Prediction and Analysis of Higher Vocational Physical Education Achievements Based on Support Vector Machine

Xiaoyan Hu(✉)

Quality Education Center, Yunnan Land and Resources Vocational College, Kunming, China
278778593@qq.com

Abstract. In recent years, people's material life has been greatly improved. However, the physical quality of college students has declined to varying degrees. The results of physical education curriculum reflect the physical fitness of students, and accurately predict the results of physical education curriculum, which is conducive to the scientific and reasonable opening of college physical education curriculum to improve the physical fitness of students. In this paper, machine learning algorithm is used to predict the performance of college students' physical education courses. The prediction model uses support vector machine, and particle swarm optimization algorithm is used to optimize the parameters of support vector machine. The method is applied to the actual data, and the results show that the method has high prediction accuracy and good universality for college students' scores of multiple physical education courses. Teachers can use this method to improve teaching methods and improve teaching quality; The teaching management department can use this method to reasonably plan the physical education curriculum in colleges and universities.

Keyword: support vector machine university sports score particle swarm optimization algorithm score prediction

1 Introduction

Although people's living standards have been improving in recent years, college students' material life has become richer, but this has resulted in a decline in physical activity levels, and the physical fitness of some college students has decreased to varying degrees [1, 2]. Therefore, physical education for college students plays an important role in improving their physical fitness. Predicting the physical performance of college students can help the university sports management department to plan and develop relevant courses, as well as establish a more scientific training mechanism. As a result, there has been increasing attention paid to developing a high-precision sports performance prediction model [3].

To predict college sports performance, statistical methods are used, but there are many factors that can affect performance, making it a complex calculation process. As a result, accuracy is often low [3, 4]. To address this, researchers have developed

X. Yuan et al. (Eds.): ICEKIM 2023, AHCS 13, pp. 1983–1989, 2023.
https://doi.org/10.2991/978-94-6463-172-2_220

automatic prediction models using information technology and artificial intelligence [5]. These models can be classified as linear or nonlinear regression models. The most used linear model is the multiple linear regression model, which works well when there are few influencing factors [6]. Nonlinear models, such as neural networks, have greater potential for capturing the unpredictability of university sports achievement [7]. However, they require many samples, which can be difficult to obtain due to limited sports achievement history [8]. University sports achievement is influenced by multiple factors, making it highly intricate and unpredictable. Linear models are inadequate for capturing this unpredictability, whereas nonlinear models show promise.

To enhance the accuracy of predicting college students' sports performance, this study proposes a prediction model based on support vector machine (SVM). SVM is a relatively new machine learning algorithm that operates on the principle of structural risk minimization, which does not require a large sample size like neural networks. In fact, it has shown remarkable fitting capability even with small sample sizes. We employ the particle swarm optimization algorithm to optimize the model parameters of SVM, and then utilize the model for predicting sports performance in a university. The findings demonstrate that SVM overcomes the limitations of traditional models and significantly improves the prediction accuracy of college sports performance.

2 Prediction Model and Algorithm of College Sports Performance

2.1 Support Vector Machine

Support vector machine (SVM) are a machine learning algorithm that have fewer restrictions and are not prone to overfitting compared to other traditional algorithms. They are particularly suitable for working with small sample sizes and nonlinear data. Based on these advantages, SVMs were selected in this study to model and predict college students' sports performance.

Set the sample set of college students' sports achievements as $X = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}, i = 1, 2, \cdots, n$. The expression form of support vector machine is:

$$f(x) = w \cdot \phi(x) + b \tag{1}$$

where w and b are parameters of support vector machines.

In order to establish the prediction model of college students' sports performance, the most reasonable sum value must be found. Therefore, according to the principle of structural risk minimization, it is transformed into the following form:

$$\min \frac{1}{2}||w||^2 + C\frac{1}{k}\sum_{i=1}^{k}\varepsilon(f(x_i) - y_i)$$
$$s.t.\, \varepsilon(f(x_i) - y_i) = \begin{cases} |f(x_i) - y_i| - \varepsilon, & |w \cdot \phi(x) + b - y_i| \geq \varepsilon \\ 0, & |w \cdot \phi(x) + b - y_i| < \varepsilon \end{cases} \tag{2}$$

where: ε is regression error; C is the penalty parameter of the error.

