International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-19

https://doi.org/10.70917/ijcisim-2026-0403
Article

Research on College Students' Physical Education Teaching and Training Load Control Methods Empowered by Intelligent Algorithms

Rongchao Zou [1,*]

[1] Institute of Marxism, Guangzhou Institute Of Technology, Guangzhou, Guangdong, 510075, China
* Correspondence author: zrc711@126.com

Abstract: With the improvement of people's living standards, sports and health have become the focus of attention, and sports monitoring systems are increasingly favored. In this paper, we design the training load control method for college students' sports teaching, propose the linear fitting sports data feature fusion method and the weighted matrix decomposition recommendation algorithm WSVDLFM based on the implicit semantic model to realize the intelligent monitoring of sports and personalized sports guidance recommendation. The experimental results show that the sports monitoring system based on the proposed data feature fusion method can effectively realize the instant sampling and fusion optimization of sports training data, and the data collection accuracy is always maintained above 92%. Meanwhile, the WSVDLFM algorithm is able to realize the accurate recommendation of sports guidance, and outperforms other comparative algorithms in both HR and NDCG. In addition, the proposed sports training method "Fun Interval Running" can maximize the efficiency of students' sports training under the premise of controlling students' training load through real-time monitoring of heart rate and pace, which is of great significance for college students' sports teaching.

Keywords: linear fitting; data feature fusion; implicit semantic modeling; weighted matrix decomposition recommendation; training load control

1. Introduction

At present, the overall physical fitness of the Chinese people has shown a downward trend, especially the college students' physical fitness decline is more serious [1]. For a long time, some colleges and universities often lack effective attention to load content validity in physical education, resulting in some load training of students does not play any role, and even cause a certain degree of damage to their bodies, thus increasing the incidence of risk [2-5]. And most of the sports training programs and methods are developed by physical education teachers based on their subjective will, which makes it difficult to develop reasonable and effective training load control methods according to the students' own actual situation, thus leading to a lower effect of sports training [6-8]. In this case, it is more urgent to explore the intelligent teaching decision-making system suitable for college students' sports training load control.

From the results of a large number of practices, students can better realize the effective combination of speed and endurance only if they continuously make adjustments to the training intensity according to the changes in their body response within the appropriate load intensity interval [9-11]. This requires teachers to make a comprehensive consideration of the hidden connotations of the training load, establish a comprehensive scientific assessment system, and develop more personalized training methods for different trainers [12]. With the continuous progress of science and technology, people's understanding of the organism is also increasing, and at the same time, the load training in the field of sports also has a certain theoretical basis, and analyzes the nature of physical training from a more comprehensive perspective [13-14]. As the new generation of digital technology is continuously embedded in the field of sports, the advantages of algorithmic regulation of the training process, refined monitoring of the training process, and scientific enhancement of the training efficiency have become the key to solving the obstacles of the reality of physical training for college students [15-17]. Only by combining intelligent algorithms with physical education can we effectively improve the lack of load control dimension in traditional training methods and cultivate more talents with comprehensive development of morality, intelligence, physical fitness and aesthetics [18-20].

The purpose of load training is to better improve the conditions of the organism, and to make the human organism produce a benign response by controlling the load, so as to enhance the athletic ability of the organism [21-22]. At present, there exists a large number of studies in the academic world on the methods of controlling the load of training for athletes, and by analyzing the condition of the human organism and carrying out load training that matches it, a better training effect can be achieved. Literature [23] shows that optimizing the training plan including training load control and recovery process development can help athletes to play their best competitive state at a specific time, and proposes a training load analysis model for athletes based on radial basis neural network (RBFNN), which can provide valuable references for the next step of training by accurately determining the state of their exercise load. Literature [24] developed a semi-automated processing tool to process athletes' heart rate monitoring data to quantify their exercise training load. Literature [25] conducted further research on the heart rate monitoring method of training load, and proposed to use wavelet transform technology to filter out the background noise of the athlete's heartbeat during exercise, so as to extract more accurate heart rate fluctuation signals, and lay the data foundation for the design of the training load control method. Literature [26] points out that external training load monitoring methods using average heart rate as an indicator are not reliable under complex training conditions, while the determination of athletes' internal training load, although it can accurately measure their athletic ability, often requires laboratory tests for calibration, so training load assessment methods based on wearable sensors have more advantages. Literature [27] constructed a machine learning model to predict the physical changes of athletes in soccer training, used 18Hz global positioning system to collect motion variables to reflect the level of athletes' training loads, and based on the assessment results of the proposed machine learning model, a more scientific training program can be developed for soccer training. With the continuous improvement of the sports training load prediction model supported by intelligent algorithms, it can better enable the rapid improvement of all physical qualities of trainers.

The sports training scenarios of college students are more complex, and the physiological conditions of individual students, injury conditions, and external factors such as the training environment, field equipment, etc., all bring more uncertainty to physical training, which brings challenges to the methods of college students' sports teaching and training load control. Literature [28] established a sports training load prediction model based on BP neural network, using the average heart rate of students' sports training as an indicator, which provides a reference for developing sports teaching and training intensity suitable for students' physical fitness. Literature [29] combined genetic algorithm and BP neural network to predict students' sports training load, which still showed high prediction accuracy in complex sports environment, providing theoretical guidance for designing personalized sports training program and improving students' physical fitness. Literature [30], in response to the local convergence and random assignment of the traditional BP model, optimized the initial weights and thresholds of the model by integrating adaptive genetic algorithm, thus improving its prediction accuracy in the task of training load prediction, which is of guiding significance for improving the training effect of athletes. It is not difficult to find that the existing studies do not consider the environmental factors of college students' training adequately, so it becomes an urgent problem to develop more scientific and reasonable intelligent training methods and reasonable prediction of students' physical loads to ensure that the students can achieve the best training effect under the premise of safety.

This paper combines the data feature fusion method based on linear fitting and intelligent algorithms such as the weighted matrix decomposition recommendation algorithm based on the hidden semantic model to realize the design of the training load control method in college sports teaching. Firstly, the feature extraction model is constructed, and through the monitoring system, the updating rules of data feature extraction are obtained, the extracted data features are counted and denoised, and finally they are quantified by regression analysis and linear fitting methods to complete the sports data feature fusion. Secondly, the hidden semantic model is applied to the SVD matrix decomposition recommendation algorithm, and the idea of weighting is introduced, so as to realize the intelligent recommendation of sports instruction. Finally, the “fun interval running” sports training method was designed to realize the effective control of training load.

2. Sports Data Feature Fusion and Intelligent Recommendation Modeling in Sports Teaching and Learning

This chapter proposes a sports data feature fusion method based on linear fitting and a weighted matrix decomposition recommendation algorithm based on a hidden semantic model for sports instruction recommendation based on the optimization of sports data feature fusion to provide intelligent algorithmic support for the design of training load control methods in sports teaching.

2.1. Feature Fusion Methods for Sports Data

The process of sports data feature extraction is highly ambiguous and weakly correlated, unlike other data, sports data present a linear correlation process between multiple indicators, but this feature is not utilized by current classical fusion algorithms, resulting in low coverage after fusion.In this paper, we propose a fusion method based on the linear fitting of sports data features to complete the sports data acquisition and feature extraction, and use regression analysis and fusion method to achieve the optimization of sports data feature extraction.

2.1.1. Sports data feature extraction

In order to realize the automatic extraction of sports data, parameters such as maximum oxygen uptake $VO_{2\max}$ and ventilation per minute VE are used as constraint indicators, and ZigBee networking sensors are used as the monitoring method to construct a monitoring system for sports data, and the collection of real-time physiological index data of sports is realized through the combination of wearable devices and the monitoring system. The process of sports data acquisition is shown in Figure 1.

