

Research Article

Prediction of Students' Performance Based on the Hybrid IDA-SVR Model

Huan Xu 

Department of Public Teaching, Hefei Preschool Education College, Hefei 230013, China

Correspondence should be addressed to Huan Xu; xhhjy001@126.com

Received 28 June 2021; Revised 25 October 2021; Accepted 26 October 2021; Published 18 February 2022

Academic Editor: Murari Andrea

Copyright © 2022 Huan Xu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Students' performance is an important factor for the evaluation of teaching quality in colleges. The aim of this study is to propose a novel intelligent approach to predict students' performance using support vector regression (SVR) optimized by an improved duel algorithm (IDA). To the best of our knowledge, few research studies have been developed to predict students' performance based on student behavior, and the novelty of this study is to develop a new hybrid intelligent approach in this field. According to the obtained results, the IDA-SVR model clearly outperformed the other models by achieving less mean square error (MSE). In other words, IDA-SVR with an MSE of 0.0089 has higher performance than DT with an MSE of 0.0326, SVR with an MSE of 0.0251, ANN with an MSE of 0.0241, and PSO-SVR with an MSE of 0.0117. To investigate the efficacy of IDA, other parameter optimization methods, that is, the direct determination method, grid search method, GA, FA, and PSO, are used for a comparative study. The results show that the IDA algorithm can effectively avoid the local optima and the blindness search and can definitely improve the speed of convergence to the optimal solution.

1. Introduction

In recent years, computer technology has been widely used in the field of education. The prediction of students' academic performance has always been an important part of education. At present, students' performance is still the main standard to measure students' level of knowledge acquisition, and an important factor to judge the teaching quality of schools and teachers. With the increase of enrollment scale, the growth of teachers and the number of students is out of proportion, which affects the teaching quality and students' performance. Therefore, it is very important to accurately predict students' performance in education management. The prediction of students' performance can guide teachers to adjust students' learning behavior in time and improve students' performance.

The common performance prediction methods can be divided into two categories. The first is to establish statistical models, such as multivariate linear regression model and sparse factor analysis model. Sravani and Bala [1] predicted the students' performance based on a linear regression

model. The second is based on data-driven performance prediction methods, such as logistic regression (LR) [2], Naive Bayes (NB) [3], decision tree (DT) [4], artificial neural network (ANN) [5], support vector regression (SVR), and so on. These methods do not require the participation of professionals but only extract the model from the relevant data. Table 1 shows the comparison of common machine learning methods. Table 2 shows the time and space complexity of common machine learning methods. As shown in Table 2, in small samples, SVR performs well in both time and space complexity. Borkar and Rajeswari [6] used education data mining and artificial neural network to predict the students' academic performance. Ghorbani and Ghousi [7] compared the performance of various machine learning such as random forest, k-nearest neighbor, support vector machine, and decision tree in predicting students' performance.

The students' performance is influenced by a variety of studying behaviors and varies greatly from individual to individual. Therefore, the traditional statistical model may be ineffective at some time. The data-driven approach

TABLE 1: Comparison of common machine learning methods.

Model	Advantages	Disadvantages
Logistic regression	1. Simple calculation and fast speed 2. Avoid overfitting through regularization	The performance is poor when faced with the multivariate or nonlinear decision boundary
Naive Bayes	Perform well on small-scale data	Very sensitive to the expression of input data
Decision tree	1. Able to apply to samples with missing attribute values 2. Strong robustness to outliers	Easy to overfit
Artificial neural networks	Perform well on nonlinear data	1. Long training time 2. The computational complexity is proportional to the network complexity
Support vector regression	1. Strong generalization ability 2. Can apply to high-dimensional nonlinear data 3. Low computational complexity	Sensitive to the selection of parameters and kernel function

TABLE 2: The time and space complexity of common machine learning methods.

Model	Time complexity	Space complexity
Logistic regression	$O(n * m)$	$O(m)$
Naive Bayes	$O(n * m * c)$	$O(m * c)$
Decision tree	$O(n * \log(n) * d)$	$O(p)$
Artificial neural networks	$O(t * \sum n_i n_2 + n_2 n_3 + \dots)$	$O(t * \sum (n_1 n_2 + n_2) + (n_2 n_3 + n_3) + \dots)$
Support vector regression	$O(n^2)$	$O(n_{sv})$

attempts to predict the students' performance directly from the student behavior data. The establishment of the data-driven students' performance prediction model only needs to collect enough performance data. Second, among the common data-driven models, SVR is more suitable for analyzing the student behavior data from Tables 1 and 2. Therefore, SVR is selected to predict the students' performance in this paper.

Note: n is the number of the training set, m is the dimensions of the sample, c is the number of categories of Naive Bayes, p is the number of nodes in the tree, n_i is the number of neurons in i^{th} layer, d is the maximum depth of the decision tree, t is the training times, p is the number of interneurons, and n_{sv} is the number of the support vectors in SVR.

Recent advances achieved in common machine learning methods are as follows. Zhou et al. [8] proposed a novel graph-based ELM (G-ELM) for imbalanced epileptic EEG signal recognition. Zhang et al. [9] combined deep learning-based image recognition methods and serological specific indicators for diagnosis of atrophic gastritis (AG). Yan et al. [10] proposed an improved early distinctive shapelet classification method for early classification on time series. Bai et al. [11] classified time series based on multifeature dictionary representation and ensemble learning. Ramanan et al. [12] developed a learning algorithm based on functional-gradient boosting methods for logistic regression, and the empirical evaluation on standard data sets demonstrated the superiority of the proposed approach over other methods for learning LR. Zhang et al. [13] proposed attribute and instance weighted Naive Bayes (AIWNB), and the experimental results validated that it indeed improved the prediction accuracy of NB. Schidler and Szeider [14]

proposed the SAT-based decision tree method by combining heuristic and exact methods in a novel way, which successfully decreased the depth of the initial decision tree in almost all cases. Khoo et al. [15] applied artificial neural networks to solve parametric PDE problems, and the simplicity and accuracy of the approach are demonstrated through notable examples. Cheng and Lu [16] developed an adaptive Bayesian support vector regression model for structural reliability analysis, and the proposed method outperformed other methods for medium-dimensional problems. Each method has its own advantages in a specific area. Through the analysis of Tables 1 and 2, support vector regression has better learning performance for the problem in this paper. It overcomes the requirement of traditional methods for large samples and can solve the problem of small sample and nonlinearity. In this paper, SVR is used to predict the students' performance.

