Review Article
The Application Method of Big Data of Data Mining Algorithm in College Basketball Teaching

Yu OuYang, 1 YanHong Zhang, 1,2 and Youliang Li 1

1 School of Physical Education, Nanchang Institute of Science and Technology, Jiangxi, Nanchang 330108, China
2 Shaanxi Normal University, Shaanxi, Xi’an 710062, China

Correspondence should be addressed to YanHong Zhang; zhangyanhong@peihua.edu.cn

Received 7 June 2022; Revised 29 June 2022; Accepted 5 July 2022; Published 23 August 2022

1. Introduction

Today is an era of informatization, and also an era of the knowledge economy. As one of the most active and important factors, talents should be highly valued. As a base for cultivating high-quality talents, colleges and universities should pay more attention to the quality of talent training, provide high-quality talents for various constructions of our society, and realize the high-quality development of my country’s economy [1]. In addition, the competition among countries is becoming more and more fierce. The essence of the competition is the competition of talents. If my country wants to occupy a more favorable position in the world, it needs to strengthen the training of talents. The cultivation of talents depends on education, and the quality of basketball teaching is of great significance to the quality of talent cultivation. Colleges and universities should fully recognize the importance of basketball teaching quality, not only to take effective measures to improve the teaching quality but also to take effective methods to monitor and evaluate the teaching quality [2–4], to ensure that the teaching quality of colleges and universities meet the needs of students development requirements.

With the continuous popularization of the Internet, basketball news information, and basketball events have been widely disseminated on the network platform, and the public’s enthusiasm for basketball is rising. College basketball teaching is facing many challenges and opportunities. In the online teaching platform, teachers can integrate and disseminate rich teaching resources, exchange and share teaching experiences with other experienced teachers, and even communicate with distant teachers through remote interaction to improve teachers’ teaching levels. Under the background of big data, the assessment and management of basketball teaching are more professional and informative, which standardizes the evaluation model of basketball teaching and training, and improves the effectiveness of basketball teaching; colleges and universities can also use digital assessment and management to improve basketball training and teaching. Carry out assessment and management, change the traditional and vague evaluation methods, and continuously optimize the relevant standards and norms of basketball training and teaching, so that basketball learning can achieve the goal of skill, professionalism, and healthy exercise.

There are many methods for monitoring and evaluating the quality of basketball teaching in colleges and universities,
among which data mining technology [5, 6] is a more effective method and has strong practical value. Relevant personnel should conduct in-depth research on data mining technology [7] so that it can play a greater role in the monitoring and evaluation system of basketball teaching quality in colleges and universities so that colleges and universities in our country can cultivate higher-quality talents. The guiding role of data mining [8] in basketball teaching is shown in Figure 1.

Data mining technology can perform data classification and prediction, cluster analysis and association analysis, etc., and can conduct deep data mining, which is an important research field to improve analysis and decision-making capabilities [9–11]. The introduction of data mining technology into teaching quality monitoring, objective analysis of existing performance data, and mining of valuable information will undoubtedly help improve teaching measures and improve teaching quality. At present, there have been related research on the application of data mining technology to teaching management [12]. For example, the literature studies the course association classification model and student achievement prediction algorithm based on frequent pattern spectral clustering; the literature proposes a K-nearest neighbor based local optimal The reconstructed incomplete data imputation method combined with the random forest model realizes the prediction of grades; the literature studies how to use the undergraduate grade data to infer the students’ performance during the postgraduate period by means of various prediction and statistical methods. On the basis of learning from previous research experience, this paper uses factor analysis to comprehensively evaluate and analyze the performance of students majoring in computer science and proposes an improved decision tree method to predict student performance. A detailed comparative analysis was carried out, and a method that could better promote the monitoring of teaching quality was found.

2. Related Work

2.1. Data Mining and Knowledge Discovery. The superficial meaning of data mining technology means that when faced with a pile of data [13], it can be well processed, analyzed, and screened over and over again among these many data to select the most useful data. Databases are generally very complex. In this complex database, resources are continuously utilized. Through the study and mastery of data mining technology, the long-term accumulated data can be processed, and some data are random. If it is processed manually, it may increase the difficulty of the work of the staff and make the data cluttered. Users must learn to extract the data so that the data can be better used in people’s lives. In the process of learning data mining technology, it is necessary to continuously learn new knowledge, discover new knowledge, follow the local characteristics and rules of different systems, and the data can be updated in time, and the knowledge in the database must be analyzed in detail; this requires A lot of information, the patterns or concepts and laws of this kind of information are different, each has its own level, has a future prediction direction for the data, and makes decisions at the same time; another point is that data mining technology should be based on different The information of the students should be updated in a timely manner, and the information of the students should be understood in time, so as to have a good guiding effect on the course.

