Research Article

An Improved Grey Wolf Optimization Strategy Enhanced SVM and Its Application in Predicting the Second Major

Yan Wei,[1] Ni Ni,[2] Dayou Liu,[3,4] Huiling Chen,[5] Mingjing Wang,[5] Qiang Li,[5] Xiaojun Cui,[1] and Haipeng Ye[1]

[1] *Wenzhou Vocational College of Science and Technology, Wenzhou, Zhejiang 325006, China*
[2] *Beijing Entry-Exit Inspection and Quarantine Bureau, Beijing 100026, China*
[3] *College of Computer Science and Technology, Jilin University, Changchun 130012, China*
[4] *Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China*
[5] *College of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou 325035, China*

Correspondence should be addressed to Huiling Chen; chenhuiling.jlu@gmail.com

Received 27 September 2016; Accepted 15 January 2017; Published 20 February 2017

Academic Editor: Dylan F. Jones

In order to develop a new and effective prediction system, the full potential of support vector machine (SVM) was explored by using an improved grey wolf optimization (GWO) strategy in this study. An improved GWO, IGWO, was first proposed to identify the most discriminative features for major prediction. In the proposed approach, particle swarm optimization (PSO) was firstly adopted to generate the diversified initial positions, and then GWO was used to update the current positions of population in the discrete searching space, thus getting the optimal feature subset for the better classification purpose based on SVM. The resultant methodology, IGWO-SVM, is rigorously examined based on the real-life data which includes a series of factors that influence the students' final decision to choose the specific major. To validate the proposed method, other metaheuristic based SVM methods including GWO based SVM, genetic algorithm based SVM, and particle swarm optimization-based SVM were used for comparison in terms of classification accuracy, AUC (the area under the receiver operating characteristic (ROC) curve), sensitivity, and specificity. The experimental results demonstrate that the proposed approach can be regarded as a promising success with the excellent classification accuracy, AUC, sensitivity, and specificity of 87.36%, 0.8735, 85.37%, and 89.33%, respectively. Promisingly, the proposed methodology might serve as a new candidate of powerful tools for second major selection.

1. Introduction

At present, most colleges and universities attach great importance to the needs of students' diversified development and provide the opportunity of categorization of professional direction for students in the high grade. Therefore, the fact that students choose the suitable major according to their own characteristics is of significant importance, which is the important prerequisite for promoting the career development in the future. However, students are apt to getting lost when they are faced with choosing the major because of a series of factors such as subjective consciousness, randomness, and blindness. Is there scientific and reasonable means to allow students to understand their own characteristics and to find the most suitable major development direction for their own? Colleges have accumulated a large number of data related to student resources during the personnel training process, and a lot of important information and knowledge is hidden in these massive data. We can extract the valuable information from these massive data by using data mining technology to construct a prediction model appropriate for major classification; therefore it can help the students to make the right decision on choosing the major.

So far, the data mining techniques in the instruction and evaluation, behavior analysis of teachers and students, scores analysis and prediction, occupation guidance, and other

aspects have more applications, which also have proposed several methods to apply to course selection, major selection, and so on. Cheng et al. [1] proposed an innovative approach that combines a student concept model and the change mining mechanism for analyzing the learning problems of students from their historical assessment data. The experimental results showed that those analysis results provided by the innovative approach were helpful to the teachers in providing appropriate instructional assistance and remedial learning materials for improving the learning achievements of the students. Elbadrawy et al. [2] proposed to predict the students' performance according to a recommendation system based on personalized analysis. The results showed that the method of multiple regression and improved matrix decomposition could be more timely and accurate to predict the students' scores in the next term compared with the traditional methods. Ognjanovic et al. [3] proposed an approach for extracting student preferences from sources available in institutional student information systems. The extracted preferences were analyzed using the analytical hierarchy process, which was used for predicting students' course selection. The results demonstrated that the accuracy was high and equivalent to that of previous data mining approaches using fully identifiable data. Campagni et al. [4] presented a data mining methodology based on clustering and sequential patterns techniques to analyze the careers of university graduated students. The results underlined that the more the students follow the order given by the ideal career, the more they get good performance in terms of graduation time and final grade. Kardan et al. [5] proposed to use neural networks to establish models for analyzing and predicting college students' network selection. Experimental results showed that the model had higher prediction accuracy than Support Vector Regression, K-Nearest Neighborhood, and Decision Tree. Thammasiri et al. [6] discussed the problem of unbalanced distribution and supported the model of vector machine and oversampling data equalization technology, which were used to predict the loss rate of new students. The results showed that this kind of data mining technology would achieve the best classification, making the overall prediction accuracy over 90%. Lee [7] performed logical regression analysis and found that a large number of computer courses learned at the middle school level had a significant impact on STEM subject selection in the US colleges. Huang and Xu [8] analyzed the evaluation theories of psychology and statistics and developed a second major selection system based on a rough set-based association rule mining algorithm. The simulation results demonstrated that their method was accurate for software engineering, networking, and programming majors.

To improve the performance of second major selection, this study proposes an improved support vector machine (SVM) based prediction system. In the proposed approach, an enhanced grey wolf optimization strategy (hereafter IGWO) was established to screen the representative features, and then SVM was employed to perform the prediction task based on the feature subset identified by the IGWO strategy. GWO is a new swarm intelligence method proposed recently by Mirjalili et al. [9]. Due to its great exploration capacity, it has been successfully applied to many practical problems,

such as load frequency control of interconnected power system [10], combined heat and power dispatch [11], design of castellated beams [12], and the two-stage assembly flow shop scheduling problem [13]. However, the initial population of original GWO is generated randomly, which may lead to the fact that the grey wolves in search space are lacking diversity. Many studies [14–18] have shown that population initialization may affect the global convergence speed and also the quality of the final solution for swarm intelligence optimization. Moreover, the initial population with good diversity is helpful to promote the performance of optimization algorithm. Inspired by this idea, we use particle swarm optimization (PSO) to generate a diverse initial population and then construct a binary version of GWO to execute feature selection task. One main reason that we have chosen the PSO strategy lies in that the PSO approach is a simple approach with fast search speed and high efficiency compared with others. As shown, the experimental results have shown that PSO have indeed improved the quality of initial population for GWO. At the same time, it is also important to choose an effective and efficient classifier for evaluating the most discriminative features. In this study, the SVM classifier is used to compute the fitness value due to its good generalization capability and excellent performance in many classification tasks [18–23].

