times l$, the space required to store the newly generated glowworm in the execution diffusion mechanism is $N_c \times l$,

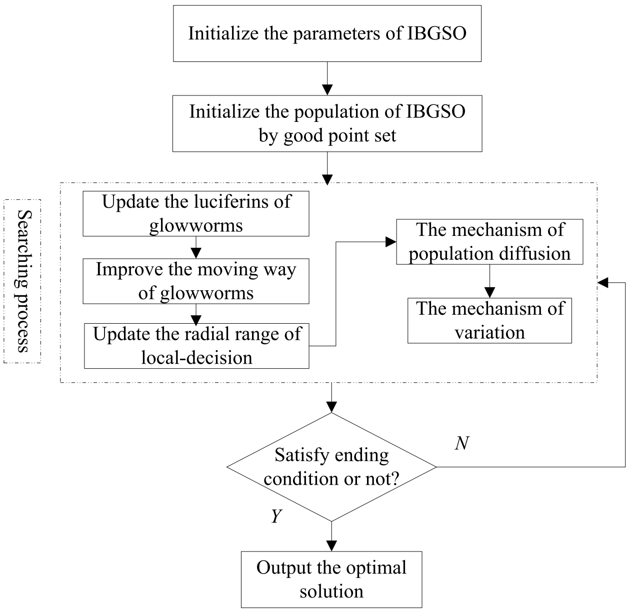

FIGURE 5. The architecture of IBGSO.

and the space required for other parameters is constant, to sum up, the total space complexity of IBGSO is $O(n \times l)$.

III. PREDICTION METHOD BASED ON IBGSO AND MFD FOR P2P LENDING INVESTMENT RISK

A. MULTI-FRACTAL DIMENSION (MFD)

Fractal theory was firstly proposed by Mandelbrot in 1983 [37], and has been applied in many fields. Traina Jr. *et al.* [20] indicated that most datasets have fractal features, and the fractal dimension is suitable as an evaluation criterion for feature selection. To eliminate redundant attributes and reduce the computational complexity, GA [26], [27], ACO [28], PSO [29], AFSA [30], [31] and Routing Algorithm [38]–[40] can be utilized as search strategies to improve computational efficiency of fractal dimension.

For an original P2P lending dataset, it involves lots of attributes such as interest rate, loan amounts, repayment period, credit rating, age, education background, and the like. However, there are many attributes are irrelevant to the P2P lending investment risk, which may lead to a low accuracy with respect to the prediction of P2P lending investment risk. The system of P2P lending has fractal characteristics, and the P2P lending dataset has high-dimensional and nonlinear characteristics. Additionally, MFD can perform well when it comes to dealing with high-dimensional and nonlinear datasets [41]. Hence, MFD can describe the P2P lending dataset from different aspects, and find the intrinsic attributes of P2P lending dataset. It is suitable as the evaluation criteria in this work. The calculation method of MFD is as follows.

$$D_q = \begin{cases} \lim\limits_{r\to 0} \dfrac{1}{q-1} \times \dfrac{\log \sum p_i^q}{\log r}, & q \neq 1 \\ \lim\limits_{r\to 0} \dfrac{\log p_i \sum p_i}{\log r}, & q = 1, \quad r \in [r_1, r_2] \end{cases} \tag{11}$$

Algorithm 1 IBGSO

Inputs: the parameters of IBGSO.
Outputs: the optimal glowworm X_{opt}, and its fitness value Y_{opt}.

Initialize the parameters.
N glowworms are generated using the good-point set theory, and compute their fitness values.
$X_{opt} \leftarrow maxfitness\ (X_1, X_2, \cdots, X_N)$, $Y_{opt} \leftarrow \max\{Y_1, Y_2, \cdots, Y_N\}$.
$t \leftarrow 1$.
While $t \leq t_{\max}$ **do**
  Update the luciferin values of N glowworms.
  for $i \leftarrow 1$ to N **do**
    Calculate the glowworms whose luciferin values are better than that of X_i in its decision domain to form the neighborhood set $N_i(t)$.
    Select the objective glowworm X_j from the neighborhood set $N_i(t)$.
    Move a step to X_j using equation (6)-(7).
    Update the radial range local-decision domain r_d^i.
    Implement the population diffusion mechanism to create new glowworms, and update the current glowworm X_i.
    if $rand < r_1$ **then**
      Implement the mutation mechanism to create the new glowworm, and update the current glowworm X_i.
    end if
  end for
  $X_{opt} \leftarrow maxfitness(X_1, X_2, \cdots, X_N)$, $f_{opt} \leftarrow \max\{f_1, f_2, \cdots, f_N\}$.
end while
return X_{opt} and Y_{opt}.

---

where p_i denotes the probability that a data point falls into the i-th grid, r demonstrates the grid size, $[r_1, r_2]$ indicates the scale-free interval of the dataset, and q is an integer. If $q < 0$, D_q describes the spatial distribution. If $q > 0$, D_q illustrates the aggregation degree.

B. CONSTRUCT THE OBJECTIVE FUNCTION

Compared with a single fractal dimension, MFD can illustrate the distribution of a dataset more comprehensively and accurately. Hence, MFD is regarded as an evaluation criterion of feature subsets in this work, and the objective function is demonstrated as below.

$$f = \sqrt{\sum_q (frac_q - D_q)^2} \tag{12}$$

where $frac_q$ and D_q respectively express the q-th fractal dimension of feature subset and original dataset. We define the difference of MFD between the attribute subset and the original dataset as the objective function. The smaller the value of the objective function, the better the solution. D_q can be specified with D_2, D_3, D_4, D_5, D_6 [41].

C. IBGSOMFD

In IBGSOMFD, IBGSO combined with MFD can be utilized to search for the optimal feature subset of an original dataset. In the first step, to find the optimal feature subset from an original P2P lending dataset composed of attributes, glowworms are generated employing the good-point set theory. The j-th member of a glowworm is composed of "1" or "0", and if it is "1", then the j-th attribute of the original dataset is selected; otherwise, it is not selected. Hence each glowworm can express a subset of the original dataset. In the second step, we calculate the MFD value of each glowworm, and then we update the luciferin values of all the glowworms according to their values of MFD. In the third step, the glowworms with higher luciferin values in the decision domain of the current glowworm are selected to construct the set of neighbors, and the probabilities of moving to the members in the set of neighbors are attained. The current glowworm can move a step to the objective individual selected by the roulette method, and it performs the mechanism of population diffusion and variation to move its search forward. All the glowworms can do a search in a similar fashion. Finally, the optimal glowworm can be achieved after multiple iterations, and it corresponds to the optimal subset of the original dataset. The optimal subset can be used to predict the P2P lending risk. The main steps of the IBGSOMFD construction are as follows:

Step 1: Extract the relevant variables from the P2P lending dataset.

