Research Article

Physical Fitness Evaluation of College Students at the Stage of Physical Exercise Behavior Based on Bayesian and Data Mining

Lei Wang¹ and Mei Yang²

¹School of Journalism & Communication, Wuhan Sports University, Wuhan 430079, Hubei, China
²School of International Education, Wuhan Sports University, Wuhan 430079, Hubei, China

Correspondence should be addressed to Mei Yang; yangmei@whsu.edu.cn

Received 22 December 2021; Revised 12 January 2022; Accepted 15 January 2022; Published 26 April 2022

Academic Editor: Tongguang Ni

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

It is difficult for the traditional physical fitness evaluation methods to dig useful information from massive data, and the accuracy of physical fitness evaluation is low. Therefore, this study proposed a physical fitness evaluation method in the stage of physical exercise behavior based on Bayesian and data mining for the college students. The purpose was to set the association rules of exercise behavior stage, to mine the association mapping relationship in the data set by using frequent itemsets, to build a regression model and to select the best physical variables in the exercise behavior stage. The frequent itemset was used to eliminate the redundancy of physical fitness data, the causal relationship was used to sort the physical exercise behavior stage, and the transformation of physical fitness evaluation index system was realized through Bayesian network topology to realize the physical fitness evaluation in the physical exercise behavior stage. The experimental results showed that the accuracy of this method was as high as 97.62%, and the recall rate of fitness data evaluation was as high as 99.3%. At the same time, the effect of fitness evaluation was better in a short evaluation time.

1. Introduction

It is easy to understand the connotation of physical fitness literally. It mainly refers to the quality of the human body. The quality of physical fitness is affected by many factors, among which it is determined by innate inheritance and acquired basis [1]. To judge whether a person’s physical fitness is good or bad mainly means that the shape of the body has the development level, the level of physiological function, athletic ability, physical quality development, and psychological development. These five aspects determine people’s physical fitness level [2–5]. Relevant scholars put forward the theory of stage change, believing that behavioral change consists of four factors: stage of change, balanced decision, change process, and self-efficacy [4, 5]. By applying the theory to physical exercise behavior, it is concluded that the change of physical exercise behavior goes through five stages, namely, pre-expectation stage, expectation stage, preparation stage, action stage, and maintenance stage.

According to the different behavioral stages of the college students’ physical exercise, a comparative research was made to prove that the physical exercise assessment put forward by this study was effective.

Cheng and Wang built a physical health policy attitude assessment model to carry on an empirical analysis on physical health policy, policy satisfaction, the relations between and among policy behavior, and attitude toward policy model by applying the method of structural equation model. The method made an effective evaluation of the physical state. The accuracy was high, but the efficiency was poor [8]. He et al. designed a kind of Chinese urban adolescent physical fitness environmental assessment method [9], the application of standard index quantitative and qualitative screening based on the Delphi method, and analytic hierarchy process was used to index weight assignment. This method could improve the physical environment qualitative measurement accuracy and the content, but the poor efficiency of environmental assessment was low.
Hu et al. proposed the physique classification and evaluation method based on k-medoids method [10] and conducted psychological tests on the students, extracting psychological factors by principal component analysis and evaluating the physique state by a three-level fuzzy evaluation model. This method could reasonably classify the physique, but the efficiency of physique classification and evaluation of the college students was not satisfactory. Zhang and Yun put forward the body posture evaluation method of the students through big data developed a set of scientific and reasonable body posture evaluation system to help the schools, teachers, and students understand their body posture and health status correctly through big data, scientifically and to guide the schools and teachers to formulate curriculum contents conducive to the students’ physical and mental development and effectively evaluated their body posture. However, the evaluation accuracy was low [11].

In the modern society, with the highly developed information, the traditional statistical methods are difficult to find useful information from massive data. Data mining has become an important means of knowledge discovery based on big data. Therefore, this study proposed a physical fitness evaluation method at the stage of physical exercise behavior based on Bayesian and data mining. The specific research ideas were as follows:

Firstly, association rules of exercise behavior stage were set, association mapping relationship of data set was mined by frequent itemsets, and regression model was constructed.

