Research Article

Application of Decision Tree in PE Teaching Analysis and Management under the Background of Big Data

Yong Che,1 Kaixuan Che,2 and Qinlong Li2

1Department of Public Education, Anyang Preschool Teachers College, Anyang, Henan 456150, China
2Exercise Science School, Beijing Sport University, Beijing 100084, China

Correspondence should be addressed to Qinlong Li; 2019210175@bsu.edu.cn

Received 5 May 2022; Revised 8 June 2022; Accepted 17 June 2022; Published 8 July 2022

Academic Editor: Vijay Kumar

Copyright © 2022 Yong Che et al. 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.

With the continuous development of physical education reform, the defects and deficiencies of physical education teaching in colleges and universities are increasingly exposed. The reform of the original physical education teaching thought, education system, teaching mode, and method has achieved little. At present, the research on the prediction of physical education teaching achievements is mainly aimed at the prediction of athletes and physical education teaching achievements or the prediction of the past data of college students. This paper studies the physical education under the decision tree under the background of big data and constructs the physical education management system. When the number of tests reaches 40, the qualified rate of long-distance running is 65.2%, that of basketball is 68.1%, and that of volleyball is 68.2%. The quality of physical education teaching determines the lifeline of the development of school physical education teaching. In the process of collecting and selecting teaching materials, this paper enriches teaching materials and teaching reflection, cultivates one’s own understanding, and improves the artistic appeal and creativity of teaching.

1. Introduction

As the grassroots management of schools, the purpose of PE teaching management is to maximize the teaching efficiency. Improving the efficiency of PE teaching has become the starting point and destination of PE teaching management. Through modern means, the PE teaching management can be standardized, automated, and networked. The reform of the original PE teaching thought, education system, teaching mode, and method has little effect [1, 2]. At present, in China’s education, teachers are the center of high schools in most areas to strictly manage students, while students are the center after college, which requires students to have a high degree of consciousness [3]. This gap has led to a significant decline in some students’ performance, and the performance prediction system has become the main means to prevent and solve this problem [4, 5]. However, due to the influence of many factors on university students’ academic performance, the shortage of manpower in the prediction system, and the influence of human factors on the prediction results, the traditional comprehensive performance prediction strategy of university students has some problems such as low efficiency and poor prediction accuracy [6]. The teaching mode of PE teaching is very urgent and necessary and has a wide range of practical and guiding significance. The research on the teaching mode of PE teaching is essentially the research on the overall and reasonable procedure of the teaching process of PE teaching [5, 7].

At present, the research on the prediction of PE teaching achievements is aimed at the prediction of athletes and PE teaching achievements or the prediction of college students’ past data. It gives the specific implementation process to realize the prediction of PE teaching performance [8, 9]. The decision tree algorithm under big data determines the attribute with the best classification of the training set among the conditional attributes [10].

As an important technical means in big data, the decision tree algorithm integrates the technologies of machine
learning, data statistics, intelligent database, neural network, and so forth. It mines the potential data of related influencing factors through the algorithm and provides an ideal model of modern teaching evaluation database for educational management and decision-making [11, 12]. By considering the factors that affect PE teaching, the decision tree is constructed, the comprehensive evaluation of the influencing indicators is realized, and targeted solutions to improve the instructional quality are given [13]. Applying the decision tree under big data to explore the modern PE teaching evaluation system, providing the ideal model hidden in the modern PE teaching evaluation database for PE teaching administrators and decision makers has certain theoretical significance and application value for promoting the information management of PE teaching and improving the quality of school PE teaching [14]. The decision tree technology in big data is introduced into the evaluation system of PE instructional quality, with the aim of promoting the theoretical research of PE instructional quality evaluation and improving the quality of PE teaching.

