

Research Article

Application of Decision Tree Algorithm Based on DM in the Development of Preschool Children’s PE

Qian Ning,1 Jamalsafria Bin Saibon,1 and Zhaoxia Guo2

1Universiti Sains Malaysia, School of Educational Studies, Minden, Penang 11800, Malaysia
2Taiyuan University of Technology, Sports Institute, Taiyuan, Shanxi 030027, China

Correspondence should be addressed to Qian Ning; ningqian827@student.usm.my

Received 23 February 2022; Revised 23 March 2022; Accepted 1 April 2022; Published 30 June 2022

Academic Editor: Chia-Huei Wu

The degree and level of early childhood education are an important guarantee to show whether a country’s overall planning mechanism is reasonable, whether the education system is perfect, whether the people’s livelihood is harmonious, and whether the national quality can develop sustainably. The American Sports and PE Association, which is at the leading level of PE in the world, has formed an advanced high-quality children’s PE system on the basis of its long-term professional evolution. The association provides scientific and appropriate PE activities for children, which is not only a necessary premise to ensure the normal physical development and physical and mental health development of children but also an important aspect of kindergarten teaching. Therefore, a lot of data has been accumulated in the development of children’s PE, and it is of practical significance for promoting the construction of children’s PE to dig out the knowledge hidden behind these data and provide help for people’s decision-making. As early as 1988, at the 10th World Congress of the World Federation of Future Studies, experts from all over the world discussed the results of “children’s future” and generally believed that “children are basic human resources,” and creating a world that is conducive to their development should be recognized as a regional and global priority measure. In this paper, the decision tree algorithm in DM is studied, and combining with the characteristics of data in educational management information, the mining of classification rules of educational information and ID3 algorithm are analyzed.

1. Introduction

Kindergarten sports activities are not only an organic part of children’s comprehensive and harmonious development education, but also one of the important contents of kindergarten health education [1]. A large amount of data has been accumulated in the teaching and management of kindergartens for many years. At present, these data have not been effectively used, but just a “treasure” to be developed [2]. Kindergarten sports can not only promote the development of children’s body and cognition but also promote the formation of children’s good personality and improve children’s social adaptability [3]. The selection of physical education (PE) activities and items can basically meet the needs of children’s physical and mental development. However, there is a certain randomness and shallow dominance, which is not conducive to children’s exercise [4]. Therefore, data mining (DM) can extract valuable information hidden behind data from a large number of data, which has been applied in more and more fields, achieved good results, and provided great help for people to make correct decisions. Therefore, cultivating excellent talents with both political integrity and ability is the inevitable requirement of modernization [5]. In addition, with the rapid development of social economy, people’s material life has been greatly improved, but it also makes our life rhythm faster and faster and the competitive environment more intense [6]. After the completion of the PE course examination, teachers should convert the nonpercentile scores recorded in minutes, seconds, meters, etc. into percentile scores according to the national standards for students’ body health, and at the same time, each teacher should log his
converted scores into the educational administration management software of the academic affairs office [7]. A large amount of data and tedious data processing and conversion work, as well as data processing and analysis problems, increase the workload of PE teachers and restrict their work efficiency [8]. Therefore, in the face of the in-depth development of education reform, it is urgent to formulate corresponding countermeasures.

Therefore, decision tree learning algorithm plays a very important role in DM technology [9]. As an important technical means in DM, decision tree algorithm integrates machine learning, data statistics, intelligent database, neural network, and other technologies [10]. Through the algorithm, mining the potential data of relevant influencing factors and the ideal modern teaching evaluation database model can be provided for education management and decision-making [11]. The decision tree can process multidimensional data, and the discovered rules are easy to understand [12]. In the past, data information was only used for conventional data recording and statistical classification, but now DM technology is not only used for daily student information statistics, teaching task recording, and data storage but also used to manage data, process and analyze this massive information, find out the relationship between data, and provide strong teaching strategies for teaching management [13]. However, fundamentally speaking, the necessity of kindergarten sports activities is determined by children’s physical and mental development needs [14]. Statistics and analysis of a large amount of data accumulated in the teaching process require teachers to use a certain amount of time and energy to sum up the problems existing in teaching and practice this semester in time.

