Retraction

Retracted: Prediction Analysis of College Students' Physical Activity Behavior by Improving Gray Wolf Algorithm and Support Vector Machine

Advances in Multimedia

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] M. Wang, "Prediction Analysis of College Students' Physical Activity Behavior by Improving Gray Wolf Algorithm and Support Vector Machine," *Advances in Multimedia*, vol. 2022, Article ID 4943510, 10 pages, 2022.

Research Article

Prediction Analysis of College Students' Physical Activity Behavior by Improving Gray Wolf Algorithm and Support Vector Machine

Minjian Wang [ID]

Jiaxing Vocational and Technical College, Department of Basic and Public Courses, Jiaxing, Zhejiang 314036, China

Correspondence should be addressed to Minjian Wang; wminjian@163.com

Physical exercise for college students is an important means to build a healthy standard of college students and an important way to a healthy campus. In addition to creating good physical fitness, physical exercise has significant effects on improving psychological stress and alleviating psychological problems and mental illnesses among college students. It is important to predict and analyze the physical exercise behavior of college students and explore the positive value of physical exercise for college education. In order to overcome the problem of low accuracy of traditional algorithms in prediction, this paper uses the improved gray wolf algorithm (IGWO) and support vector machine (SVM) for predictive analysis of college students' physical exercise behavior. A nonlinear decreasing convergence factor strategy and an inertia weight strategy are introduced to improve the gray wolf optimization algorithm, which is used to determine the SVM parameters for the purpose of improving the model accuracy. Then, the college students' physical exercise data are input into the model for validation. By constructing a campus behavior data set of college students and conducting experiments, the algorithm achieves 90.45% behavior prediction accuracy, which is better than that of typical prediction models. Finally, individual growth monitoring of college students is targeted to warn students with abnormal behaviors. At the same time, the higher-order information such as physical exercise behavior habits of college students is explored to provide meaningful reference for constructing personalized training.

1. Introduction

Promoting physical exercise among college students is an important part of the national health strategy and also one of the mainstream directions of current research [1]. In the field of college students' physical exercise psychology, physical exercise intention and physical exercise behavior are two common basic concepts [2]. Most researchers regard physical exercise intention as a key factor to promote physical exercise behavior, and some even use physical exercise intention as a dependent variable in physical exercise promotion studies [3]. But studies of the consistency of intention and behavior have shown that there is a gulf between intention and behavior. Therefore, the question to be answered in this study is how physical exercise intention can predict physical exercise behavior.

Exercise intention is people's decision to take a physical exercise behavior [4]. People can form a behavior or goal intention by saying "I plan to do a certain sport," "I intend to do a certain sport," "I will do a certain sport," etc. This is the result of deliberate consideration of what the individual is going to do and can indicate the level of effort that the individual will put in to achieve the desired result. Physical exercise intention was regarded as the proximal determinant of physical exercise behavior by researchers [5]. For example, when using rational behavior theory, planned behavior theory, attitude and behavior theory, and protective motivation theory to study the change of physical exercise behavior, they all agree that the most direct and important predictor of individual physical exercise behavior is individual physical exercise intention.

The analysis of college students' physical exercise behavior, especially the spatiotemporal modeling of campus data to mine the physical exercise behavior pattern, is a current research hot spot in the interdisciplinary disciplines of computer science, behavioral psychology, and education [6]. As one of the most concerned groups in the whole society, college students have more and more mental health problems [7]. The analysis of physical exercise behaviors and patterns of college students is one of the important topics dealt with in university education. It is an important guideline for college students to regulate their own study and life, for colleges and universities to make teaching plans and regulations, and for the state to make guidelines and policies of higher education [8]. The prediction of college students' physical exercise behavior has important theoretical and application values in the aspects of students' physical health control, mental health education, abnormal behavior detection, education, and talent cultivation. The intelligent education mode supported by information technology has become the development trend of modern education informatization [9]. The physical exercise behaviors of college students are continuously recorded by various devices, forming a large amount of multisource fusion data reflecting students' physical health information. The physical activity data of college students are highly nonlinear and correlated [10]. By extracting the activity characteristics, analyzing the correlations, and mining the spatiotemporal process of their activity habits, we can predict the physical activity behavior and reason about the abnormal behavior on campus.

The prediction of college students' physical activity behavior is a typical problem of spatiotemporal data prediction [11]. On the one hand, the campus data record students' daily physical exercise activities by time. On the other hand, from the trajectory of students' physical exercise activities, the trajectory data has strong semantic information, reflecting the trajectory location association of different physical exercise activities in a certain time series [12]. Therefore, it is helpful to improve the accuracy of spatiotemporal behavior prediction by fully considering the connectivity, time series, and semantic meaning among trajectory locations. Figure 1 shows the physical exercise behavior data of college students expressed in time slices according to spatiotemporal correlation. In the temporal dimension, the content of physical exercise behaviors of college students at different moments will have a certain influence on the physical exercise behaviors of that node in future moments. In the spatial dimension, different location nodes in the campus road network have strong semantic relationships and a series of mutual influences between nodes. At the same time, location nodes will also have influence on their associated nodes at different moments. It can be seen that college students' physical exercise activities have strong correlations in both temporal and spatial dimensions, and different location points have unique semantic association relationships with each other [13]. It is a great challenge to mine behavioral characteristics from these complex, nonlinear, and semantic spatiotemporal data to achieve behavioral prediction.