The data used in the prediction includes two attributes. One is the students' previous performance, and the other is the students' basic behavior attributes, including the students' age, gender, attendance rate, self-study frequency, library access records, and so on. Bunkar et al. [17] used students' class test scores, seminar scores, homework scores, class attendance, and lab work to predict students' scores at the end of the semester. The second attribute often contains many redundant features, which may have a bad effect on the computational complexity and prediction accuracy of the model. Therefore, it is necessary to remove the redundant information before detecting product quality, and feature selection is a crucial method to deal with such a problem. Feature selection is an important preprocessing step for many high-dimensional quality classification problems [18]. With the increase of the number of features,

the search space of the feature subset grows exponentially. Most traditional feature selection algorithms are of low efficiency, so many scholars turn to using intelligent algorithms with stronger search ability, such as genetic algorithm [19], particle swarm optimization [20], and so on.

However, to obtain satisfactory prediction accuracy, it is not only related to the input characteristics of SVR but also closely related to the selection of SVR model parameters. The empirical method and the grid search method are the common SVR parameter selection methods. The empirical method is too subjective and the grid search method is time-consuming. In addition, the above two methods can only modify the parameters individually, and cannot achieve collaborative optimization among the parameters. At present, more and more scholars have applied the intelligent optimization algorithm to parameter selection of the SVR model. Luo et al. [21] proposed a novel artificial intelligence approach to predict the vertical load capacity of driven piles in cohesionless soils using SVR optimized by genetic algorithm (GA). Huang et al. [22] applied SVR to predict the strength of steel fibre reinforced concrete and used the firefly algorithm (FA) to tune the hyperparameters of SVR. Liu et al. [23] analyzed surface acoustic wave (SAW) yarn tension sensor's measured data by SVR and used the PSO algorithm to optimize the hyperparameters of SVR. Intelligent algorithms have been proved to be effective in solving parameter optimization problems. It does not depend on the specific domain of the problem and has strong robustness to the various types of problem. DA algorithm is an effective global optimization algorithm. After the duel between individuals, individuals continue to evolve and get closer to the optimal solution to the problem. Therefore, the DA algorithm is selected to optimize the SVR model parameters and the features collaboratively in this paper.

The rest of the paper is organized as follows. In Section 2, the IDA-SVR model is established. In Section 3, a real example about the academic performance of students is given to illustrate the proposed model. Finally, we summarize this paper and put forward future research directions in Section 4.

2. Methodology

This section will introduce the necessary background knowledge and the proposed model. First, the SVR model and DA algorithm are elaborated. Next, the proposed IDA-SVR model is described in detail.

2.1. Support Vector Regression. SVR is the application of a support vector machine (SVM) in regression learning. Suppose $(x_1, y_1), \dots, (x_n, y_n)$, $x_i \in R^m, y_i \in R$, are the sample data. Such a linear function, namely SVR function, is as follows:

$$f(x) = \omega^T \varphi(x) + b, \quad (1)$$

where $\omega = (\omega_1, \omega_2, \dots, \omega_m)^T$ is a vector normal to the maximum-margin hyperplane and b is the deviation. $\varphi(\cdot)$ is a nonlinear mapping.

The problem can be treated as the following optimization problem:

$$\begin{aligned} \min_{\omega, b, \varepsilon} & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i^+ + \xi_i^-) \\ \text{s.t.} & \begin{cases} \omega^T \varphi(x) + b - y_i \leq \varepsilon + \xi_i^+, \\ y_i - \omega^T \varphi(x) - b \leq \varepsilon + \xi_i^-, \\ \xi_i^+, \xi_i^- \geq 0, \quad i = 1, 2, \dots, l, \end{cases} \end{aligned} \quad (2)$$

where C is the regularization factor and ξ_i^- and ξ_i^+ are slack variables representing lower and upper constraints on the outputs of the model. ε is a positive constant. Errors are calculated only if the deviation between $f(x)$ and y_i is greater than ε .

The above problem is a quadratic problem with linear constraints, so the Karush–Kuhn–Tucker (KKT) optimal conditions are necessary and sufficient. The solution, which can be obtained from the dual problem, is a linear combination of a subset of sample points denominated support vectors (s.v.) as follows:

$$\omega = \sum_{\text{s.v.}} \beta_i \varphi(x_i) \Rightarrow f_{\omega, b}(x) = \sum_{\text{s.v.}} \beta_i \langle \varphi(x_i), \varphi(x) \rangle + b. \quad (3)$$

Let $\kappa(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$, which is called the kernel function. It can map points from low- to high-dimensional space. Then, equation (3) can be rewritten as follows:

$$\omega = \sum_{\text{s.v.}} \beta_i \varphi(x_i) \Rightarrow f_{\omega, b}(x) = \sum_{\text{s.v.}} \beta_i \kappa(x_i, x) + b. \quad (4)$$

Kernel selection is one of the key technologies to improve the ability of SVR. This paper uses the radial basis function as shown in the following equation:

$$\kappa(x_i, x_j) = e^{-\sigma \|x_i - x_j\|^2}. \quad (5)$$

The prediction accuracy of the SVR model depends on the good settings of the hyperparameters C and ε and the kernel parameter σ . Therefore, the selection of the parameters is an important issue. Next, we will introduce the IDA algorithm to optimize SVR parameters.

2.2. DA Algorithm. Duelist algorithm (DA) is a new algorithm based on a genetic algorithm proposed by Biyanto [24] from the perspective of human combat and learning ability. The process of the DA algorithm is shown in Figure 1.

2.2.1. Encoding. In this paper, the encoding of the DA algorithm is composed of parameters (C, ε, σ) and feature subsets, as shown in Figure 2.