Therefore, we evaluate the classroom teaching in PE class and find out the factors that affect the quality of PE classroom teaching, so as to improve the teaching in PE class and improve the level of PE classroom teaching. Modern preschool health science research shows that preschool children’s bodies are still in the first high-speed period of physiological development. In addition to the general increase in height and weight, various tissues and organs have also made great progress in their functions [15]. At the same time, the social requirements for talents have also changed, requiring people not only to have superb intelligence, but also to have healthy and coordinated physical quality, psychological quality, and good social adaptability. In addition, the physique of Chinese teenagers is showing a downward trend. This fact has attracted the great attention of our government and relevant leaders. Therefore, we should combine the teaching data information of the school, introduce advanced network information technology, and apply it to school education management, teaching evaluation, paperless examination, teaching records, students’ inductive learning, and teachers’ teaching technology analysis. This huge amount of data, through DM technology, analyzes the kindergarten teaching management and teaching activities, so as to improve the school teaching management and optimize the teaching process, so as to improve the quality of education and teaching.

2. Related Work

Literature [16] discusses the principal component analysis method used in the evaluation of students’ comprehensive scores. Literature [17] discusses the concept of sports teaching methods and sports dance teaching and puts forward the establishment of public sports dance teaching method system in kindergartens. Literature [18] uses the classification mining method based on decision tree to analyze the data in student achievement database and applies the algorithm to student achievement analysis to build the professional ability decision tree model, so that teachers and school education decision makers can have an insight into the problems existing in teaching, so as to use the information provided by achievement to optimize the planning and decision-making of education and teaching. Literature [19] puts forward the theory of “periodic cycle” by arranging PE teaching materials horizontally, dividing PE teaching items and contents vertically according to the characteristics of students’ physical and mental development, and establishing a new system of PE practice teaching contents. Literature [20, 21] introduce psychological research methods into public PE teaching research in kindergartens. Literature [22] puts forward a multistrategy design idea, which combines DM technology with statistical analysis, adopts classification mining method based on decision tree, analyzes the data in student achievement database, and generates student achievement decision tree, which can intuitively show the position of a certain achievement in different grades of calculation methods and provide evaluation information for teaching departments. Literature [23,24] introduce the classic Apriori algorithm of association rules and the famous decision tree algorithm ID3 and uses the association rule algorithm to mine the impact of the excellence of a course on other courses. Literature [25] uses children’s performance as characteristic data, reduces the dimension of characteristic data through principal component analysis, and realizes classification direction prediction through Bayesian nearest neighbor algorithm.

3. DM and Decision Tree Generation Method

3.1. Methods of DM. DM technology involves the fields of machine learning, pattern recognition, statistics, intelligent database, knowledge acquisition, expert system, data visualization, and high-performance computing. The DM process is shown in Figure 1.

Kindergarten PE teaching activity is a planned, purposeful, and organized activity organization form, which takes children’s physical exercise as the main content, stimulates children’s interest in participating in activities, develops children’s basic activity ability, improves children’s physical quality and increases children’s physique, and pays attention to promoting children’s cognitive, personality, and social development. Kindergarten PE teaching activities mainly include PE teaching objectives, teaching contents, teaching organization forms, teaching methods, teaching evaluation, teaching effect, and health supervision [26]. DM is the most critical step in the process. In practical
application, DM and these two terms are often indistinguishable [27]. To some extent, a DM model can be regarded as a relational table composed of many columns of different data types, some of which are input columns and others are prediction columns [28]. Different and targeted classroom contents are designed for different children’s individual development levels. Age is related to movement development, but it is not an absolute standard [29]. Classification is mainly carried out according to the attributes and definitions of objects, so as to establish class groups [30]. It is not only the basis for the selection, organization, and implementation of teaching activities, but also the reference standard for the evaluation of kindergarten PE. However, DM model is different from relational table because it does not store row data. On the contrary, it stores the patterns found by DM algorithm in relational table. It is necessary to make effective observation according to the past value of the relevant attribute of an object, so as to evaluate the future value of the attribute.