Knowledge discovery (KDD): knowledge discovery is a broader term for the so-called “data mining,” which is to obtain knowledge according to different needs from information represented by various media. The purpose of knowledge discovery is to shield users from the tedious details of the original data, extract meaningful and concise knowledge from the original data, and report directly to the users. There is still confusion between database-based knowledge discovery (KDD) and data mining [14], and the two terms are often used interchangeably. KDD represents the process of transforming low-level data into high-level knowledge. KDD can be simply defined as KDD is the specific process of identifying valid, novel, potentially useful, and basically understandable patterns in data. Data mining, on the other hand, can be thought of as the extraction of patterns or models from observational data, which is a general interpretation of data mining. Although data mining is at the heart of the knowledge discovery process, it is usually only a part of KDD (roughly 15% to 25%). Therefore, data mining is only one step of the whole KDD process, and there is no exact definition of how many steps and which steps must be included in the KDD process. However, a general process should accept raw data input, select significant data items, reduce, preprocess, and enrich data sets, transform data into appropriate formats, find patterns in data, and evaluate and interpret findings.

Because the quality of the data mining algorithm [15, 16] will directly affect the accuracy of the knowledge found by KDD, and most of the current KDD research focuses on data mining algorithms and applied technologies, people often do not strictly distinguish between data mining and knowledge discovery in databases both use each other. Generally, it is

![Figure 1: The role of data mining in basketball teaching.](image-url)
called KDD in the field of scientific research, and it is called data mining in the field of engineering [17].

2.2. Data Mining Process. The KDD process is shown in Figure 2. The KDD process can be summarized into three parts: data preprocessing, data mining, and interpretation and evaluation of results.

2.2.1. Data Preprocessing. Data preprocessing refers to the set of techniques implemented on a database to remove noise, and missing and inconsistent data. The different data preprocessing techniques involved in data mining are data cleaning, data integration, data reduction, and data transformation. The need for data preprocessing stems from the fact that real-time data and many times database data are often incomplete and inconsistent, which can lead to incorrect and inaccurate data mining results [18]. Therefore, in order to improve the quality of the data to be observed and analyzed, it can be processed through these four steps of data preprocessing. The more data you improve, the more accurate observations and predictions will be. Figure 3 shows the steps of data preprocessing.

2.2.2. Data Mining. Data mining technology refers to a process of automatically retrieving from a large amount of complex data and automatically sorting out relevant information. To be precise, data mining is the “eye” for discovering knowledge base data, that is, to regularly search for chaotic data in massive data, so as to sort out the information that people need in an orderly manner. In fact, the process of data mining is fully automated, but many experts point out that only 80% of the time and experience in the data mining process is spent in the preprocessing stage. According to objective facts, many preparations need to be done before data mining technology. Even so, data mining technology is very convenient and can maximize work efficiency in practical applications [19]. In particular, teaching quality management using data mining technology in colleges and universities can assist teachers and college administrators in teaching analysis and other work, continuously develop and improve teaching systems and mechanisms, and promote the scientific development of schools. Accelerate the modernization of higher education with Chinese characteristics in the new era.

2.2.3. Interpretation and Evaluation of Results. The patterns discovered in the data mining stage may have redundant or irrelevant patterns after evaluation, and then they need to be eliminated; it is also possible that the patterns do not meet the user’s requirements, then it is necessary to fall back to the previous stage of the discovery process, such as reselecting. The data adopts new data transformation methods, sets new parameter values, and even changes a mining algorithm. Also, since KDD is ultimately intended for human users, it may be necessary to visualize the patterns found, or convert the results into another representation, that is, understandable to the user. Data mining is just one step in the overall process. The quality of data mining has two influence factors: one is the effectiveness of the data mining techniques used, and the other is the quality and quantity of the data used for mining (the size of the data). If the wrong data or inappropriate attributes are selected, or the data is inappropriately transformed, the mining results will not be good [20]. The whole mining process is a continuous feedback process. For example, the user found that the selected data was not very good during the mining process. Or the mining technique used.