2. Related Work

Wang et al. [15] proposed that there is a lack of clear teaching objectives in PE teaching. Although, in the process of continuous promotion of instructional reform in recent years, education on physical fitness and moral quality has been added in PE teaching curriculum, there is no clear positioning in the setting of specific talent training objectives [15]. Silverman [17] put forward that PE teaching is a very complex professional business activity, involving many factors, such as teachers, students, managers, and teaching conditions [16]. Vercruysse et al. proposed that the application of computer and network integrated technology can reduce the labor intensity of PE teaching teachers and managers in school PE teaching management, improve work efficiency, and promote the reform and development of school PE teaching. Therefore, it is more and more urgent to apply sports technology to network teaching [17]. Ward et al. [18] proposed that the development of society is inseparable from the promotion of talents. As the cradle of cultivating talents, it is imperative to speed up the reform of PE teaching mode under the guidance of professional physical fitness, but there are some difficulties in this process. Therefore, it is necessary to constantly strengthen the research on PE teaching mode under the guidance of professional physical fitness [18]. Meendering et al. [19] proposed that the management of students in PE teaching is mainly carried out through teachers’ PE teaching classes and related activities. This includes the management of students’ organization, compliance with teaching conventions, attendance, physical condition, learning performance, coordination, and control in the teaching process [19]. Curran and Standage [20] put forward new ideas, as well as the overall instructional reform and experiment under its guidance, including PE teaching objectives, teaching contents, teaching methods, organizational forms, and teaching means, and gradually established one group or even several groups, in line with the spirit of reform and the teaching reality of colleges and universities in China’s teaching procedures with strong operability and systematization, corresponding structure and function of PE teaching curriculum and teaching process structure, and framework or type [20]. Gu and He [21] put forward that there are still many problems in the current PE teaching in higher vocational colleges, and the teaching ideas cannot keep up with the development requirements of the times. When carrying out PE teaching, they are greatly influenced by the traditional teaching ideas, and the teaching methods are mainly classroom teaching, focusing on the teaching of knowledge points in textbooks, but there is a lack of teaching on physical skills [21]. Li et al. [22] put forward that the PE teaching process is a process of gradual improvement, and the management of this process should be continuous, uninterrupted, and disjointed, and some routine cannot be changed casually. From the time point of view, although it is divided into different links and stages, the teaching management cannot be relaxed, and it is necessary to persevere [22]. Wang et al. [23] put forward that expanding the enrollment scale of colleges and universities is an important strategy for the transformation of which has a very positive significance. However, it is undeniable that it also has negative aspects that affect the quality of higher education. Under this environment, the decline of instructional quality in college PE teaching is more serious [23]. Fang L et al. put forward that the teaching content of PE teaching is unscientific, and teaching is to cultivate applied talents. Therefore, the teaching content of the course must be targeted in terms of occupation. However, it is obvious that this point is ignored in the current PE teaching process, and the setting of teaching content is unscientific, which has a great impact on the cultivation of students’ sports consciousness [24]. At present, many schools pay close attention to management, and each school has strict requirements on its teaching methods, classroom discipline, and teaching arrangements and even constantly updates them. The campus system is filled in to strengthen management and it is hoped to achieve the education standard with the help of the seriousness and meticulous management. However, in such strict management, the teaching effect of some schools is sometimes criticized because it cannot achieve the desired effect or even backfire. It can be seen that, under the strict teaching management, there are still many problems in principle which we need to find out.

This paper studies the analysis and management of PE teaching under the background of big data. Because of its asynchronous teaching characteristics of rich teaching materials and unlimited learning time, PE teaching can undertake learning outside the PE teaching classroom and there can be an extension of the PE teaching classroom to expand students’ learning content and learning time. Teachers can also guide students’ after-school physical exercise and learning. In this paper, the construction of sports information management system is based on the security of the network, which is fully considered by the operators. The PE teaching management system under the decision tree mainly includes teaching guiding ideology, teaching objectives, teaching structure, teaching system, teaching content, syllabus, teaching organization form, teaching methods, teaching means, teaching evaluation, and related conditions.
3. Research on Decision Tree Algorithm under the Background of Big Data

3.1. Algorithm and Model of Decision Tree. Because of the huge dataset, it is impossible to process all the data in one time when calculating in memory, and a lot of data needs to be temporarily stored in disk. Because the decision tree algorithm needs to read and write data, the huge data makes the reading and writing speed slow. The training process of decision tree is iterative splitting and growing from the root node to the bottom, layer by layer. Regulating the construction of the whole tree is mainly used to train the tree model of big data. The controller can be involved in the computing cluster, and it is an independent computer, which performs distributed computing through MapReduce. When designing the development system architecture, on the one hand, we should combine the user’s needs and select the appropriate architecture according to the requirements definition; on the other hand, when selecting the system architecture, we should consider the maturity and foresight of the current architecture.

