Advances in Educational Technology and Psychology (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/aetp.2023.071816
ISSN 2371-9400 Vol. 7 Num. 18

Reform of Early Childhood Physical Education Development Based on Big Data Technology

Haoduo Yang[1], Jingxian Li[2]

[1]*International Education Management, Woosong University, Daejeon Metropolitan, South Korea*
[2]*Northwest Institute of Nuclear Technology Kindergarten, Northwest Institute of Nuclear Technology, Xi'an, Shaanxi, China*

Keywords: Early Childhood Physical Education, Big Data Technology, Decision Tree, Data Mining Technology

Abstract: Early childhood is a critical period for a person's quality development, and their own learning thinking, ability, and quality are greatly affected. Physical education is a course that enhances students' body quality and physical fitness level. In the current environment of deepening reform, it is necessary to optimize teaching methods, content, and system to achieve the goal of strengthening the quality and physical fitness level of young children, and to some extent, promote students to form good physical exercise habits and physical fitness at the current stage. In order to carry out a new reform in the development of early childhood education, this article attempted to collect data on various indicators of physical training for young children (physical fitness testing, motor skills, and physical fitness indicators) through big data technology for data analysis, in order to implement personalized physical training for young children. In the experiment, when the sample data was between 1 and 10×104, the mean square error of data analysis in data mining technology was less than or equal to 0.12076, which was lower than other algorithms.

1. Introduction

Children are the future and hope of a country. To promote the comprehensive development of children's physical fitness, it is necessary to strengthen the scientific and moderate management of early childhood physical education. Therefore, early childhood education should keep up with the pace of the times, strengthen advanced physical education teaching methods, and improve the efficiency of early childhood physical training. Moreover, personalized education should be provided to young children based on their physical condition and the teaching objectives of physical exercise, combining traditional education with high-tech, promoting the reform of school physical education, and laying a solid foundation for cultivating outstanding talents in the new era. Therefore, it is of great significance to effectively improve students' physical fitness, enhance their physical fitness, and promote their healthy growth. It is essential to reform kindergarten physical education, strengthen the kindergarten physical education faculty, cultivate students' interest in physical education, and promote the concepts of lifelong physical education and joyful physical education in the context of the new era.

The combination of early childhood physical education and modern technology has long been studied by relevant experts. In order to improve the effectiveness of intelligent processing of physical education resources, Li C combined genetic algorithm with unsupervised learning to deeply process physical education teaching images, so as to preserve the original adjacent structural relationships as much as possible. To minimize the local information loss of the sample as much as possible, he established a graph model that integrates compressed data and input. In response to the problems of hash algorithm, he proposed an intelligent physical education teaching resource processing scheme based on the combination of genetic algorithm and unsupervised learning, and established corresponding functional modules. Finally, he tested the proposed algorithm through experiments. Through experimental comparison, it has been proven that the algorithm is an effective method [1]. To explore the application of artificial intelligence (AI) technology in physical education teaching and evaluate its practical significance, Zhang Y proposed a continuous monitoring system for mobile people based on wireless sensors. The purpose of coaches is to utilize AI technology and their athletic expertise to enable players to achieve optimal performance in different sports events. He used Bayesian deep classifiers and combined them with adaptive optimization methods of deep neural networks to solve them. He conducted rigorous testing and analysis on the designed QoS (Quality of Service) indicators, such as communication conflicts, data loss rate, etc. Research found that as the number of packets increased, the packet loss rate increased while the collision rate remained basically unchanged. In addition, the application results of the system in universities showed that 14.4% of students were very satisfied with the use of the new course; 24.5% were satisfied with the new course; 30.1% were satisfied with the training method, and 17.9% were dissatisfied [2]. In response to the problems of poor noise removal, low recognition efficiency, and low recognition accuracy in the current pose recognition technology for table tennis players, Yin J used genetic algorithm for pose recognition of table tennis players and proposed a multi-objective keyframe extraction method based on genetic algorithm. He used the method of Kalman filtering to eliminate the noise between keyframes, effectively suppressing the impact of noise on recognition, thereby reducing the system's identification speed. Experiments showed that this algorithm had good noise reduction effect, high recognition efficiency, and high recognition accuracy, with a maximum recognition rate of 98.7% [3]. The above algorithm is only applicable to a certain motion and cannot be promoted in different types of motion.

The utilization of big data can both help advance early education and effectively address the shortcomings of traditional education. Teachers should continuously improve their professional level during the teaching process, give full play to the advantages of big data, and reform the traditional teaching methods in order to improve the teaching effect. This article explored the sports indicators of early childhood physical training based on data mining technology, and constructed a prediction of student sports indicators through data mining technology. The experiment has proven that the model has high accuracy and strong analytical ability.