In order to simplify the solution process and reduce the computational complexity of modeling, the other factors ε and ε^* are introduced to obtain the quadratic programming form as follows:

$$\min_{w,b,\xi_s,\xi_i^*} \frac{1}{2}||w|| + C\sum_{i=1}^{l}\left(\xi_i + \xi_i^*\right)$$

$$\text{s.t.} \begin{cases} y_i - w \cdot \phi(x) - b \le \varepsilon + \xi_i, \;\; \varepsilon_i \ge 0; i = 1, 2, \cdots, n \\ w \cdot \phi(x) + b - y_i \le \varepsilon + \xi_i^*, \; \varepsilon_i^* \ge 0; i = 1, 2, \cdots, n \end{cases} \tag{3}$$

Using Lagrange multiplies a_i and a_i^* further transforming Eq. (3), we can get:

$$\min_{a^{(*)}\in R^{2l}} \frac{1}{2}\sum_{i,j=1}^{n}\left(a_i^* - a_i\right)\left(a_j^* - a_j\right)k\left(x_i, x_j\right) + \varepsilon\sum_{i=1}^{n}\left(a_i^* + a_i\right) - \sum_{i=1}^{n}y_i\left(a_i^* - a_i\right) \tag{4}$$

where $k\left(x_i, x_j\right)$ represents the kernel function.

The regression function of support vector machine can be described as:

$$f(x) = \sum_{i=1}^{n}\left(a_i - a_i^*\right)(\phi(X_i), \phi(X)) + b \tag{5}$$

The kernel function is selected as RBF function, and its expression is:

$$k\left(x_i, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \tag{6}$$

where σ represents the parameters of RBF.

2.2 Particle Swarm Optimization

The current optimal solution of particle swarm optimization algorithm is pbest, and the optimal solution of the current population is gbest. The fitness function to describe the degree of individual particle is:

$$\text{fitness} = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{D}\left(y_{ij} - t_{ij}\right)^2 \tag{7}$$

$x_{i,d}^k$ and $v_{i,d}^k$ are the speed and position of i th particle at the k th iteration. Their update method is:

$$v_{id}^{k+1} = \varpi v_{id}^k + c_1\text{Rand}\left(p_{id} - x_{id}^k\right) + c_2\text{Rand}\left(p_{\text{gbest}}^k - x_{id}^k\right) \tag{8}$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^k \tag{9}$$

where: c_1, c_2 is the learning factor; Rand is a random number; ϖ is the inertia weight.

In order to solve the problem of determining the parameters of support vector machines, particle swarm optimization algorithm is used to optimize the training process of support vector machines online, determine the optimal parameters C and σ, according to the optimal parameters C and σ improve the prediction accuracy of college students' sports performance.

2.3 Prediction Steps of College Sports Performance

We use the following steps to predict college sports performance.

(1) Collect the historical data of a sports project and process it.
(2) Initialize particle swarm. The position vector of each particle includes a set of parameters C and σ.
(3) According to each set of parameters C and σ the training samples of college students' sports performance, support vector machines are used to learn, and the fitness function value of each particle is calculated.
(4) Sort all particles according to the fitness function value, and replace the current optimal solution pbest and the current population optimal solution gbest.
(5) Update the speed and position of the particle swarm.
(6) The number of iterations of particle swarm optimization algorithm increases.
(7) If the number of iterations exceeds the preset maximum, the algorithm execution will be terminated, otherwise, return to step (3) to continue the optimization operation.
(8) The optimal parameters C and σ values of support vector machines are obtained according to the group optimal solution gbest.
(9) According to C and σ re-training college students' sports performance, a prediction model of college students' sports performance based on support vector machine is established.
(10) Test and analyze the performance of college students' sports performance prediction model by using college students' sports performance test samples.

3 Prediction Model Application and Result Analysis

3.1 College Students' Sports Performance Data

To comprehensively analyze the performance of SVM, Yunnan Vocational College of Land and Resources was chosen as the subject of this experiment. The 50m run results were selected as the first test, with a total of 400 results shown in Fig. 1. Among these results, 370 50m results were used to establish the college students' sports performance prediction model, while the remaining 30 were utilized to test the model's generalization performance.

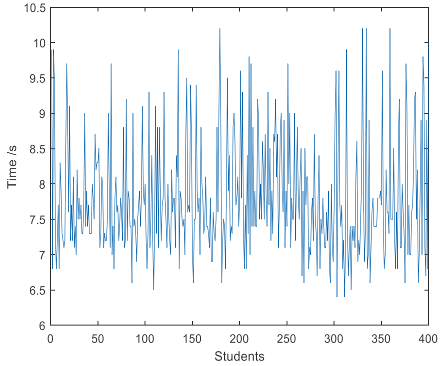

Fig. 1. Boys' 50-m running results

Table 1. Advantages of different models

Models	Prediction accuracy /%
Model in [3]	82.30
Model in [8]	89.68
This paper	97.21

3.2 Model Superiority Test

To compare the prediction model of college students' sports performance based on machine learning algorithms, we conducted comparative experiments using the linear regression model described in literature [3] and the neural network described in literature [8]. We calculated the prediction accuracy of their boys' 50m running performance, which is presented in Table 1. We analyzed and compared the results in Table 1 to draw the following conclusions: the linear regression model, among all the prediction models of college students' sports performance, demonstrated the lowest prediction accuracy; the neural network has improved the prediction accuracy of college students' performance to a certain extent; the SVM has shown higher prediction accuracy for college students' sports performance. As a result, SVM has clear advantages over the other models.