2.3. Methods and Technologies of Data Mining. There are many ways to operate data mining technology. First of all, the genetic algorithm will be introduced. This method is based on the laws of nature, survival of the fittest, survival the fittest, combining different data, evolving, and finally merging together so that it becomes new data information, and the newly established data still needs to have the ability to select the overall situation, integrate the data, and finally become a data system, and then use the data, so that it will be very convenient to use [21]. It makes the arrangement of information more convenient; the second method is a decision tree, which is to first organize and summarize all the data, and then classify the data, organize the data information for branch processing, and search from the middle. The most valuable information, and then clean up the data that does not meet the conditions, because this process is like the process of growing a sapling, so it is called a decision tree. The biggest advantage of this method is that the operation is very simple. The process is also very smooth, and the most important thing is that the work efficiency is very high. This method is very suitable for a large amount of data, which will reduce the pressure and burden on the staff.

The first method is to clarify the management and decision-making problems. If there are problems in the process of education and teaching management, we must constantly summarize and summarize the emergence of such decision-making management problems, and at the same time identify them. Established in time, so that specific data can be turned into goals so that it can be redefined; the second method is to extract the original data, and the data can be managed and customized according to different goals, and because the teaching management information system The establishment of the database and other related teaching functions require the support of a lot of data [22]. In the process of extracting from the database, the noise data must be excluded, and it cannot be interfered with by the vacant data. These data are integrated and transformed for accurate processing; the third point is to design data and continuously mine data. Because of the different goals, for the completion of data mining tasks, various data algorithms should be used to establish a data algorithm. Processing model; the fourth is the refinement of data. With the huge database and the screening of value metrics, this data will have a corresponding mining mode, and at the same time, it must be processed and integrated according to different teaching management needs.
3. Explore the Application of Data Mining Technology in College Basketball Teaching

This chapter mainly explores the application of data mining classification technology in college basketball teaching, puts forward the implementation plan of data mining technology in college basketball teaching application, and introduces the implementation process of the plan by taking the analysis of students’ performance in college basketball teaching as an example.

3.1. The Implementation Process of Classification Mining. Data mining is a decision support process and a deep-level data information analysis method. It is undoubtedly very beneficial to apply data mining techniques to the evaluation of teaching. It can comprehensively analyze the hidden internal relationship between test results and various factors connection. For example, through the analysis of the school’s student performance-related database system, data mining tools can answer similar questions such as “what factors may have an impact on students’ academic performance,” which cannot be achieved by traditional evaluation methods? Through data mining and analysis, the evaluation results can bring unprecedented gains and surprises to teaching.

In the past, the database query method was usually used to process a large amount of data information in the teaching process. Here, the author proposes a classification algorithm in data mining, which can convert a large amount of data into classification rules, so as to better analyze these data. Figure 4 is a flow chart of classification implementation.

3.2. Data Collection. In this example, the author can discuss the basic learning situation of students (such as knowledge base, classroom learning effect, students’ interest in the course, homework completion, time spent after class, and learning methods used), what factors have an impact on academic performance, and what are the reasons why students’ academic performance is excellent or their academic performance is unsatisfactory, and expect to use the obtained analysis results to guide future teaching work [23].

Student achievement analysis is all about finding functional relationships between two or more attributes. To analyze the causes of students’ academic performance, we need data from the following aspects:

3.2.1. Basic Information about Students. The data structure is as follows: student number, name, gender, place of origin, department, major, and class. This information is available through the school’s student management information system.
3.2.2. Student Survey Information. The content includes the degree of love for the major, the course, the mastery of preschool knowledge, the effect of classroom learning, and learning methods [24]. This information is mainly generated by students filling out surveys. In the past, these tasks usually required the production of questionnaires. After students filled out the questionnaires, teachers spent a lot of time and energy collecting these data. Because this work is very tedious, it will take up a lot of time. Therefore, many teachers are reluctant to do this work, thus making it impossible to complete the very important work of mastering the basic information of students.

3.2.3. Grade Database. The score database includes students’ usual homework scores and course test scores. This database is generated by teachers during the teaching process.

In this example, the method adopted by the author is: that all assignments require students to upload the written examinations to the submission system in the form of electronic documents in the prescribed format and prescribed file name [25]. The basketball movement standard test is conducted in accordance with the prescribed movements and prescribed methods, and the results are registered.

3.3. Data Pre-Processing

3.3.1. Data Integration. It is merging data from multiple data sources together. In this study, the multiple database files obtained by data collection are used to generate the basic database of student achievement analysis by database technology, as shown in Figure 5.