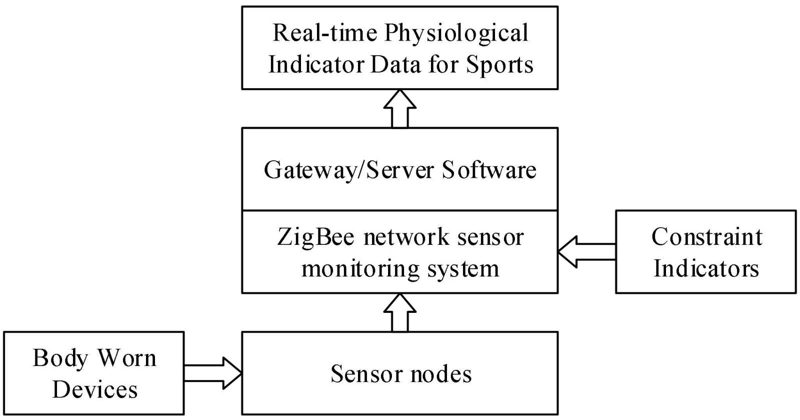


Figure 1. Process of sports data collection

According to the general structure of Fig. 1, the FIFO RAM buffer output command is used as the writing device of state parameters in the monitoring of real-time physiological indicators of sports. The Sync/Trigger triggering method is used to realize the loading of control commands during the monitoring of real-time physiological indicators of sports. The wireless sensor network device is used to construct the reset circuit, clock circuit and sleep control circuit, and the sports data state parameter recognition is realized in the asynchronous standard transmission interface device to extract the sports data characteristics.

According to the sports data, combined with the fuzzy spatial scheduling model $P(n_i)=\{p_k \mid pr_{kj}=1, 1\le k\le m\}$ for the joint control of real-time monitoring of sports, then its control function is expressed as follows:

$$s_{PPM}=\lim_{x\to\infty}\frac{a_i(t)}{\sqrt{w\Phi(x_i)-b}}-\lim_{x\to\infty}\frac{d_j}{P(n_i)+T_s} \tag{1}$$

Where, $a_i(t)$ is the collected real-time physiological indicator monitoring data set of the sport. w is the fuzzy weighting coefficient. T_s is the maximum sampling interval. d_j is the regression coefficient. $\Phi(x_i)$ is the fuzziness coefficient. b is the association rule function of feature extraction. Let the finite dataset of real-time physiological index monitoring of sports be:

$$
\begin{aligned}
E_{r_{\min}} &= \lim_{x\to\infty}\sqrt{\frac{\alpha_i}{K\left(x_i,x_j\right)}} \\
i &\geq 1,2,\cdots,n,\alpha_i \subseteq [0,A]
\end{aligned}
\tag{2}
$$

where α_i is the LOADmin minimum sampling factor. $K\left(x_i,x_j\right)$ is the fuzzy kernel function. j and i represent the number of nodes for data detection. Based on the above analysis, the sports data feature extraction model is constructed:

$$
\begin{aligned}
&\max imize\left(\sqrt{2w}-C\left(\xi_i-\xi_i^*\right)\xi_i\right) \\
&subject\ to\ \alpha_i+\left(\frac{w'x'(t)}{b}\right)\geq \varepsilon+\xi_i \\
&\left(w'x'(t)-b\right)+\alpha_i \geq \varepsilon+\xi_i^* \\
&\xi_i,\xi_i^* \geq 0, 1\leq i\leq n; C>0
\end{aligned}
\tag{3}
$$

where C is the state parameter of the motion process. ξ_i is the environment variable. w' is the inertia coefficient. $x'(t)$ is the change function of sports body indicators.

2.1.2. Constructing a big data fusion cluster analysis model

The above collected sports data mainly includes human cardiorespiratory function and heart rate data under different exercise frequencies, and the next step is to construct a big data fusion clustering analysis model based on the feature extraction of real-time physiological indicator monitoring data of sports to obtain the real-time monitoring feature output of the sports data.

First of all, let the fuzzy feature sequence $w(t)$ of the real-time physiological indicator monitoring data feature extraction of sports collected be $\left\{x^n-1\right\}$, according to the collection time range, the distributed fusion output of the real-time physiological indicator monitoring data features of sports is obtained as:

$$
\begin{cases}
K_0 : x'(t)+w(t) & 0<t\leq \dfrac{T}{2} \\
K_1 : E^2 s'(t)-w(t) & \dfrac{T}{2}<t\leq T
\end{cases}
\tag{4}
$$

where E represents the energy parameter of real-time physiological indicators of the sport. $s'(t)$ is the set of physiological indicator state feature distribution. T is the maximum acquisition time length. Under the control of pulse 02P, the set of conditionally distributed control state parameters for real-time physiological indicator monitoring in sports is obtained as:

$$
ML_{i'}=\frac{2L_i}{New_{i'}}, i\geq 1 \tag{5}
$$

Where $New_{i'}=\left(e_{i'1},e_{i'1}e_{i^2 2},e_{i'1}e_{i'2}e_{i3},\cdots,e_{i'D}\right)$ represents the fuzzy state parameters of real-time physiological indicators monitoring in sports, and L_i is the a priori parameters of data feature extraction. The asynchronous transmission control of the spectral feature components of the state feature parameters of real-time physiological indicators of sports is carried out to obtain the big data fusion clustering analysis model of data feature extraction for real-time physiological indicators monitoring of sports, which is expressed as follows:

$$h_w(n) = \frac{s'(t)}{K_0 f_s(n\Delta t)} + \frac{ML_i}{K_1 f_s(n\Delta t)} \quad (6)$$
$$0 \le n \le M+1$$

where n is the frame sequence number for data detection. Δt is the sampling time interval of the sports data. f_s is the time sampling frequency. Combined with the frame segment scanning technology to realize the real-time physiological index characterization of sports, the real-time sports data expression is obtained as:

$$P_{ij}(k) = \frac{l_i(k) + l_i(k)}{\eta_{ij}(k)} * \sqrt{l_j(k) + l_i^2(k) + l_i(k)\eta_{ij}^2(k)} \quad (7)$$

Where, j and i denote the number of nodes for data detection. $l_j(k)$ is the sample set of real-time physiological indicators of sports. $\eta_j(k)$ is the fuzzy degree indicator.

2.1.3. Fusion optimization based on linear fitting

The following update rules for data feature extraction were obtained through the real-time physiological indicator monitoring system for sports:

$$\hat{k}_\mu(t-1) = \sqrt{\frac{1}{Q(t-1)}} - \left[\partial \hat{F}_\mu (\partial S \partial t)^2\right] \quad (8)$$

Where, $Q(t-1)$ represents the similarity of real-time physiological indicator monitoring in sports. $\hat{F}_\mu$ is the degree of adaptation of the distribution of real-time physiological indicators data in sports. $\partial S \partial t$ is the adaptive weighting coefficient. Then the extracted data features are statistically analyzed by the statistical analysis model for the feature extraction of real-time physiological indicators monitoring data of sports, and the model is as follows:

$$s_{PAM}(t) = \sqrt{\frac{F(z)}{K(v)}} \quad (9)$$

Where $F(z)$ is the weighting function and $K(v)$ is the state distribution function.

If the sports data features fractal wavelet conversion as an arbitrary signal, when the original data x_k' and the feature extracted data x_k the larger the mean square deviation of the two, it means that the de-noising effect of fractal wavelet coding is better. Among them, the denoising of optimal fractal wavelet coding is mainly achieved by reconstructing the sports feature signals through x_k and x_k' to achieve the denoising effect, and its expression is:

$$y = MESf'(x) + x_k \times x_k' \quad (10)$$

where MES is the mean square error of the original data x_k' and the data x_k after feature extraction. $f'(x)$ is the feature of real-time physiological index data of sports. The denoising process is carried out using the above equation to ensure that a high degree of accuracy can be achieved when feature extraction is carried out subsequently.