2. Data Mining Technology and Early Childhood Physical Education

2.1. Data Mining Technology

DM (data mining) is an interdisciplinary field that integrates multiple disciplines such as statistics, AI, and fuzzy analysis. It is a system that obtains knowledge through data analysis of specific problems. Based on specific issues, data mining techniques can be divided into the following:

(1) Statistical analysis method

It is a method of using mathematical methods such as regression analysis and correlation analysis to correlate the internal connections between things and uncover hidden internal connections [4].

(2) Genetic algorithm

It is a computer algorithm that simulates the process of biological evolution and is based on the mechanism of biological evolution [5-6]. This algorithm uses simulation evolution to generate a series of solution sets, and then, based on a certain screening criterion, removes the solution sets with low correlation, and repeats until the best solution is obtained. Especially in many fields such as industrial engineering, transportation, and economics, it has important applications.

(3) Decision tree method

It is an analytical prediction model that represents existing data in a tree like form and constructs a decision tree based prediction method. This method helps decision-makers understand and analyze the problem [7].

(4) Visualization technology

This method makes the connections between existing data more intuitive and vivid through graphical processing, which is of great significance for data mining.

(5) Bayes network

Bayes networks focus on uncertain information problems and connect them in the form of graphs to achieve solutions under incomplete and uncertain conditions [8].

(6) Concept tree method

The concept tree method belongs to an abstract data processing method that can classify and organize data, and is a data preprocessing method.

2.2. Integrating Mining Technology with Early Childhood Physical Education

Combining mining technology with physical training can change the current chaotic state of data mining technology. With systematic learning of relevant knowledge, many sports analysis tools introduce the relevant data used on the basis of comprehensive application principles, such as maturity calculation and skill testing of mining technology in warehouses. The integration of multiple mining technologies greatly improves the level of sports training and greatly saves students' physical strength [9-10]. The preferred condition for sports training is the venue, which should be ensured to be spacious. Some sports training research should be conducted in a relatively relaxed environment to help students improve their sports level [11-12]. Of course, the current mining technology supports more thinking on the progress of mining technology at the decision-making level, and is committed to developing a data technology that can transform students' basic physical training into more advanced data technology to meet the needs of various projects, thereby achieving the sports goals that China has not yet achieved [13-14].

3. Model of Sports Training Indicators

3.1. Data Preprocessing Based on Rough Sets

This article proposes a new method that removes redundant information from data by evaluating the approximate expression of knowledge, thereby obtaining more accurate and reliable information. Traditional rough set methods can only evaluate and process existing classification data, while for deeper processing, discretization methods must be used, which leads to data loss. The attribute reduction algorithm for motion training indicators based on the nearest neighbor rough set method uses environmental factors as output parameters for evaluating the quality of motion training [15-16]. The article proposes a classification level for sports training based on triples, which is a rough set algorithm established based on existing knowledge bases.

$$D_{nt} =< U, A, \quad D > \tag{1}$$

Among them: $U = \{x_1, x_2, \cdots, x_n\}$ is the dataset; D is the classification level of sports training; A is the attribute set [17].

On this basis, the experiment treats the simplified initial value set as an empty set and calculates the significance parameters of the other attributes in this motion training index. If it is not 0, it would be included in the reduction set first. This process can be summarized as:

Step 1: $\forall a \in A$ calculates the proximity matrix of each attribute.

Step 2: The attribute rough set is initialized. R_{ED} is empty, and $\varphi \to R_{ED}$ is set.

Step 3: All attributes in attribute A that are not included in R_{ED} are traversed, and the importance of each attribute parameter is calculated, that is, $\forall a \in A - R_{ED}$.

$$S_{ig}(a_i, R_{ED}, D) = \gamma_{RED \cup a}(D) - \gamma_{RED}(D) \tag{2}$$

Step 4: The attribute with the greatest importance is selected, namely:

$$S_{ig}(a_i, R_{ED}, D) = S_{ig}(a_i, R_{ED}, D)_{max} \tag{3}$$

Step 5: If $a_i > 0$, it is added to R_{ED} and $R_{ED} \cup_{a_i} \to R_{ED}$; otherwise, it would be redirected to step 3 until the loop termination condition is met.

3.2. Data Mining Processing

The processing and analysis of data can be divided into three stages: data selection, data processing, and data transformation. Data selection involves extracting data from the database to form the target data. Pre processing is the process of processing the extracted data to meet the required conditions. Data conversion is the process of dimensionality reduction on data. The initial characteristic function is as follows:

$$m_i I_i N_i = \frac{v_i \theta_i^2}{(E - E_1 - E_2) l_2} \tag{4}$$

Among them, m is the data feature variable; I represents data variability; N is the target data; v is the calculation quantity; E is data mining, and E_1 is initial condition mining. E_2 refers to the i-level data of work state mining.