The main reason why decision tree algorithm is so popular in data analysis and data mining applications is that the construction of decision tree does not need any domain knowledge, which is very suitable for exploratory knowledge mining, and it can process high-dimensional data. Among numerous data mining and statistical analysis algorithms, the greatest advantage of decision tree is that it produces a series of rules from root to branch (or leaf), which can be easily understood by analysts and business personnel. Moreover, these typical rules even need to be sorted out slightly, which are ready-made business optimization strategies and business optimization paths that can be applied. In addition, decision tree technology is very tolerant to the distribution or even lack of data and is not easily affected by extreme values.

The architecture of PE teaching management system based on decision tree algorithm is shown in Figure 1.

Using the basic idea of decision tree algorithm, the attribute with the largest information gain rate is selected as the splitting attribute to form branches, and then the algorithm is called recursively for each branch until it cannot split a new branch. The calculation formula of information gain rate is as follows:

\[
\text{GainRatio}(A) = \frac{\text{Gain}(A)}{\text{Split}(A)}
\]

In the above formula, the information gain \( \text{Gain}(A) \) represents the information gain of attribute \( A \), which can be calculated according to the information gain in the decision tree algorithm. \( \text{Split}(A) \) is called split information, and its calculation method is as follows:

\[
\text{Split}_A(D) = -\sum_{i=1}^{v} \frac{|e_i|}{D} \times \log_2 \left( \frac{|e_i|}{D} \right).
\]

The training dataset \( D \) has \( m \) classes \( C_i \) \((i = 1, 2, \ldots, m)\). If \( A \) is a split attribute and has \( v \) values \( v_1, v_2, \ldots, v_v \), \( D \) is divided into \( v \) subsets \( \{e_1, e_2, \ldots, e_v\} \), and the number of tuples belonging to class \( C_j \) in subset \( e_i \) is \( |e_i| \); then the probability \( p_{ij} \) belonging to class \( C_j \) is \( \frac{|e_i|}{|D|} \).

For a training sample, the error between the actual output and the expected output can be defined as

\[
\text{Error} = \frac{1}{2} \sum_{k=1}^{c} (T_k - O_k)^2,
\]

where \( c \) is the number of output layer units, \( T_k \) is the expected output of output layer unit \( k \), and \( O_k \) is the actual output of output layer unit \( k \).

Let \( S \) be a set of \( s \) data samples, the class attribute has \( m \) different classes \( C_i \), and \( s_i \) is the number of samples contained in \( C_i \) class, and the information entropy required to classify a given sample is given by the following formula:

\[
I(s_1, s_2, \ldots, s_m) = -\sum_{i=1}^{m} p_i \log_2 (p_i),
\]

where \( p_i \) is the probability that any sample belongs to \( C_i \), and it is estimated by \( s_i/s \).

The law of data is changing over time. New data may not match the old parameters well, and the model is also changing over time. A single model is difficult to fully describe the data. Therefore, how to update the decision-making model according to the trend of data change is also the development direction of decision tree algorithm model in the future. In order to solve the problems of limited memory and training efficiency faced by the streaming data classification algorithm, the most effective way is to use the divide-and-conquer method to decompose the original computing task into several identical subtasks to deal with, so that each computer node can balance the load. Through the established decision tree model, each instance in the test data stream is predicted in real time, and the prediction results are matched with the class labels printed in advance, and then the matching results are inserted into the adds window to detect the concept drift. The detailed growth process of the decision tree algorithm is shown in Figure 2.