The validity of information requires that the mined data should be carefully checked before mining. Only by ensuring the validity of information or data can the validity of the mined information be guaranteed. It can be considered that effectiveness is the core component of teaching, which plays the role of guidance, regulation, and evaluation in teaching activities. In the traditional data analysis methods, it is necessary to analyze related problems manually, while DM can automatically analyze and find predictive information based on large databases. The most important thing is that the information obtained is analytic; that is, the information or knowledge is effective, practical, and realizable for the business or research field under discussion. Because the goal of teaching activities is the goal associated with specific teaching activities and the lowest goal in the goal system, it should be specific and operable.

### 3.2. Generation Process of Decision Tree

By analyzing and testing the training set data, the data classification model of the studied problem is established. In the process of data classification by decision tree, firstly, it is necessary to analyze and test the obtained data training set, establish the data classification model of the problem to be solved according to the obtained data categories, and then classify the unknown data according to the established decision tree. At first, all data are at the root node, and then recursively the data are divided into pieces. The second is tree pruning, that is, removing some data that may be noise or abnormal. Figure 2 briefly describes the process of decision tree generation.

Children’s learning and development are promoted in a holistic way, sports is integrated into other fields, and children’s cognition, emotion, and sociality are developed through sports experience. Certain data is classified through differences and similarities. The main purpose is to classify the data belonging to the same category into one category, so as to reduce the similarity of different categories of data as much as possible. Decision tree technology is a kind of “greedy” search, which uses greedy algorithm. It tries to add each attribute value to the left subtree in turn. If you can find greater information gain, then add this attribute value to the left subtree; otherwise, return it to the right subtree. In the process of building a tree, because of the noise in the dataset, many branches of the decision tree reflect the anomalies in the training set, and these branches often have the problem of overfitting. The solution to this problem is to prune the built decision tree. According to the number of samples obtained, the attributes in the set are tested or by one, and the data training sets are classified by attribute categories to obtain several word training sets, in which each sub-training set is used as a nonleaf node, and recursively and circularly executed, until the set conditions are met, the leaf nodes of the tree are formed, and the calculation is terminated. The basic idea of the application of decision tree method in kindergarten public PE practice teaching is to build a decision tree composed of the attribute values of each index through the decision tree method, sort the given index set, and find out the most important factors affecting children’s PE practice teaching, so as to provide reference for children’s PE practice teaching.

Learning algorithm based on decision tree has many advantages, such as fast establishment speed, high precision, understandable rules that can be generated, relatively small amount of calculation, continuous value and discrete value attributes that can be handled, and the ability of clearly displaying which attributes are more important. Besides, users do not need to have a lot of background knowledge in the learning process; so long as the training examples can be expressed by attribute-conclusion formula, we can use this algorithm to learn. Therefore, the formed branches reflect the abnormal situation in the training set, which easily leads to the problem of overfitting.
Therefore, it is necessary to prune the constructed decision tree to remove the abnormal branches and ensure the accuracy of the classification results.

4. Mining Educational Information Classification Rules and ID3 Algorithm Analysis

4.1. Classification Rule Mining. DM is to use relevant algorithms and technologies to make predictions based on existing data, so as to make proactive and knowledge-based decisions. Data preprocessing is mainly to form data training set tuples by means of data cleaning, integration, selection, transformation, and attribute concept stratification. The information entropy of the sample classification is

\[ I(x_1, x_2, \ldots, x_n) = - \sum_{i=1}^{n} p_i \log_2 p_i. \]  

(1)

The 300 samples in the training set tuple are randomly sampled by setting random seeds, and the effective information is divided into training samples and test samples by randomly selecting the proportion of 95% and 5%. The information gain and information gain rate of each attribute are shown in Figure 3.