Foreign scholars have also conducted empirical studies on the relationship between the two. In the literature [14], it has been shown that the duration and frequency of exercise are negatively associated with the level of anxiety and depression in college students and that the level of anxiety and depression decreases as the duration of physical activity increases. In a clinical trial, the literature [15] showed that exercise-based treatment can provide relief for patients with severe or mild depression. The literature [16] pointed out that whether or not to participate in physical activity and the frequency of participation in it will cause a significant effect on the level of psychological subhealth of our population. The literature [17] pointed out that gray system model does not map well the complex nonlinearity of physical activity behavior. Literature [18] pointed out that Elman suffers from the problem of randomly determining weights and thresholds and requires enough training data to predict physical activity behaviors well. For short data and information-poor physical exercise behavior data, support vector machine (SVM) is chosen for prediction [19].

To address the above problems, in order to better analyze the physical exercise behavior prediction of college students, this paper first defines the hierarchical mathematical relationships between physical exercise activities, behaviors, and behavior prediction of college students. By combining the strong semantic relationships and spatiotemporal correlations among behavioral data locations, semantic road networks and specific cycle segments are constructed. Then, a nonlinear decreasing convergence factor and an inertia weighting strategy based on Euclidean distance are introduced to improve the basic gray wolf algorithm. An improved gray wolf algorithm (IGWO) and support vector machine (SVM) are used to build a prediction model of college students' physical activity. The graph structure representation of college students' physical activity is established, and the output of multiple time segments is convolved and fused to obtain the behavior prediction results. Finally, we collected and constructed a data set of college students' physical activity behaviors, analyzed the behavioral prediction experiment results, identified the abnormal behaviors of college students on campus, and mined the information of personal activity habits.

The innovations and contributions of this paper are listed below.

(1) This paper combined education data mining and spatial-temporal behavior analysis to study the prediction and analysis of college students' physical exercise behavior.

(2) This paper proposes a twofold strategy to improve the basic gray wolf algorithm.

(3) SVM is optimized with the improved gray wolf algorithm.

This paper consists of five main sections: Section 1 is the introduction, Section 2 is the state of the art, Section 3 is the methodology, Section 4 is the result analysis and discussion, and Section 5 is the conclusion.

Time Segment 1   Time Segment 2   Time Segment 3   Time Segment 4   Time Segment 5

- 🟡 Sports activity area   🟠 Classroom
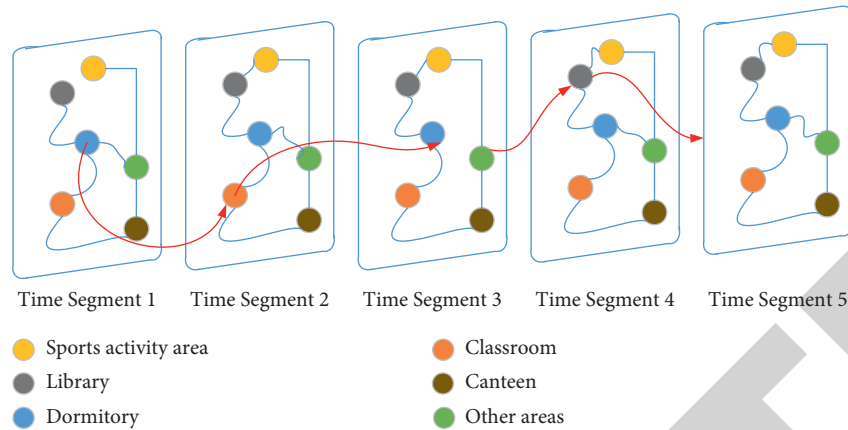- ⚫ Library   🟤 Canteen
- 🔵 Dormitory   🟢 Other areas

FIGURE 1: Diagram of spatial and temporal correlation of campus daily behavior data.

2. State of the Art

The research content of this paper is related to educational data mining, and the core is spatiotemporal data prediction. Therefore, this section introduces relevant work from two aspects: education data mining and spatiotemporal behavior prediction.

2.1. Education Data Mining. At present, with the development of education informatization and exponential growth of education big data, education data mining (EDM) has become a new interdisciplinary research field. It aims to use data mining technology to explore unique data in different educational environments, better understand students and their learning environment, and improve students' learning process and education management. Due to limited data sources, most studies on educational data mining focus on the data generated by students' online learning activities, modeling and analyzing students' online learning process and learning performance. At present, the technology and field of educational data mining are relatively single and lack multi-technology and cross-field integration. At the same time, the understanding of education data is not deep, with only some simple applications.

With the construction of smart campus and multisource collection of campus big data, analysis of students' daily behaviors on campus has become a research trend. Based on students' basic information data, literature [20] uses artificial neural network, support vector machine, and other basic prediction models to achieve the prediction of students' academic performance. Literature [21] analyzes students' complex behavioral data such as campus swipe card, WiFi access, and trajectory change; builds a multimode and multi-label learning model for privacy protection; and effectively predicts the subsidy level to be given to students. Literature [22] uses the data of students' campus check-in behavior to encode the correlation among students' individual, interest points and activities based on heterogeneous graph method. These researches mostly analyze and forecast the data of a specific problem or a kind of phenomenon but lack the prediction and mining of students' daily behaviors on campus.

2.2. Spatiotemporal Behavior Prediction. The exploration of the laws of human behavior in time and space has always been the research direction of various disciplines such as nature, economy, and society. As more and more behavioral data have been accurately recorded, scholars have been able to quantitatively analyze the temporal and spatial patterns of human behavior and its dynamics, thus improving the traditional understanding of human behavior. Although the spatiotemporal characteristics of the data are extracted, the input is limited to standard structured data, so it cannot be used to predict the behavior of graph structure.