Step 2: Calculate the MFD of the original P2P lending dataset A, denoted as D, and D is called to round up. Then the number of attributes in the preliminary subset $m'(m' = \lceil D \rceil, D = max(D_q))$ is achieved, the objective function is shown in equation (12).

Step 3: Search the optimal attribute subset A^* with the minimal objective function value employing IBGSO.

Step 4: Divide the optimal attribute subset A^* into non-crossed k folds randomly, that is, $A_i^*, i = 1, 2, \cdots, k$.

Step 5: Select one of k folds in order as the test set, and the remaining $k-1$ folds is used as the training set.

Step 6: Calculate the classification accuracy of each test set using the SVM based on the grid search method.

Step 7: Loop the steps 5-6 until each dataset is considered as a test set, and calculate the average accuracy.

A novel prediction method based on IBGSOMFD is expressed in Algorithm 2. The framework of P2P lending investment risk prediction is exhibited in Figure 6.

Algorithm 2 IBGSOMFD

Inputs: the initial parameters of IBGSO, P2P lending datasets, and MFD computing system.
Outputs: the predictive result *pre*.
Construct the index system of P2P lending investment risk prediction, to obtain the P2P lending datasets.
Initialize the parameters, and N glowworms are generated using the good-point theory, and calculate their MFD f using equation (12).
$X_{opt} \leftarrow maxfitness(X_1, X_2, \cdots, X_N)$,
$f_{opt} \leftarrow \max\{f_1, f_2, \cdots, f_N\}$.
$t \leftarrow 1$.
while $t \leq t_{\max}$ **do**
Update the luciferin values of N glowworms.
for $i \leftarrow 1$ to N **do**
  Calculate the neighborhood set $N_i(t)$ of X_i.
  Move a step to X_j, and update its radial range local-decision domain r_d^i.
Implement the population diffusion mechanism to create new glowworms, and update the current glowworms.
if $rand < r_1$ **then**
  Implement the mutation mechanism to create the new glowworms, and update the current glowworms.
end if
end for
$X_{opt} \leftarrow maxfitness(X_1, X_2, \cdots, X_N)$,
$f_{opt} \leftarrow \max\{f_1, f_2, \cdots, f_N\}$.
end while
Attain the preliminary attribute subset A^* which corresponds to X_{opt}.
Divide the optimal attribute subset A^* into k folds randomly, namely, $A_i^*, i = 1, 2, \cdots, k$.
Select one of k folds as the testing set, and the retaining $k-1$ folds is used as the training set.
Train SVM model based on the grid search method on the training set, and calculate the classification accuracy of each test set, $pre_1, pre_2, \cdots, pre_k$.
Achieve the mean classification accuracy
$pre = \frac{1}{k}\sum_{i=1}^{k} pre_i$.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of IBGSOMFD, the experiments are realized with Matlab 2017a. IBGSOMFD is tested on a computer running 64-bit Windows 7 with Intel(R) Core(TM) i7 6700 CPU 3.40GHz, and 12.00GB memory. The experimental parameters are set as below: the population size $n = 25$, the maximal number of iterations $t_{max} = 40$, luciferin volatile factor $\rho = 0.4$, luciferin renewal rate $\gamma = 0.6$, dynamic decision domain update rate $\beta = 0.08$, neighborhood threshold $n_t = 5$, the remaining parameters are analyzed in section IV. C. All experimental results are obtained by means of 20 independent repeated tests.

The experiments are implemented in two steps: (1) To verify the validity and universality of IBGSOMFD, the performance test is carried out on 6 UCI benchmark datasets. (2) The IBGSOMFD is applied to the prediction of P2P lending investment risk in practical problems, so as to realize accurate prediction, and provide effective support for the investment decision making of P2P lending investors.

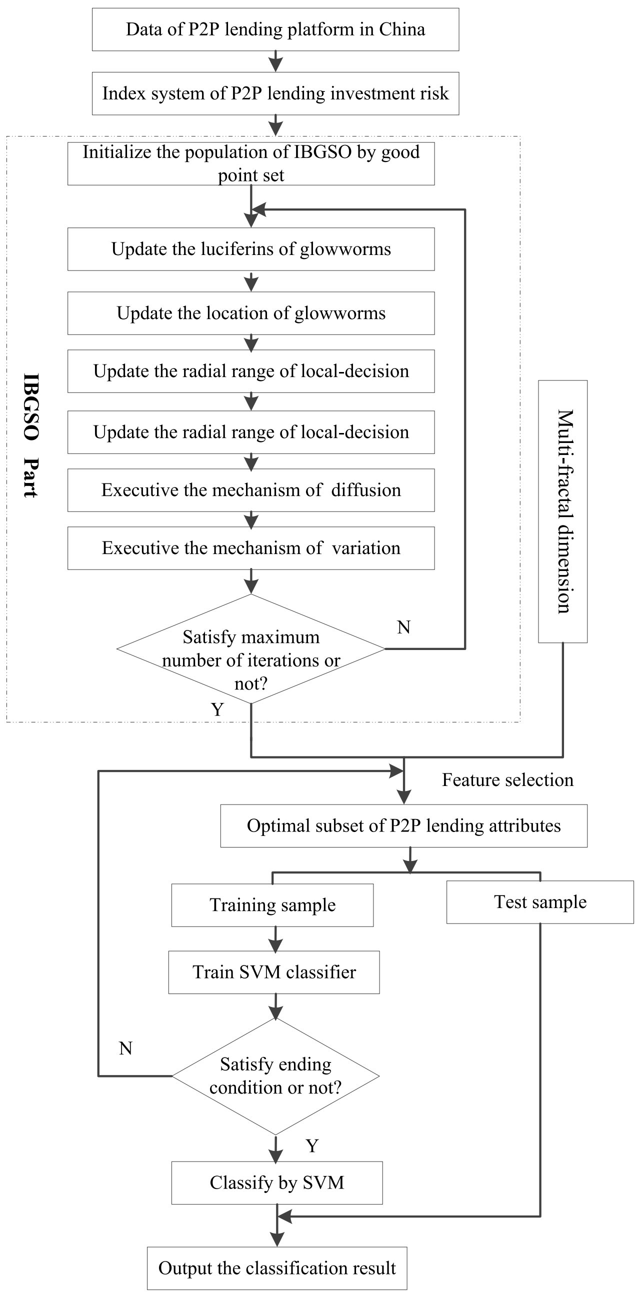

FIGURE 6. **The framework of P2P lending investment risk prediction model based on IBGSO and MFD**

A. PERFORMANCE TESTING ON THE UCI DATASET

The performance of IBGSOMFD is tested on 6 UCI datasets, which are shown in table 1. In this section, IBGSO is combined with MFD to eliminate the redundant attributes. In addition, MFD is regarded as an evaluation criterion, and IBGSO is taken as a search strategy. As shown in table 2, the reduction rates on the 6 UCI datasets are higher than 76%. The redundant and irrelevant attributes have been significantly reduced in the original datasets, which provides high-quality datasets for subsequent prediction.