$b_C^1 \sim b_C^{n_C}$, $b_\varepsilon^1 \sim b_\varepsilon^{n_\varepsilon}$, $b_\sigma^1 \sim b_\sigma^{n_\sigma}$, and $b_f^1 \sim b_f^{n_f}$ are the binary strings of parameters C, ε, σ , and features, respectively. $n_C, n_\varepsilon, n_\sigma$, and n_f are the numbers of binary digits of C, ε, σ , and features, respectively.

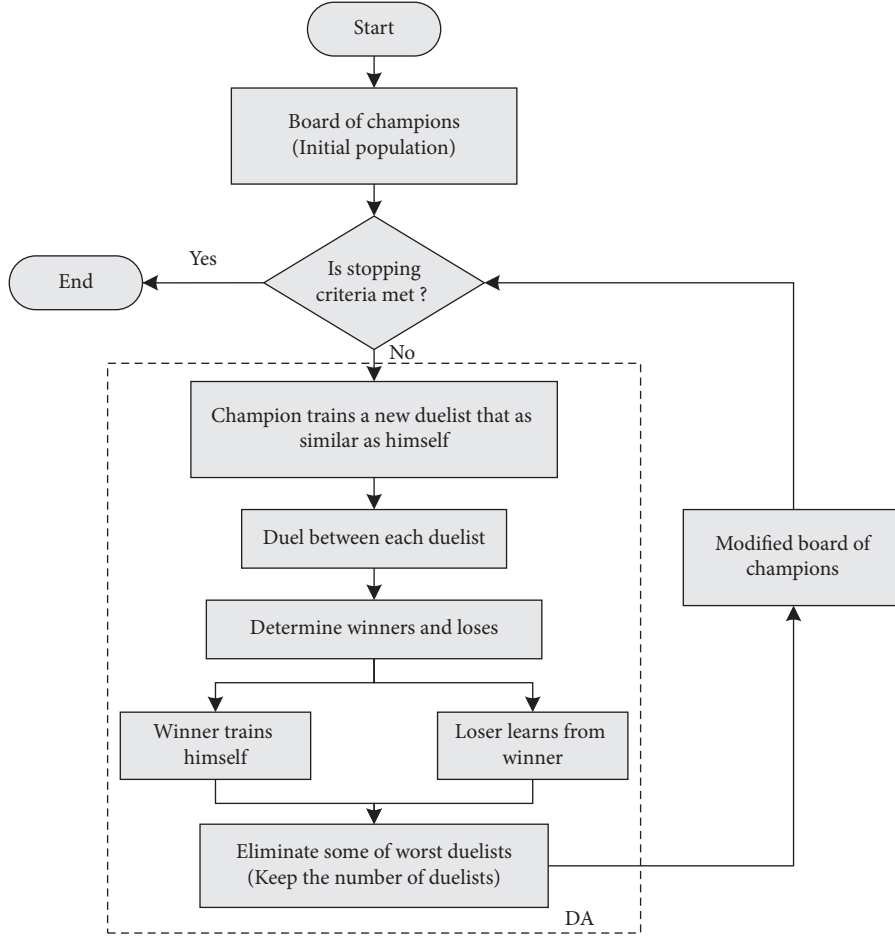


FIGURE 1: Flowchart of DA algorithm.

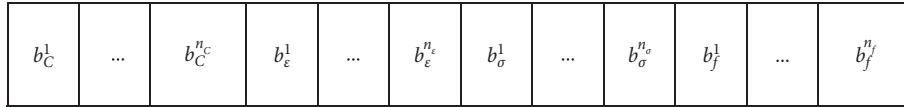


FIGURE 2: Encoding of DA-VNS algorithm.

2.2.2. Fighting Capability Function. In this paper, we take the mean squared error (MSE) as the fighting capability. Let y_i be the observed values and \hat{y}_i be the predicted values, then

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (6)$$

Here, n is the number of samples. A smaller MSE value indicates a better fighting capability.

2.2.3. Duel Scheduling between Duelists. DA algorithm optimizes the solution by one-to-one dueling between duelists. The pseudocode of the duel process is shown in Algorithm 1.

2.2.4. Duelist's Improvement. In this step, both the loser and the winner need to improve their fighting capabilities. The

pseudocode of duelist's improvement is displayed in Algorithm 2.

2.3. IDA-SVR Model. After in-depth analysis, it is found that the DA algorithm has four shortcomings.

- (1) The value of the initial solution is generated randomly. The random process cannot guarantee the uniform distribution of the initial population and the quality of the individual. Some of the solutions are far away from the optimal solution.
- (2) By analyzing the whole process of the DA algorithm, we can conclude that the luck coefficient has a great impact on the performance of the algorithm. The larger the luck coefficient, the more random the new individual. It follows that the fitness fluctuation of the solution becomes larger, and the speed of convergence to the optimal solution becomes slower. On

the contrary, the smaller the luck coefficient is, the weaker the randomness of the new individual will be, which leads to the slower speed of obtaining the optimal solution. Therefore, the setting of the luck coefficient is crucial to the effectiveness of the algorithm.

- (3) Each duelist is categorized into winner and loser after the duel. To improve duelists' fighting capability, each loser is trained by learning from the winner, and winners evolve on their own. Therefore, it can be seen that the loser's improvement is based on the information exchange between two individuals, which will lead to the slow convergence of the algorithm.
- (4) Like other swarm intelligent optimization algorithms, the DA algorithm is also prone to local optimization and low search accuracy in the search process.

In view of the shortcomings of the DA algorithm, this paper has made improvements in the following aspects:

- (1) The chaotic sequence is used to initialize the population. It not only improves the diversity of the population by using chaos but also does not change the randomness of the optimization algorithm during initialization. There are various mathematical models for generating chaotic sequences. In this paper, a logistic equation is used to construct chaotic sequences as follows:

$$x(t+1) = \mu x(t)(1-x(t)), \quad t = 0, 1, 2, \dots, \quad (7)$$

where μ is the control parameter. When $0 < x(0) < 1$ and $\mu = 4$, the logistic equation is in a complete chaotic state. In this case, $x(t)$ is chaotic and in the interval $(0, 1)$. Given the initial value $x(0) \in (0, 1)$, the time series $x(1), x(2), \dots$, can be generated.