The efficacy of the resultant method, the IGWO-SVM based prediction system, was rigorously compared against SVM without feature selection, GWO based SVM (hereafter GWO-SVM), genetic algorithm based SVM (hereafter GA-SVM), and particle swarm optimization-based SVM (hereafter PSO-SVM) on the real-life dataset collected from Wenzhou Vocational College of Science and Technology. The classifiers were compared with respect to the classification accuracy (ACC), sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC) criterion. The experimental results demonstrated that the proposed IGWO-SVM approach achieved much better performance than other competitive counterparts.

The rest of this paper is organized as follows. Section 2 offers brief background knowledge on SVM and binary version of GWO and PSO. Section 3 presents the detailed implementation of the proposed method. Section 4 describes the experimental design. Section 5 presents the experimental results and discusses the proposed approach. Finally, Section 6 summarizes the conclusions and recommendations for future work.

2. Background Knowledge

2.1. Support Vector Machine (SVM). SVM is a kind of classification algorithm, which is devoted to improving the generalization ability by seeking structural risk minimum of the learning machine. The core idea lies in it is the maximum margin strategy, which can be finally transformed into solving a convex quadratic programming problem. Thanks to the good property, SVM has found its applications in a wide range of fields [19, 20, 23–32].

In a binary classification task, the samples are separated with a hyperplane $w^T x + b = 0$, where w is a d-dimensional coefficient vector that is normal to the hyperplane and b is

the offset from the origin and x are data points. The main task of SVM is to get the results of w and b. In linear case, w can be solved by introducing Lagrangian multipliers. The data points on the maximum border are called support vectors. As a result, the solution of w takes the following form: $w = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$, where n is the number of SVs and y_i are the labels corresponding samples x. After then b can be derived from $y_i(\mathbf{w}^T \mathbf{x}_i + b) - 1 = 0$, where $\mathbf{x}_i$ are SVs. After w and b are determined, the linear discriminant function can be given by

$$g(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b\right). \tag{1}$$

In nonlinear cases, a general idea of kernel trick is introduced. And then the decision function can be expressed as follows:

$$g(\mathbf{x}) = \text{sgn}\left(\sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right). \tag{2}$$

Generally, any positive semidefinite functions that satisfy Mercer's condition can be used as kernel functions [33], such as the polynomial kernel ($K(x, x_i) = ((x^T x_i) + 1)^d$) and the Gaussian kernel ($K(x, x_i) = \exp(-\gamma\|x - x_i\|^2)$).

This section gives a brief description of SVM. For more details, one can refer to [34, 35], which provides a complete description of the SVM theory.

2.2. Binary Grey Wolf Optimization.

Grey wolf optimization (GWO) is a metaheuristic algorithm proposed by Mirjalili et al. [9] in 2014. It mimics the social leadership and hunting behavior of grey wolves in nature. In every iteration of GWO, there are three fittest candidate solutions assumed as alpha, beta, and delta to produce that lead the population toward promising regions of the search space. The rest of grey wolves are named as omega and required to assist alpha, beta, and delta to encircle, hunt, and attack prey, that is, to find better solutions.

In order to mathematically simulate the encircling behavior of grey wolves, the following equations are proposed:

$$\vec{D} = \left|\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)\right|,$$
$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D}, \tag{3}$$

where t indicates the current iteration, $\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}$, $\vec{C} = 2\vec{r}_2$, $\vec{X}_p$ is the position vector of the prey, $\vec{X}$ is the position vector of a grey wolf, $\vec{a}$ is linearly decreased from 2 to 0, and $\vec{r}_1$ and $\vec{r}_2$ are random vectors in $[0, 1]$.

In order to mathematically simulate the hunting behavior of grey wolves, the following equations are proposed:

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right|,$$
$$\vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right|, \tag{4}$$
$$\vec{D}_\delta = \left|\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}\right|,$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha,$$
$$\vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta, \tag{5}$$
$$\vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta,$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}. \tag{6}$$

In this work, a new binary GWO (IGWO) is proposed for the feature selection task. Figure 1 presents a flowchart of the proposed IGWO. In the IGWO, each grey wolf has a flag vector, whose length is equal to the total number of features in the dataset. When the position of a grey wolf was updated by (6), the following equation is used to discrete the position.

$$\text{flag}_{i,j} = \begin{cases} 1 & X_{i,j} > 0.5 \\ 0 & \text{otherwise}, \end{cases} \tag{7}$$

where $X_{i,j}$ indicates the jth position of the ith grey wolf.

2.3. Binary Particle Swarm Optimization (PSO).

Particle swarm optimization (PSO) was first developed by Kennedy and Eberhart [36]. The basic idea of PSO algorithm is to find the optimal solution through collaboration and information sharing among individuals in a group. The advantage of PSO is simple and easy to implement and has no many parameters to adjust. It has been widely used in function optimization, combinatorial optimization, neural network training, fuzzy system control, and other applications [37]. In PSO, each individual is treated as a particle in d-dimensional space, and each particle has a position and velocity. The position vector of the ith particle is represented as $X_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,d})$, and its according velocity is represented as $V_i = (v_{i,1}, v_{i,2}, \ldots, v_{i,d})$. The velocity and position are updated as follows:

$$v_{i,j}^{n+1} = w \times v_{i,j}^n + c_1 \times r_1 \left(p_{i,j}^n - x_{i,j}^n\right) + c_2 \tag{8}$$
$$\times r_2 \left(p_{g,j}^n - x_{i,j}^n\right),$$

$$x_{i,j}^{n+1} = x_{i,j}^n + v_{i,j}^{n+1}, \quad j = 1, 2, \ldots, d, \tag{9}$$

where c_1 and c_2 are acceleration coefficients to better balance the search space between global exploration and local exploitation. In addition, r_1 and r_2 are random numbers in (8) generated uniformly in the range of $[0, 1]$. The velocity $v_{i,j}$ is restricted to the range $[-v_{\max}, v_{\max}]$.