3.3 Model Validity Test

The results of predicting the performance of 30 50m runs in the test sample are presented in Fig. 2. These results reveal that SVM can also provide accurate predictions for 30m runs.

3.4 Generality Analysis

The research aimed to evaluate the generalizability of SVM-based prediction model for college students' sports performance by testing its accuracy in predicting the results

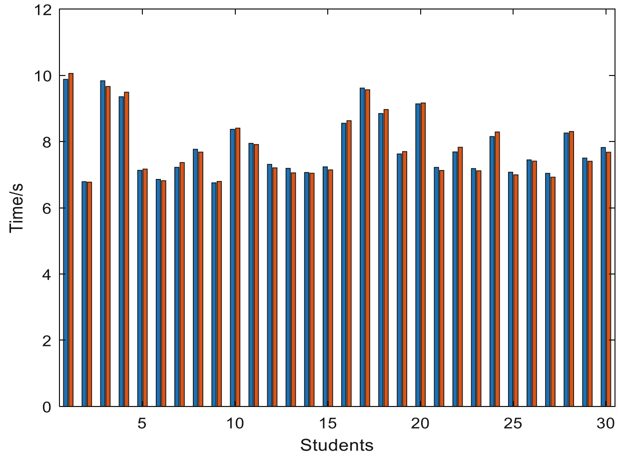

Fig. 2. Prediction results of boys' 50m performance

Table 2. Universality test of college students' sports performance prediction model

type	Prediction accuracy /%
Male triple jump	92.81
Boys' 1000m run	92.95
Boys pull up	94.81
Girls' 800m run	97.02
Girls standing long jump	96.10

of various sports events, including boys' triple jump, boys' 1000m run, boys' pull-up, girls' 800m run, and girls' standing long jump, at Yunnan Vocational College of Land and Resources. The prediction accuracy of the model was assessed and presented in Table 2. The prediction accuracy for all college students' sports achievements exceeded 92%.

4 Conclusion

Physical training is crucial for college students to improve their physical fitness. To create effective training programs for universities, it is essential to predict physical performance accurately. This study proposes a machine learning-based prediction model to improve the accuracy of existing models for predicting college students' sports performance. The model uses support vector machine, which is the top-performing machine learning algorithm, to fit the sports performance data of college students. The model shows robust performance when applied to specific examples, indicating its potential for various college sports performance predictions. Moreover, the model generates highly dependable predictions, providing valuable insights for college sports training.

References

1. Xu Jing. Competitive characteristics and strength analysis of China's top men's decathlon athletes [J] Journal of Guangzhou Institute of Physical Education, 2008, 28 (5): 61 - 64
2. Zhao Bo. Research on the Analysis and Prediction of the Performance of the World's Top Men's Decathlon Athletes [J] Sports Culture Guide, 2013 (3): 76 - 79
3. Zhao Bingjun. Research on the establishment of training model of Chinese men's long jump by multiple regression method [J]. Journal of Xi'an Institute of Physical Education, 2001, 18 (1): 81 - 82
4. Zhou Wei. Grey correlation analysis of the triple jump technique of Chinese and foreign elite women triple jumpers [J]. Journal of Jilin Institute of Physical Education, 2004, 20 (3): 64 - 65
5. Deng Meilan. Research on the grey prediction model of the top three men's decathlon results in the Olympic Games [J]. Mathematical Practice and Understanding, 2009 (18): 44 - 49
6. Wei Chunling. Factor analysis and grey system regression prediction of decathlon perfor-mance in China [J]. Sports Science, 2004, 24 (11): 66 - 69
7. Sun Qun, Liu Guobi, Cheng Wei, et al. Prediction of Liu Xiang's 110m hurdle performance based on fuzzy neural network [J]. Journal of Chongqing University of Science and Tech-nology (Natural Science Edition), 2011, 10 (2): 104 - 107
8. Wang Zongping, Sun Guang. Empirical study on prediction of sports performance using BP neural network algorithm [J]. Journal of Nanjing Institute of Physical Education, 2006, 20 (4): 109 - 111