Then the level $X^{(0)}$ of real-time physiological index monitoring data feature extraction in sports is divided into N levels for $X^{(1)}, X^{(1)}X^{(2)}, X^{(1)}X^{(2)}X^{(3)}, \cdots, X^{(N)}$, that is, $X^{(0)} = \bigcup_{i=1}^{N} X^{(i)}$. Regression analysis and linear fitting methods are used to realize the quantitative regression analysis of the basic indexes, such as VO_2, CO_2 discharge VCO_2, and heart rate HR, in the sports data of different feature levels, and to realize the optimization of the feature extraction of real-time physiological index monitoring data in sports. The regression analysis model for real-time physiological detection of sports is:

$$\partial^{n+1}\left(d_{\gamma_0}\right) = \lim_{t \to \infty} \frac{\sqrt{f(t)}}{d_{\gamma_0}^*(t)} \quad (11)$$

where $f(t)$ is the frequency component and $d^{*}_{\gamma_0}(t)$ is the joint characteristic distribution function. The characteristic linear fit of the real-time physiological index monitoring data for sports is given by:

$$pid_{new} = \begin{cases} \dfrac{pid - m}{\sqrt{\left(X_{\min} + P_{ij}(k)\right)}} & 0 < m \le \dfrac{M}{2} \\ \dfrac{pid - m}{\sqrt{\left(X_{\max} + P_{ij}(k)\right)}} & \dfrac{M}{2} < m \le M \end{cases} \tag{12}$$

where $X_{\max}$ and $X_{\min}$ are the maximum and minimum state detection thresholds, respectively. pid is the set of joint probability distributions. M is the embedded maximum dimension of the data feature reconstruction.

2.2. Weighted Matrix Decomposition Recommendation Algorithm Based on Hidden Semantic Modeling

Based on the sports data feature fusion method, this paper proposes a Weighted Matrix Decomposition Recommendation Algorithm based on Hidden Semantic Model (WSVDLFM) by introducing the idea of weighting in order to realize the accurate sports instruction recommendation in sports teaching.

2.2.1. Hidden Semantic Modeling

From the perspective of machine learning, the Implicit Semantic Model (LFM) [31] takes the user correlation matrix and the object correlation matrix as the parameters of the model, uses the historical scoring data as the dataset, and takes the square of the difference between the predicted score and the actual score as the loss function of the model, optimizes the loss function through the stochastic gradient descent algorithm (SGD)[32], and obtains the optimal solution of the class of the hidden attributes of the user and the object after training the model.

The implicit semantic model assumes that if two exercise instructions are viewed by multiple users at the same time, some degree of similarity between the two exercise instructions is recognized.

It is assumed that matrix R is the information about the users' evaluation of the exercise guidance, i.e., R_{ui} represents the specific rating values of the users u for the exercise guidance i and R is defined as the rating matrix. Matrix P is the effectiveness of the user's guidance for each type of exercise guidance, i.e., P_{uf} represents the degree of preference of the user u for exercise guidance in the hidden category f . Matrix Q is the weight of each exercise instruction i in the hidden category f , i.e., Q_{if} represents the percentage of exercise instruction i in the hidden category f .

Using the Funk-SVD algorithm [33], the rating matrix R is decomposed into two low-dimensional matrices multiplied together:

$$\hat{R} = PQ^{T} \tag{13}$$

where $P \in R^{u \times f}$ and $Q \in R^{i \times f}$ are two matrices after dimensionality reduction, i.e., the user u is linked to the exercise guidance information i through the exercise guidance hidden category f in the matrix. As a result, the implicit semantic model divides the sports instruction into F categories, $P(u,f)$ denotes the degree of effectiveness of the user u instruction for the f th sports instruction category, $Q(i,f)$ denotes the weight of the sports instruction i in the f th sports instruction category, $1 < f \le F$. Then the predictive value $\hat{R}(u,i) = \hat{r}_{ui}$ of the user u rating of the sports instruction i is computed by the following formula:

$$\hat{r}_{ui} = \hat{R}(u,i) = P_u Q_i^{T} = \sum_{f=1}^{F} p_{u,f} q_{f,i} \tag{14}$$

where $p_{u,f} = P(u,f)$, $q_{f,i} = Q(i,f)$.

Assuming that the difference between the user's true score and the predicted score follows a

Gaussian distribution, the corresponding loss function is given, and then the loss function is optimized by the stochastic gradient descent (SGD) algorithm, and the existing historical scoring data is used to calculate and solve for the parameter values in matrix P and matrix Q, so as to obtain the corresponding matrices P and Q. The loss function $C\left(p,q\right)$ represents the root-mean-square error between the user u true rating of the exercise guide i and the predicted rating, as follows:

$$C(p,q) = \sum_{(u,i)\in Train} \left(r_{ui} - \hat{r}_{ui}\right)^2 = \sum_{(u,i)\in Train} \left(r_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i}\right)^2 \tag{15}$$

In order to avoid the learning overfitting that may result from direct optimization, it is also necessary to add the overfitting prevention term $\lambda\left(\|p_u\|^2 + \|q_i\|^2\right)$, where λ is the regularization parameter, which needs to be obtained by repeated trials according to the specific application scenario, where λ takes the value of 0.01, thus it can be obtained:

$$C(p,q) = \sum_{(u,i)\in Train} \left(r_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i}\right)^2 + \lambda\left(\|p_u\|^2 + \|q_i\|^2\right) \tag{16}$$

(1) The loss function is optimized using the stochastic gradient descent (SGD) algorithm, and the fastest descent direction is determined by taking the partial derivatives of parameters $p_{u,f}$ and $q_{f,i}$:

$$\frac{\partial c}{\partial p_{u,f}} = -2\left(r_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i}\right) q_{f,i} + 2\lambda p_{u,f} \tag{17}$$

$$\frac{\partial c}{\partial q_{f,i}} = -2\left(r_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i}\right) p_{u,f} + 2\lambda q_{f,i} \tag{18}$$

(2) Iterative computation continuously optimizes the parameters until they converge:

$$p_{u,f} = p_{u,f} + \alpha\left(\left(r_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i}\right) q_{f,i} - \lambda p_{u,f}\right) \tag{19}$$

$$q_{f,i} = q_{f,i} + \alpha\left(\left(r_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i}\right) p_{u,f} - \lambda q_{f,i}\right) \tag{20}$$

where α is the learning rate, which represents the speed in the gradient descent process. α also needs to be obtained by repeated trials based on specific application scenarios, where α takes the value of 0.02.

2.2.2. Method of calculating weights

In order to be able to better utilize the implicit feedback data, make the exercise recommendation technology more perfect, and realize the user-accurate push of exercise guidance, this chapter will make corresponding improvements to the shortcomings of the matrix decomposition recommendation algorithm in the traditional Funk-SVD.

Before improvement, we first introduce three dimensional features, namely user personal attributes, user sports characteristics and user adoption degree, and use the normalization method to construct the user data behavior weighting matrix, so as to intuitively reflect the proportion of each dimensional feature in the weighting matrix of user data behavior, which interprets the user's preference of the exercise guidance from different angles.

(1) User personal attributes

The user's personal attributes include the user's age age, gender sex and sports level $grade$ attribute information. Users of different ages, genders, and sports levels will inevitably show different interest preferences in the process of using the sports monitoring system, so it is necessary to define the user's personal attribute feature weights.

Definition 1: Firstly, the user's age age is divided into K subgroups, and secondly, the K subgroups that have been divided are again subdivided into $2K$ subgroups according to the user's gender sex, then each user will belong to a unique subgroup, i.e., the user's age and gender can be represented by a matrix P_{uf} of order $K \times 2$ in size, and each subgroup belonging to each subgroup corresponds to an element with a value of 1, and the others with 0. Similarly, the user's sport level and gender can be represented by a matrix Q_{if} of size $W \times 2$ order. Then, the user's personal attribute feature weights are calculated as:

$$a_{ui} = P_{uf} Q_{if}^T \tag{21}$$

(2) User movement characteristics

The user's sports characteristics are mainly obtained by analyzing the sports data collected by the smart bracelet worn on the user. If the frequency of a certain sport is high, the user is considered to have a high interest preference for this type of sport guidance information.