Secondly, the best physical fitness variable in exercise behavior stage was chosen. Frequent itemsets were used to eliminate the redundancy of physical fitness data, and causality was used to rank the stages of physical exercise behavior.

Next, the physical fitness evaluation index system was transformed by Bayesian network topology to realize physical fitness evaluation at the stage of physical exercise behavior.

Finally, through the accuracy and efficiency of physical fitness assessment, the effectiveness of this method in physical fitness assessment of college students at the stage of physical exercise behavior was verified and a conclusion was drawn.

2. Data Mining of Physical Fitness of College Students at the Stage of Physical Exercise Behavior

2.1. Association Rules in Data Mining of College Students’ Physical Fitness Data. Each test object determined a unique identification (TID), then each TID contained the following information: physical exercise behavior stage (field 1), gender (field 1), and physical fitness test index (field 11). The height and weight in the test index derive BMI, $BMI = \frac{weight \ (kg)}{height \ (m)}$, $VO2max$, and vital capacity were divided into weight composition (relative $VO2max$ and relative vital capacity) respectively to form database D. Clementine12.0 data mining software requires type consistency and dimensionless processing of the data of input variables and output variables when modeling [12]. The type of input field of association rule data mining was generally required to be number, which could realize the digitization of physical fitness test data according to the national physical fitness measurement standard, that is, the test results could be scored as 5, 4, 3, 2, and 1 according to their advantages and disadvantages. In association rule data mining, the type of output field required to be set was character type (string), which meant that the output field “physical exercise behavior” was divided into five types as follows: pre-expectation stage, expectation stage, preparation stage, action stage, and maintenance stage. Association rule data mining was to find the minimum support min specified by the user in the transaction database $D_\_ Sup$ and min_ All association rules of conf [13]. Set min as required_ Sup, “individual index score of physical fitness test” was taken as the input field and set min_ Conf, establishing the “association mining rules between physical exercise behavior and single indicators” model to explore the impact of physical exercise behavior on physical fitness and providing decision support for the college students’ physical health with different physical exercise behaviors.

Since there were collinearity problems in the college students’ physical health indicators, such as body weight and vital capacity, the principal component regression analysis was adopted. Principal component regression analysis transformed a group of multiple related data into a group of linearly unrelated variables through orthogonal changes, and the transformed group of variables was called the main component [14]. The main method of principal component analysis was F1 (the first comprehensive index), that is, the greater the variance of F1, the more information it contained. Therefore, the variance of the first comprehensive index selected in all linear combinations became greater. If the information contained in the first index selected was not enough to include all indexes, F2 shall be included on the basis of the first index until it was accepted. The input indicators could represent the information of all the original indicators [15]. The second stage was to construct a regression model from the principal components included. The constructed regression model was of great significance for the screening of regression variables. In a word, the principal component regression model could use a set of fewer variables to obtain the selection of the best variables [16]. The specific process was as follows:

Since each index was different in units, the value was also different, so the data must be standardized, and the processing method adopted $Z$ change:

$$X_i = \frac{x_2 - x_1}{\gamma}$$  \hspace{1cm} (1)

In formula (1), $x_1$ and $x_2$, respectively, represent the physical health data of the college students were measured twice, $\gamma$ represents the number of indicators, and $n$ represents the standardized function of data. For principal component suitability test, such as Bartlet Test statistics and KMO test, if the $p$ value obtained by Bartlet Test was less
than 0.05 or the value of KMO test was greater than 0.5, the principal component analysis could be performed.

The correlation coefficient matrix between each index was calculated, and the correlation formula is as follows:

\[ r_{ij} = \frac{1}{n} \sum_{k=1}^{k} x_{ik}x_{jk}. \] (2)

The matrix of correlation coefficient was changed by \( R \) to make its variance reach the maximum value, and on this basis, the eigenvalue of \( R \) and \( I \) and their corresponding eigenvector were obtained.

The principal component formula was constructed according to the component coefficient matrix. It should be noted that the variables in the constructed expression were not original variables but standardized variables.

The construction of principal component regression equation required it to take the extracted principal components as independent variables and physical fitness score as dependent variables to establish a principal component regression model and select meaningful factors [17].