- (2) According to the statistics principle, there are more chances to search for more optimal solutions around the optimal solution. That is, we can set the luck coefficient a little bigger at the beginning. Then the solution will be more random and easier to find the optimal solution. When close to the optimal solution, a small luck coefficient allows the algorithm to search for more optimal solutions around it. Therefore, based on the above analysis, we define the adaptive luck coefficient c :

$$c = \frac{i_{\max}}{\lambda(i_{\max} + i + 1)}. \quad (8)$$

Here, i_{\max} is the total number of iterations, i is the current iteration number, and λ is the adjustment coefficient of step length, which is determined according to the feasible regions of different optimization problems.

- (3) For the loser's improvement, each loser is trained by learning from one of the winners after a duel. The

roulette method is used to determine the winner that the loser will learn from.

- (4) The chaotic sequence search method is used to generate the neighborhood solutions. The randomness and ergodicity of chaotic variables can make the algorithm jump out of the local optimization. In this way, the global searching ability of the algorithm is improved. First, the chaotic sequence is generated by equation (7) based on the optimal position currently. Then the optimal position of the chaotic sequence is used to replace the position of a duel. Through the above steps, neighborhood solutions of the local optimal solution can be generated in the iteration, which can help the current solution escape from the local optimal solution.

The four strategies are to improve the algorithm in different steps without overlapping. Strategy (a) is an improvement on the initial value. Strategy (b) makes the lucky coefficient adjust adaptively and enables the algorithm to converge to the optimal solution faster. Strategy (c) is to increase the diversity of solutions in the duelist improvement step. Strategy (d) can make the newly generated solution jump out of the local optimum. The above four strategies guarantee the prediction accuracy and the speed of convergence to the optimal solution of the algorithm together.

Then, we will use the improved DA algorithm to optimize the parameters of SVR. Figure 3 shows the flowchart of the hybrid IDA-SVR model developed in this research work.

3. Experimental Study

In this paper, the mathematical performance data of 240 students from five classes in grade two in a vocational college in Hefei are selected as the research object. Among them, 180 samples are used as the training data and the remaining 60 samples are used as the testing data. Each sample contains 18 features, as shown in Table 3.

3.1. Data Preparation. To eliminate the influence of the different dimensions on the numerical values, further normalization of data is needed. The normalization formula is as follows:

$$a'_{ij} = \frac{a_{ij} - a_{i\min}}{a_{i\max} - a_{i\min}}, \quad (9)$$

where a_{ij} is the initial sample data to be normalized and $a_{i\min}$ and $a_{i\max}$ are the minimum and maximum values in the column sample values, respectively.

3.2. Experimental Study. All experiments are run on Intel Core i5-1035 8 GB, the Microsoft Windows 10 operating system, and the development environment of Python 3.6.6, PyCharm 2021.1. The parameter settings are shown in Table 4.


```

(1) Duelist A and B, Luck_coeff
(2) FC = Fighting capability; LC = Luck_coeff
(3) A(Luck) = A(FC) * (LC + (random(0, 1) * LC));
(4) B(Luck) = B(FC) * (LC + (random(0, 1) * LC));
(5) if ((A(FC) + A(Luck)) >= (B(FC) + B(Luck)))
(6) A(Winner) = 1;
(7) B(Winner) = 0;
(8) else
(9) A(Winner) = 0;
(10) B(Winner) = 1;
(11) end if

```

ALGORITHM 1: Determination of the winner and the loser.

```

(1) Duelist A and B, Duelist_length, Prob_innovate; Prob_learn
(2) if A(Winner) = 1
(3) for i = 1:(Duelist_length)
(4) r = random(0,1)
(5) if r < Prob_innovate
(6) if A[i] = 1
(7) A[i] = 0
(8) else A[i] = 1
(9) end if
(10) end if
(11) end for
(12) else
(13) for i = 1: (Duelist_length)
(14) r = random(0, 1)
(15) if r < Prob_learn
(16) B[i] = A[i]
(17) end if
(18) end for
(19) end if

```

ALGORITHM 2: Duelist's Improvement.

We apply the IDA-SVR model to predict the students' performance, and the results are shown in Figure 4 and Table 5.

By analyzing the students' performance prediction model based on SVR, it is found that the number of support vectors used in the prediction model is 153. It can be seen that only the data of 153 students is needed to realize the prediction of performance in the sample set composed of 240 students. Due to the limited length of the article, two examples are listed below for analysis.

Take the example of a support vector with index number 36, which is a student with a performance of 90. In the model, the top five important learning behavior features are the number of assignments submitted, the average number of hands raised, the time of study this course outside the class, the frequency of distraction, and the number of absenteeism. The corresponding feature weight vector of this sample is [1.0335, 0.8327, 0.8133, -0.7415, -0.5448]. It can be seen that the number of assignments submitted has a great influence on the performance. Next, we will look at a student with low performance.

Now we analyze the support vector with index number 228, which is a student with lower performance of 72. In the model, the top five important learning behavior features are work as a student cadre, the number of absenteeism, the frequency of distraction, the number of assignments submitted, and the average number of hands raised. The corresponding feature weight vector of this sample is [1.2814, -0.9738, -0.7122, 0.5415, 0.4388]. Compared with the student with index number 36, this student has a lower weight in the number of assignments submitted and the average number of hands raised. Therefore, the above two features have a significant impact on students' achievement. In addition, absenteeism is one of the main reasons for a student's poor performance.

Based on the above model analysis, the following teaching suggestions are proposed. Teachers should explore targeted measures to improve students' performance based on the results of experimental analysis and the actual condition. To improve students' performance, teachers should reasonably use teaching methods in the teaching process to stimulate students' learning engagement and motivation.