Definition 2: First assume that the system has a total of N exercise guide, A exercise guides $df_i = \{df_1, df_2, df_3, \cdots, df_a\}$ for each exercise i, and B exercise guides $tf_{ui} = \{tf_{11}, tf_{12}, tf_{13}, \cdots, tf_{1b}, tf_{21}, tf_{22}, tf_{23}, \cdots, tf_{2b}, \cdots\}$ covering exercise i for each user u, then the user exercise feature weights are calculated as:

$$b_{ui} = tf_{ui} \times \log \frac{N}{df_i} \tag{22}$$

(3) User adoption level

In order to differentiate between campaign guides that a user has viewed and those that have not, data on the dimension of user adoption degree is introduced in the matrix decomposition.

Definition 3: Matrix c_{ui} represents the user's user adoption level weights in the matrix decomposition. v is used to control the proportion of the weight, which is treated as a constant and generally takes the value of 0.3. t_{ui} represents the proportion of the length of time that the user u has viewed the campaign guide i and the length of time that the campaign guide i has been viewed by all users. The user adoption degree weight is calculated by the formula:

$$c_{ui} = \frac{1}{1 + \log\left(t_{ui} \times 10^v\right)} \tag{23}$$

(4) User data behavior weighting matrix

In order to better shape the personalized part of the user and enhance the accuracy of the recommendation, this paper will introduce the user data behavior weighting matrix, which integrally considers the influence of the features of the three dimensions on the user's personalization degree.

Definition 4: The data behavior weighting matrix is a matrix of overall weights obtained by normalizing the features of the three dimensions of user personal attributes, user movement characteristics and user adoption degree, and the default is $\alpha + \beta + \gamma = 1$. The specific calculation formula is:

$$w_{ui} = \alpha \frac{a_{ui}}{\|a_{ui}\|^2} + \beta \frac{b_{ui}}{\|b_u\|^2} + \gamma \frac{c_{ui}}{\|c_u\|^2} \tag{24}$$

2.2.3. Weighted Matrix Decomposition Recommendation Algorithm

With the help of user data behavior weighting matrix, this section will propose a weighted matrix decomposition recommendation algorithm based on the matrix decomposition of hidden semantic model. The core idea of this algorithm is to decompose the user rating matrix R into two low-dimensional sub-matrices P and Q, and then calculate the missing ratings in the rating matrix, and complete the recommendation based on the calculated predicted ratings.

The specific flow of the algorithm is as follows:

Input: user data behavior weighting matrix W and user rating matrix R.

Output: user feature matrix P and motion guidance feature matrix Q.

Step1: Initialize the relevant parameters λ, α and $iterations$.

Step2: Select a set of data, calculate the partial derivatives of parameters $p_{u,f}$ and $q_{f,i}$ to determine the direction of stochastic gradient descent:

$$\frac{\partial c}{\partial p_{u,f}} = -2\left(r_{ui} + w_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i} \right) q_{f,i} + 2\lambda p_{u,f} \tag{25}$$

$$\frac{\partial c}{\partial q_{f,i}} = -2\left(r_{ui} + w_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i} \right) p_{u,f} + 2\lambda q_{f,i} \tag{26}$$

Step3: Constantly update parameters $p_{u,f}$ and $q_{f,i}$ using iterative equations (27) to (28):

$$p_{u,f} = p_{u,f} + \alpha\left(\left(r_{ui} + w_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i} \right) q_{f,i} - \lambda p_{u,f} \right) \tag{27}$$

$$q_{f,i} = q_{f,i} + \alpha\left(\left(r_{ui} + w_{ui} - \sum_{f=1}^{F} p_{u,f} q_{f,i} \right) p_{u,f} - \lambda q_{f,i} \right) \tag{28}$$

Step4: By calculating the difference between the two iterations before and after is less than a certain value or by determining whether the number of iterations reaches the value set at the beginning to decide whether to terminate, if the condition is not met, then continue to Step3 iteration operation. Otherwise, return P and Q.

3. Experiments on feature fusion of sports data and sports instruction recommendation

3.1. Experiments on feature fusion of sports data

In this section, based on the proposed sports data feature fusion optimization method, the sports training instant data acquisition system is constructed, and sports data acquisition and feature fusion experiments are conducted to verify the effectiveness of the proposed method. The classification fusion and correlation analysis of the physical function data, athletic skill data and sports performance in the sports training instant data in the experiment get the output of sports training instant data acquisition and fusion as shown in Table 1.

Table 1. Real-time data acquisition and fusion output of sports training

Test index	Normalized index	Degree of fuzzy matching	Degree of fusion	Information entropy
Oxygen ventilation equivalent	0.121	0.253	0.271	0.187
Oxygen uptake VO_{2AT}	0.111	0.246	0.262	0.181
CO_2 emissions VCO_2	0.113	0.252	0.247	0.195
Respiratory entropy RQ	0.115	0.254	0.258	0.173

According to the instantaneous data acquisition and fusion results of sports training in Table 1, the instantaneous data acquisition and analysis of sports training is carried out, and the acquisition results are obtained as shown in Fig. 2. For the sports training data such as oxygen ventilation equivalent, oxygen uptake VO2AT, CO2 excretion VCO2 and respiratory entropy RQ, the system in this paper can effectively achieve instant sampling and fusion optimization.

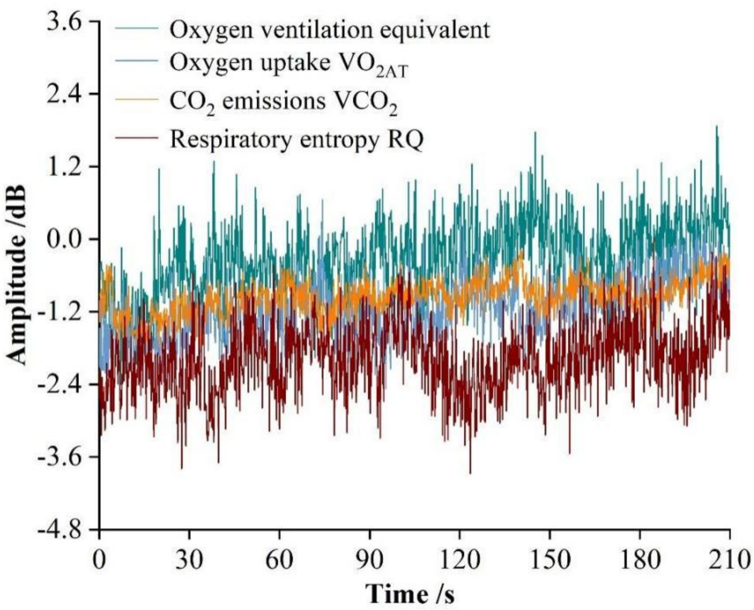


Figure 2. Real-time data acquisition results of sports training

The accuracy of the collected data is tested, and the accuracy rate of the instant data collection for sports training is one of the important indexes to measure the accuracy of data collection. The data acquisition accuracy rate refers to the degree of agreement between the collected data and the actual situation, usually expressed as a percentage, and the accuracy rate calculation formula is as follows:

$$A = \left(\frac{a_i}{a_j} \right) \times 100\% \tag{29}$$

Where, a_i indicates the total amount of data collected by the system in accordance with the actual situation, and a_j indicates the total amount of experimental data.

The total amount of experimental data used during this experimental test is 20,000, and the total amount of data collected by the three systems that are in accordance with the actual situation are counted, and some of the data collected by the three systems are shown in Table 2.