3.3. Decision Tree Based Data Mining Model,

The decision tree model has been widely used in the field of data mining due to its simple and easy to understand characteristics. The decision tree represents the final classification result as a tree.

$$E_0 = \sum_{i=1}^{n} (a_i - \bar{a}) \left(f_i - \bar{f}\right) \cdot \frac{1}{\sum_{i=1}^{n} (e_i - \bar{e})} \tag{5}$$

In the formula: E_o is the theoretical expression function; N is the calculated length; A is the range of element records; F is a discrete indicator; E is the range of indicators.

Decision tree is a method of partitioning data using a series of rules, which can infer a classification rule based on a set of irregular elements. Usually, this algorithm uses a top-down recursive method to compare the attribute values of each node, and then branches each node down according to different attributes, with leaf nodes as the categories to be segmented. In this way, the path from the root node to the leaf node corresponds to a corresponding classification criterion. The decision tree structure includes three aspects: decision nodes, branch nodes, and leaf nodes. Each node corresponds to an unclassified attribute, and each node represents a classification. The center node of a tree is usually a rectangle, while the leaf node is an ellipse. A significant characteristic of decision trees is their numerous branches and wide coverage. However, when the eigenvalues are continuous, their performance is not ideal. A significant drawback of decision tree methods is that they only focus on selecting features as the best choice for partitioning data. However, existing research has shown that selecting the current optimal feature solely based on "maximizing information entropy increase" cannot fundamentally solve this problem.

To solve the above problem, an improved decision tree algorithm is proposed here. The algorithm can be divided into two parts: one is learning, and the other is testing. During the learning process, the parameters are trained through top-down recursive methods. On this basis, the established model and parameters are introduced into the actual system for verification and optimization. This algorithm is divided into two steps: the first step is to establish a spanning tree; the second step is to trim the tree and remove data containing noise and anomalies. The following formula is used to eliminate noise and abnormal data.

$$L_n(x) = \coprod_{\substack{i=0 \\ i\neq j}}^{n} \frac{x-x_i}{x_j-x_i} \tag{6}$$

In the formula: L_n represents the noise removal function; x_i is the i-th result of the decision tree, and similarly, x_j is the jth result of the decision tree; n represents the range of search criteria.

4. Experiment on the Model of Sports Training Indicators

The correctness of this method was verified through experiments. The developed data collection device for training behavior evaluation mode and the device related to traditional training behavior evaluation mode were respectively installed at the same location on the training ground. The basic information of the subjects was input into two modes. This exam lasted for 4 days. During the exam, the team members continued their daily training [18-19]. The changes in indicators such as heart rate, maximum oxygen uptake, blood lactate threshold, body composition, exercise technique evaluation, training load, and error rate of the tested athletes were compared [20]. The other algorithms for comparative analysis are BPNN (back propagation neural network) algorithm and Yggdrasil.

4.1. Precision Evaluation

In order to test the changes in the accuracy of BPNN during the entire data increase process, the experiment used the dataset HEPMASS to control the number of instances and recorded the changes in classification accuracy using three methods as the number of instances increased.

From Figure 1, it can be seen that as the number of samples increased, the accuracy of Yggdrasil was always lower than that of BPNN, while the accuracy of data mining techniques remained above 0.87. The main reason for this difference is that during the experimental process, each experiment randomly selects the same number of samples from the HEPMASS database to test the three algorithms, resulting in differences in the effectiveness of the three methods.

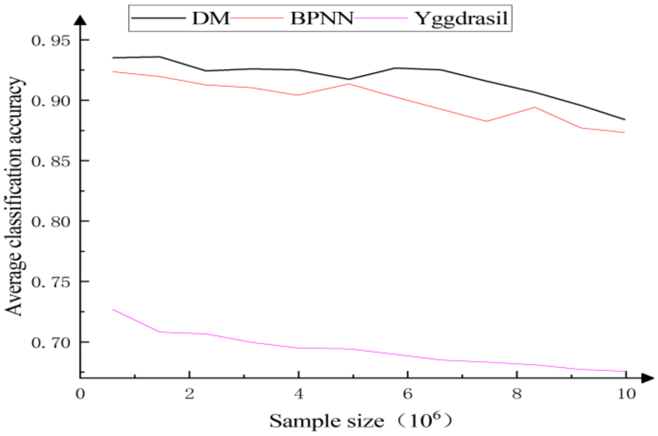

Figure 1: Average classification accuracy of different methods (randomly selected training and testing datasets)