To verify the effectiveness of IBGSOMFD, SVM based on the grid search method and 5-folds cross validation technique is employed in this work. It is used to compare the classification accuracies of 6 UCI datasets before and after feature selection as shown in table 3. The optimal parameters of SVM based on the grid search method are $c \in [2^{-5}, 2^{15}]$ and $g \in [2^{-15}, 2^{3}]$ [34]. The table 3 demonstrates that the classification accuracies of optimal datasets after feature selection is slightly lower than that of the original dataset on 6 UCI datasets, which is within the acceptable range. On the *Heart* dataset, the optimal classification accuracy is higher than that of the original dataset. It illustrates that the classification ability of the optimal dataset is nearly equal to that of the original dataset. In addition, the running time of finding the optimal subset is significantly reduced, and the computational efficiency is improved, which expresses the effectiveness and credibility of IBGSOMFD.

TABLE 1. **Datasets of UCI.**

Datasets	Instances	Attributes	classes
Wine	178	13	3
Heart	270	13	2
Wdbc	569	30	2
Forest	523	27	4
Vehicle	846	18	4
German	1000	20	2

TABLE 2. **Feature selection results of UCI datasets.**

Dataset	Number of original attributes	Number of optimal attributes	Optimal attributes	Reduction rate
Wine	13	2	7, 10	86.62%
Heart	13	3	4, 5, 8	76.92%
Wdbc	30	2	4, 14	93.33%
Forest	27	3	1, 12, 25	88.89%
Vehicle	18	3	4, 7, 13	83.33%
German	20	2	4, 10	90.00%

From table 4, we can see that the performance of IBGSOMFD is compared with the following methods: LDA [42], KPCA [43], IAFSARS [44], FSANRS [45], BGSOFD [46] and CEAFSAMFD [31]. LDA [42] is linear discriminant analysis, projecting high-dimensional pattern samples into the optimal discriminant vector space. It achieves the effect of extracting the dimension of feature space and ensures the optimal separability of the projected pattern samples in the space; KPCA [43] is a kernel principal component analysis, which maps data into a high-dimensional space through the nonlinear mapping. Then the covariance matrix of data matrix is calculated in high dimensional space. After that, the eigenvalues and eigenvectors are obtained. We select a feature vector with a large feature value and multiple features. Finally, it achieves a new data matrix to realize the dimensionality reduction of data features. In IAFSARS [44] and FSANRS [45], feature selection is carried out utilizing rough set. The rough set is regarded as the evaluation criterion of attribute subset. As a supplement, artificial fish swarm algorithm is adopted as the search strategy. In BGSOFD [46], fractal dimension is regarded as the evaluation criterion, and the improved glowworm swarm optimization is used as the search strategy. In CEAFSAMFD [31], the multi-fractal

TABLE 3. Classification accuracy is compared before and after feature selection on UCI datasets.

	Classification accuracy of original dataset (%)			Classification accuracy of optimal dataset (%)		
	Optimal	Average	Run time(s)	Optimal	Average	Run time(s)
Wine	**100**	**89.3511**	1.2043	97.1220	88.0092	**0.1837**
Heart	87.5660	**76.9580**	3.5347	**88.6885**	76.6794	**1.5815**
Wdbc	**99.1736**	**94.6101**	1.9986	98.0803	94.2390	**0.7568**
Forest	**92.1923**	**86.6320**	4.8773	91.9192	86.4964	**1.9342**
Vehicle	**77.8235**	**71.9335**	23.3771	75.4321	69.3938	**6.2818**
German	**80.1328**	**73.6500**	79.4052	78.6070	72.6027	**24.7555**

TABLE 4. Classification comparison analysis with other state-of-the-art approaches.

Datasets	Methods	Optimal subset of attributes	Optimal classification accuracy (%)	Average classification accuracy (%)	Variance	Run time (s)
Wine	LDA	—	97.3500	86.9815	28.5308	0.2134
	KPCA	—	93.4286	78.4603	66.7051	0.2300
	IAFSARS	5,13	93.1111	81.2466	53.7665	0.5359
	FSANRS	4,10	92.2941	84.7783	37.1478	0.4781
	BGSOFD	3, 9	94.4194	80.9148	64.6901	0.8545
	CEAFSAMFD	4, 13	97.4118	86.4046	37.6626	0.5750
	IBGSOMFD	7,10	**97.5220**	**88.0092**	**16.8127**	**0.1837**
Heart	LDA	—	88.3333	76.4392	31.7305	1.6516
	KPCA	—	84.5918	71.5563	30.9860	2.7853
	IAFSARS	1,4,5	87.9245	73.9911	39.5914	1.9845
	FSANRS	1,5,9	87.7778	75.3840	38.7681	1.8292
	BGSOFD	1,4,8	85.7273	72.5539	28.2627	1.6137
	CEAFSAMFD	1,5,8	86.4706	76.1610	41.7094	1.6317
	IBGSOMFD	4,5,8	**88.6885**	**76.6794**	**26.0871**	**1.5815**
Wdbc	LDA	—	99.0219	93.8115	5.1653	0.7625
	KPCA	—	93.8	90.2927	6.4790	0.7717
	IAFSARS	3,4,21	98.0803	92.1729	6.7198	0.9184
	FSANRS	4,6,22	97.1651	90.2046	5.5119	1.0319
	BGSOFD	3, 22	95.0820	91.1655	7.7307	0.8831
	CEAFSAMFD	14, 24	94.4954	93.4054	6.7658	0.7596
	IBGSOMFD	4, 14	**99.1228**	**94.2390**	**5.0509**	**0.7568**
Forest	LDA	—	90.9715	86.3820	8.5579	2.1196
	KPCA	—	91.4444	86.0744	15.7777	2.4874
	IAFSARS	6,10,21,26	84.4660	78.3390	13.8895	2.4140
	FSANRS	1,7,25	73.8636	69.4278	9.6901	2.7556
	BGSOFD	6,7,10	87.2881	80.6030	11.0717	2.3596
	CEAFSAMFD	1, 14, 16	89.5238	84.2615	9.0736	2.7646
	IBGSOMFD	1, 12, 15	**91.9192**	**86.4964**	**8.2563**	**1.9342**
Vehicle	LDA	—	75.0829	69.2715	10.5107	6.9804
	KPCA	—	72.4532	63.1906	10.3420	7.8740
	IAFSARS	5,9,13,15	73.2530	66.6720	10.6824	8.3459
	FSANRS	4,11,15	75.1716	62.1448	18.1274	9.3771
	BGSOFD	4,7,11	**75.8017**	68.5276	10.7542	8.3620
	CEAFSAMFD	3,4,13	73.6364	67.4077	11.4310	7.7157
	IBGSOMFD	4,7,13	75.4321	**69.3938**	**10.2382**	**6.2818**
German	LDA	—	78.5217	72.5554	8.1881	25.7736
	KPCA	—	75.1381	70.6402	8.8394	36.1703
	IAFSARS	1,3,7,10	78.1119	71.6893	11.4253	29.0671
	FSANRS	2,6,7,10	78.1794	72.5437	10.6227	28.8610
	BGSOFD	2,10	77.9744	71.5617	9.7970	25.8872
	CEAFSAMFD	2,10	77.9744	71.5617	9.7970	25.8872
	IBGSOMFD	4,10	**78.6070**	**72.6027**	**7.1971**	**24.7555**