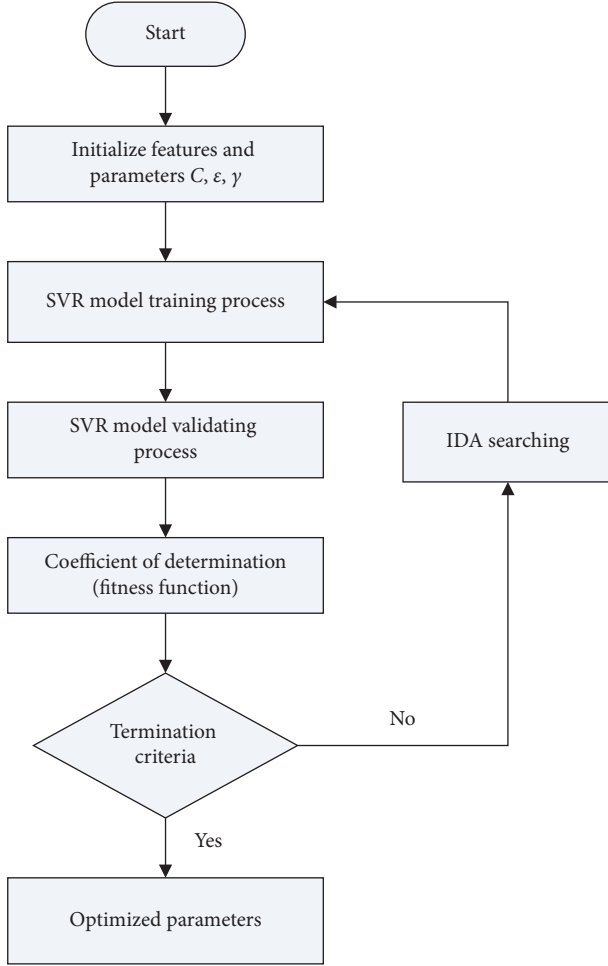


FIGURE 3: Flowchart of the IDA-SVR model.

Figure 5 shows the accumulative Pearson's correlation coefficient between the actual and predicted values. As shown in Figure 5, with the number of predicted students' performance increases, we can see a strong correlation that exists between the actual and predicted values. The Pearson's correlation coefficient of the total testing data is 0.9214. Figure 6 describes the degree of fluctuation of errors between the actual and predicted values. As shown in Figure 6, the errors are between -0.2 and 0.2 . In addition, there are almost no outliers. From the comprehensive analysis of Figures 4–6, we can conclude that the proposed method has a good performance in predicting students' performance. (All the data are normalized.)

To demonstrate the superiority of the IDA-SVR model, four other models, basic SVR, PSO-SVR, decision tree (DT), and artificial neural networks (ANN), are selected to compare with the IDA-SVR model. The constructed ANN model includes three hidden layers with the ReLU activation function. In the decision tree algorithm, the maximum depth of the decision tree is set to 5. Table 6 recapitulates the results of the proposed method and the comparative methods on the students' performance prediction problem. We find that the proposed model outperforms the selected comparative methods in prediction accuracy.

Note: pop is the population size of the intelligent algorithm and iter is the number of the iteration times of the intelligent algorithm.

As shown in Table 6, the proposed IDA-SVR model has the highest time complexity and running time, which can be said that it trades time for accuracy. How to reduce the time complexity and running time is a future research direction for us. First, the runtime may be further reduced by exploring more computational-efficient SVR algorithms and a faster parameter tuning mechanisms. Moreover, parallelization techniques and methods are worth exploring and utilizing to improve learning performance and reduce the computational cost in the model.

3.3. Comparative Experiment. We design a set of comparative experiments to evaluate the performance of the IDA algorithm in optimizing the parameters of SVR for students' performance prediction problem. Parameter optimization methods, that is, the direct determination method [25], grid search method [26], genetic algorithm (GA) [21], firefly algorithm (FA) [22], and particle swarm optimization (PSO) [23], with SVR are used to compare with the IDA-SVR. The results of the direct determination method, grid search method, GA-SVR, FA-SVR, and PSO-SVR are reported in Table 7.

Table 7 displays the results of the prediction performance of six algorithms. The direct determination method has the worst classification accuracy among these algorithms. The possible reason is that it requires high data quality, and it is not suitable for students' performance data. The grid search method is to divide the parameters to be optimized into grids in a certain spatial range and then search for the optimal parameters by traversing all points in the grid. It has a good effect in a small interval, but a poor effect in a large interval or multiparameter case. GA, FA, PSO, and IDA are all heuristic algorithms. To investigate the efficacy of IDA algorithms in optimizing the SVR parameters, we compare the running results after 500 iterations of GA-SVR, FA-SVR, PSO-SVR, and IDA-SVR, respectively. The results are shown in Figures 7 and 8.

As shown in Figure 7, in general, the solutions obtained by IDA-SVR is superior to GA-SVR, FA-SVR, and PSO-SVR. The solutions obtained by PSO-SVR are mostly clustered around the optimal solution. In addition, the solutions obtained by GA-SVR and FA-SVR are more concentrated than those obtained by DA and DA-VNS algorithms. In other words, the search scope of the GA and FA algorithms is smaller than the other two algorithms. The IDA algorithm has the largest search range, which can effectively avoid the local optima and the blindness search.

Figure 8 shows the evolution of the minimum MSE for GA-SVR, FA-SVR, PSO-SVR, and IDA-SVR over 500 iterations. From Figure 8, we can see that the PSO and IDA algorithms can make the solutions converge to the optimal solution continuously. Compared with PSO, the speed of convergence to the optimal solution of IDA is faster. In

TABLE 3: The features of each student sample.

Category	No.	Feature	Value
Basic information	1	Sex	{1,2}
	2	Native place	{1,2,3,4,5}
	3	Semester	{1,2}
	4	Education level of parents	{1,2,3,4,5}
	5	Work as a student cadre	{0,1}
Interest	6	Interest in the course	{1,2,3,4}
	7	The degree of keeping up with the class	{1,2,3,4}
	8	Learning initiative	{1,2,3,4}
Behavior in class	9	Number of absenteeism	[0,25]
	10	Frequency of distraction	[5,47]
	11	Average number of hands raised	[0,12]
	12	Number of questions answered	[0,9]
	13	Number of assignments submitted	[0,20]
	14	Number of interactions between teachers and students	[0,17]
	15	Number of group discussions attended	[0,10]
Behavior outside class	16	Time of study this course outside the class	[5,65]
	17	Study extracurricular material time	[0,12]
	18	Online viewing time	[0,60]

TABLE 4: List of preset parameters in DA-VNS.