Table 2. Partial data collection of the three systems

Number of experiments	Text system	Contrast system 1	Contrast system 2
30	18436	13144	11661
60	19142	12509	9188
90	18931	14927	13683
120	18593	13499	12977
150	19414	15200	13436
180	18811	16181	12534
210	19669	16872	13570

The total amount of data collected by the three systems that matched the actual situation and the total amount of experimental data were brought into Eq. (29), and the comparison results of data collection accuracy were obtained as shown in Fig. 3.

Analyzing the simulation results, we know that the instantaneous data acquisition accuracy of the exercise training of the comparison system 1 and the comparison system 2 varies in the intervals of [62.55%,84.36%] and [45.94%,68.42%], respectively, and the acquisition accuracy curves of the two systems are always below the system of this paper and the changes are unstable, which indicates that the data acquisition accuracy of the two systems is low and the stability is poor. Compared with these two systems, the instantaneous data acquisition accuracy of sports training of this paper's system always stays above 92%, which indicates that the data acquisition accuracy of this system is higher and the acquisition results are more accurate, and it also proves the practicability of the sports data feature fusion method proposed in this paper.

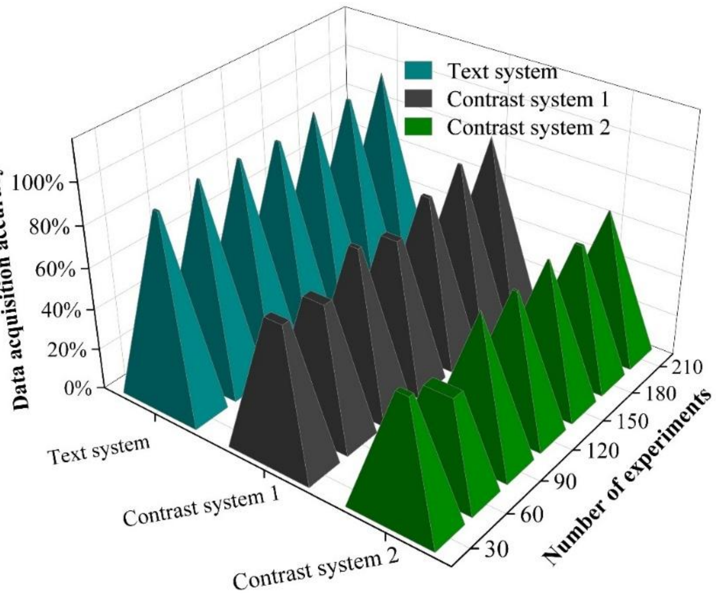


Figure 3. Comparative test of data acquisition accuracy

3.2. Experiment on Precision Recommendation for Sports Instruction

In this paper, the WSVDLFM algorithm is used to realize the precise recommendation of sports instruction in college students' physical education. In order to test the effectiveness of the algorithm, the WSVDLFM algorithm and its decomposition algorithms WSVD and SVDLFM are evaluated experimentally respectively.

3.2.1. Performance evaluation of the WSVD algorithm

(1) Experimental setup

The dataset used in this experiment is derived from the real-time physiological index data of sports and exercise collected by ZigBee networking sensors, and is optimized by data cleaning as well as data feature fusion based on linear fitting.

Meanwhile, the experiment uses hit rate (HR) and normalized discounted cumulative gain (NDCG) as the criteria for evaluation.HR indicates whether the data in the test set appears in the sorted list.NDCG indicates the position in the sorted list of the data in the test set that has appeared in the sorted list. The higher the position of the relevant element in the sorted list, the more it contributes to the final result. The formula for its calculation is as follows:

1) Hit Rate (HR):

$$HR@n = \frac{Number\_of\_hits@n}{Ground\_Truth} \tag{30}$$

where $Ground\_Truth$ (GT) denotes the correct value, i.e., the data in the test list. $Number\_of\_hits$ denotes the number of hits in the recommended list.

2) Normalized Discounted Cumulative Gain (NDCG):

$$DCG@n = \sum_{i=1}^{n} \frac{2^{rel_i} - 1}{\log_2(i+1)} \tag{31}$$

$$NDCG@n = \frac{DCG@n}{IDCG@n} \tag{32}$$

Where, rel_i denotes the relevance level of the motion guidance in the i nd position and the query, expressed in binary, with a value of 1 for a hit and 0 otherwise. DCG denotes the discounted cumulative gain per user, which reduces the contribution of the relevant motion guidance in a lower position in the rankings. $IDCG@n$ denotes the discounted cumulative gain of the motion guidance that ideally ranks in the n th position. The final evaluation result of the experiment i.e. the mean value of NDCG for all users.

(2) Comparison of Algorithms

Two representative implicit feedback matrix decomposition recommendation algorithms are selected for the experiments in this section: alternating least squares (ALS) [34] and randomized block coordinate descent (RCD) [35].

1) ALS: a matrix decomposition recommendation algorithm based on implicit feedback with high time complexity.

2) RCD: An efficient implicit feedback based matrix decomposition algorithm characterized by being based on full data, which in this context means treating all missing data as negative samples. The algorithm uses a linear search to determine the learning step of the feature vector.

3) Comparison of experimental results

The comparison results of HR and NDCG are shown in Fig. 4 and Fig. 5, respectively. When WSVD, RCD and ALS all enter the steady state, their accuracy generally does not fluctuate greatly and stabilizes in their respective intervals. And WSVD achieved the best performance with HR and NDCG of about 0.607 and 0.162, respectively.The reason is that the WSVD algorithm makes full use of the missing data, and constructs the user data behavioral weighting matrix by calculating the missing data of sports training scores and combining the three feature dimensions of the user's personal attributes, the user's sports characteristics, and the user's degree of adoption with normalized weighting. As can be seen from the figure, the prediction accuracy of WSVD is better than that of RCD and ALS, but the convergence speed of WSVD is smaller than that of RCD and ALS.