Literature [23] proposed a general graph convolution framework, which transforms the Laplace matrix eigenvector of a graph into the spectral domain to achieve approximate solution of the problem. Literature [24] first proposed the spatiotemporal graph convolution network model and solved the time series prediction problem by using the time-domain and spatiotemporal convolution structures with fewer parameters. By designing the convolution network model of spatiotemporal synchronization graph, the spatiotemporal correlation of complex data is captured effectively by using synchronous modeling mechanism. The validity of semantic trajectory extraction method for semantic enhancement of spatiotemporal behavior data is studied.

3. Methodology

3.1. Improved Gray Wolf Algorithm and Simulation Experiment. Gray wolf optimization algorithm is a new meta-heuristic swarm intelligence algorithm based on gray wolf social behavior. Compared with the traditional algorithm, the basic gray wolf algorithm is easy to fall into the disadvantage of local optimization although it has a faster search speed. Therefore, this paper proposes two strategies to improve it.

3.1.1. Nonlinear Decreasing Strategy of Convergence Factors. The coordination between global search performance and local search performance exists in swarm intelligence algorithms, and gray wolf algorithm is no exception. Therefore, balancing local and global search performance can

improve the optimization ability of gray wolf algorithm. In the process of optimization, the cooperative coefficient vector G value affects the exploration and development of the algorithm to a certain extent. G is affected by the value of the convergence factor g. In the standard gray wolf algorithm, a decreases linearly from 2 to 0. This linear decreasing convergence factor cannot meet the balance of global search and local search performance. Therefore, this paper proposes a nonlinear convergence factor based on sine form, as shown in the following formula:

$$g = \sin\left(\frac{\pi n}{t} + \frac{\pi}{2}\right) + 1, \tag{1}$$

where n is the current number of iterations. t is the maximum number of iterations. The variation of the convergence factor a with the number of iterations is shown in Figure 2.

At the early stage of convergence, the value of the improved convergence factor g decreases relatively slowly compared to the standard algorithm. Therefore there will be a larger value of G. Expanding the search range in the early stage while ensuring the diversity of the total population of the algorithm can avoid falling into local optimum. In the late convergence stage, the improved convergence factor g decreases faster compared to the convergence factor of the standard algorithm. This also results in a smaller G value, which focuses more on local search and makes the algorithm convergence more accurate.

3.1.2. Dynamic Weight Strategy Based on Euclidean Distance.
In the standard gray wolf algorithm, α, β, and γ wolves have the same guidance for ω wolves. This will cause the algorithm to converge too slowly and make the algorithm easily fall into local optimum. In fact, the wolves in the gray wolf algorithm are based on social hierarchy. In order to make the wolves of the 3 leadership levels play their due contribution to accelerate the convergence of the algorithm and enhance the global optimality finding ability of the algorithm, this paper uses a dynamic weighting strategy based on Euclidean distance. This can improve the position update formula in the algorithm.

According to the positions I_1, I_2, I_3 updated by ω wolves under the guidance of the leading gray wolves in the iterative process, the learning rates of ω wolves to α, β, and γ wolves are obtained as m_1, m_2, and m_3, respectively, as shown in (2)–(4). Then, the position update formula becomes the form of (5) to improve the standard gray wolf algorithm.

$$m_1 = \frac{|I_1|}{|I_1| + |I_2| + |I_3|}, \tag{2}$$

$$m_2 = \frac{|I_2|}{|I_1| + |I_2| + |I_3|}, \tag{3}$$
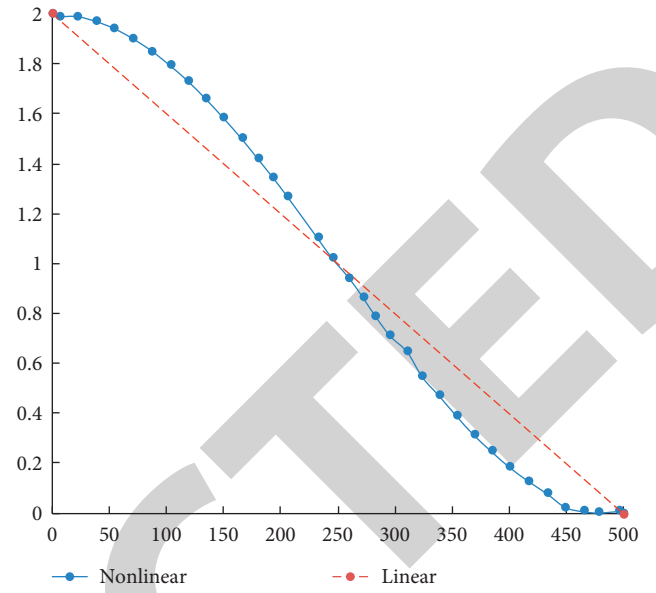
$$m_3 = \frac{|I_3|}{|I_1| + |I_2| + |I_3|}, \tag{4}$$

Figure 2: Convergence factor comparison diagram.

$$I(n+1) = \frac{m_1 I_1 + m_2 I_2 + m_3 I_3}{3}. \tag{5}$$

3.1.3. Simulation Experiment.
In order to verify the superior performance of the improved gray wolf algorithm, 2 single-peak functions and 2 multi-peak functions are used as test functions to conduct simulation experiments. In order to ensure the objective and reliable results, 20 independent experiments were conducted, and the test function is shown in Table 1.