Parameters	Value
Population size	100
Iteration times	500
Search range of penalty parameter C	$[10^{-3}, 10^3]$
Allowable error ε	$[10^{-3}, 10^{-1}]$
Search range of kernel width σ	$[10^{-6}, 10^1]$
Luck coefficient (LC)	0.1

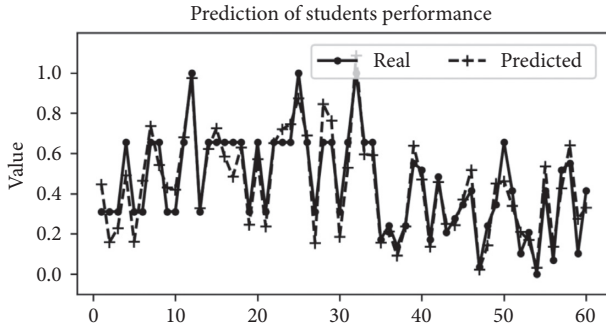


FIGURE 4: The prediction results of students' performance based on the IDA-SVR model.

TABLE 5: The results of IDA-SVR in students' performance prediction.

Model	MSE	Selected features
IDA-SVR	0.0092	6

addition, among the four algorithms, the IDA algorithm takes the least number of iterations to find the optimal solution. It successfully finds the global optimal solution of

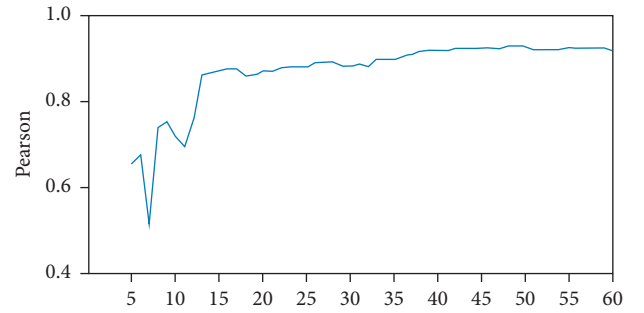


FIGURE 5: The accumulative Pearson's correlation coefficient between the actual value and the predicted value.

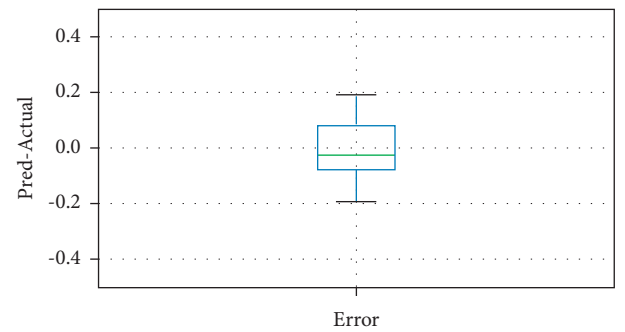


FIGURE 6: Boxplot representing the difference between the actual value and the predicted value.

input feature set and hyperparameters within 200 iterations when GA uses 380 iterations, FA 338 iterations, and PSO 372 iterations. After the above analysis, the IDA algorithm can definitely improve the speed of convergence to the optimal solution.

TABLE 6: Comparison of performance between IDA-SVR and other models.

Model	MSE	Time complexity	Running time (second)
SVR	0.0251	$O(n^2)$	0.33
PSO-SVR	0.0117	$O(\text{pop} * \text{iter} * n^2)$	90.58
DT	0.0326	$O(n * \log(n) * d)$	1.2
ANN	0.0241	$O(t * \sum n_1 n_2 + n_2 n_3 + \dots)$	3.59
IDA-SVR	0.0092	$O(\text{pop} * \text{iter} * n^2)$	105.17

TABLE 7: Comparison of performance between IDA-SVR and other methods.

Method	MSE
Direct determination method [20]	0.227
Grid search method [21]	0.203
GA-SVR [21]	0.0168
FA-SVR [22]	0.0149
PSO-SVR [23]	0.0117
IDA-SVR	0.0092