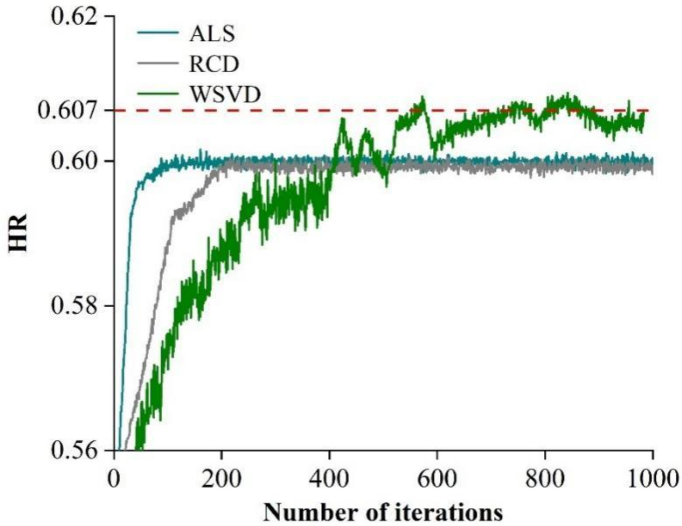


Figure 4. HR comparison results

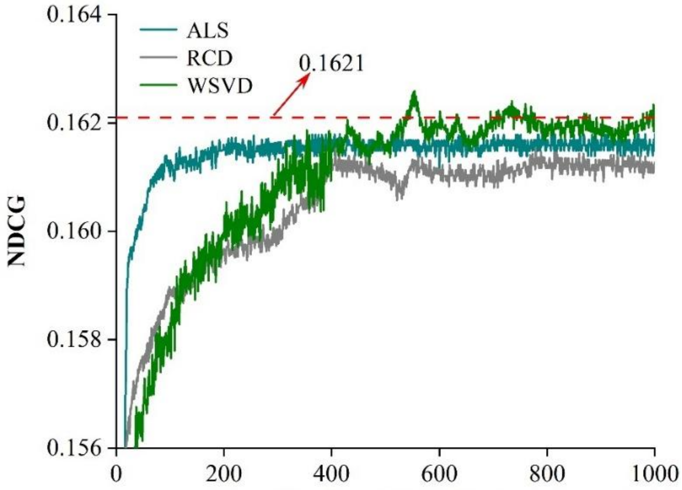
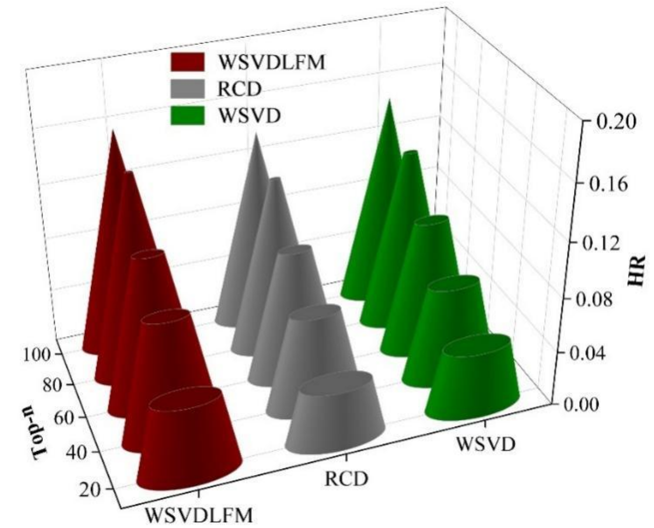


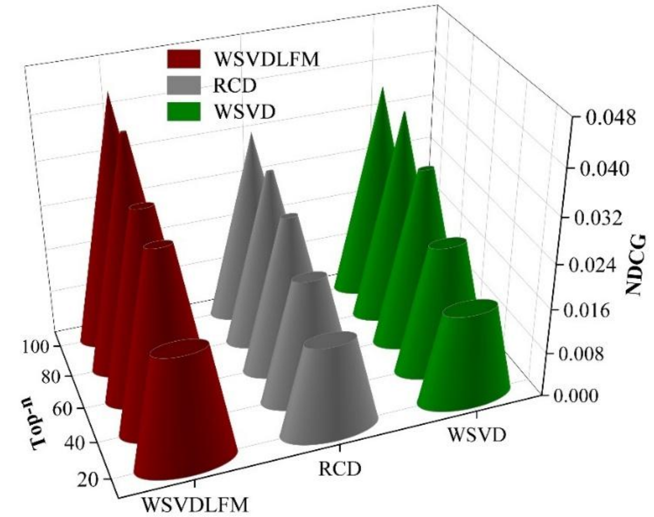
Figure 5. Comparison results of NDCG

3.2.2. Impact of Top-n values on algorithm performance

The data collection and data feature fusion optimization were re-conducted to construct a new dataset of physiological indicators of sports training to continue the experiment. The effects of different Top-n values on the recommendation accuracy of the algorithm are shown in Fig. 6, where the value of Top-n indicates the length of the recommendation list, and (a) and (b) indicate the effects of Top-n on HR and NDCG, respectively. As the value of Top-n increases, the prediction accuracy of WSVD, RCD and WSVDLFM are all improved. The HR and NDCG values of the WSVDLFM algorithm reached 0.1603 and 0.0440, respectively, when Top-n=100.This phenomenon suggests that increasing the value of Top-n is beneficial for improving the performance of the algorithms. Meanwhile, the performance gap between the WSVD algorithm and the RCD algorithm is gradually obvious with the increase of the Top-n value, and this dominance indicates that the WSVD algorithm is superior to the RCD algorithm on the real-time physiological index dataset of the sports, while the WSVDLFM algorithm consistently outperforms the RCD algorithm and the WSVD algorithm. This phenomenon indicates that the WSVDLFM algorithm based on the multidimensional feature normalized weighting strategy and the hidden semantic model is a high performance algorithm.



(a) HR



(b) NDCG

Figure 6. Top-n Indicates the impact of the value on recommendation accuracy

3.2.3. Performance comparison experiments of WSVDLFM algorithm

The comparison results between different algorithms under the two evaluation criteria of HR and

NDCG are shown in Fig. 7 and Fig. 8, respectively. Where, the horizontal coordinate represents the number of iterations and the vertical coordinate represents the evaluation metrics.

The WSVD, SVDLFM and WSVDLFM algorithms, all show the spiking phenomenon when the number of iterations on the dataset is small, i.e., they obtain the peak value when the number of iterations is very small, and then start to decline. In this paper, we hypothesize that the phenomenon may be related to certain data patterns in the dataset.

First, this paper analyzes the performance comparison between WSVD algorithm and RCD algorithm on the dataset. As can be seen from the figure, the performance of both WSVD and RCD is improved as the number of iterations increases. The HR values of both are similar and the gap of NDCG with the increase of iterations is significant. This phenomenon suggests that WSVD and RCD have a better recommendation for WSVD with similar hit rates.

Secondly, WSVD outperforms SVDLFM in the same experimental environment.However, it can be seen from the figure that the convergence speed of SVDLFM is significantly higher than that of WSVD and RCD.Therefore, combining the matrix decomposition recommendation with the Hidden Semantic Model will help to improve the performance of the algorithm.

Finally, the performance of WSVDLFM is analyzed in comparison with other algorithms. As can be seen from the figure, the WSVDLFM algorithm outperforms other algorithms in both convergence speed and prediction accuracy. This fully demonstrates that the use of multidimensional feature normalization weighting strategy and hidden semantic model optimization method can effectively improve the SVD algorithm.

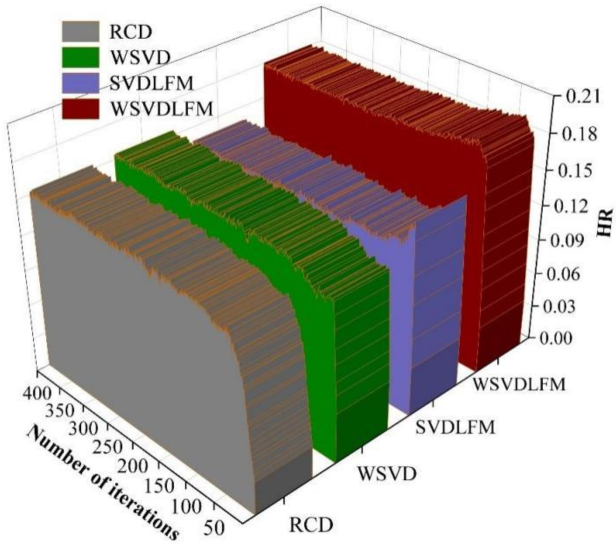


Figure 7. HR comparison results

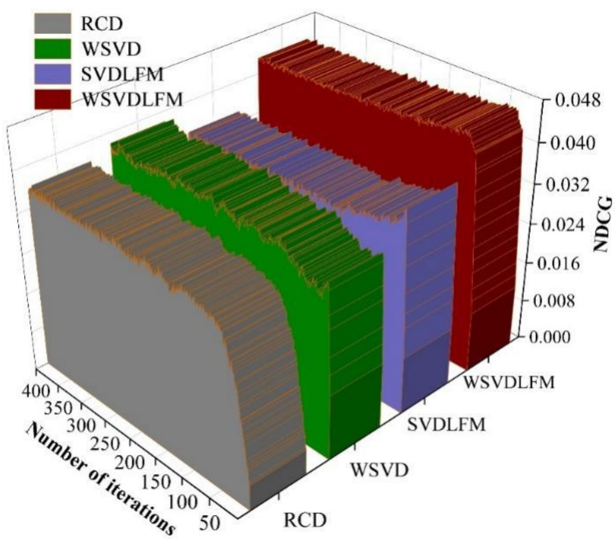


Figure 8. Comparison results of NDCG

4. Design of methods for controlling physical training loads for university students

4.1. Methods of physical training load control - "fun interval running"

Based on the fusion of sports data features and sports guidance recommendations, this paper designs a sports training load control method and names it "Fun Interval Running".