For the above test function, set the number of wolves to 35 and the maximum iteration to 600. The CPU of the simulation experiment computer is Intel core i5–5400U, with main frequency of 4 GHz, 16 GB RAM, and Microsoft Windows 12 64-bit operating system. The computing environment is Matlab 2018(a). Through the simulation experiment, the optimization results of the four test functions are shown in Table 2. As can be seen from Table 2, IGWO algorithm has obtained good global optimal solutions in all the four test functions. For Sphere and Rastrigin functions, IGWO algorithm converges to the theoretical optimal value 0. In the optimization process of Schwefel and Ackley functions, compared with the basic GWO algorithm, IGWO algorithm has more accurate optimization results. By comparing the standard deviations of the optimal values of the four functions, it can be found that the IGWO algorithm is superior to the basic GWO algorithm. In summary, the IGWO algorithm in this paper is superior to the basic GWO algorithm in solving accuracy and stability.

3.2. Optimizing the Behavior Prediction of SVM

3.2.1. Support Vector Machines.
Support vector machine (SVM) is a machine learning algorithm based on statistical theory. Based on VC dimension theory, the introduction of

TABLE 1: Benchmark functions.

Function		The formula	Dimension	A search scope	The minimum value
Unimodal function	Sphere	$f_1(i) = \sum_{x=1}^{t} i_x^2$	35	(−150, 150)	0
	Schwefel	$f_2(i) = \sum_{x=1}^{t} i_x + \prod_{x=1}^{t} |i_x|$	35	(−15, 15)	0
	Rastrigin	$f_3(i) = \sum_{x=1}^{t} [i_x^2 - \cos(2\pi i_x) + 10]$	35	(−5.14, 5.14)	0
Multimodal function	Ackley	$f_4(i) = -20 \exp\left(-0.2\theta\sqrt{\sum_{x=1}^{t} i_x^2}\right)$ $-\exp\left(\sqrt{1/t \sum_{x=1}^{t} \cos(2\pi i_x)}\right) + 20 + e$	35	(−35, 35)	0

TABLE 2: Test result functions of different algorithms.

Function	GWO algorithm		IGWO algorithm	
	The average	The standard deviation	The average	The standard deviation
Sphere	1.59×10^{-27}	2.65×10^{-27}	0	0
Schwefel	8.81×10^{-17}	9.41×10^{-17}	1.27×10^{-208}	
Rastrigin	2.14	2.61	0	0
Ackley	1.07×10^{-13}	1.53×10^{-14}	4.46×10^{-15}	2.36×10^{-30}

structural risk minimization has obvious advantages in solving small sample problems. SVM mainly maps input vector to high-dimensional space through preselected nonlinear mapping and constructs regression estimation function in high-dimensional space, as shown in the following formula:

$$f(i) = \left[m^N \cdot \Phi(i)\right] + h, \tag{6}$$

where h is the threshold value, m is a high-dimensional feature space, and $\Phi(i)$ is a nonlinear mapping.

Due to the introduction of the structural risk minimization principle, it can be translated into the following formula:

$$\left\{ \min \frac{1}{2}\|m\|^2 + \frac{1}{2}E \sum_{x=1}^{t} \xi_x^2, s, n, j_x = \Phi^N(i) + h + \xi_x, \right. \tag{7}$$

where E is the penalty parameter and ξ is the slack variable.

To solve the optimization problem, the Lagrange operator g is introduced and the Lagrange function is constructed. The partial derivative of the Lagrange function at the saddle point is 0. By transforming the quadratic planning problem into a new pairwise problem and introducing the kernel function, the prediction model can be obtained as follows:

$$J = \sum_{x=1}^{t} \left[g_x Z(i, i_x)\right] + h. \tag{8}$$

3.2.2. IGWO-SVM Prediction Model. The operation steps of IGWO-SVM model for college students' physical exercise behavior prediction are as follows. The flowchart is shown in Figure 3.

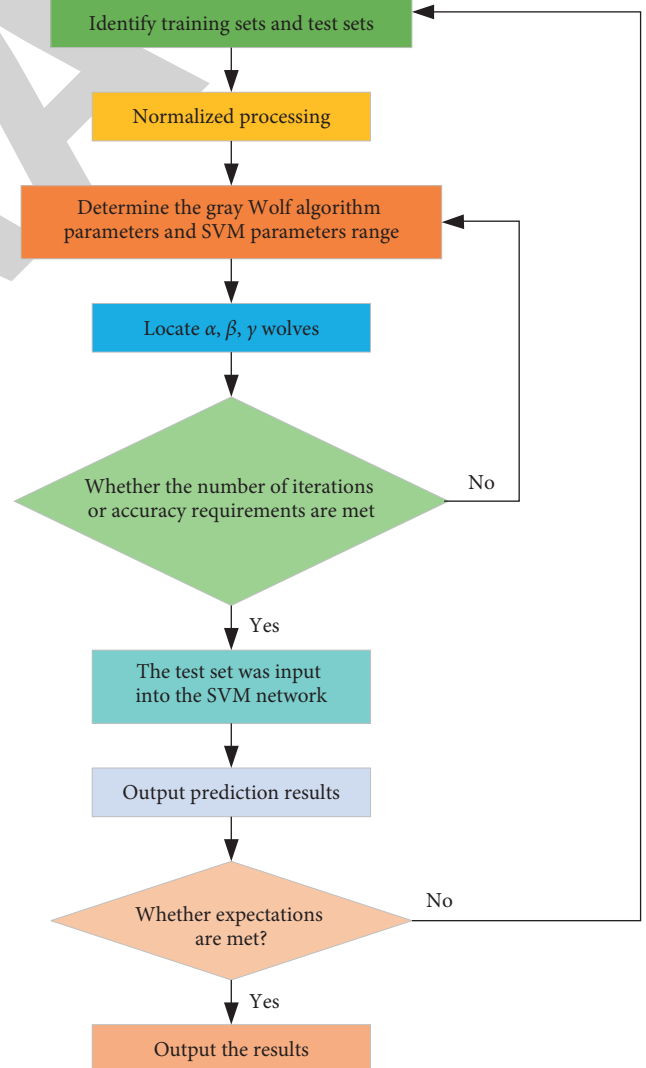(1) Determine the training set and test set, and normalize the processing.