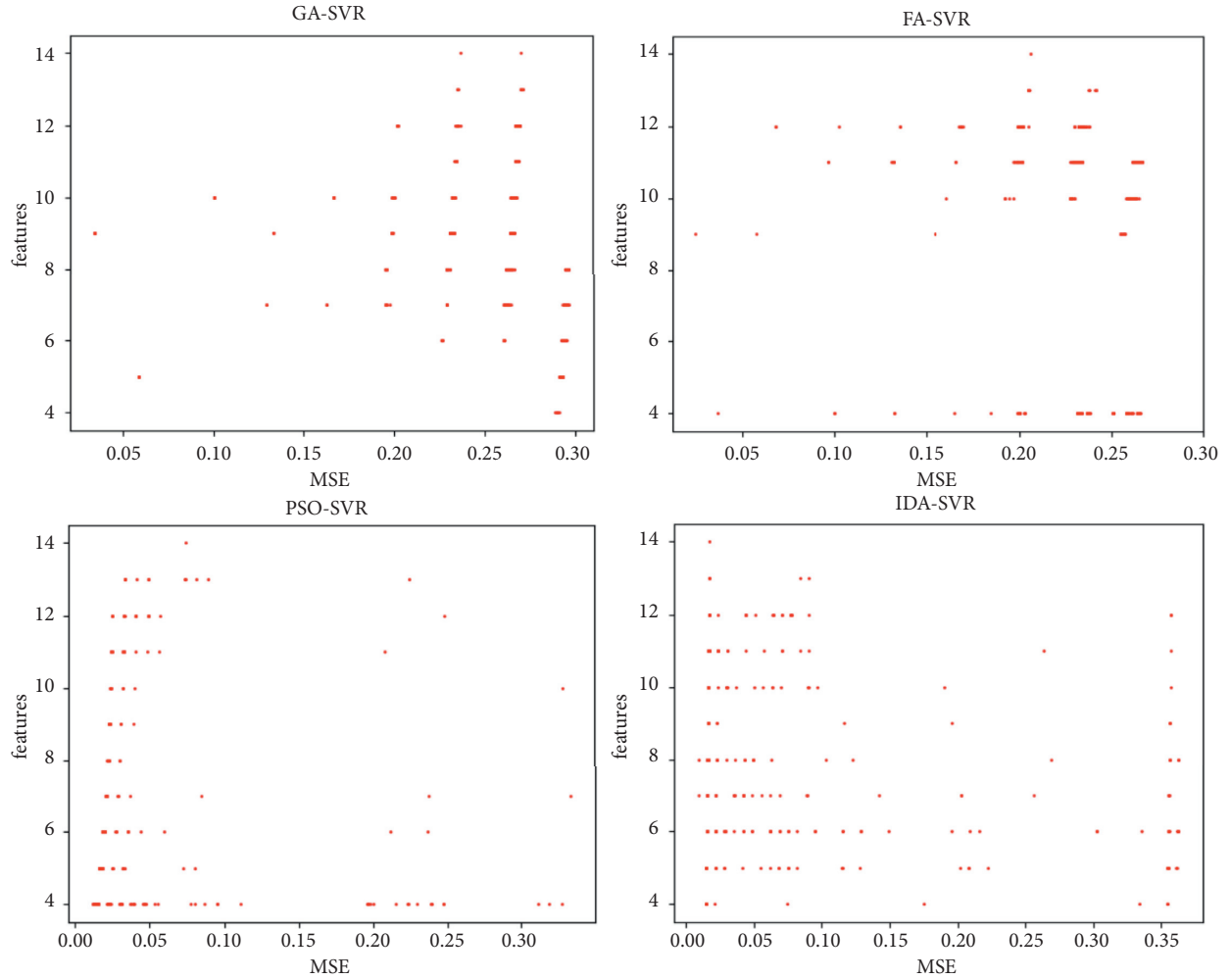


FIGURE 7: Running results after 500 iterations of GA-SVR, FA-SVR, PSO-SVR, and IDA-SVR.

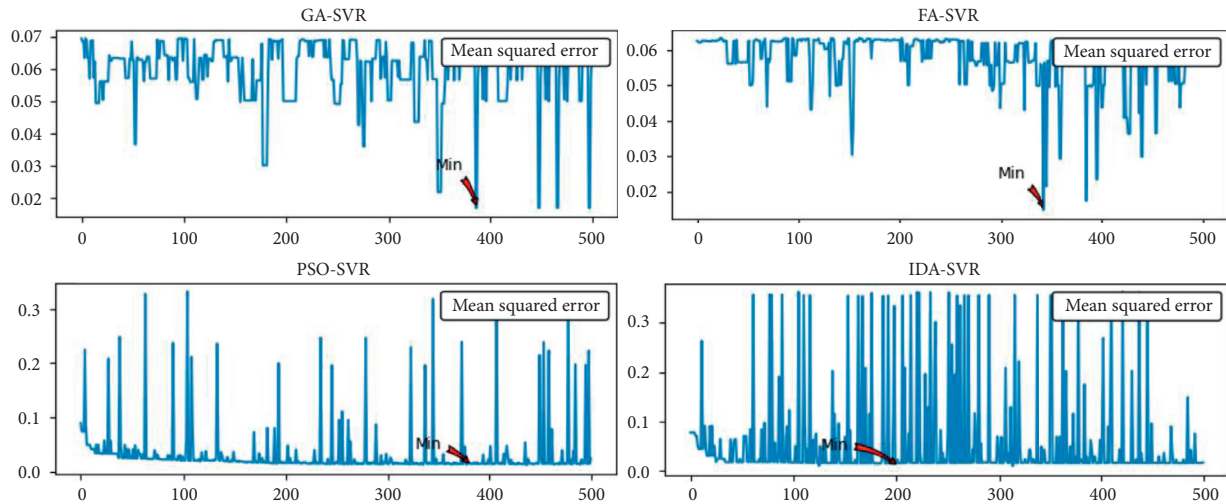


FIGURE 8: The evolution of the minimum MSE for GA-SVR, FA-SVR, PSO-SVR, and IDA-SVR over 500 iterations.

4. Conclusions

In this paper, a hybrid IDA-SVR model is proposed to predict the students' performance. The main contributions of this paper can be summarized as follows: (1) a novel intelligent approach is proposed to predict students' performance based on student behavior using support vector regression. Some experiments are conducted using the mathematics score data of students. The experimental results show that the proposed model has excellent performance in solving the prediction problem of students' performance. (2) The improved dual algorithm is designed to optimize the kernel parameters of SVR and select the features. Compared with other parameter optimization methods, the IDA algorithm can effectively avoid the local optima and the blindness search and can definitely improve the speed of convergence to the optimal solution.

The method proposed in this paper aims at solving students' performance prediction problem. However, it also can be applied to other problems in other fields. Because the proposed hybrid method is essentially a prediction algorithm for the small sample data with labels. It is applicable to any field that meets the above point, such as prediction of some economic indicators, environmental indicators, abnormal detection of ECG signals, diagnosis of circuit failures, and so on.

Although the proposed model performs well among many models, it still has some limitations. First, the improved DA algorithm has some instability. For example, the initial values of the parameters to be optimized are given randomly, and different initial values will have different effects on the results. In addition, even though the improved DA algorithm provides the possibility of global search, it cannot ensure that it converges to the global best. Second, SVR can get much better results than other algorithms on a small sample training set. But when the sample dimension is large, the time complexity of SVR will increase, which will greatly reduce the efficiency of the predictor. Third, the improved DA algorithm optimizes the parameters of SVR by

training individuals on the training set and evaluating the scores on the testing set. The more iterations of optimization, the higher the accuracy. In other words, the proposed model trades time for accuracy to a large extent.