For the purpose of feature selection, the binary PSO introduced by Kennedy and Eberhart [38] was employed. In this version of PSO, a sigmoid function is applied to transform the velocity from continuous space to probability space:

$$\text{sig}(v_{i,j}) = \frac{1}{1 + \exp(-v_{i,j})}, \quad j = 1, 2, \ldots, d. \tag{10}$$

The velocity updating scheme in (8) keeps unchanged except that $x_{i,j}$, $p_{i,j}$, and $p_{g,j} \in \{0, 1\}$ and in order to
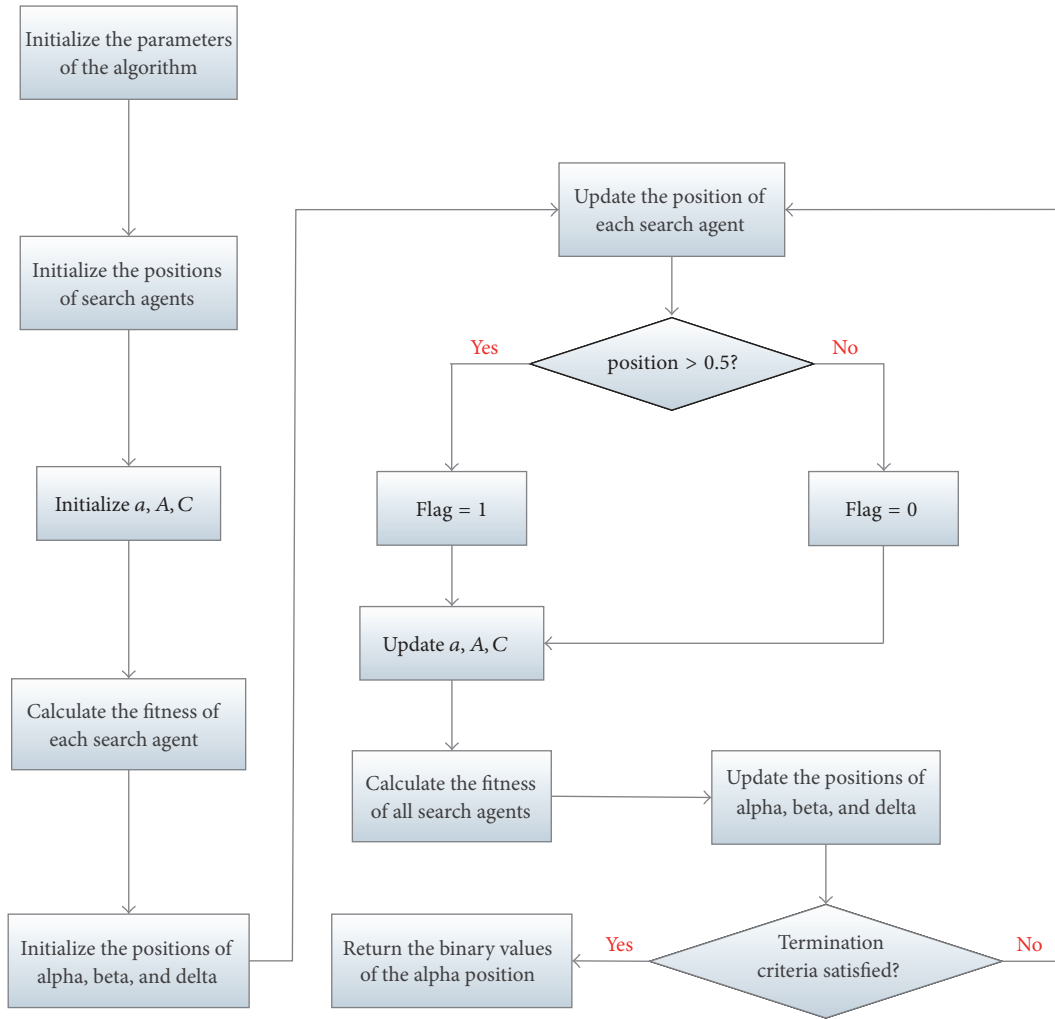
Figure 1: Flowchart of IGWO.

ensure that bit can transfer between 1 and 0 with a positive probability, $v_{\max}$ was introduced to limit $v_{i,j}$. The new particle position is updated using the following rule:

$$x_{i,j}^{n+1} = \begin{cases} 1, & \text{if } 0.5 < \text{sig}\left(v_{i,j}\right), \\ 0, & \text{if } 0.5 \geq \text{sig}\left(v_{i,j}\right), \end{cases} \quad j = 1, 2, \ldots, d, \quad (11)$$

where $\text{sig}(v_{i,j})$ is calculated according to (10).

3. Proposed Methodology

In this section, we briefly describe the proposed system for second major selection. The proposed IGWO-SVM methodology consists of two big steps. In the first step, PSO is firstly used to generate the initial positions of population, and then IGWO is used to select best feature combination by searching for the feature space adaptively. In the second step, the SVM classifier is carried out to predict the classification accuracy based on the optimal feature subset. The flowchart of the proposed IGWO-SVM approach is presented in Figure 2. The IGWO is mainly used to adaptively search the feature space

for best feature combination. The best feature combination is the one with maximum classification accuracy and minimum number of selected features. The fitness function used in IGWO to evaluate the selected features is the average classification accuracy over the 10-fold cross validation scheme. The pseudocode of the feature selection procedure is presented as shown in Algorithm 1.