Interval Running: Interval running is integrated into the game process, combining full-speed curve running, straight-line jogging and resting.

Bend run on the way: the right arm swings forward and backward, the left arm swings close to the body, the swing is forward and backward, the swing is forward and backward. Support leg hip, knee, ankle full stirrup extension, and finally toes off the ground.

Rules of the game: the head of the row of two rock-paper-scissors, the winner of the first from the inside lane (the negative from the outside lane) to run to the game area placed at the 1 sign bucket, and then along the straight line jogging back to the starting point of the end of the queue queue. The first group to connect the barrels in the game area in a straight line wins, and the group that loses does 5 Bobi Jumps.

Organization: The starting point and the play area are at the ends of the curves, after the starting point (in the straight part) split into two columns. According to the class and the boys' school number single and double numbers are divided into four groups. In the first round, the single and double numbered groups play against each other. In the second round, the winner's group and the loser's group play each other. Each group has 3 buckets of the same color. After the 3 buckets have been placed, the students in the next round need to move the buckets once until they are in a straight line. Teacher whistles to direct and words to motivate students.

4.2. Training results and analysis

The study chose undergraduate students of college A as the experimental subjects for physical training in the form of fun interval running, the students were aged 17-20 years old, with body BMI in the range of 18.5-24.0 and in good physical condition.

The trend of heart rate loading in fun interval running is shown in Figure 9. Where BPM denotes heart rate and PL denotes body load. During the fun interval run, the lowest average heart rate of the students was 91 beats/min, the overall average heart rate was 128 beats/min, and the peak of the heart rate reached 162 beats/min. the students had a certain exercise load in general, and at the same time, they also had time to take a rest.

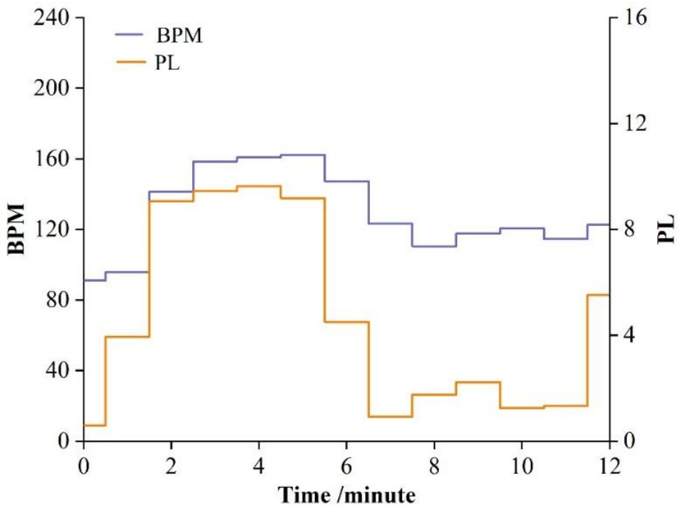


Figure 9. Heart rate load trend of fun interval running

Meanwhile, the combined distribution of internal and external loads of the students and the rhythmic changes of heart rate velocity are shown in Figures 10 and 11, respectively. In Figure 10, the body load PL and heart rate load indicate the students' exercise volume and body response, respectively. The more to the right the point indicates a greater body response, and the more to the top indicates a greater amount of exercise. To summarize, the more to the upper left corner, it is the more exercise and less body response, which indicates that he still has a lot of potential. The more to the lower right corner, is the exercise volume is small but the body reaction is big, that is the physical capacity is not enough or fatigue manifestation, need to pay extra attention.

It can be seen that the physiological load and heart rate of the students changed during the fun interval run. The values of students' per capita exercise load and body reaction were 57.31 and 18.52 respectively.The intensity of this training period is moderate, and through high-intensity physical training, it is conducive to the formation of muscle memory in this stage and helps students to consolidate technical movements. Combined with the changes in students' heart rate load, it can be seen that the training load control method designed in this paper is feasible.

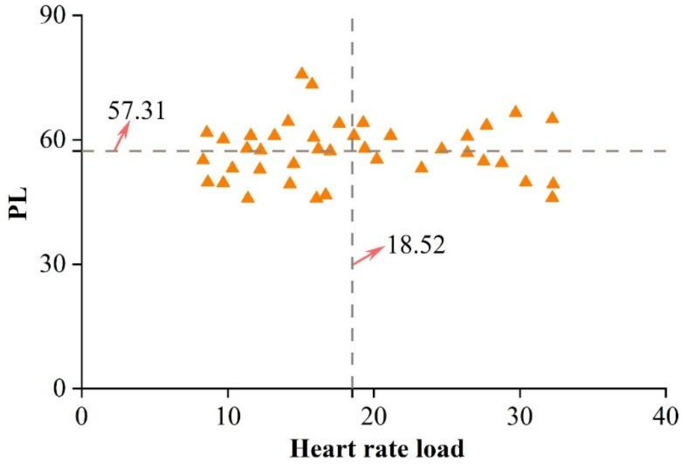


Figure 10. Comprehensive distribution of internal and external load of fun interval running

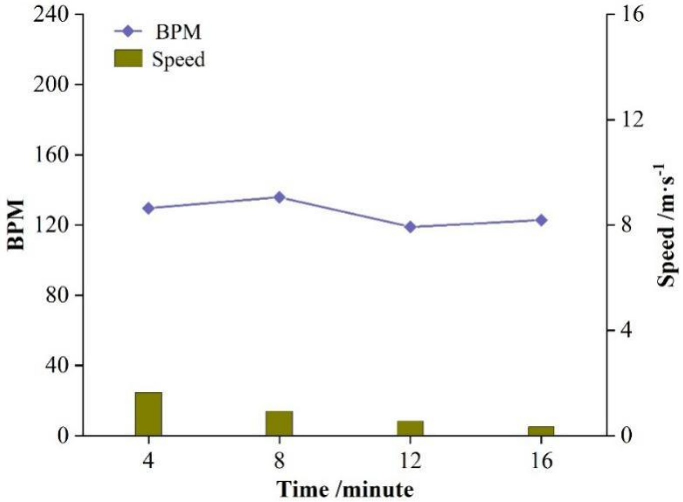


Figure 11. Changes of heart rate, speed and rhythm during fun interval running

5. Conclusion

In this paper, we have successfully designed a training load control method for college students' physical education teaching by empowering intelligent algorithms such as data feature fusion method based on linear fitting and weighted matrix decomposition recommendation algorithm based on hidden semantic model.

For the data of physiological characteristics of sports such as oxygen ventilation equivalent, oxygen uptake VO_{2AT}, CO_2 excretion VCO_2, and respiratory entropy RQ, the system in this paper can effectively realize instant sampling and fusion optimization, and the collection accuracy is always optimal and stable compared with other systems, which is above 92%, proving the practicability of the fusion of characteristics of sports data proposed in this paper.

The prediction accuracies of WSVD, RCD and WSVDLFM are all improved with increasing Top-n values, while the performance gap between the three algorithms is gradually obvious. The WSVD algorithm outperforms the RCD algorithm on the real-time physiological index dataset, and the WSVDLFM algorithm proposed in this paper always outperforms the RCD algorithm and the WSVD algorithm. When Top-n=100, the HR and NDCG values of the proposed WSVDLFM algorithm reach 0.1603 and 0.0440, respectively. In addition, the WSVDLFM algorithm outperforms the other comparative algorithms in terms of convergence speed and prediction accuracy, which verifies the feasibility of the SVD algorithm using multi-dimensional feature normalization and weighting strategy and the hidden semantic model.