2.2. Redundancy Elimination of College Students’ Physical Fitness Data Based on Frequent Itemsets. The process of physique data mining in the college students’ physical exercise stage is shown in Figure 1.

Frequent Itemsets with support were greater than the minimum support threshold. The enumeration method could be used to enumerate all possible \( K \) itemsets, and then calculate the support of each itemset. A dataset with items could produce a set of items, and a set of items that met the support threshold could be small [18]. Obviously, enumeration was not an effective method when the data set was large. The frequent itemset enumeration lookup process was shown in Figure 2. As seen from Figure 2, there were altogether 15 data sets with 4 data sets.

In order to improve the efficiency of finding Frequent Itemsets, itemsets that were unlikely to reach the support
threshold were deleted. If the itemset was frequent, its subset must also appear frequently. Conversely, if a child itemset was infrequent, its parent set was not a frequent itemset. The redundancy elimination of frequent itemset data is shown in Figure 3, and the shaded itemset in the figure indicates that it is cropped out [19].

2.3. Physical Fitness Evaluation of College Students at the Stage of Physical Exercise Behavior Based on Bayes Theorem. Bayes theorem is a theorem for calculating conditional probability proposed by Bayes in the 18th century [20]. The main contents of the theorem include assuming that a complete event contains n mutually exclusive events, each mutually exclusive event is represented by $H_1, H_2, \ldots, H_n$ and its probability is $P(H_i), i = 1, 2, \ldots, n$. It is found that event $A$ occurs randomly with each mutually exclusive event, and when the conditional probability of event $A$ when each mutually exclusive event occurs is known, that is $P(A|H_i)$, then the probability of $P(H_i|A)$ is as follows:

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{\sum_{j=1}^{n} P(H_j)P(A|H_j)}.$$ (3)

Bayesian network (BN) is also called causal probability network and reliability network. The core of this method is to construct the Bayesian network topology between variables based on causal relationship and analyze the occurrence probability of variables through structure learning, parameter learning, and reasoning analysis. The Bayesian network topology learning method based on the ranking principle of causality and the principle of reason pointing to result is adopted to make the constructed Bayesian network topology clear [21–23]. The transformation of physical fitness evaluation index system of college students in physical exercise stage is realized through the initial topology of Bayesian network, as shown in Figure 4.

**Figure 3**: Frequent itemset data redundancy elimination.

**Figure 4**: Initial topology of Bayesian network.
Through the transformation of the index system, the Bayesian network topology lacks the causality between the evaluation indexes. In order to increase the causality between indexes in the network topology and make the network structure more reasonable, this study adopted the social network analysis method to optimize the structure.

The Bayesian network topology was transformed through the index system, which lacked the causal relationship between evaluation indexes. In order to increase the causality between indexes in the network topology and make the network structure more reasonable, the logical reasoning method was used to identify the causal relationship between nodes. Based on the Bayesian network topology, the key risk factors and key risk relationships in the social network were analyzed, and the middle centrality of points was used for physical fitness evaluation [24]. The greater the middle centrality of the node, the stronger its physique evaluation ability was. Its calculation formula is as follows [25–27]:

\[
S_r = \sum_{j \neq k \neq r} \sum_{k} O_{jk}(r),
\]

where \(j \neq k \neq r\) and \(j < k\), and \(O_{jk}(r)\) represents the ability of point \(r\) to control the communication between point \(j\) and point \(k\). Assuming that there were \(W_{jk}\) lines in the path between point \(j\) and point \(k\), and \(W_{jk}(r)\) lines passing through point \(r\) in the path, then

\[
O_{jk}(r) = \frac{W_{jk}(r)}{W_{jk}}.
\]

The middle centrality theory of line referred to the degree to which the relationship between two nodes in the network structure was located at the center of the whole social network structure [28]. The greater the middle centrality of the line between two nodes, the stronger the risk transmission capacity was. The calculation formula is as follows:

\[
S_{x \rightarrow y} = \sum_{j \neq k \neq r} \sum_{k} O_{jk}(x \rightarrow y),
\]

where \(j \neq x \neq y \neq k \neq r\) and \(j < k\), and \(O_{jk}(x \rightarrow y)\) represents the ability of relationship \(x \rightarrow y\) to control the communication between point \(j\) and point \(k\) [29]. Assuming that there were \(W_{jk}\) bars in the path between point \(j\) and point \(k\), and \(W_{jk}(x \rightarrow y)\) bars in the path passing through the relationship \(x \rightarrow y\), then

\[
O_{jk}(x \rightarrow y) = \frac{W_{jk}(x \rightarrow y)}{W_{jk}}.
\]