4. Experimental Designs

4.1. Data Description. The data used for this study was acquired from Wenzhou Vocational College of Science and Technology. The set contained 402 students that majored in Digital Media Technology, which includes graphic design and video production. At the end of the second term, 195 of the students decided on the graphic design major, and 207 students selected the video production. The 11 factors include gender, type of college entrance applications, whether they came from the south of Zhejiang, whether they were science students, whether they volunteered to major in digital media, basic course scores, participation in after-class activities,
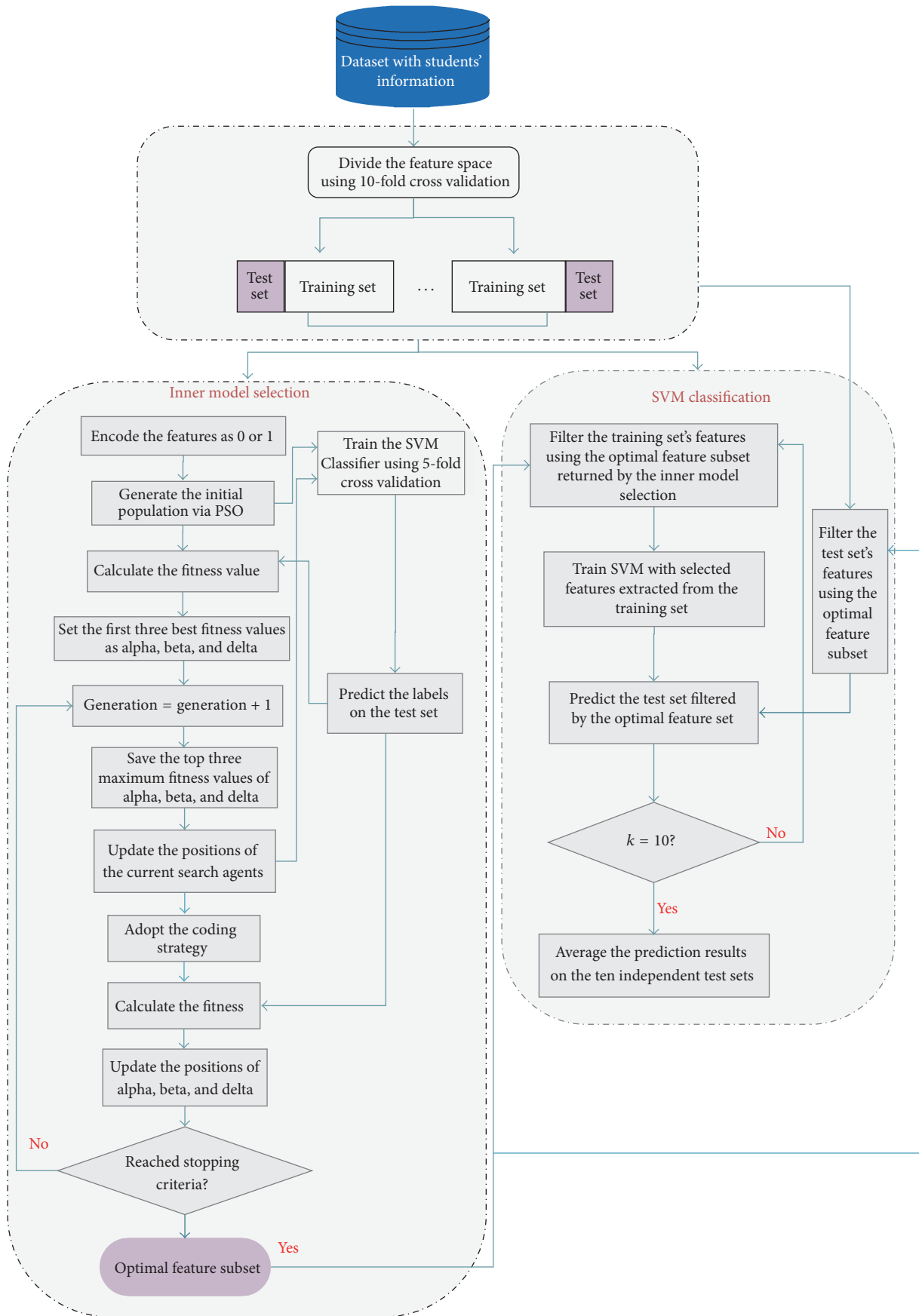
FIGURE 2: Flowchart of the proposed IGWO-SVM based system.

Begin
Initialize the parameters *popsize*, *maxiter*, *n*, *pos*, and *flag* where
popsize: size of population,
maxiter: maximum number of iterations,
n: total number of features,
pos: position of grey wolf,
flag: mark vector of features;
Generate the initial positions of grey wolves using binary PSO;
Initialize a, $\vec{A}$, and $\vec{C}$;
for $i = 1 : popsize$
  for $j = 1 : n$
    if $pos(i, j) > 0.5$
        $flag(j) = 1;$
    else
      $flag(j) = 0;$
    end if
  end for
end for
Calculate the fitness of grey wolves with selected features;
alpha = the grey wolf with the first maximum fitness;
beta = the grey wolf with the second maximum fitness;
delta = the grey wolf with the third maximum fitness;
while $k < maxiter$
  for $i = 1 : popsize$
    Update the position of the current grey wolf;
  end for
  for $i = 1 : popsize$
    for $j = 1 : n$
      if $pos(i, j) > 0.5$
        $flag(j) = 1;$
      else
        $flag(j) = 0;$
      end if
    end for
  end for
  Update a, $\vec{A}$, and $\vec{C}$;
  Calculate the fitness of grey wolves with selected features;
  Update *alpha*, *beta*, and *delta*;
  $k = k + 1;$
end while
Return the selected features of *alpha* as the optimal feature subset;
End

ALGORITHM 1: Pseudocode of the IGWO algorithm.

whether they enrolled in self-study undergraduate courses, and the scores of the basic course relevant to their majors. Table 1 shows a detailed description of these 11 factors.

4.2. Experimental Setup. To verify the proposed approach, several methods including SVM without feature selection, PSO-SVM, GA-SVM, and GWO-SVM were employed for rigorous comparison. For SVM, LIBSVM implementation developed by Chang and Lin [39] was utilized. Data was first scaled into the range $[-1, 1]$ before classification. The empirical experiment was conducted on an AMD Athlon 64 X2 Dual Core Processor 5000+ (2.6 GHz) with 4 GB of RAM, running Windows 7.

In order to gain an unbiased estimate of the generalization accuracy, the k-fold cross validation (CV) was used to evaluate the classification accuracy [40]. This study set k as 10; that is, the dataset is divided into ten subsets. Each time, one of the 10 subsets is used as the test set and the remaining 9 subsets are put together to form a training set. Then the average error across all 10 trials is computed. The advantage of this method is that all of the test sets are independent and the reliability of the results could be improved. It should be pointed out that only one repetition of the 10-fold CV will not generate enough classification accuracies for comparison due to the arbitrariness partition of the dataset. So the 10-fold CV will be repeated and averaged over 10 runs for accurate evaluation.

TABLE 1: Description of the dataset.