dimension is employed as the evaluation measurement criterion, and artificial fish swarm is served as the search strategy. The parameters of the above six methods are implemented as described in their literature, and the experimental results are the average values of 20 independent runs, as shown in table 4.

In table 4, LDA and KPCA map the data matrix in high-dimensional space to low-dimensional space using linear or nonlinear mapping, so as to achieve the goal of dimensionality reduction. Additionally, the above two methods only extract data features, and cannot conduct attribute selection, so the optimal attribute subset column is illustrated by "—". IBGSOMFD is slightly lower than BGSOFD of the optimal classification accuracy on *Vehicle* dataset. The best classification accuracies on the remaining 5 UCI datasets and the average classification accuracies on the 6 datasets are significantly higher than other methods. It denotes that the classification performance of IBGSOMFD is better than the other 6 methods as a whole. The variance of 6 datasets is better than that of other methods, which indicates that IBGSOMFD has a good stability. The running time for 6 datasets is less than that of other state-of-the-art methods. It shows that the optimal attribute subset selected by IBGSOMFD is more suitable for classification. IBGSOMFD reduces a lot of computational burdens, and improves computational efficiency, so as to achieve efficient prediction. The optimal classification accuracy on the *Vehicle* dataset is slightly lower than that in BGSOFD, because different parameter combinations selected by the grid search method will lead to major differences in classification. Nevertheless, the average classification accuracy can be regarded as the criteria for the overall classification ability assessment. In summary, the classification performance of IBGSOMFD is generally better than that of the other 6 methods, and its stability and significance.

B. EXPERIMENTAL RESULTS OF PREDICTION FOR P2P LENDING INVESTMENT RISK

1) DATASET OF P2P LENDING

Renrendai, Paipaidai and Yilongdai platforms are three of the earliest P2P lending information intermediary service platforms in China, which have been steadily operating since its establishment. They have been ranked in the top internet companies in China. In consequence, we employ the P2P lending datasets of Renrendai, Paipaidai and Yilongdai as the empirical data in practical prediction [47]–[49]. The outliers and flow-marked orders in the Renrendai, Paipaidai and Yilongdai datasets are removed, and then three experimental datasets are obtained. A binary classification is used in this work. From the three platforms, 1000 transaction orders that have completed repayment are randomly selected as samples. Among them, 500 cases are successfully repaid, and 500 loans are in default. The successfully repaid sample is recorded as "+1", and it shows there is no investment risk; the sample in default is marked as "−1", and it indicates there is investment risk. According to the relevant knowledge and research of Internet finance of P2P lending investment risk prediction [50]–[52], excluding irrelevant and severely lost attributes, 17 attributes are selected as model variables in this work. The first 5 features express orders information, including interest rate, loan amounts, repayment period, numbers of investors and payment method. The remaining 12 attributes, called borrower information, are credit rating, age, education background, marriage, income level, historical borrowings, numbers of historical overdue, house property, car property, occupation, scale of company and order status.

TABLE 5. Experimental results of Renrendai dataset.

Targets of solution	Original dataset	Optimal dataset
Number of attributes	17	4
Attribute subset	$H_1, H_2, \ldots, H_{17}$	H_1, H_4, H_{10}, H_{11}
Optimal accuracy (%)	91.3750	91.9255
Average accuracy (%)	85.8332	85.7632
Run time(s)	16.9334	5.8034

TABLE 6. Experimental results of Paipaidai dataset.

Targets of solution	Original dataset	Optimal dataset
Number of attributes	17	4
Attribute subset	$H_1, H_2, \ldots, H_{17}$	H_1, H_4, H_{10}, H_{11}
Optimal accuracy (%)	92.7578	93.1419
Average accuracy (%)	87.3789	87.3181
Run time(s)	15.5993	5.7473

TABLE 7. Experimental results of Yilongdai dataset.

Targets of solution	Original dataset	Optimal dataset
Number of attributes	17	4
Attribute subset	$H_1, H_2, \ldots, H_{17}$	H_1, H_4, H_{10}, H_{11}
Optimal accuracy (%)	92.5549	92.5789
Average accuracy (%)	86.3114	86.3025
Run time(s)	15.1539	5.2469

2) RESULTS ANALYSIS

The experimental results of P2P lending investment risk prediction are illustrated in Table 5 to 7. In the three P2P lending datasets, the features selected by IBGSOMFD are interest rate, number of investors, historical times of borrowing and historical times of default respectively. The selected features are in good agreement with the actual situation: (1) On the premise that market is efficient, the price fully reflects the market information. It is extended to the P2P lending market, and the interest rate can partly reflect the default risk [48]. (2) Psychology of risk aversion exists among investors in P2P lending industry. The higher the risk of order default, the less individual investors choose to invest. In other words, more investors are required to participate in the bidding if the order financing is successful [49]. (3) The historical information of the borrower is an important factor influencing the investment risk. The specific performance is the number of historical borrowings and historical defaults. The higher the repayment rate of historical lending on schedule, the smaller the ratio of historical overdue to historical lending. It denotes that the borrower has conveyed a message of market trust and welcome to investors, with less risk compensation [52]. The reduction rate of the dataset is over 76.47% in this section. IBGSOMFD significantly reduces redundant and irrelevant

TABLE 8. Comparison of the prediction results of 7 methods in Renrendai dataset.