On the above constraints, \(n\) experts in relevant fields were invited to judge the occurrence probability of the risk level of the node. The nth expert judged the probability of the risk level \(i, i = [1, 2]\) of the node \(A_i, B_i, t = [0, 1, 2]\) according to the seven-level language variables, transformed it according to the corresponding relationship, and calculated the triangular fuzzy probability \(PTI\), \(n\) of the evaluation index with the transformed triangular fuzzy number.

\[
P_{ti,n} = (x_{ti,n}, y_{ti,n}, z_{ti,n}).
\]

The obtained triangular fuzzy probability was averaged. Averaging was the arithmetic average of the evaluation results of experts, which aimed to rationalize the final probability value [30]:

\[
P_{ti,FAM} = \frac{P_{ti,1} \oplus P_{ti,2} \oplus \cdots \oplus P_{ti,n}}{n} = (x_{ti}, y_{ti}, z_{ti}).
\]

The mean triangular fuzzy probability was defuzzified and calculated by the face value mean method:

\[
P_{ti,AM} = \frac{x_{ti} + 2y_{ti} + z_{ti}}{4}
\]

Normalize the probability value so that the sum of the final probability value was 1:

\[
P_{ti} = \frac{P_{ti,AM}}{\sum_{i=0}^{5} P_{ti,AM}}
\]

According to the calculated root node prior probability distribution and child node conditional probability distribution and with the help of GENIE 2.3 software, the physical parameters of the college students in the stage of physical exercise were studied. Then to carry out the initial assignment to each node, the calculated probability value was imported according to the triangular fuzzy number, the subnode probability was calculated with the maximum likelihood estimation method built in the software, and finally, the probability value of each node was obtained, so as to realize the physical exercise behavior stage.

3. Experiment

3.1. Experimental Design. Determining the evaluation index was important in the comprehensive evaluation of college students’ health. The intelligent health promotion service system focused on the internal functions of the body and used intelligent instruments to directly measure individual vital capacity, body composition, and cardiovascular and balance ability, including height, weight, vital capacity, and balance ability. The specific forms of physical exercise for the college students included running, core strength training, and vital capacity.

According to the National Student Physical Health Standard (2014), physical fitness test was conducted for 3969 students. The test indicators included height, weight, 50 m running, vital capacity, standing long jump, and sitting posture forward flexion. In addition, male students needed to measure 1000 m running and one-minute pull up, while female students needed to measure 800 m running and one-minute sit-ups. The BMI of the college students was divided into four grades: 17.2 (female)/17.9 (male)~23.9 is normal, \(\leq 17.1\) (female)/17.8 (male) is low weight, 24.0~27.9 is overweight, and \(\geq 28\) is obesity. SPSS17.0 and R language 3.4.4 software were used to analyze age, height, weight, and physical health indicators, including 50 m running, vital capacity, standing long jump, sitting body flexion and
1000 m running (male), pull-up (male), 800 meters running (female), and sit-ups (female). The original data of 3969 students were carefully reviewed to delete the missing value data and retain the data of 3960 students. According to the needs of data mining, a physical database was created in the form of Excel 2010 according to gender, height, weight, and vital capacity, including sitting and lying precursor, standing long jump, sit-ups, pull up, 50 m running, and endurance (1000 m or 800 m). The original data were processed according to the standard.