Let the dataset \( S \) contain \( s \) data samples, and let the category attributes take \( m \) values, corresponding to \( m \) different categories \( c_i \), \( i = 1, 2, \ldots, m \). If \( s_i \) is set as the number of samples in category \( c_i \), then the amount of information required to classify a given sample is

\[
I(s_1, s_2, \ldots, s_m) = -\sum_{i=1}^{m} p_i \log_2 (p_i),
\]
Start

Initialize root node

Leaf nodes and substitution subtrees

Add to window

Release sample

The process triggered when the number of samples reaches

End

The number of samples in the alternative subtree reaches the trigger process

Conceptual generation

The node is replaced by a new replacement subtree

The algorithm reads the input test data stream

Figure 2: Flow chart of decision tree algorithm.
Computational Intelligence and Neuroscience

\[
E(A) = \sum_{j=1}^{n} \frac{(s_{1j} + s_{2j} + \cdots + s_{mj})}{s} I(s_{1j}, s_{2j}, \ldots, s_{mj}) \times \frac{1}{s} \sum_{j=1}^{m} \frac{s_{1j} + s_{2j} + \cdots + s_{mj}}{s} \quad (6)
\]

A attributes the information gain obtained by the set partition of the current branch node.

\[
gain(A) = I(s_1, s_2, \ldots, s_m) - E(A). \quad (7)
\]

According to the knowledge of limit theory, when \(0 \leq x \leq 1\), \(n\) tends to infinity, and, with the increase of the power degree in the above formula, the result after the third term will become smaller and smaller. Compared with the first two terms, the latter terms are approximately 0, and the final function \(f(x)\) is simplified to

\[
\ln(1 + x) = x - \frac{x^2}{2}.
\]

From the definition in the previous decision tree algorithm, the calculation formula of information amount is

\[
I(s_1, s_2, \ldots, s_n) = -\sum_{i=1}^{n} \frac{s_i}{s} \log_2 \frac{s_i}{s} \quad (9)
\]

The above formula is further simplified as

\[
I(s_1, s_2, \ldots, s_n) = -\sum_{i=1}^{n} \frac{s_i}{s} \log_2 \left(1 + \left(\frac{s - s_i}{s}\right)\right). \quad (10)
\]

Through the input training sample, the method of inserting decision leaf node is called to insert the training sample from the root node to the corresponding leaf node, and the statistical information of the leaf node is updated. If all the arrived samples belong to the same class, the method of splitting the leaf node is called, in which the number of samples arrived by the leaf node is artificially set to avoid continuous calculation. If the calculation results meet the constraints and other corresponding requirements, split is executed, and the number of newly added leaf nodes is the number of the decision attributes of the leaf nodes.

3.2. Realization of PE Teaching Management System. At present, most public PE teaching courses in universities in China adopt the option teaching method. Although this teaching mode has very positive significance compared with the previous unified teaching, there are still some problems in the setting of teaching content, the arrangement of teaching hours, and the guidance of after-school PE teaching. This paper constructs a PE teaching management system under the BIG DATA background decision tree. In terms of teaching guiding ideology, it pays attention to imparting basic sports knowledge, basic technology, and skills, as well as enhancing students’ physique, and the teaching organization form is mainly the original teaching class. Such curriculum can relatively meet students’ interest needs and meet the development trend of modern education, but there are still some problems that are difficult to solve in practical operation, such as the division of classes at the beginning of the semester and the submission of students’ scores by departments at the end of the semester. From the perspective of structure in the teaching management system, the data types include unstructured, semi-structured, and structured. In order to unify and integrate multichannel data and facilitate later BIG DATA processing, it is necessary to normalize it and preprocess it and finally store it in the BIG DATA teaching management system. The pretreatment, management, and collection of teaching status data in colleges and universities need to compare the connotation of data. Check the data sources one by one and provide data support materials for important data. At present, the teaching status data acquisition system in both undergraduate colleges and higher vocational colleges uses the stand-alone version, which has the advantages of good data confidentiality and simple operation.