Methods	Optimal attribute subset	Optimal classification accuracy (%)	Average classification accuracy (%)	Variance	Run time (s)
LDA	—	86.7955	81.2554	6.4908	6.5469
KPCA	—	85.6234	79.9182	12.9370	6.3744
IAFSARS	H_1, H_3, H_7, H_{11}, H_{16}	88.1988	81.7183	11.2890	9.8850
FSANRS	H_4, H_7, H_{11}, H_{15}	87.3086	79.4176	12.4599	8.2646
BGSOFD	H_1, H_4, H_7, H_{16}	87.7143	81.7222	7.1882	8.7116
CEAFSAMFD	H_1, H_4, H_7, H_{11}	88.1250	83.8420	6.3777	6.5087
IBGSOMFD	H_1, H_4, H_{10}, H_{11}	**91.9255**	**85.7632**	**3.1868**	**5.8034**

TABLE 9. Comparison of the prediction results of 7 methods in Paipaidai dataset.

Methods	Optimal attribute subset	Optimal classification accuracy (%)	Average classification accuracy (%)	Variance	Run time (s)
LDA	—	86.8235	82.4578	6.2605	6.6724
KPCA	—	87.7166	81.2882	10.1885	6.4792
IAFSARS	H_1, H_3, H_4, H_7, H_{16}	88.7500	83.4865	9.6470	8.3978
FSANRS	H_4, H_7, H_8, H_{12}	88.2716	79.8880	13.6140	7.9087
BGSOFD	H_1, H_4, H_7, H_8	87.7419	83.2900	5.9862	6.5068
CEAFSAMFD	H_1, H_4, H_7, H_{12}	89.1553	85.2575	5.7059	6.3967
IBGSOMFD	H_1, H_4, H_{10}, H_{11}	**93.1419**	**87.3181**	**4.1371**	**5.7473**

TABLE 10. Comparison of the prediction results of 7 methods in Yilongdai dataset.

Methods	Optimal attribute subset	Optimal classification accuracy (%)	Average classification accuracy (%)	Variance	Run time (s)
LDA	—	87.7654	81.9497	6.8019	6.3731
KPCA	—	86.5375	80.0997	10.4682	6.2467
IAFSARS	H_1, H_3, H_7, H_{12}, H_{16}	89.1250	82.7028	10.3544	8.5687
FSANRS	H_4, H_5, H_7, H_{11}	88.1815	79.5713	12.1539	8.2469
BGSOFD	H_1, H_4, H_7, H_{13}	88.5000	81.4122	7.5413	6.7911
CEAFSAMFD	H_1, H_4, H_7, H_{11}	89.2716	83.7040	5.6147	6.4457
IBGSOMFD	H_1, H_4, H_{10}, H_{11}	**92.5789**	**86.3025**	**3.2858**	**5.2469**

attributes and retains key attributes, providing high-quality datasets for the subsequent P2P lending risk prediction.

In table 5 to 7, IBGSO combined with MFD is used to reduce the redundant and irrelevant attributes, and then SVM is applied for subsequent prediction. After that, the predictive performance of the attribute subset is comparable to that of original dataset, and the optimal classification accuracies are slightly improved in Renrendai, Paipaidai and Yilongdai datasets. To a large extent, the redundant and irrelevant attributes of P2P lending datasets are eliminated, which saves running time and predicts the investment risk of P2P lending accurately and efficiently.

Table 8 to 10 show the comparative analysis of the predictive results of 7 methods in Renrendai, Paipaidai and Yilongdai datasets. In table 8 to 10, compared with the other 6 methods, the optimal prediction accuracies of IBGSOMFD are improved by 3-7 percentage points, and the average prediction accuracies increase by 1-7 percentage points. It also denotes that the attributions of P2P lending investment risk selected by IBGSOMFD are more reasonable. The variance and running time of IBGSOMFD are less than the other 6 methods, indicating the stability and effectiveness of IBGSOMFD. In summary, the overall performance of IBGSOMFD is significantly superior to that of the other 6 state-of-the-art methods.

C. PARAMETERS ANALYSIS

In IBGSOMFD, IBGSO is employed as the search strategy. Considering the space limitation, the parameters are analyzed on the representative Renrendai dataset. The parameters includes iteration number T_{max}, population size n, initial decision domain radius $r_d^i(0)$ and maximal decision domain radius r_s, In addition to analyzing T_{max}, the maximal number of iterations for each running of the algorithm is 50.

Figure 7 indicates the relationship between objective function value and iteration, IBGSO is compared with IDGSO [52], DGSO [53] and BGSO [54]. Furthermore, the difference of MFD between the selected attribute subset and the original Renrendai dataset is regarded as the objective function. The smaller the function value, the better the result. In Figure 7, with the increase of iterations, the objective function value of the 4 algorithms decreases continuously. When the number of iterations reaches a certain value, the performance of the 4 algorithms gradually tends to be stable. Meanwhile,

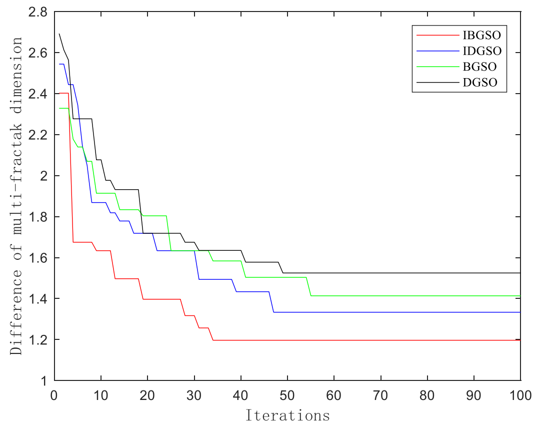

FIGURE 7. Iteration analysis.