The algorithm is used to mine the dataset obtained through the data preprocessing process. Data integration is to combine the basic information of students and the data items in the questionnaire into a dataset for analysis. For the classification system, assuming that \( x \) is the data sample set, \( x \) is the number of \( X \), \( C \) is the attribute category variable, \( n \) is the total number of categories, and \( X_i \) is the number of samples contained in \( C_i \), then the probability of any sample belonging to \( C_i \) is

\[ p_i = \frac{x_i}{x}. \]  

(2)

In the traditional sense, data summarization is to sum, average, and perform variance and other statistics of related fields in the database and to present the calculated or statistical results in a vivid and intuitive way such as histogram and pie chart. At the level of \( a = 0.05 \), three statistically significant variables are screened out from 32 analysis variables to build the decision tree model of factors affecting kindergarten public PE practice teaching. Figure 4 shows the comparison of training sets of analysis variables of different samples.

DM is mainly concerned with discussing data summarization from the perspective of data generalization. By applying DM technology, an effective evaluation system of PE teaching is established, the teaching evaluation is analyzed, and the shortcomings in PE teaching are found, so as to change the teaching plan and improve the teaching quality. By using the query and retrieval functions of database relational system, combined with statistical analysis processing, the statistical analysis data for decision-making reference can be obtained. At the same time, the DM model must describe the mining algorithm related to it and any existing parameter list. Taking database parameters as classification attributes, set correlation threshold and classification threshold. The mining results are classification rules in the form of if-then. The mining results are shown in Figure 5.

Evaluation is an important element of DM model, which needs to be carried out continuously, and its standard is consistent with the state and national PE standards. We should monitor and strengthen children’s sports learning through regular evaluation and combine formative evaluation and summative evaluation. Each evaluation unit should be conducive to the realization of high-quality PE. Users can choose different datasets to mine student achievement information, student basic information, and student comprehensive information. Data is stored in a single data unit in the database, and a single data cannot express meaning.
Clustering is to summarize and merge these scattered data to form a meaningful subset. When building the model, enter the column to select automatic number, and enter the column to select Gender, Age, PE Course, Physical Function Score, Body Shape Score, Physical Quality Evaluation Items, and Physical Quality Evaluation Grade as “Achievement.” Students’ score mining data has been divided by semester. After the user selects the table and classification attributes to be mined in the interface, the correlation analysis in the improved algorithm will remove the attributes less than the correlation threshold according to the set correlation threshold, and the user can also select the unclassified fields in the table as unclassified attributes for mining in the interface. Let the attribute $y$ have $m$ different values; then, the information entropy of $Y$ is

$$E(Y) = \sum_{j=1}^{m} \frac{x_{ij} + x_{ij} + \ldots + x_{nj}}{x}$$

(3)

This is mainly because in the database, the storage of data is the most primitive and cannot express specific information. The whole DM model is like a fallen tree, with roots on the left and branches on the right. This hierarchy created by if-then rule can intuitively describe nodes. EIDT-DM algorithm calculates the correlation between nonclassified attributes and classified attributes according to the set correlation closed value and removes the attributes less than the correlation threshold, so as to improve the operation efficiency. In order to have a high-level understanding and grasp of these data, we need to carry out the abstract operation of data at different levels, so as to meet the needs of different users for browsing and processing data or images.

4.2. Analysis of Decision Tree ID3 Algorithm. ID3 algorithm constructs a decision tree according to a set of given data rows or data objects whose category attributes are known and then uses the decision tree to classify data of unknown categories. By selecting the attribute with the largest information gain as the test attribute of sample division, branches are established by different values of this attribute, and then the next level nodes and branches of the decision tree are recursively established by using this method for instance subsets of each branch until the instances in a subset belong to the same class. In the process of mining association rules, the minimum support and confidence are determined to be 0.2 and 0.5, and approximate accurate rules are generated. The support and confidence of different factors approximate the association rules as shown in Figures 6 and 7.