Label	Attribute	Description
F_1	Gender	Male or female, denoted by 1 and 2, respectively. There were 233 males and 169 females in the collected data.
F_2	Type of college entrance applications	Students from rural areas applying to college for the first time were denoted as 1, students from urban areas applying to college for the first time were 2, students from rural areas who have applied to college before were 3, and students from urban areas who have applied to college before were 4. In the collected data, there were 317, 74, 6, and 5 students fitting into these 4 categories, respectively.
F_3	Whether they come from the south of Zhejiang	Whether or not they came from the south of Zhejiang including Wenzhou, Taizhou, Lishui, Jinhua, and Quzhou was denoted by 1 and 2, respectively. In the collected data, there were 192 students who came from the south of Zhejiang and 210 students who did not come from the south of Zhejiang.
F_4	Whether they are science students	Whether they were arts or science students was denoted by 1 and 2, respectively. In the collected data, there were 212 arts students and 190 science students.
F_5	Whether they specify in the college entrance application that they want to major in digital media	If yes, 1; otherwise, 2.
F_6	Scores of the basic courses	Average scores of the public basic courses that they took at the freshman year.
F_7	Participation in the after-class graphic design activities	If they participated in multimedia design competitions, design studios, or the design projects of "tutor + project + team." If yes, 1; otherwise, 2.
F_8	Participation in the after-class video production activities	If they participated in microfilm production, visual communication design, (3D) animation design, video production studios, and the video production projects of "tutor + project + team." If yes, 1; otherwise, 2.
F_9	Whether they have enrolled in self-study undergraduate courses	If they enrolled in self-study undergraduate courses at their freshman year, 1; otherwise, 2.
F_{10}	The scores of the basic course relevant to graphic design	Average scores of the graphic design courses, such as image processing and advertisement design.
F_{11}	The scores of the basic course relevant to video production	Average scores of the video production courses, such as 2D animation design and video production basics.

TABLE 2: The settings of parameters.

Parameter	Setting
Population size	8
Iterations	250
Dimensions of problem	11
c_1 in PSO	2.0
c_2 in PSO	2.0
w in PSO	1.0

Considering the two key parameters of SVM, penalty factor C and kernel width γ are both set as $\{2^{-5}, 2^{-4}, \ldots, 2^4, 2^5\}$. By the method of trial and error, when $C = 2^5$ (32) and $\gamma = 2^{-3}$ (0.125), SVM achieved the best performance. Therefore, C and γ for SVM are set to 32 and 0.125 in the experiments. The detailed parameter settings are outlined in Table 2.

4.3. Measure for Performance Evaluation. To evaluate the performance of the second major selection by the IGWO-SVM approach, we mainly examined four metrics: classification accuracy (ACC), the area under the receiver operating characteristic (ROC) curve (AUC) [41], sensitivity (SE), and specificity (SP).

ACC is the proportion of the total number of predictions that were correct. It is determined as follows:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\%. \tag{12}$$

SE is the proportion of positives that were correctly classified, as calculated using the following equation:

$$\text{SE} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%. \tag{13}$$

SP is the proportion of negatives that were correctly classified, as calculated as follows:

$$\text{SP} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%, \tag{14}$$

where TP is the number of true positives, which means that major "A" is correctly classified as the corresponding one; FN is the number of false negatives, which means that major "A" is classified as the major "B"; TN is the number

TABLE 3: The detailed results obtained by SVM.

Fold	ACC	AUC	SE	SP
#1	0.8000	0.7995	0.7895	0.8095
#2	0.8780	0.8774	0.8500	0.9048
#3	0.8537	0.8536	0.8500	0.8571
#4	0.8750	0.8750	0.9000	0.8500
#5	0.8000	0.8000	0.8000	0.8000
#6	0.7250	0.7250	0.7000	0.7500
#7	0.8500	0.8521	0.8947	0.8095
#8	0.8250	0.8233	0.7895	0.8571
#9	0.8000	0.7995	0.7895	0.8095
#10	0.8000	0.8020	0.8421	0.7619
Avg.	0.8207	0.8207	0.8205	0.8210
Dev.	0.0460	0.0460	0.0595	0.0470

Avg. and dev. means the average value and standard deviation of the 10-fold CV results.

of true negatives, which means that major "B" is correctly classified as the corresponding one; and FP is the number of false positives, which means that major "B" is classified as the major "A." The ROC curve is a graphical display that measures the predictive accuracy of a logistic model. The curve displays the true and false positive rate. The AUC is the area under the ROC curve, which is one of the best methods for comparing classifiers in two-class problems.

5. Experimental Results and Discussion

In this experiment, we firstly evaluated the effectiveness of the SVM model on the original feature space. Table 3 lists the detailed results obtained by the SVM model with the whole 11 variables. From the table we can see that SVM have achieved promising classification performance with 82.07% ACC, 0.8207 AUC, 82.05% sensitivity, and 82.10% specificity. In addition, it can be observed that the values of C and γ can be specified adaptively by the grid search strategy for each fold of the data. The range of the related parameters C and γ in the RBF kernel varied between $C = \{2^{-5}, 2^{-3}, \ldots, 2^5\}$ and $\gamma = \{2^{-5}, 2^{-3}, \ldots, 2^5\}$. The optimal parameter pair (C, γ) was employed to construct the predictive model. Figure 3 depicts the contour diagram of the two control parameters obtained by the grid search strategy on the 1st fold data for the SVM classifier. The 3D view of the parameter searching procedure for the 1st fold data can be also shown in Figure 4. From the above figures, we can see that the performance of SVM is sensitive to the two parameters.

In order to further improve the performance of the SVM model, we evaluated the proposed IGWO strategy enhanced SVM approach on the same data. Table 4 lists the detailed classification results obtained by the IGWO-SVM approach. We can see that the IGWO-SVM has achieved a promising result with 87.32% ACC, 0.8722 AUC, 86.58% sensitivity, and 87.86% specificity. Compared with original SVM method, IGWO-SVM has lifted 5.29%, 5.28%, 3.32%, and 7.23% in terms of ACC, AUC, sensitivity, and specificity, respectively. It indicates that there are certain redundant and irrelevant
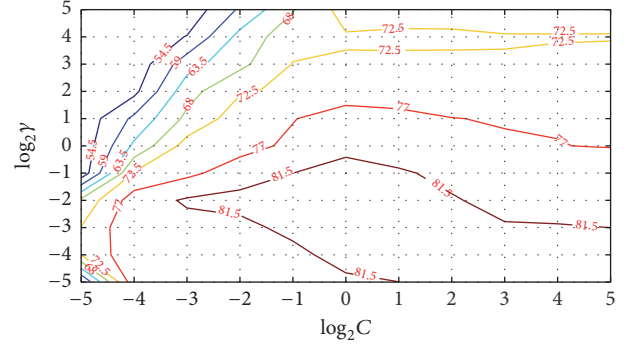