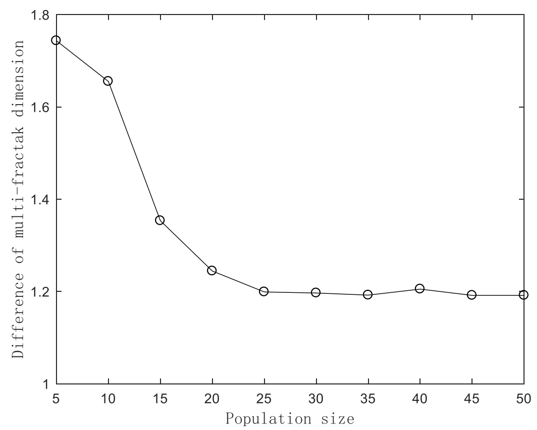

FIGURE 8. Population size analysis.

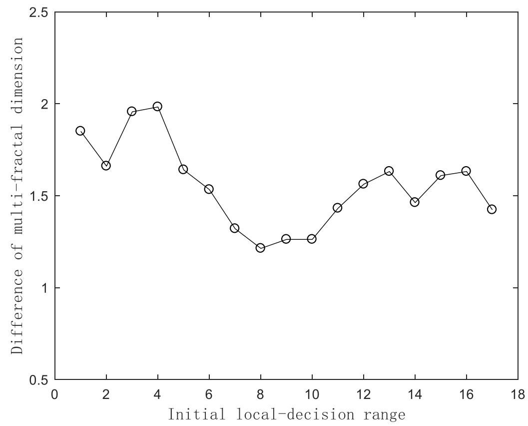

FIGURE 9. Initial local-decision range analysis.

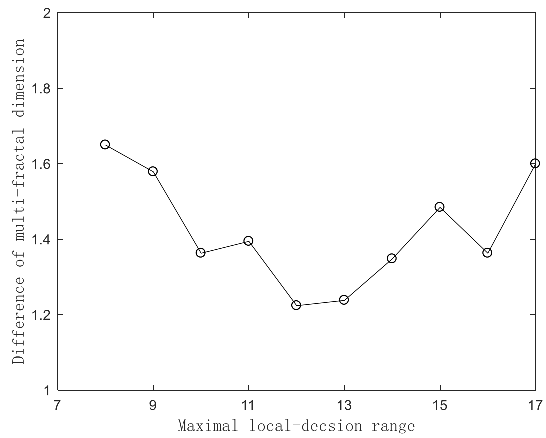

FIGURE 10. Maximal local-decision range analysis.

the convergence speed and accuracy of IBGSO are markedly better than that of IDGSO, BGSO and DGSO. When the maximum number of iterations reaches 40, performance tends to be stable, and the maximal number of iteration is suggested to be set at 40.

In Figure 8, with the increasing of population size, the objective function value of IBGSO keeps decreasing, and the performance of IBGSO keeps improving. When the glowworms population size reaches 25, the performance of IBGSO tends to be stable. If the population continues to increase, the performance of IBGSO can be improved very little and the computational complexity will increase greatly. Therefore, the population size is advised to be set at 25.

Figure 9 denotes the influence of the initial local-decision range $r_d^i(0)$ on the performance of IBGSO. If $r_d^i(0)$ is too small, the less number of glowworms in the decision domain at the initial stage is achieved. It is difficult for glowworms to find objective glowworms with better performance, which affects the global convergence speed of IBGSO. If $r_d^i(0)$ is too large, most glowworms will gather together, and it is easy to fall into a local optimal solution. Since there are 17 attributes in the Renrendai dataset, the maximal distance between any two solutions is 17, and the initial local-decision range varies within [1, 17]. In the early stage of IBGSO, the overall performance of IBGSO indicates an upward trend with the increase of $r_d^i(0)$.When $r_d^i(0)$ is 8, the performance of IBGSO is the best, and then decreases with the increase of $r_d^i(0)$. Due to a larger local-decision range is adopted at early stage of IBGSO, the number of glowworms in the decision domain will increase. The individuals move towards the optimal glowworm rapidly, which improves its convergence speed. After $r_d^i(0)$ reaches 8, $r_d^i(0)$ is increased continuously. It will bring glowworms gather together, and affect the convergence accuracy. So the initial local-decision ranges should be set at 8.

It is clearly observed from Figure 10 that, the influence of the maximal local-decision range r_s on the performance of IBGSO. r_s must be greater than or equal to $r_d^i(0)$, so the variation range of r_s is [8, 17]. Considering the density of glowworms in the solution space is constantly changing, if the individual density of glowworms is low, r_s should be increased. If the density is high, r_s needs to be decreased. However, it is necessary to set proper r_s to prevent the radius of the decision domain from being too large and avoid falling into a local optimum. According to Figure 10, when the change range is [8, 12], the objective function values keep decreasing, and the performance of IBGSO keeps improving.

When the variation range is [12, 17], the performance of IBGSO tends to decline. In Figure 10, when the change interval is [8, 12], the objective function values keep decreasing, and the performance of IBGSO keeps improving. When the change interval is [12, 17], the algorithm performance shows a downward trend. When r_s is 12, IBGSO performs at its best. Therefore, r_s is set at 12.

V. CONCLUSION

To predict the investment risk of P2P lending accurately and efficiently, IBGSOMFD is proposed by combining IBGSO and MFD. GSO is regarded as a search strategy, and MFD is employed as an evaluation criterion. The redundant and irrelevant attributes of P2P lending datasets are reduced using IBGSOMFD, so as to realize the improving of the computational efficiency of subsequent prediction. SVM is used to construct the prediction model, and the accurate and efficient predictive results of P2P lending investment risk are achieved. Experimental results on 6 UCI benchmark datasets demonstrate that the validity and effectiveness of IBGSOMFD, then it is applied in the real P2P lending datasets of Renrendai, Paipaidai and Yilongdai platforms in China, and good predictive results are attained. In addition, the proposed IBGSO 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 SVMs to predict the investment risk of P2P lending. We believe that the promising results can be obtained, which can provide new research ideas for the investment risk prediction of P2P lending.

REFERENCES

[1] X. Fu, S. Zhang, J. Chen, T. Ouyang, and J. Wu, "A sentiment-aware trading volume prediction model for P2P market using LSTM," *IEEE Access*, vol. 7, pp. 81934–81944, 2019.

[2] Z. Geng, D. Shang, Y. Han, and Y. Zhong, "Early warning modeling and analysis based on a deep radial basis function neural network integrating an analytic hierarchy process: A case study for food safety," *Food Control*, vol. 96, pp. 329–342, Feb. 2019.

[3] C. Xu, S. Zhao, and F. Liu, "Sensor fault detection and diagnosis in the presence of outliers," *Neurocomputing*, vol. 349, pp. 156–163, Jul. 2019.