To solve the above limitations, our study can be extended in the following future research directions. With the development of computer technology, the number of layers of neural networks that can handle is increasing, and the performance of deep learning methods has surpassed machine learning in many fields. To improve the performance of SVR, it is necessary to improve the objective function, constraint conditions, and kernel function of the SVR model based on the problem itself.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] B. Sravani and M. M. Bala, "Prediction of student performance using linear regression," in *Proceedings of the 2020 International Conference for Emerging Technology (INCET)*, pp. 1–5, IEEE, Belgaum, India, June 2020.
- [2] A. C. J. W. Janssens, Y. Deng, G. J. J. M. Borsboom, M. J. C. Eijkemans, J. D. F. Habbema, and E. W. Steyerberg, "A new logistic regression approach for the evaluation of diagnostic test results," *Medical Decision Making*, vol. 25, no. 2, pp. 168–177, 2005.
- [3] I. Rish, "An empirical study of the naive bayes classifier," *Journal of Universal Computer Science*, vol. 1, no. 2, p. 127, 2001.
- [4] J. R. Quinlan, "Induction on decision tree," *Machine Learning*, vol. 1, 1986.
- [5] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application

- in pharmaceutical research,” *Journal of Pharmaceutical and Biomedical Analysis*, vol. 22, no. 5, pp. 717–727, 2000.
- [6] S. Borkar and K. Rajeswari, “Attributes Selection for predicting students’ academic performance using education data mining and artificial neural network,” *International Journal of Computer Applications*, vol. 86, no. 10, pp. 25–29, 2014.
 - [7] R. Ghorbani and R. Ghousi, “Comparing different resampling methods in predicting students’ performance using machine learning techniques,” *IEEE Access*, vol. 8, pp. 67899–67911, 2020.
 - [8] J. Zhou, X. Zhang, and Z. Jiang, “Recognition of imbalanced epileptic EEG signals by a graph-based extreme learning machine,” *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 5871684, 12 pages, 2021.
 - [9] J. Zhang, J. Yu, S. Fu, and X. Tian, “Adoption value of deep learning and serological indicators in the screening of atrophic gastritis based on artificial intelligence,” *The Journal of Supercomputing*, vol. 77, pp. 8674–8693, 2021.
 - [10] W. Yan, G. Li, Z. Wu, S. Wang, and P. S. Yu, “Extracting diverse-shapelets for early classification on time series,” *World Wide Web*, vol. 23, no. 6, pp. 3055–3081, 2020.
 - [11] B. Bai, G. Li, S. Wang, Z. Wu, and W. Yan, “Time series classification based on multi-feature dictionary representation and ensemble learning,” *Expert Systems with Applications*, vol. 169, Article ID 114162, 2021.
 - [12] N. Ramanan, G. Kunapuli, and T. Khot, “Structure learning for relational logistic regression: an ensemble approach,” *Data Mining and Knowledge Discovery*, vol. 35, pp. 2089–2211, 2021.
 - [13] H. Zhang, L. Jiang, and L. Yu, “Attribute and instance weighted naive Bayes,” *Pattern Recognition*, vol. 111, Article ID 107674, 2021.
 - [14] A. Schidler and S. Szeider, “SAT-based decision tree learning for large data sets,” in *Proceedings of the AAAI 21, the Thirty-Fifth AAAI Conference on Artificial Intelligence*, Vancouver, Canada, 2021.
 - [15] Y. Khoo, J. Lu, and L. Ying, “Solving parametric PDE problems with artificial neural networks,” *European Journal of Applied Mathematics*, vol. 32, no. 3, pp. 421–435, 2021.
 - [16] K. Cheng and Z. Lu, “Adaptive bayesian support vector regression model for structural reliability analysis,” *Reliability Engineering & System Safety*, vol. 206, Article ID 107286, 2021.
 - [17] K. Bunkar, U. K. Singh, B. Pandya, and R. Bunkar, “Data mining: prediction for performance improvement of graduate students using classification,” in *Proceedings of the 2012 Ninth International Conference on Wireless and Optical Communications Networks (WOCN)*, pp. 1–5, IEEE, Indore, India, September 2012.
 - [18] B. Frénay, G. Doquire, and M. Verleysen, “Is mutual information adequate for feature selection in regression,” *Neural Networks*, vol. 48, no. 6, pp. 1–7, 2013.
 - [19] S. Li, H. Wu, D. Wan, and J. Zhu, “An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine,” *Knowledge-Based Systems*, vol. 24, no. 1, pp. 40–48, 2011.
 - [20] N. Alias, M. N. Ismail, and K. de Silva, “A hybrid gini PSO-SVM feature selection: an empirical study of population sizes on different classifier,” in *Proceedings of the International Conference on Artificial Intelligence*, Kota Kinabalu, Malaysia, December 2013.
 - [21] Z. Luo, M. Hasanipanah, H. Bakhshandeh Amnieh, K. Brindhadevi, and M. M. Tahir, “GA-SVR: a novel hybrid data-driven model to simulate vertical load capacity of driven piles,” *Engineering with Computers*, vol. 37, no. 2, pp. 823–831, 2021.
 - [22] Y. Huang, J. Zhang, F. Tze Ann, and G. Ma, “Intelligent mixture design of steel fibre reinforced concrete using a support vector regression and firefly algorithm based multi-objective optimization model,” *Construction and Building Materials*, vol. 260, Article ID 120457, 2020.
 - [23] S. Liu, P. Xue, J. Lu, and W. Lu, “Fitting analysis and research of measured data of SAW yarn tension sensor based on PSO-SVR model,” *Ultrasonics*, vol. 116, Article ID 106511, 2021.
 - [24] T. Biyanto, H. Fibrianto, E. Listijorini, and T. Budiati, “Duelist algorithm: an algorithm inspired by how duelist improve their capabilities in a duel,” in *Proceedings of the Seventh International Conference on Swarm Intelligence*, Bali, Indonesia, 2015.
 - [25] V. Cherkassky and Y. Ma, “Practical selection of SVM parameters and noise estimation for SVM regression,” *Neural Networks*, vol. 17, no. 1, pp. 113–126, 2004.
 - [26] Y. Sun, S. Ding, Z. Zhang, and W. Jia, “An improved grid search algorithm to optimize SVR for prediction,” *Soft Computing*, vol. 25, no. 7, pp. 5633–5644, 2021.