FIGURE 3: Contour diagram obtained by the grid search for the SVM classifier on the 1st fold data.
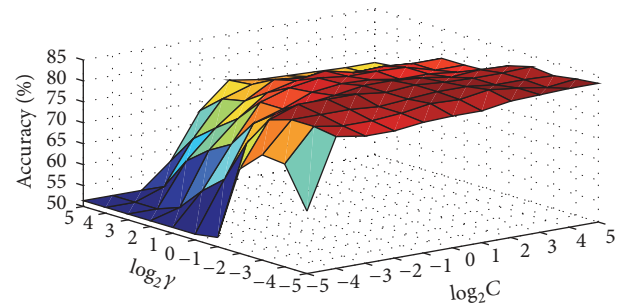

FIGURE 4: Training accuracy surface of SVM with parameters obtained by the grid search for the 1st fold data.

features existing in the data. In addition, it is interesting to find that the standard deviation obtained by the IGWO-SVM is smaller than that of the original SVM, which indicates that the IGWO-SVM can offer more robust and stable prediction results.

With the purpose of verifying the property of the constructed IGWO-SVM, the other metaheuristics based SVM methods including GA-SVM, PSO-SVM, GWO-SVM, and the original SVM for comparison were rationally proposed. The mean ACC and standard deviation obtained by each method via the 10 runs of 10-fold CV are vividly exhibited in Figure 5. It can be seen from this figure that the IGWO-SVM surpasses original SVM, GA-SVM, PSO-SVM, and GWO-SVM in terms of the ACC, AUC, sensitivity, and specificity, which means that the improved GWO based on PSO initialization can improve the performance of model selection of GWO to a certain extent. According to the standard deviation abstracted from Figure 5, a fact can be conducted that the standard deviation of the IGWO-SVM was much smaller than the other four competitors in terms of the four evaluation metrics including ACC, AUC, sensitivity, and specificity. Hence, the SVM constructed based upon IGWO can lead to a more stable classification result for selecting a second major than the others. In short, the IGWO can not only increase the property of SVM in terms of the other competitors but also construct a more stable classifier for selecting a second major. The comparison results indicate that IGWO-SVM is the most stable and robust method for

TABLE 4: The detailed results obtained by IGWO-SVM.

Fold	Selected feature subset	ACC	AUC	SE	SP
#1	$\{F_1, F_3, F_6, F_7, F_8, F_{10}\}$	0.9268	0.9262	0.9000	0.9524
#2	$\{F_1, F_3, F_5, F_6, F_7, F_8, F_{11}\}$	0.8750	0.8759	0.8947	0.8571
#3	$\{F_1, F_2, F_3, F_7, F_8, F_{10}, F_{11}\}$	0.8780	0.8774	0.8500	0.9048
#4	$\{F_1, F_3, F_7, F_8, F_{10}, F_{11}\}$	0.8500	0.8500	0.8000	0.9000
#5	$\{F_1, F_3, F_4, F_7, F_8, F_{10}, F_{11}\}$	0.8750	0.8750	0.8500	0.9000
#6	$\{F_1, F_2, F_5, F_7, F_8\}$	0.8537	0.8536	0.8500	0.8571
#7	$\{F_1, F_3, F_7, F_9, F_{10}, F_{11}\}$	0.9024	0.9024	0.9000	0.9048
#8	$\{F_1, F_3, F_6, F_8, F_{10}\}$	0.8500	0.8500	0.8500	0.8500
#9	$\{F_1, F_3, F_7, F_9, F_{10}, F_{11}\}$	0.8750	0.8750	0.8000	0.9500
#10	$\{F_3, F_7, F_8, F_{10}, F_{11}\}$	0.8500	0.8496	0.8421	0.8571
Avg.	N/A	0.8736	0.8735	0.8537	0.8933
Dev.	N/A	0.0253	0.0252	0.0364	0.0378

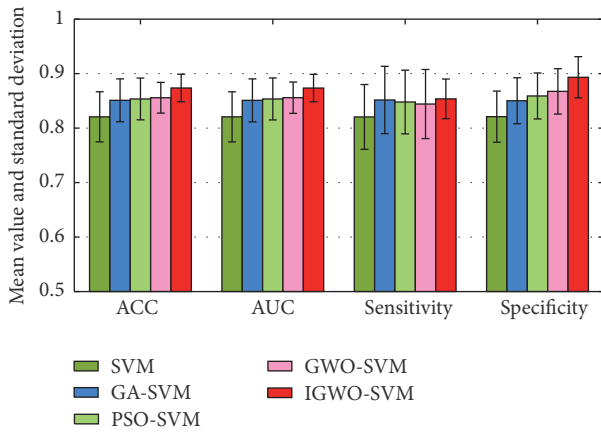Avg. and dev. means the average value and standard deviation of the 10-fold CV results.

FIGURE 5: The classification performance obtained by the five methods in terms of ACC, AUC, sensitivity, and specificity.

FIGURE 6: The average results of best fitness during the training stage in one run of 10-fold CV obtained by the four methods.

second major selection, followed by original SVM, GA-SVM, PSO-SVM, and GWO-SVM.

To explore the optimization procedures of the meta-heuristic optimization methods including IGWO-SVM, GWO-SVM, PSO-SVM, and GA-SVM, we have recorded the evolutionary process of the four methods. Figure 6 vividly expresses the evolutionary process of the involved SVM methods based on various strategies. It can be seen that the SVM model trained via IGWO exhibits an obvious advantage over the other SVM methods in terms of not only convergence rate, but also the quality of solution, and it takes only 22 iterations to achieve the best fitness. However, the GWO-SVM, PSO-SVM, and GA-SVM need 62, 76, and 61 iterations to reach to the maximum fitness, respectively. In addition, it is interesting to find that the curve of IGWO-SVM ended at the 220th iteration since the fitness cannot be improved for 200 consecutive iterations which were preset as the termination condition. Another special fact also can be found that the GA-SVM has a better convergence rate than PSO-SVM, although its maximum fitness is smaller than PSO-SVM,
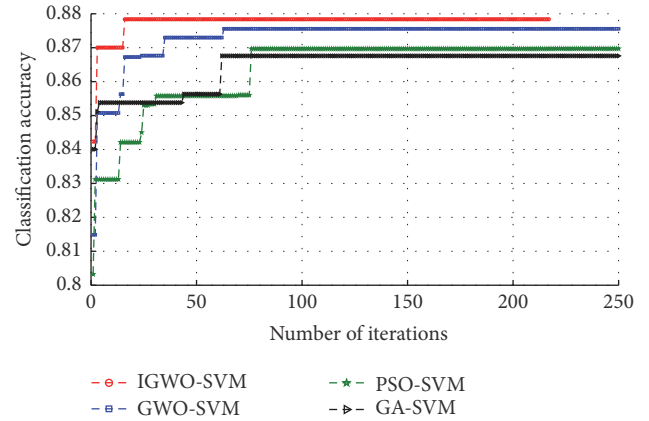
which means that GA has superior performance than PSO in terms of the convergence rate. In short, the IGWO can significantly improve not only the convergence rate but also the SVM's classifier fitness compared to the original GWO, PSO, and GA. The main reason may be the good initialization population produced by the PSO algorithm.