[4] D. Jiang, L. Huo, Z. Lv, H. Song, and W. Qin, "A joint multi-criteria utility-based network selection approach for vehicle-to-infrastructure networking," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3305–3319, Oct. 2018.

[5] D. Jiang, W. Wang, L. Shi, and H. Song, "A compressive sensing-based approach to end-to-end network traffic reconstruction," *IEEE Trans. Netw. Sci. Eng.*, to be published.

[6] D. Jiang, L. Huo, and Y. Li, "Fine-granularity inference and estimations to network traffic for SDN," *PLoS ONE*, vol. 13, no. 5, May 2018, Art. no. e0194302.

[7] J. Hu, L. Pan, Y. Yang, and H. Chen, "A group medical diagnosis model based on intuitionistic fuzzy soft sets," *Appl. Soft Comput.*, vol. 77, pp. 453–466, Apr. 2019.

[8] J. H. Cheon, D. Kim, Y. Kim, and Y. Song, "Ensemble method for privacy-preserving logistic regression based on homomorphic encryption," *IEEE Access*, vol. 6, pp. 46938–46948, 2018.

[9] B. Choubin, E. Moradi, M. Golshan, J. Adamowski, F. Sajedi-Hosseini, and A. Mosavi, "An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines," *Sci. Total Environ.*, vol. 651, pp. 2087–2096, Feb. 2019.

[10] W. Cai, L. Dou, M. Zhang, W. Cao, J.-Q. Shi, and L. Feng, "A fuzzy comprehensive evaluation methodology for rock burst forecasting using microseismic monitoring," *Tunnelling Underground Space Technol.*, vol. 80, pp. 232–245, Oct. 2018.

[11] S. Zhang, Z. Liu, and W. Xiao, "A hierarchical extreme learning machine algorithm for advertisement click-through rate prediction," *IEEE Access*, vol. 6, pp. 50641–50647, 2018.

[12] C.-X. Zhang, S. Xu, and J.-S. Zhang, "A novel variational Bayesian method for variable selection in logistic regression models," *Comput. Statist. Data Anal.*, vol. 133, pp. 1–19, May 2019.

[13] R. E. Frank, W. F. Massy, and D. G. Morrison, "Bias in multiple discriminant analysis," *J. Marketing Res.*, vol. 2, no. 3, p. 250, Aug. 1965.

[14] Y.-W. Du, S.-S. Wang, and Y.-M. Wang, "Group fuzzy comprehensive evaluation method under ignorance," *Expert Syst. with Appl.*, vol. 126, pp. 92–111, Jul. 2019.

[15] X. Ma, X. Chen, and X. Zhang, "Non-interactive privacy-preserving neural network prediction," *Inf. Sci.*, vol. 481, pp. 507–519, May 2019.

[16] B. S. Raghuwanshi and S. Shukla, "Generalized class-specific kernelized extreme learning machine for multiclass imbalanced learning," *Expert Syst. Appl.*, vol. 121, pp. 244–255, May 2019.

[17] S. Benítez-PeÑa, R. Blanquero, E. Carrizosa, and P. Ramírez-Cobo, "Cost-sensitive Feature Selection for Support Vector Machines," *Comput. Oper. Res.*, vol. 106, pp. 169–178, Jun. 2019.

[18] X. Wang and Y. Xing, "An online support vector machine for the open-ended environment," *Expert Syst. Appl.*, vol. 120, pp. 72–86, Apr. 2019.

[19] X.-J. Lu, T.-T. Hu, Y. Zhang, and B. Fan, "Robust clustered support vector machine with applications to modeling of practical processes," *IEEE Access*, vol. 6, pp. 75143–75154, 2018.

[20] C. Traina, Jr., A. Traina, L. Wu, and C. Faloutsos, "Fast feature selection using fractal dimension," *J. Inf. Data Manage.*, vol. 1, no. 1, pp. 158–171, 2010.

[21] D. Jiang, Y. Wang, Z. Lv, S. Qi, and S. Singh, "Big data analysis based network behavior insight of cellular networks for industry 4.0 applications," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 1310–1320, Feb. 2020, doi: 10.1109/tii.2019.2930226.

[22] D. Jiang, L. Huo, and H. Song, "Rethinking behaviors and activities of base stations in mobile cellular networks based on big data analysis," *IEEE Trans. Netw. Sci. Eng.*, to be published.

[23] J. Liu, Y. Lin, M. Lin, S. Wu, and J. Zhang, "Feature selection based on quality of information," *Neurocomputing*, vol. 225, pp. 11–22, Feb. 2017.

[24] A. Ferone, "Feature selection based on composition of rough sets induced by feature granulation," *Int. J. Approx. Reasoning*, vol. 101, pp. 276–292, Oct. 2018.

[25] K. Mukherjee, J. K. Ghosh, and R. C. Mittal, "Variogram fractal dimension based features for hyperspectral data dimensionality reduction," *J. Indian Soc. Remote Sens.*, vol. 41, no. 2, pp. 249–258, Jun. 2013.

[26] S. Jadhav, H. He, and K. Jenkins, "Information gain directed genetic algorithm wrapper feature selection for credit rating," *Appl. Soft Comput..*, vol. 69, pp. 541–553, Aug. 2018.

[27] H. Dong, T. Li, R. Ding, and J. Sun, "A novel hybrid genetic algorithm with granular information for feature selection and optimization," *Appl. Soft Comput..*, vol. 65, pp. 33–46, Apr. 2018.

[28] S. Tabakhi and P. Moradi, "Relevance–redundancy feature selection based on ant colony optimization," *Pattern Recognit.*, vol. 48, no. 9, pp. 2798–2811, Sep. 2015.

[29] Y.-Y. Wang, W.-X. Peng, C.-H. Qiu, J. Jiang, and S.-R. Xia, "Fractional-order Darwinian PSO-based feature selection for media-adventitia border detection in intravascular ultrasound images," *Ultrasonics*, vol. 92, pp. 1–7, Feb. 2019.

[30] L. Zou, H. Li, W. Jiang, and X. Yang, "An improved fish swarm algorithm for neighborhood rough set reduction and its application," *IEEE Access*, vol. 7, pp. 90277–90288, 2019.

[31] X. H. Zhu, Z. W. Ni, M. Y. Chen, J. M. Li, F. F. Jin, and L. P. Ni, "Haze prediction method based on multi-fractal dimension and co-evolution discrete artificial fish swarm algorithm," *Syst. Eng.-Theory Pract.*, vol. 37, no. 4, pp. 999–1010, 2017.