FIGURE 3: Flowchart of IGWO-SVM prediction model.

(2) Determine the gray wolf algorithm parameters (including the number of wolves and the maximum number of iterations) and the SVM hyperparameter range (upper and lower limits of penalty parameter C and kernel parameter A).

(3) Determine the initial α, β, γ wolf position and iteration. MSE was selected as fitness function, and the training set was input into SVM network to calculate the updated position of ω wolves under the guidance of gray wolves. The learning rate of ω wolves to α, β, and γ wolves was updated by (2)–(4). Track and kill the prey according to (5), use the improved gray wolf algorithm to determine the C and A values, and obtain the fitness values of C and A that meet the requirements.

(4) If the number of iterations or accuracy requirements are met, the prediction model can be obtained. Otherwise, repeat the process.

(5) Bring the test set into the trained SVM network and output the prediction results.

(6) Compare the predicted result with the expected value. Relative error and absolute error were used to evaluate the accuracy of IGWO-SVM slope stability prediction model.

4. Result Analysis and Discussion

4.1. Behavioral Data Set Construction and Data Preprocessing. In order to verify the effectiveness of the proposed method in the analysis of physical activity behavior on campus, an experiment was conducted using real data. The data used in this paper were obtained from the data of students in a central university in China. All the student numbers in this data are generated by the privacy-protected student number codes. In addition, considering the uniqueness of some individuals' spatial and temporal patterns of behavior, the identifiability of the data was further reduced in the data construction stage. For example, the specific information was anonymized, the specific room number was blurred, and the location was regionalized. After the processing of raw data, it is difficult to reidentify individual students accurately. Examples of specific student campus daily activity data are shown in Table 3.

According to this paper, the formulaic definition of physical exercise behavior is shown in (9) and (10). Firstly, the physical exercise data are clustered to construct the data set of campus physical exercise behavior, which lays a foundation for semantic information fusion on campus. Field information such as activity time, region, location, and description in the database is used to obtain the campus behavior data set with accurate behavior meaning, as shown in the last column in Table 3. This provides the basis for data preprocessing and network input. This article is created from active field information. While constructing the data set, the activity records with clear meaning are screened, and the activity data without practical meaning is deleted.

$$g_x = (p_x, n_x, l_x, d_x), \tag{9}$$

$$G = \{g_1, g_2, \ldots, g_n\}, \tag{10}$$

where $p_x \in P$ represents the performer of physical exercise, n_x represents the time of physical exercise, $l_x \in L$ represents the place of physical exercise, $d_x \in D$ represents the specific description of physical exercise, and G represents the collection of all physical exercise in the data.

In the preprocessing of data set, this paper is divided into three steps.

(1) Digitize the data labels. The specific step is to digitize the category marking features in the original data. For example, the regional category labels in the data of students' campus physical exercise behavior are numerically analyzed. This method keeps the basic characteristics of data and does not change the original data information. However, the singularities in the data will greatly affect the experimental accuracy, so the data should be further processed.

(2) Standardize the data. In data analysis, it is generally required that the original data meet normal distribution. In order to ensure the accuracy of data analysis and reduce the computational complexity, the consistency of the influence degree of different range features on the results should be ensured. The formula for data standardization is as follows:

$$k = \frac{(i - \mu)}{\sigma}\sigma, \tag{11}$$

where μ and σ are the mean and standard deviation of the features in the original data set, respectively. For example, the exercise duration, exercise items, and other numerical characteristics of students' campus physical exercise behavior data can be standardized to make the data standardized and more conducive to data analysis.

(3) Normalize data: after the numerical value of the category tag features in the original data, the linear function is normalized, as shown in the following formula:

$$I_{norm} = \frac{(I - I_{\min})}{(I_{\max} - I_{\min})}, \tag{12}$$

I is the original data. $I_{\max}$ and $I_{\min}$ are the maximum and minimum values in the original data, respectively. I_{norm} is the characteristic result after normalization. In this way, equal scale of data can be achieved to avoid the influence of outliers on data analysis.

During the experiment, the data of 396 students from a college from September 1, 2020, to January 20, 2021, totaled about 140 days, 20 weeks, and 5 months, and a total of 7605741 activity records were used. In addition, when constructing the semantic network of college students' physical exercise, in order to better connect the association

TABLE 3: Active data instances and behavior labels in the database.