Table 5 presents the frequencies of the selected features by the GA-SVM, PSO-SVM, GWO-SVM, and IGWO-SVM within the 10-fold CV procedure. It is interesting to find out that the average frequencies of six selected features including the 1st feature (F_1), the 3rd feature (F_3), the 7th feature (F_7), the 8th feature (F_8), the 10th feature (F_{10}), and the 11th feature (F_{11}) are more than five by the four methods. It indicates that the four methods are highly consistent to pick out the most important features. It also suggests that these features should be paid more attention in the decision-making process. The frequency of each feature in the course of the feature selection by the four methods can be vividly shown in Figure 7. It can be observed from the figure that the frequencies of F_1, F_3, F_7, F_8, F_{10}, and F_{11} selected by the GWO-SVM and IGWO-SVM are higher than the other two counterparts,

TABLE 5: Average frequencies of the selected features by the four methods.

Feature	Average selected frequencies			
	GA-SVM	PSO-SVM	GWO-SVM	IGWO-SVM
F_1	8	8	9	9
F_2	4	3	3	2
F_3	8	9	9	10
F_4	3	2	2	1
F_5	2	2	1	1
F_6	5	5	4	3
F_7	8	8	8	9
F_8	7	8	8	8
F_9	3	3	2	2
F_{10}	6	7	7	8
F_{11}	6	6	7	7
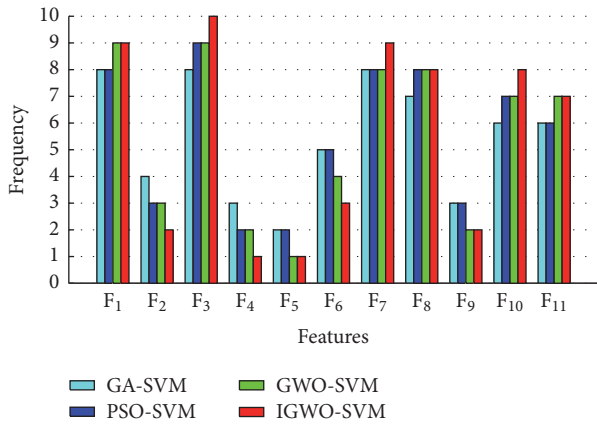
FIGURE 7: The frequency of each feature chosen by IGWO-SVM over the 10-fold CV procedure.

while F_2, F_4, F_5, F_6, and F_9 chosen by the GWO-SVM and IGWO-SVM are lower than the other two counterparts. And between the GWO-SVM and IGWO-SVM methods, IGWO-SVM performs even better than the other one. From the table, we can see that the frequencies of the most important features chosen by the IGWO-SVM approach are F_3, F_1, F_7, F_8, and F_{10} ranked from highest to lowest. On the one hand, it indicates that GWO has better capability to remove the redundant and relevant features from the data. On the other hand, it also suggests that the IGWO strategy can pick out the most discriminative features from the dataset compared to the other three comparative counterparts.

From the above analysis, we can find that the most important features include whether they come from the south of Zhejiang (F_3), gender (F_1), third classroom which includes participation in the after-class graphic design activities (F_7) and participation in the after-class video production activities (F_8), and the basic course result of the classified major which includes the scores of the basic course relevant to graphic design (F_{10}) and the scores of the basic course relevant to video production (F_{11}); the influence of these characteristics on the choice of classified major is relatively prominent. Previous data show that the impact of whether they come from the south of Zhejiang on the choice of classified major is prominent, because most students return to hometown for work after graduation, and they will take into account the resources of the employment according to the classified major, such as Wenzhou, Taizhou, Lishui, Jinhua, and Quzhou where the electronic commerce and printing industry are prosperous and the students are more inclined to choose graphic design. In Hangzhou, Shaoxing, Jiaxing, Huzhou, Zhoushan, Ningbo, and other places where the film and animation industry are prosperous, the students are more apt to choose video production. There are great differences in the major adaptability between male and female students, which always results in the differences in the process of employment. Under the influence of many factors such as needing more effort on the shooting, the weight of film, and television equipment being relatively heavy, when the students who major in video production are faced with selecting the second major, the girls are less inclined to choose video production, while the boys have stronger adaptability. The third classes mainly include the students who participate in creative studio, professional skills competition, and student projects. Differing from the students' subjective understanding of their own interest, these cases from the participation of students objectively can reflect students' major interest. Thus it is also an important factor when the students are faced with selecting the second major. The scores of basic courses associated with the specialized direction also have big impact on the selection of the second major. The scores of basic courses can objectively reflect the students' major knowledge, so it can better guide the students to choose the most suitable major according to their own knowledge structure. According to the previous students who experience choosing the second major, we can know that the students who have achieved higher scores for basic courses such as graphic design and the basis of digital graphic form are more inclined to choose the major of graphic design, while the students who achieve high scores for basic courses such as basic video and basic photography are more inclined to choose the major of video production.

6. Conclusions and Future Work

In this study, we have examined an improved SVM for effectively predicting the students' major selection. To exploit the maximum potential of SVM, an improved GWO strategy was established to search for the optimal feature subset for classification. The experimental results have demonstrated that the developed approach has achieved more superior classification performance over other advanced machine learning approaches in terms of the ACC, AUC, sensitivity, and specificity. Therefore, it can be safely concluded that the developed intelligent system can serve as a promising alternative decision support system for students' second major selection. In the future work, we plan to implement our approach in a parallel manner to further improve the computational cost. In addition, collecting more data samples to ameliorate the prediction performance of the proposed system is the other work that should be done in the future.