4. Experimental Results and Analysis

25 students and 5 PE teaching teachers were randomly selected to participate in the system performance evaluation test. The PE teaching and learning information of 25 students was completely entered into the system. Students and teachers used three kinds of systems for routine operation and generate intelligent teaching decision-making scheme. In the test, the response time required for users to log in to the main function interface of the system is recorded. After the test, the satisfaction generated by students and teachers using the system and the evaluation of the rationality of intelligent teaching decision-making are understood in the form of questionnaire. Through the first step of operation, the students’ PE teaching scores are expressed by “pass” or “fail.” On this basis, the numbers of passing students and failing students in each sports test item are counted. The results are shown in Tables 1 and 2.

As shown in Tables 1 and 2, the average response time of the system entering the main function interface is 0.31 s, indicating that the system has no obvious stuck problem, and the response speed is faster than those of the other two systems, and the sensitivity of the system entering the main function interface is higher. Students’ and teachers’ satisfaction levels with the system reached 96.5% and 95.7%, respectively, which were excellent. However, the distributed intelligent management system of PE teaching in colleges and universities achieved only 87.8% satisfaction from teachers, which showed that the system interface design layout, decision-making accuracy, function setting, and other aspects did not meet the requirements of teachers.

In order to accurately record the changes of utilization and the number of users, score entry, score import, and score query are carried out to record the changes of indicators in each stage. The test records the operation process of performance management. The changes of test data, utilization, and the number of users are shown in Figure 3.

As can be seen from Figure 3, when the number of users is 5000, the utilization rate of score entry is 15.5%, the utilization rate of score import is 19.3%, and the utilization rate of score query is 18%. Only with the increase of the number of users is the utilization rate reached at the time of
Nevertheless, the utilization rate is within the allowable range.

In this experiment, students’ scores of Class 1, Class 2, and Class 3 were selected as training sets by the method of overall sampling. Long-distance running, basketball, and volleyball were tested, with a total of 150 records. The scores of these three classes were copied to the training example worksheet. Using the function, the number of people who passed and failed in a single subject in the training set was found out. The experimental results are shown in Figure 4.

As can be seen from Figure 4, when the test number reaches 40, the qualified rate of long-distance running is 65.2%, that of basketball is 68.1%, and that of volleyball is 68.2%. The quality of PE teaching determines the lifeline of school PE teaching development. To have a team of high-quality PE teaching teachers, we must have a complete evaluation system of PE instructional quality.

In order to determine the research direction and feasibility of this topic, Delphi method was used to issue three rounds of consultation questionnaires to experts in education and sports. According to the survey statistics, "very important," "important," "relatively important," "average," and "irrelevant" were assigned to 5, 4, 3, 2, and 1, respectively. According to the evaluation results of the first round of expert consultation questionnaire, the indicators with lower average value were excluded item by item and sorted out, a questionnaire was made and returned to each expert to evaluate the recovery of the expert consultation questionnaire again, as shown in Table 3. Experts evaluated the structural validity and content validity, as shown in Tables 4 and 5.

From the results of expert consultation and evaluation, the structural validity is 85.74% and the content validity is 88.33%, indicating that this questionnaire has high structural validity and content validity and is considered a formal questionnaire. The satisfaction of college PE teaching intelligent management system based on data warehouse is less than 85%, which is also not ideal.

In order to accurately record the relationship between each performance index and the number of people, the changes of indexes in each stage are recorded by grade entry, grade import, and grade inquiry. The test records the operation process of performance management, and the changes of test data, average response time, and number of users are shown in Figure 5.