Student id	Date	Time	Area	Description	Behavior
Abc120	2021/9/6	6 : 30	Playground	Physical exercise	Physical activity behavior
Abc121	2021/9/6	8 : 07	Dining area	Rice congee	Diet behavior
Abc122	2021/9/6	8 : 11	Dormitory area	Rest	Rest behavior
Abc123	2021/9/6	9 : 52	The library district	I like the book	Borrowing behavior
Abc124	2021/9/6	10 : 15	Teaching and research area	Fundamentals of education	A class act
Abc125	2021/9/6	11 : 39	Teaching and research area	Fundamentals of education	A class act
Abc126	2021/9/6	12 : 30	Dining area	Dumplings	Diet behavior
Abc127	2021/9/6	13 : 00	Dormitory area	Rest	Rest behavior
Abc128	2021/9/6	14 : 30	Teaching and research area	Higher mathematics	A class act
Abc129	2021/9/6	18 : 40	Basketball court	Play basketball	Physical activity behavior

between time-series data, according to the behavioral characteristics of physical exercise data set, the semantic network was constructed as a graph data structure containing time points and spatial positions, that is, 9 campus exercise areas and 30 time points. The dimension of adjacency matrix of graph structure is 168×168. Based on this, the convolution operation of physical exercise behavior characteristic map of three time segments is realized.

4.2. Experimental Methods and Analysis. Since there are few studies on the prediction of students' campus physical exercise behavior, this paper adopts several typical machine learning prediction models such as literature [25], literature [26], literature [27], and literature [28] to conduct comparative analysis of experimental results with the method in this paper.

The network structure of the model in this paper includes three time segments, which are daily segment, weekly segment, and monthly segment. Each segment has one level of standard two-dimensional convolution of time dimension and two levels of spatial convolution. Experiments were carried out on data sets; 70% of the physical exercise behavior data were set as training set, and 30% of the physical exercise behavior data were set as verification set. The main evaluation indexes include accuracy (Acc), precision (Pre), recall rate (Rec), and $F1$. The calculation formulas are as follows:

$$\text{Acc} = \frac{TP + TN}{TP + FP + FN + TN},$$

$$\text{Pre} = \frac{TP}{TP + FP},$$

$$\text{Rec} = \frac{TP}{TP + FN},$$

$$F1 = \frac{2 \times \text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}}.$$

(13)

TP represents the number of samples whose positive class behavior is predicted by the model. FP represents the number of samples whose negative class behavior is predicted to be positive by the model. FN represents the number of samples whose positive class behavior is predicted to be negative class by the model. TN represents the number of samples whose negative class behavior is predicted to be negative.

The Acc and $F1$ values predicted by the different model behaviors in the experiments of this paper are shown in Table 4. The accuracy and $F1$ of the models in this paper are higher than those in the literature [25–28]. This proves that the model in this paper is more effective in analyzing the data after considering the temporal period correlation and spatial semantic correlation together. Compared with four typical machine learning prediction models in the literature [25–28], this model has obvious advantage, which proves that using this model to solve the problem of campus road network data is more consistent with the basic structure of the data. Compared with the literature [25–28], this paper's model has the highest prediction accuracy for college students' physical exercise behavior. This is because the model in this paper considers the campus semantic features at the spatial level. Meanwhile, the periodic features of daily, weekly, and monthly segments of college students' physical exercise behaviors are considered at the temporal level to achieve more accurate feature extraction in the prediction process.

Figure 4 shows the Acc of different models for different categories of physical exercise behavior prediction. The model in this paper outperformed other models in different behavior categories. The highest prediction accuracy for physical exercise behavior proves that it can effectively obtain the periodicity of students' campus physical exercise behavior data and improve the accuracy of behavior prediction.

Table 5 compares the number of parameters of the above four methods and the training time for obtaining the optimal behavior prediction effect. Since the semantic characteristic information of road network is considered in the construction of adjacency matrix, the number of parameters is larger than those of the other models, and the training process is also the most time-consuming. The attention mechanism was added in [25, 26] to increase the number of references. It also improves the accuracy of predictions. Literature [28] synchronizes time and space convolution to improve algorithm efficiency. However, the model in this paper is more suitable for the prediction of students' physical exercise behavior that can construct semantic network. It can extract more valuable semantic information of physical exercise, and its accuracy is better than that of the other methods. Longer training duration is also acceptable for the mining of massive physical exercise daily behavior data.

Table 4: Experimental results of different algorithms in data sets.

Model	Acc (%)	F1 (%)
Literature [25]	84.34	82.37
Literature [26]	81.08	80.17
Literature [27]	78.62	75.35
Literature [28]	86.45	85.17
Proposed	90.45	90.27

Table 5: Comparison of training time and model parameters.

Model	Training time (s)	Number of parameters
Literature [25]	956.84	7451
Literature [26]	1030.65	7583
Literature [27]	979.16	7918
Literature [28]	757.65	7649
Proposed	1075.69	10263

Table 6: Analysis of students' daily behavioral habits.