Competing Interests

Acknowledgments

References

[1] L.-C. Cheng, H.-C. Chu, and B.-M. Shiue, "An innovative approach for assisting teachers in improving instructional strategies via analyzing historical assessment data of students," *International Journal of Distance Education Technologies*, vol. 13, no. 4, pp. 40–61, 2015.

[2] A. Elbadrawy, A. Polyzou, Z. Ren, M. Sweeney, G. Karypis, and H. Rangwala, "Predicting student performance using personalized analytics," *Computer*, vol. 49, no. 4, Article ID 7452320, pp. 61–69, 2016.

[3] I. Ognjanovic, D. Gasevic, and S. Dawson, "Using institutional data to predict student course selections in higher education," *Internet and Higher Education*, vol. 29, pp. 49–62, 2016.

[4] R. Campagni, D. Merlini, R. Sprugnoli, and M. C. Verri, "Data mining models for student careers," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5508–5521, 2015.

[5] A. A. Kardan, H. Sadeghi, S. S. Ghidary, and M. R. F. Sani, "Prediction of student course selection in online higher education institutes using neural network," *Computers and Education*, vol. 65, pp. 1–11, 2013.

[6] D. Thammasiri, D. Delen, P. Meesad, and N. Kasap, "A critical assessment of imbalanced class distribution problem: the case of predicting freshmen student attrition," *Expert Systems with Applications*, vol. 41, no. 2, pp. 321–330, 2014.

[7] A. Lee, "Determining the effects of computer science education at the secondary level on STEM major choices in postsecondary institutions in the United States," *Computers & Education*, vol. 88, pp. 241–255, 2015.

[8] C. Huang and J. Xu, "Second major selection system based on data mining," *Journal of Information and Computational Science*, vol. 6, no. 6, pp. 2333–2340, 2009.

[9] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[10] D. Guha, P. K. Roy, and S. Banerjee, "Load frequency control of interconnected power system using grey Wolf optimization," *Swarm and Evolutionary Computation*, vol. 27, pp. 97–115, 2016.

[11] N. Jayakumar, S. Subramanian, E. B. Elanchezhian, and S. Ganesan, "An application of grey wolf optimisation for combined heat and power dispatch," *International Journal of Energy Technology and Policy*, vol. 11, no. 2, pp. 183–206, 2015.

[12] A. Kaveh and F. Shokohi, "Application of Grey Wolf Optimizer in design of castellated beams," *Asian Journal of Civil Engineering*, vol. 17, no. 5, pp. 683–700, 2016.

[13] G. M. Komaki and V. Kayvanfar, "Grey Wolf Optimizer algorithm for the two-stage assembly flow shop scheduling problem with release time," *Journal of Computational Science*, vol. 8, pp. 109–120, 2015.

[14] R. L. Haupt and S. E. Haupt, *2. The Binary Genetic Algorithm*, John Wiley & Sons, 2004.

[15] R. S. Rahnamayan, H. R. Tizhoosh, and M. M. A. Salama, "Opposition-based differential evolution," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 1, pp. 64–79, 2008.

[16] S. Rahnamayan, H. R. Tizhoosh, and M. M. A. Salama, "A novel population initialization method for accelerating evolutionary algorithms," *Computers & Mathematics with Applications*, vol. 53, no. 10, pp. 1605–1614, 2007.

[17] W.-F. Gao, L.-L. Huang, J. Wang, S.-Y. Liu, and C.-D. Qin, "Enhanced artificial bee colony algorithm through differential evolution," *Applied Soft Computing*, vol. 48, pp. 137–150, 2016.

[18] Q. Li, H. Chen, H. Huang et al., "An enhanced grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis," *Computational and Mathematical Methods in Medicine*, vol. 2017, Article ID 9512741, 15 pages, 2017.

[19] H.-L. Chen, B. Yang, J. Liu, and D.-Y. Liu, "A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis," *Expert Systems with Applications*, vol. 38, no. 7, pp. 9014–9022, 2011.

[20] H.-L. Chen, D.-Y. Liu, B. Yang, J. Liu, and G. Wang, "A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11796–11803, 2011.

[21] H.-L. Chen, B. Yang, G. Wang, S.-J. Wang, J. Liu, and D.-Y. Liu, "Support vector machine based diagnostic system for breast cancer using swarm intelligence," *Journal of Medical Systems*, vol. 36, no. 4, pp. 2505–2519, 2012.

[22] H. L. Chen, B. Yang, S. j. Wang et al., "Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy," *Applied Mathematics and Computation*, vol. 239, pp. 180–197, 2014.

[23] L. Shen, H. Chen, Z. Yu et al., "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Systems*, vol. 96, pp. 61–75, 2016.

[24] H. L. Chen, B. Yang, S. J. Wang et al., "Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy," *Applied Mathematics and Computation*, vol. 239, pp. 180–197, 2014.

[25] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 6, pp. 637–646, 1998.

[26] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999.

[27] H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1048–1054, 1999.

[28] S. Hua and Z. Sun, "A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach," *Journal of Molecular Biology*, vol. 308, no. 2, pp. 397–407, 2001.

[29] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.

[30] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *International Journal of Remote Sensing*, vol. 23, no. 4, pp. 725–749, 2002.

[31] Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu, "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decision Support Systems*, vol. 37, no. 4, pp. 543–558, 2004.

[32] J. Mourão-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data," *NeuroImage*, vol. 28, no. 4, pp. 980–995, 2005.

[33] B. Schölkopf, C. J. C. Burges, and A. J. Smola, *Advances in Kernel Methods: Support Vector Learning*, The MIT Press, Cambridge, Mass, USA, 1998.

[34] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.

[35] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, 2000.

[36] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, Perth, Western Australia, December 1995.

[37] Y. Zhang, S. Wang, and G. Ji, "A comprehensive survey on particle swarm optimization algorithm and its applications," *Mathematical Problems in Engineering*, vol. 2015, Article ID 931256, 38 pages, 2015.

[38] J. Kennedy and R. C. Eberhart, "Discrete binary version of the particle swarm algorithm," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, vol. 5, pp. 4104–4108, Orlando, Fla, USA, October 1997.

[39] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, article no. 27, 2011.

[40] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th international Joint Conference on Artificial Intelligence (IJCAI '95)*, C. S. Mellish, Ed., vol. 2, pp. 1137–1143, Morgan Kaufmann, Montreal, Canada, 1995.

[41] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.