3.3.2. Data Cleaning. The main job of data cleaning is to fill in missing data values.

In the basic database of student achievement analysis, we see that there are some attributes that we are interested in missing attribute values. For these vacancies, data cleaning techniques can be used to fill them.

For these vacancies, data cleansing techniques can be used to fill them. There are many ways to fill in blank values for properties:

1. Ignore tuples: this is usually done when the class label is missing or the tuple has multiple attributes with missing values.

2. Fill in missing values manually: generally speaking, this method is time-consuming and may not work when the dataset is large and many values are missing.

3. Fill in the missing value with a global constant: replace the missing property value with the same constant (e.g., “Unknown”). But if the missing values are all replaced with “Unknown,” the mining program might mistake them for an interesting concept, since they all have the same value—“Unknown.” Therefore, although the method is simple, it is not recommended here.

4. Fill in the empty values with the average value of the attribute.

5. Use the mean of all samples that belong to the same class as the given tuple.

6. Fill gaps with the most probable values which can be determined by regression methods, Bayesian methods, or decision tree induction.

In this example, the method of ignoring tuples is used to delete records that have not taken the test or that have a large number of vacancies in the online survey data filled out by students. For other individual vacancies, because the total number of records is not too many, and the vacancy values are few, other individual vacancies are filled manually. The filling principle is to use other attribute values of the record as filter conditions to filter in

---

**Figure 4:** The implementation process of classification mining.
the database. After filtering, use most of the attribute values to fill the vacancy.

3.3.3. Data Conversion. Data transformation is mainly to normalize the data.

This paper uses the concept stratification technique. Continuous-valued attributes can be converted to discrete-valued attributes (i.e., discretized). Histogram analysis is a relatively simple discretization method, which is divided into two categories: equal-width binning and equal-depth binning. Equal-width binning divides attribute values into equal parts or intervals. In equal-depth bins, the values are divided so that each part contains as many samples as possible. Here, using equal-depth binning for discretization, all values of the usual grade attribute are divided into three categories: the grades from 0 to 70 belong to “poor,” 70 to 85 belong to “average,” and above 85 belong to “good.”

3.3.4. Data Reduction. The purpose of data reduction is to reduce the size of the data to be mined, but it will not affect (or basically not affect) the final mining results. Here, the method of dimensionality reduction is adopted, that is, the really useful feature attributes are found from the initial feature attributes to reduce the number of feature attributes or variables to be considered in data mining.

Since there are many attribute fields in the student information table, in order to facilitate the establishment of the decision tree model, this paper selects the after-school practice time, the degree of understanding of the course before learning, the classroom learning situation, and the usual practice situation. The total grade attribute is used as the basis for establishing the total grade classification decision tree model to analyze the student’s learning situation.

3.4. Data Classification Mining. The purpose of classification mining is to establish a decision tree model for performance analysis.

Based on the characteristics of the dataset, in this classification mining research, in order to make the generated rules easy to understand, the decision tree method is chosen. Since the training set is not too large, you can choose the GongD3 or C4.5 algorithm for classification and mining, and here the GongD3 algorithm is selected for classification.

3.4.1. ID3 Algorithm. The most famous algorithm in the decision tree algorithm is the ID3 algorithm proposed by Quinlan. The ID3 algorithm starts with all training samples at the root node of the tree, selects an attribute to distinguish these samples, and produces a branch for each value of the attribute. Move the corresponding sample subset of branch attribute values onto the newly generated child nodes. This algorithm is applied recursively to each child node until all samples on a node are assigned to a certain class. ID3 algorithm is a greedy algorithm. It uses a top-down, divide-by-recursion approach to construct a decision tree.

Let S be a set containing s data samples, and the category attribute can take m different values, corresponding to m different categories C (1 = 1' · · · m). Suppose s1 is the number of samples in category C1, then, the amount of information required to classify a given data object is

$$I(s_1, s_2, \ldots, s_m) = -\sum_{i=1}^{m} p_i \log_2(p_i).$$  \hspace{1cm} (1)
where $p_i$ is the probability that any data object belongs to a category $C_i$.