[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," *Appl. Soft Comput.*, vol. 58, pp. 104–114, Sep. 2017.

[33] Z. W. Ni, C. Zhang, and L. P. Ni, "Haze forecast method of selective ensemble based on glowworm swarm optimization algorithm," *Pattern Recognit. Artif. Intell.*, vol. 29, no. 2, pp. 143–153, 2016.

[34] Z. W. Ni, H. W. Xiao, Z. J. Wu, and Y. J. Xue, "Attribute selection method based on improved discrete glowworm swarm optimization and fractal dimension," *Pattern Recognit. Artif. Intell.*, vol. 26, no. 12, pp. 1169–1178, 2013.
[35] J. M. Li, Z. W. Ni, X. H. Zhu, and Y. Yu, "Drought prediction model based on GPSGSO-BPNN parallel ensemble learning algorithm," *Syst. Eng.-Theory Pract.*, vol. 38, no. 5, pp. 1343–1353, 2018.
[36] H. Y. Sang and Q. K. Pan, "A discrete invasive weed optimization algorithm for the integrated lot-streaming flow shop scheduling problem," *Control Theory Appl.*, vol. 32, no. 2, pp. 246–250, 2015.
[37] B. B. Mandelbrot, "The fractal geometry of nature," *Amer. J. Phys.*, vol. 51, no. 3, pp. 286–2879, 1983.
[38] D. Jiang, P. Zhang, Z. Lv, and H. Song, "Energy-efficient multi-constraint routing algorithm with load balancing for Smart City applications," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1437–1447, Dec. 2016.
[39] D. Jiang, W. Li, and H. Lv, "An energy-efficient cooperative multicast routing in multi-hop wireless networks for smart medical applications," *Neurocomputing*, vol. 220, pp. 160–169, Jan. 2017.
[40] F. Wang, D. Jiang, and S. Qi, "An adaptive routing algorithm for integrated information networks," *China Commun.*, vol. 16, no. 7, pp. 195–206, 2019.
[41] C. Zhang, Z. Ni, L. Ni, and N. Tang, "Feature selection method based on multi-fractal dimension and harmony search algorithm and its application," *Int. J. Syst. Sci.*, vol. 47, no. 14, pp. 3476–3486, Oct. 2016.
[42] H. Kim, B. L. Drake, and H. Park, "Multiclass classifiers based on dimension reduction with generalized LDA," *Pattern Recognit.*, vol. 40, no. 11, pp. 2939–2945, Nov. 2007.
[43] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," *Appl. Soft Comput.*, vol. 18, pp. 178–184, May 2014.
[44] 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, Jan. 2016.
[45] Y. Chen, Z. Zeng, and J. Lu, "Neighborhood rough set reduction with fish swarm algorithm," *Soft Comput.*, vol. 21, no. 23, pp. 6907–6918, Dec. 2017.
[46] J. M. Li, Z. W. Ni, Y. Xu, and X. H. Zhu, "Research on attribute selection method based on binary glowworm swarm optimization algorithm," *J. Syst. Sci. Math. Sci.*, vol. 37, no. 2, pp. 407–424, 2017.
[47] J. Y. Hu and W. S. Song, "The rational sense and balanced behavior of P2P investors: Empirical analysis based on transaction data from renrendai.Com," *J. Financial Res.*, vol. 7, pp. 86–104, 2017.
[48] L. Liao, M. R. Li, and Z. W. Wang, "The intelligent investor: Not-fully-marketized interest rate and risk identify-evidence from P2P lending," *Econ. Res. J.*, vol. 49, no. 7, pp. 125–137, 2014.
[49] L. Liao, J. Xiang, and Z. W. Wang, "Wisdom of crowds from P2P lending investors," *Chin. J. Manage. Sci.*, vol. 26, no. 10, pp. 30–40, 2018.
[50] 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," *Decis. Support Syst.*, vol. 89, pp. 113–122, Sep. 2016.
[51] Y. Guo, W. Zhou, C. Luo, C. Liu, and H. Xiong, "Instance-based credit risk assessment for investment decisions in P2P lending," *Eur. J. Oper. Res.*, vol. 249, no. 2, pp. 417–426, Mar. 2016.
[52] 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," *Electron. Commerce Res. Appl.*, vol. 11, no. 5, pp. 469–483, Sep. 2012.
[53] Y. Zhou, Z. Huang, and H. Liu, "Discrete glowworm swarm optimization algorithm for TSP problem," *Acta Electronica Sinica*, vol. 40, no. 6, pp. 1164–1170, 2012.
[54] M. Li, X. Wang, Y. Gong, Y. Liu, and C. Jiang, "Binary glowworm swarm optimization for unit commitment," *J. Mod. Power Syst. Clean Energy*, vol. 2, no. 4, pp. 357–365, Dec. 2014.

PINGFAN XIA was born in Wuhu, China, in 1994. She received the M.S. degree in financial engineering from the Anhui University of Finance and Economics, in 2018. She is currently pursuing the Ph.D. degree in management science and engineering with the Hefei University of Technology. Her research interests include machine learning, evolution computation, and internet finance.

ZHIWEI NI was born in Anqing, China, in 1963. He received the B.E. and M.S. degrees in computer software and theory from Anhui University, and the Ph.D. degree from the University of Science and Technology of China, in June 2002. Since 2002, he has become a Professor and Ph.D. Supervisor with the Hefei University of Technology, where he has presided over or participated in more than 20 national and provincial projects. From 2010 to 2019, he has served as the Director of the Intelligent Management Institute, Hefei University of Technology, where he has authored three books and more than 40 articles. His research interests include artificial intelligence, machine learning, and cloud computing.

XUHUI ZHU was born in Fuyang, China, in 1991. He received the B.E. degree from the School of Mathematics, Hefei University of Technology, and the Ph.D. degree in management science and engineering from the Hefei University of Technology. Since 2018, he has become a Lecturer with the Hefei University of Technology, where he has authored more than ten articles. His research interests include evolution computation and machine learning.

JIAPENG ZHANG was born in Bozhou, China, in 1994. He received the B.S. degree in electronic commerce from the Hefei University of Technology, in 2017, where he is currently pursuing the master's degree in management science and engineering. His research interests include artificial intelligence and data mining.

YUANYUAN JIN was born in Wuhu, China, in 1998. She received the B.E. degree in management science from Anhui University, in 2019. She is currently pursuing the master's degree in management science and engineering with the Hefei University of Technology. Her research interests include internet finance and data mining.