Instead, create a leaf node to store the subset and the class distribution of the subset samples. Then, in the process of tree pruning, for the leaf node created as a substitute, find out the category with the largest sample number of classification attributes in the subset as the value of the classification attributes of the leaf node. The attribute with the largest information gain is considered as the attribute with the largest resolution in the current dataset. The algorithm calculates the information gain of each attribute, and the attribute with the highest information gain is selected as the test attribute of a given $S$. The information gain brought by $Y$ attribute is

$$\text{Gain} = I(x_1, x_2, ..., x_n) - E(Y).$$

(4)

In the process of decision tree creation, due to the existence of noise in the training set, there are abnormal branches in the training data. In order to generate an easy-to-understand decision tree, it is necessary to prune the decision tree wrapped with noise branches to solve the problem of overadaptation of some branches. The information entropy of the residual splitting attribute is calculated and the decision tree is established. In the process of establishing the decision tree, the function of calculating the pass rate is added. The comparison results of information entropy before and after adding are shown in Figure 8.

Click different nodes to analyze various relationships. The first pruning method is that, in the process of establishing the decision tree, for the abnormal branches, the establishment of child nodes is stopped directly, so that abnormal branches are not formed. This attribute is used to
construct a node of the decision tree, and all values of the attributes it represents are tested at this node to obtain each branch of the node, which divides the original dataset into several subdatasets. Therefore, the final formula for calculating information entropy is

\[
H = -\sum_{i=1}^{n} p_i \log_2 (p_i)
\]

If the data rows contained in a node are of the same category, the node is the leaf node of the decision tree and is marked as the corresponding category. The postpruning method is to remove some abnormal branches after establishing the decision tree and use leaf nodes to reestablish a new decision tree. This decision tree construction process is repeated until all nodes do not need to continue branching. The expected information required for a given node classification is given by the following formula:

\[
H(s_1, s_2, ..., s_m) = -(p_1 \cdot \log_2 (p_m))
\]

In the DT-EIDM system, there are three Boolean quantization methods, namely, equal quantization within a fixed range, equal quantization with max and min of the mined data instead of the upper and lower limits of the fixed range, and equal quantization of the number of people. Different quantization algorithms will affect the DM results due to different methods. To construct the decision tree as small as possible, the key is to choose the appropriate logical judgment or attribute. The size of threshold will affect the proportion of “0” and “1” in the quantized data table. The change of threshold has a great influence on the DM results. Different thresholds can be used for mining many times to get different classification rule sets, and the appropriate rule set can be selected by comparison. And create subchains of the node, and each subchain represents a unique value of the selected attribute. The postpruning method based on the minimum error principle is adopted; that is, after the decision tree is completely generated, the redundant branches are cut off, and a new decision tree is obtained by using leaf nodes. There are two options for stopping segmentation: the data on a node belongs to the same category; no attributes can be used to segment data.

5. Conclusion

Kindergarten PE activities are an integral part of kindergarten PE, and the basic organizational form to achieve the overall goal of children’s PE. Scientific and reasonable PE activities for children are necessary teaching work for kindergartens. However, in the real educational practice, it is far from enough for kindergartens to know and attach importance to PE activities. Therefore, obtaining unknown, effective, and practical information, rules, and knowledge by DM method without clear assumptions is the essential feature of DM technology which is different from traditional statistical data analysis. There are rules hidden in data in the development of preschool children’s PE. These rules can be mined by different methods, and rules can be obtained according to different conditions. DM technology and decision tree algorithm not only improve the storage of traditional information but also find the potential laws and values between data, provide a scientific basis for education and teaching management, assisting students’ learning and teaching methods, and provide a correct decision-making basis for the development of teaching. How to make
computer and network technology serve the daily PE teaching of the school is an inevitable requirement in today’s information age. Kindergartens must give appropriate PE activities according to the characteristics of children’s physical and mental development and the change law of human physiological function, so as to promote the healthy development of children’s physical and mental health. In this paper, DM and decision tree ID3 algorithm are used in the development of preschool children’s PE. Through the mining and analysis of survey data, we try to find the factors affecting the quality of children’s PE classroom teaching, so as to provide scientific suggestions for the reform of preschool children’s PE classroom teaching in the future.

Data Availability
The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest
The authors declare no conflicts of interest.

References