When the designed “fun interval running” method is applied to sports training, the training intensity is moderate, and it can maximize the training efficiency while controlling the training load of the students within a certain range, which is conducive to the formation of muscle memory and helps students consolidate the technical movements, and it is feasible.

References

1. Yang, Y. (2018). The innovation of college physical training based on computer virtual reality technology. Journal of Discrete Mathematical Sciences and Cryptography, 21(6), 1275-1280.
2. Impellizzeri, F. M., Marcora, S. M., & Coutts, A. J. (2019). Internal and external training load: 15 years on. Int J Sports Physiol Perform, 14(2), 270-273.
3. Boullosa, D., Casado, A., Claudino, J. G., Jiménez-Reyes, P., Ravé, G., Castaño-Zambudio, A., ... & Zouhal, H. (2020). Do you play or do you train? Insights from individual sports for training load and injury risk management in team sports based on individualization. Frontiers in physiology, 11, 995.
4. Vanrenterghem, J., Nedergaard, N. J., Robinson, M. A., & Drust, B. (2017). Training load monitoring in team sports: a novel framework separating physiological and biomechanical load-adaptation pathways. Sports medicine, 47, 2135-2142.
5. Bartlett, J. D., O’Connor, F., Pitchford, N., Torres-Ronda, L., & Robertson, S. J. (2017). Relationships between internal and external training load in team-sport athletes: evidence for an individualized approach. International journal of sports physiology and performance, 12(2), 230-234.
6. Zhang, R. (2022). College Sports Decision‐Making Algorithm Based on Machine Few‐Shot Learning and Health Information Mining Technology. Computational intelligence and neuroscience, 2022(1), 7688985.
7. Wei, S., Wang, K., & Li, X. (2022). Design and implementation of college sports training system based on artificial intelligence. International Journal of System Assurance Engineering and Management, 13(Suppl 3), 971-977.
8. Li, Z. (2021). Sports policy and training decision support method based on wireless sensor network. Wireless Communications and Mobile Computing, 2021(1), 1608340.
9. Egoshi, S., Kandori, T., Ueno, M., Shinya, M., Wakiji, A., & Ito, R. (2025). Exercise Load Variation at Different Subjective Intensities of High-Intensity Interval Training Using Body Weight. Cureus, 17(2).
10. Conte, D., Kolb, N., Scanlan, A. T., & Santolamazza, F. (2018). Monitoring training load and well-being during the in-season phase in national collegiate athletic association division I men’s basketball. International journal of sports physiology and performance, 13(8), 1067-1074.
11. Coyne, J. O., Gregory Haff, G., Coutts, A. J., Newton, R. U., & Nimphius, S. (2018). The current state of subjective training load monitoring—a practical perspective and call to action. Sports medicine-open, 4, 1-10.
12. Hamlin, M. J., Wilkes, D., Elliot, C. A., Lizamore, C. A., & Kathiravel, Y. (2019). Monitoring training loads and perceived stress in young elite university athletes. Frontiers in physiology, 10, 34.
13. Govus, A. D., Coutts, A., Duffield, R., Murray, A., & Fullagar, H. (2018). Relationship between pretraining subjective wellness measures, player load, and rating-of-perceived-exertion training load in American college football. International journal of sports physiology and performance, 13(1), 95-101.
14. An, P. (2024). Fuzzy Decision Support Systems to Improve the Effectiveness of Training Programs in the Field of Sports Fitness. International Journal of Computational Intelligence Systems, 17(1), 168.

15. Ma, W., & Guo, B. (2024). Construction of neural network model for exercise load monitoring based on yoga training data and rehabilitation therapy. Heliyon, 10(12).
16. Cao, L. (2022). Design and optimization of a decision support system for sports training based on data mining technology. Scientific Programming, 2022(1), 1846345.
17. Guan, X., Lin, Y., Wang, Q., Liu, Z., & Liu, C. (2021, October). Sports fatigue detection based on deep learning. In 2021 14th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1-6). IEEE.
18. Liu, K., Li, X., & Rochester, C. A. (2022). Relationship between physical training and tactical training in sports training relying on boosting and bagging algorithms. Scientific Programming, 2022(1), 8429597.
19. Curtis, M. A., Kupperman, N., Westbrook, J., Weltman, A. L., Hart, J., & Hertel, J. (2025). Neuromuscular Performance and the Intensity of External Training Load During the Preseason in National Collegiate Athletic Association Division I Men's Collegiate Basketball Players. The Journal of Strength & Conditioning Research, 39(1), 54-61.
20. Nosek, P., Andrew, M., Sormaz, M., Drust, B., & Brownlee, T. (2023). The use of principal component analysis for reduction of training load data in professional soccer. Kinesiology, 55(2), 202-212.
21. Asker, M., & Møller, M. (2018). Training load issues in young handball players. Handball Sports Medicine: Basic Science, Injury Management and Return to Sport, 583-595.
22. Duggan, J. D., Moody, J. A., Byrne, P. J., Cooper, S. M., & Ryan, L. (2021). Training load monitoring considerations for female gaelic team sports: From theory to practice. Sports, 9(6), 84.
23. Li, S., & Ma, X. (2022). An Analysis Method of Exercise Load in Physical Training Based on Radial Basis Neural Network Model. Mobile Information Systems, 2022(1), 8383448.
24. Dausin, C., Ruiz-Carmona, S., De Bosscher, R., Janssens, K., Herbots, L., Heidbuchel, H., ... & Claessen, G. (2023). Semiautomatic Training Load Determination in Endurance Athletes. Journal for the Measurement of Physical Behaviour, 6(3), 193-201.
25. Zhao, F., Sharma, A., & Samori, I. A. (2022). Heart rate monitoring of physical fitness training load based on wavelet transform. The Journal of Engineering, 2022(11), 1095-1103.
26. Foster, C., Rodriguez-Marroyo, J. A., & De Koning, J. J. (2017). Monitoring training loads: the past, the present, and the future. International journal of sports physiology and performance, 12(s2), S2-2.
27. Teixeira, J. E., Encarnação, S., Branquinho, L., Morgans, R., Afonso, P., Rocha, J., ... & Forte, P. (2024). Data mining paths for standard weekly training load in sub-elite young football players: a machine learning approach. Journal of Functional Morphology and Kinesiology, 9(3), 114.
28. Liu, D., Li, S., & You, K. (2022). Training Load Prediction in Physical Education Teaching Based on BP Neural Network Model. Mobile Information Systems, 2022(1), 4821208.
29. Cheng, X. (2022). Application of optimized bp neural network model for the training load prediction in physical education teaching. Wireless Communications and Mobile Computing, 2022(1), 7524153.
30. Liu, L., & Sheng, G. (2022). Application of Training Load Prediction Model based on Improved BP Neural Network in Sports Training of Athletes. International Journal of Advanced Computer Science and Applications, 13(10).
31. Chuyu Wang & Guanglong Zhang. (2025). In the shadows of opacity: Firm information quality and latent factor model performance. International Review of Financial Analysis103970-103970.
32. M. Yigit Turali, Ali T. Koc & Suleyman S. Kozat. (2024). Optimal stochastic gradient descent algorithm for filtering. Digital Signal Processing104731-104731.
33. Guiling Zhao, Zihao Jiang, Jinbao Wang & Shuai Gao. (2025). A robust SVD-UKF algorithm and its application in integrated navigation systems. Advances in Space Research(2),1902-1912.

34. Collin G White, Thomas M Hancewicz, Ayuba Fasasi, Junior Wright & Barry K Lavine. (2024). Alternating and Modified Alternating Least Squares Applied to Raman Spectra of Finished Gasolines. Applied spectroscopy37028241292649.
35. Ke Zhang, Chun Ting Liu & Xiang Long Jiang. (2024). A Greedy Randomized Block Coordinate Descent Algorithm with k-means Clustering for Solving Large Linear Least-squares Problems. IAENG International Journal of Computer Science(5).