3.2. Training Sample Preprocessing. For \( N \) samples \( 1, 2, \ldots, n \), there were \( p \) indicators \( x_{1i}, x_{2i}, \ldots, x_{pi} \) to represent the \( i \)-th physical fitness data and the \( j \) index result to obtain the physical fitness data matrix:

\[
\begin{bmatrix}
X_{11} & \cdots & X_{1p} \\
\vdots & \ddots & \vdots \\
X_{ni} & \cdots & X_{np}
\end{bmatrix}
\]  

(12)

Then, the mean and standard deviation were respectively expressed as follows:

\[
X_{ni} = \frac{1}{n} \sum_{i=1}^{n} x_{ij},
\]

\[
S_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \overline{x_j})^2}.
\]

(13)

After standardization, the new data was

\[
x'_{ij} = \begin{cases} 
\frac{x_{ij} - \overline{x_j}}{S_j}, & S_j \neq 0, \\
0, & S_j = 0,
\end{cases} 
\]

(14)

Experimental verification was carried out according to the new standardized data results.

3.3. Experimental Result

3.3.1. Accuracy of Physical Fitness Assessment. In order to verify the effect of this method on physical fitness evaluation,
3.3.2. Physical Fitness Assessment Time. In order to verify the evaluation efficiency of this method, the methods of literature [8], literature [9], literature [10], and literature [11] were compared with this method, and the evaluation time results are shown in Figure 5.

By analyzing Figure 6, it could be seen that there were differences in evaluation time under different methods. For group 1, the evaluation time of literature [8] method was 47.5 min, that of literature [9] method was 45.0 min, that of literature [10] method was 48.2 min, that of literature [11] method was 43.6 min, and that of this method was 3.2 min. For group 3, the evaluation time of literature [8] method was 36.9 min, that of literature [9] method was 47.5 min, that of literature [10] method was 40.8 min, that of literature [11] method was 47.0 min, and that of this method was 3.7 min.

On the whole, in the above three groups of experiments, the time of physique evaluation by this method was no more than 5 min, which showed that the physique evaluation efficiency of this method was high due to the optimal variables of physical fitness in the exercise behavior stage were selected, the redundancy of physical fitness data was eliminated by frequent itemsets, and the causal relationship was used to rank the stages of physical exercise behavior, so as to shorten the evaluation time of the method.

3.3.3. Evaluation Result Recall Rate. In order to verify the evaluation recall rate of this method, literature [8], literature [9], literature [10], and literature [11] were compared with this method, and the evaluation recall rate results were shown in Figure 7.

Figure 7 revealed that the recall rate was different under different methods. When the number of experimental iterations was 100, the data evaluation recall rate of literature [8] method was 64.5%, the evaluation recall rate of literature [9] method was 57.2%, the evaluation recall rate of literature [10] method was 70.6%, the evaluation recall rate of literature [11] method was 63.9%, and the evaluation recall rate of this method was 98.2%; When the number of experimental iterations was 300, the evaluation recall rate of physical data of literature [8] method was 50.8%, that of literature [9] method was 42.1%, that of literature [10] method was 67.5%, that of literature [11] method was 60.2%, and that of this method was 99.3%; The evaluation recall rate of this method was significantly higher than that of other methods, indicating that this method had a higher effect on physical fitness evaluation. This was because this study used frequent itemsets to mine the association mapping relationship in the data set and constructed the regression model and frequent itemsets could effectively improve the effect of physical fitness assessment by eliminating the redundancy of physical fitness data.

4. Conclusion

This study introduced a physical fitness evaluation method for the college students at the stage of physical exercise behavior based on Bayesian and data mining. The frequent itemset was used to mine the association mapping relationship in the data set to obtain the selection of the best physical variables. The data redundancy of the frequent itemset was used to eliminate, the causal relationship was used to sort the exercise behavior stage, the transformation of the physical evaluation index system in the physical exercise stage was realized through the initial topology of Bayesian network, and the physical evaluation in the physical exercise behavior stage was realized. The experimental results showed that

(1) When the data volume was 1000 GB, the accuracy of physical fitness evaluation in the physical exercise behavior stage of this method was 97.62%, which proved the effectiveness of this method.
(2) This method took less than 5 minutes to evaluate the physique in the stage of physical exercise behavior, which showed that the physique evaluation efficiency of this method was high.

(3) When the number of experimental iterations was 300, the recall rate of fitness data evaluation in the physical exercise behavior stage of this method was 99.3%, which showed that the evaluation effect of this method was high.

However, even though some progress in the physical exercise behavior stage of physical fitness assessment was made, it needed to be further improved.

Data Availability
The dataset can be accessed upon request.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

References