Let an attribute $A$ take $V$ different values $\{a_1, a_2, \ldots, a_v\}$. Using the attribute $A$, the set $S$ can be divided into $v$ subsets $\{s_1, s_2, \ldots, s_v\}$. Among them, $s_j$ contains the number of samples belonging to category $c_j$ in subset $s_j$. Then the information required to divide the current sample set by attribute $A$ (the direct descendant) can be calculated according to the following formula:

$$E(A) = \sum_{i=1}^{v} \frac{S_{ij} + \cdots + S_{mj}}{S} I(S_{ij} + \cdots + S_{mj}).$$  \hspace{1cm} (2)

In this way, the information gain obtained by using attribute $A$ to divide the corresponding sample set of the current branch node is

$$\text{Gain}(A) = I(S_1, S_2, \ldots, S_m) - E(A).$$ \hspace{1cm} (3)

The information gain of each attribute is calculated by the above formula. Select the attribute with the highest information gain as the test attribute for the given set $S$, create a node, mark it with this attribute, create a branch for each value of the attribute, and perform sample division.

3.4.2. Using the ID3 Algorithm to Generate a Decision Tree Model. Since there are many attribute fields in the student information table, when establishing a decision tree model for whether the grades are good or not, this paper selects the after-class computer time, the degree of understanding of the course before learning, and the classroom learning situation. The attribute field of normal work status, whether the attribute is good or not is used as a category attribute. When establishing the decision tree model of whether the grade is failed or not, the attribute of failing or not is used as the category attribute.

3.4.3. Decision Tree Pruning. When a decision tree is just established since many branches are constructed from abnormal data in the training sample set (due to noise and other reasons), the decision tree is too “blooming,” which reduces the comprehensibility of the tree. At the same time, it also increases the dependence of the decision tree itself on historical data, that is to say, this decision tree may be very accurate for this historical data, but the accuracy drops sharply once it is applied to new data. The condition is called overtraining. In order to make the rules contained in the obtained decision tree have a general meaning, the decision tree must be modified. The task of branch pruning is mainly to delete one or more branches and replace these branches with leaves to simplify the decision tree, so as to improve the speed of classification and recognition in the future and the ability to classify and recognition of new data.

There are usually two methods for pruning branches, which are

(I) Prepruning Method. The method is realized by stopping the branch generation process in advance, that is, by judging whether it is necessary to continue to divide the training sample set contained in the node on the current node. Once the branch is stopped, the current node becomes a leaf node. The leaf node may contain multiple training samples of
different categories. Since this pruning is done before branching, it is called prepruning.

(2) Postpruning Method. This approach addresses the problem of overtraining from another angle. On the basis of allowing the decision tree to grow most fully, it cuts off those leaf nodes or branches that are not generally represented in the decision tree according to certain rules. After pruning, the pruned branch node becomes a leaf node, and it is marked as the category with the largest number of categories in the samples it contains.

Prepruning requires more computation time, but the resulting decision tree is more reliable. In this paper, the method of postpruning is adopted. First, the error rate of a fully grown decision tree is calculated, and a maximum allowable error rate is specified by the user. When the pruning reaches a certain depth and the calculated error rate is higher than the maximum allowable value, stop pruning immediately, otherwise, continue pruning. Figure 6 is a decision tree for the classification of failing grades after pruning.

4. Conclusion

Under the background of the increasing development of science and technology in our country, the combination of science and technology with higher education, the application, and research of intelligent information systems, is more conducive to promoting the modernization of colleges and universities. Through the application of data mining technology in the quality evaluation of basketball teaching in colleges and universities, the limitations of traditional basketball teaching management can be effectively improved, and various unfavorable factors that hinder teaching management can be found in time. In the school’s teaching, the use of data mining technology in basketball teaching can reasonably collect, analyze and summarize the relevant data of colleges and universities, and then find and solve problems in time, which not only reduces teaching accidents but also makes all aspects of colleges and universities. Management Metropolis eliminates drawbacks, coordinates the overall construction of the school, comprehensively cultivates high-quality talents and innovative ability levels, and promotes the construction of modern teaching in colleges and universities on the basis of improving the quality of teaching, laying a solid precondition for cultivating high-quality talents, and for the realization of lay the foundation for the construction of world-class universities and first-class disciplines [18–26].

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[6] Information Technology, "Data Mining: Data from iran university of science and technology advance knowledge in data mining (application of data mining techniques for the investigation of track geometry and stiffness variation)[ ]," Information Technology Newsweekly, 2020.
[12] Information Technology, "Data Mining: Research Conducted at Technical University Has provided New Information about Data Mining (Data Mining Application in Assessment of Weather-Based Influent Scenarios for a Wtwp: Getting the Most Out of Plant Historical Data)[ ]," Computers, Networks & Communications, 2019.