As can be seen from Figure 5, when the number of users increases, the response time of score query suddenly increases to about 2.7 seconds, but it does not exceed 6 seconds. This shows that even if the maximum number of users is reached, the changes of throughput and number of users, score entry, score import, and score query are carried out, respectively, to record the changes of indicators in each

### Table 1: Training set data.

<table>
<thead>
<tr>
<th></th>
<th>Long-distance run</th>
<th>Basketball</th>
<th>Short-distance run</th>
<th>Volleyball</th>
<th>Shot</th>
<th>High jump</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of passing students</td>
<td>54</td>
<td>65</td>
<td>100</td>
<td>121</td>
<td>106</td>
<td>98</td>
</tr>
<tr>
<td>Number of failing students</td>
<td>103</td>
<td>92</td>
<td>58</td>
<td>38</td>
<td>53</td>
<td>62</td>
</tr>
</tbody>
</table>

### Table 2: Performance evaluation of PE teaching management system.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>System response time</th>
<th>Student satisfaction (%)</th>
<th>Teacher satisfaction (%)</th>
<th>Management rationality (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>0.31</td>
<td>96.5</td>
<td>95.7</td>
<td>97.7</td>
</tr>
<tr>
<td>Distributed intelligent management system of PE teaching in colleges and universities</td>
<td>0.83</td>
<td>90.3</td>
<td>87.8</td>
<td>93.1</td>
</tr>
<tr>
<td>Intelligent management system of college PE teaching based on data warehouse</td>
<td>0.58</td>
<td>88.3</td>
<td>88.3</td>
<td>90.3</td>
</tr>
</tbody>
</table>

![Figure 3](image1.png)  
**Figure 3:** Changes of test data, utilization, and number of users.

![Figure 4](image2.png)  
**Figure 4:** Qualification rate of different sports.
stage. The test records the operation process of performance management. The changes of test data, throughput, and number of users are shown in Figure 6.

As can be seen from Figure 6, when there are 4,000 users, the throughput of score entry is 24, the throughput of score import is 24, and the throughput of score query is 26. When there are 6,000 users, the throughput of score entry is 254, the throughput of score import is 23, and the throughput of score query is 24. This shows that the business processing ability meets the requirements under the premise of the maximum number of users.

At present, physical education teaching generally only pays attention to the management of physical education teaching methods but lacks a clear evaluation system for the corresponding management effect and management system. This problem exists in the majority of colleges and universities. At this stage, the performance of this issue is also multifaceted. In the process of making corresponding teaching plans, the relevant departments of the school lack the cognition of corresponding teaching results and even neglect monitoring them. This experiment aims at the main problems existing in the management of PE teaching in colleges and universities, such as the shortage of teachers, the shortage of venues and equipment, and the bad teaching environment. The experimental results are shown in Figure 7.

Figure 7 shows the changes of three different factors in the teaching environment. Among the main problems existing in the management of physical education in colleges and universities, the number of teachers, the content changes of on-site equipment, and teaching environment are the main factors. It can be seen from Figure 7 that, among the main problems existing in the management of PE teaching in universities, when the number of test samples is 40, the proportion of insufficient teachers is 29.3%, the proportion of insufficient venues and equipment is 28.8%, and the proportion of poor teaching environment is 23.5%. The proportions of the three factors are almost the same. In the process of collecting and screening materials, enriching
teaching materials, and teaching reflection, their own understanding is cultivated and the artistic appeal of teaching expressiveness and creativity is improved.

5. Conclusions

The way of physical education teaching will change from “teacher centered” to “student autonomy,” reflecting the characteristics of subjectivity and modernity. This paper studies the physical education under the decision tree under the background of big data and constructs the physical education management system. When the number of tests reaches 40, the qualified rate of long-distance running is 65.2%, that of basketball is 68.1%, and that ofvolleyball is 68.2%. In the process of realizing informatization, colleges and universities should learn from each other. Guided by the spirit of the report of the 17th National Congress of the Communist Party of China, we should thoroughly implement the “Scientific Outlook on Development” and promote the reform and development of physical education teaching in colleges and universities in Henan Province. The decision tree algorithm under big data should highlight the diversity and acceptability of physical education teaching content. From the perspective of physical education curriculum, the text curriculum should be transformed into students’ own curriculum. In this paper, the decision tree technology in big data (hereinafter referred to as big data) is introduced into the PE teaching quality evaluation system, which promotes the theoretical research of PE teaching quality evaluation and improves the PE teaching quality. However, there are still some limitations in the research. The research has not evaluated the individual abilities of different athletes, so further analysis is needed in the future research.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