Category	Time											
	7	8	9	10	11	12	13	14	15	16	17	18
Other acts	0	0	0	0	0	0	0	0	0	0	0	0
Exercise behavior	0	0	0	0	0	0	0	0	0	0	0	0
Reading behavior	0	0	0	0	0	0	0	0	0	0	0	0
Self-study behavior	0	0	0	0	0	0	0	1	1	1	1	1
Classroom behavior	0	1	1	1	1	1	0	1	1	1	1	1
Eating behavior	0	1	1	1	1	1	0	1	1	1	1	1
Rest behavior	1	0	0	0	0	0	1	0	0	0	0	0

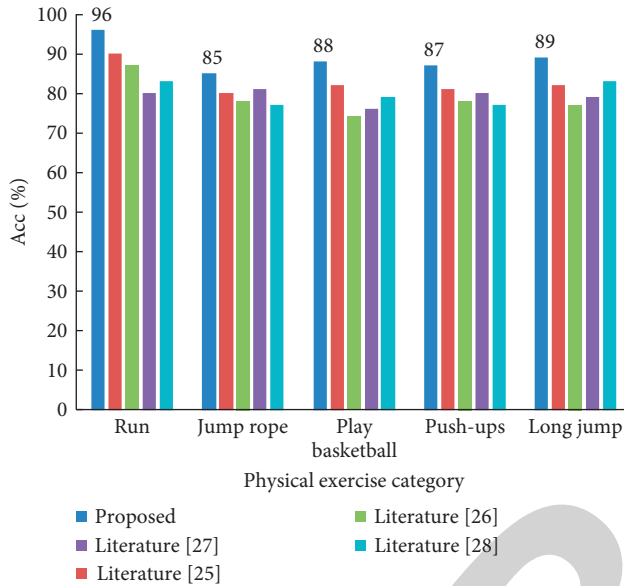*Note.* 0 means the behavior did not happen. 1 means the behavior happened.

Figure 4: Comparison of prediction accuracy of different types of behavior.

4.3. Individual Behavior and Habit Information Mining. Taking a specific student as the research object, we screened a total of 6845 campus activity data of the student from the data set. The model with better training effect was used to predict the student's behavior on the day, week, and month, and the behavior rule was found by data correlation analysis. Table 6 shows the statistical analysis of behaviors in some time periods of a day. It can be found that the behavior of the student at 7 : 00 is mainly resting behavior; the behavior between 8 : 00 and 12 : 00 is mainly class behavior and eating behavior; and the behavior between 14 : 00 and 18 : 00 is mainly class behavior, self-study behavior, and eating behavior. In the statistical analysis of behavior rules in different time periods, there are few deviating behaviors from the mean, and the whole has no influence on the statistics of the mean. From a macroscopic point of view, it is proved that the student has good daily behavior habits. Habits refer to regular behavior patterns, which are formed by the generalization of behavior sequences with repeated cycles and overlapping time and space. This paper establishes strong correlation analysis between individual habit discovery and campus behavior data. This paper deduces the nature of behavior by data-driven method and obtains the correlation between personalized behavior habits and academic growth. Therefore, suggestions and references based on daily behavior level are provided for personalized training of college students.

5. Conclusion

Aiming at overcoming the problems of insufficient data mining of college students' physical exercise behavior and insufficient consideration of temporal and spatial correlation in existing research methods, this paper proposes an improved gray wolf algorithm and support vector machine algorithm to predict and analyze college students' physical exercise behavior. The algorithm adopts two improved strategies to optimize the gray wolf algorithm to accelerate the convergence of the gray wolf algorithm and obtain a higher precision solution. By combining IGWO algorithm with support vector machine (SVM), the improved gray wolf optimization algorithm has strong global search ability and high search efficiency to determine the parameters of SVM, so that the nonlinear mapping ability of SVM is strengthened. Experiments show that the accuracy and F1 of the proposed algorithm are higher than those of other comparison algorithms. The algorithm in this paper can not only predict college students' physical exercise behavior with good accuracy, but also deduce the correlation between college students' personalized behavior habits and their academic growth according to the results of behavior prediction. Therefore, it provides suggestions and references based on daily behavior level for personalized training of college students, in order to correct abnormal behaviors of students, so that they can better complete their studies. The next step is to increase the richness and granularity of the data set. At the same time, more data preprocessing methods are tried to make the prediction more accurate.

Data Availability

The labeled data set used to support the findings of this study is available from the author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] M. Grasdalsmoen, H. R. Eriksen, K. J. Lønning, and B Sivertsen, "Physical exercise and body-mass index in young adults: a national survey of Norwegian university students," *BMC Public Health*, vol. 19, no. 1, pp. 1354–1359, 2019.

[2] S. Teleki, A. N. Zsidó, L. Lénárd, A. Komocsi, E. C. Kiss, and I Tiringer, "Role of received social support in the physical activity of coronary heart patients: the Health Action Process Approach," *Applied Psychology: Health and Well-Being*, vol. 14, no. 1, pp. 44–63, 2022.

[3] F. Rodrigues, D. S. Teixeira, L. Cid, and D. Monteiro, "Have you been exercising lately? Testing the role of past behavior on exercise adherence[J]," *Journal of Health Psychology*, vol. 26, no. 10, pp. 1482–1493, 2021.

[4] B. Behzadnia, P. J. C. Adachi, E. L. Deci, and H Mohammadzadeh, "Associations between students' perceptions of physical education teachers' interpersonal styles and students' wellness, knowledge, performance, and intentions to persist at physical activity: a self-determination theory approach," *Psychology of Sport and Exercise*, vol. 39, pp. 10–19, 2018.

[5] F. Rodrigues, D. S. Teixeira, H. P. Neiva, L. Cid, and D Monteiro, "The bright and dark sides of motivation as predictors of enjoyment, intention, and exercise persistence," *Scandinavian Journal of Medicine & Science in Sports*, vol. 30, no. 4, pp. 787–800, 2020.

[6] X. Du, M. Zhang, B. E. Shelton, and J. L. Hung, "Learning anytime, anywhere: a spatio-temporal analysis for online learning," *Interactive Learning Environments*, vol. 30, no. 1, pp. 34–48, 2022.

[7] G. Gonzales, E. Loret de Mola, K. A. Gavulic, T. McKay, and C Purcell, "Mental health needs among lesbian, gay, bisexual, and transgender college students during the COVID-19 pandemic," *Journal of Adolescent Health*, vol. 67, no. 5, pp. 645–648, 2020.

[8] E. L. Caputo and F. F. Reichert, "Studies of physical activity and COVID-19 during the pandemic: a scoping review," *Journal of Physical Activity and Health*, vol. 17, no. 12, pp. 1275–1284, 2020.

[9] J. Wang, "A review of the development of the integration strategy of information technology and education in the four countries of the United States, britain, China, and Singapore," *Science Insights Education Frontiers*, vol. 9, no. 2, pp. 1283–1303, 2021.

[10] Y. Wang, Q. Chen, D. Gan, J. Yang, D. S. Kirschen, and C Kang, "Deep learning-based socio-demographic information identification from smart meter data," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2593–2602, 2019.

[11] H. Yuan and G. Li, "A survey of traffic prediction: from spatio-temporal data to intelligent transportation," *Data Science and Engineering*, vol. 6, no. 1, pp. 63–85, 2021.

[12] H. Cao, F. Xu, J. Sankaranarayanan, Y. Li, and H. Samet, "Habit2vec: trajectory semantic embedding for living pattern recognition in population," *IEEE Transactions on Mobile Computing*, vol. 19, no. 5, pp. 1096–1108, 2020.

[13] Paul Portner, K. von Heusinger, and C. Maienborn, "Possessives and Relational nouns," *Semantics-Noun Phrases and Verb Phrases*, pp. 177–203, 2019.

[14] P. J. Puccinelli, T. S. da Costa, A. Seffrin, C. A. B. de Lira, R. L. Vancini, and P. T. Nikolaidis, "Reduced level of physical activity during COVID-19 pandemic is associated with depression and anxiety levels: an internet-based survey," *BMC Public Health*, vol. 21, no. 1, pp. 1–11, 2021.

[15] J. Á Bellón, S. Conejo-Cerón, A. Sánchez-Calderón et al., "Effectiveness of exercise-based interventions in reducing depressive symptoms in people without clinical depression: systematic review and meta-analysis of randomised controlled trials," *The British Journal of Psychiatry*, vol. 219, pp. 1–10, 2021.

[16] H. Wang, X. Dai, J. Wu, X. Wu, and X. Nie, "Influence of urban green open space on residents' physical activity in China," *BMC Public Health*, vol. 19, no. 1, p. 1093, 2019.

[17] J. Schoukens and L. Ljung, "Nonlinear system identification: a user-oriented road map," *IEEE Control Systems*, vol. 39, no. 6, pp. 28–99, 2019.

[18] Z. Li, B. Zhu, Y. Dai, W. Zhu, Q. Wang, and B. Wang, "Thermal error modeling of motorized spindle based on Elman neural network optimized by sparrow search algorithm," *International Journal of Advanced Manufacturing Technology*, vol. 121, pp. 1–18, 2022.

[19] F. Kang, J. Li, and J. Dai, "Prediction of long-term temperature effect in structural health monitoring of concrete dams using support vector machines with Jaya optimizer and salp swarm algorithms," *Advances in Engineering Software*, vol. 131, pp. 60–76, 2019.

[20] S. U. Damuluri, K. Islam, P. Ahmadi, and N. S. Qureshi, "Analyzing navigational data and predicting student grades using support vector machine," *Emerging Science Journal*, vol. 4, no. 4, pp. 243–252, 2020.

[21] Y. X. Jin and J. M. Dewaele, "The effect of positive orientation and perceived social support on foreign language classroom anxiety," *System*, vol. 74, pp. 149–157, 2018.

[22] S. Buecker, T. Simacek, B. Ingwersen, S. Terwiel, and B. A Simonsmeier, "Physical activity and subjective well-being in healthy individuals: a meta-analytic review," *Health Psychology Review*, vol. 15, no. 4, pp. 574–592, 2021.

[23] D. Cao, Y. Wang, J. Duan et al., "Spectral temporal graph neural network for multivariate time-series forecasting," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17766–17778, 2020.

[24] J. An, L. Guo, W. Liu et al., "IGAGCN: information geometry and attention-based spatiotemporal graph convolutional networks for traffic flow prediction," *Neural Networks*, vol. 143, pp. 355–367, 2021.

[25] A. Javeed, S. Zhou, L. Yongjian, I. Qasim, A. Noor, and R Nour, "An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection," *IEEE Access*, vol. 7, pp. 180235–180243, 2019.

[26] C. Manapragada, G. I. Webb, and M. Salehi, "Extremely fast decision tree[C]//," in *Proceedings of the 24th ACM SIGKDD*

International Conference on Knowledge Discovery & Data Mining, pp. 1953–1962, London, United Kingdom, 2018.

[27] S. Karatay and M. O. A. Algahani, "Marmara depremi ve güneş tutulmasının naive bayes sınıflayıcısı ile i?statistiksel analizi[J]," *Avrupa Bilim ve Teknoloji Dergisi*, vol. 2021, no. 23, pp. 643–648, 1999.

[28] G. Xiao, R. Wang, C. Zhang, and A. Ni, "Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks," *Multimedia Tools and Applications*, vol. 80, no. 15, pp. 22907–22925, 2021.