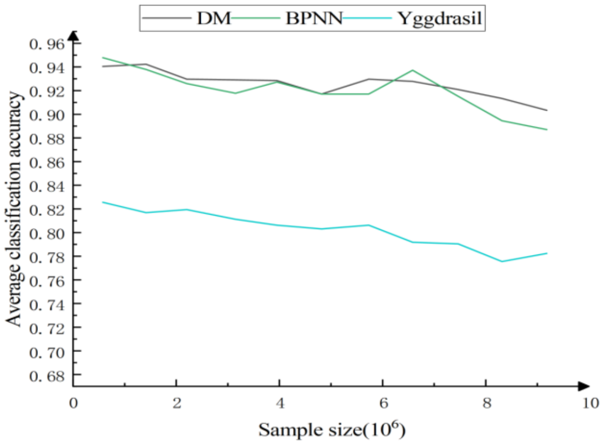

Figure 2: Average classification accuracy of different methods (same training and testing datasets)

Yggdrasil is divided based on information gain, and the attributes selected by these two partition rules are different, so the final results are also different. From Figure 2, it can be seen that after removing random factors, the accuracy of Yggdrasil is still not as good as that of BPNN or data mining techniques. Although the distance between the two has been shortened, there is still a certain distance, which is mainly due to the division criteria of the two datasets.

4.2. Training Time Evaluation

On this basis, this article studied the efficiency of learning big data with massive attributes based on three different methods such as BPNN, DM, and Yggdrasil, limiting the sample size to 10×10^6, as shown in Figure 3.

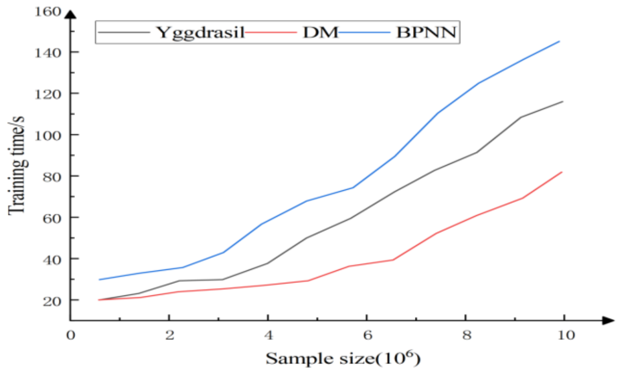

Figure 3: Changes in decision tree training time as the number of instances increases

The experimental results in Figure 3 showed that the training time of DM was shorter than that of BPNN and Yggdrasil. However, when the actual sample size reached 10×10^6, Yggdrasil's training time increased from 20 seconds to 116 seconds; BPNN's training time increased from 30 seconds to 145 seconds, and DM's improvement only increased from 20.1 seconds to 81.9 seconds. DM had higher overall learning efficiency than algorithms such as Yggdrasil and BPNN. Due to its higher overall learning efficiency and slower growth rate compared to the other two methods, DM has certain advantages in multidimensional data.

4.3. Mean Square Error

The experiment investigated the efficiency of learning big data with massive attributes using three different methods, including BPNN, DM, and Yggdrasil, as shown in Table 1.

Table 1: Mean square error of different algorithms

Sample size(10^4)	DM	BPNN	Yggdrasil
1	0.12076	0.21506	0.45759
2	0.11495	0.38071	0.45953
3	0.11302	0.32678	0.36716
4	0.10154	0.28434	0.32098
5	0.09573	0.2517	0.26124
6	0.08799	0.18616	0.22474
7	0.08232	0.14966	0.18049
8	0.07651	0.1227	0.13417
9	0.06296	0.08993	0.10915
10	0.04374	0.06296	0.07071

From Table 1, it can be seen that as the sample size increased, the mean square error of different algorithms continued to decrease. However, the error of the motion training index analysis model based on data mining technology was always lower than the other two algorithms. This indicates that the analysis accuracy of data mining technology is higher than that of other algorithms.

5. Conclusions

The era of "big data" has arrived, and its impact on production and life is becoming increasingly significant. It has penetrated into people's daily lives, and education is no exception. Due to the fact that primary school students are still very young and their physical and mental development is not yet mature enough, it is even more necessary to utilize big data for their development and improvement. Big data technology can be fully utilized in various aspects of early childhood education. Through big data analysis, three problems can be solved: sharing educational resources, understanding the current situation of children's learning and development, and understanding children's physical and mental health. Based on the above objectives, this article explored the development and reform of early childhood physical education based on big data technology, attempting to analyze the data of early childhood physical training through data mining technology, to achieve targeted training and improve the ability of early childhood sports competition.

References

[1] Li C, Liu B, Kim K. Intelligent unsupervised learning method of physical education image resources based on genetic algorithm. Neural Computing and Applications, 2023, 35(6): 4225-4242.

[2] Zhang Y, Duan W, Villanueva L E, et al. Transforming sports training through the integration of internet technology and artificial intelligence. Soft Computing, 2023, 27(20): 15409-15423.

[3] Yin J. The method of table tennis players' posture recognition based on a genetic algorithm. International Journal of Biometrics, 2021, 13(2-3): 243-257.

[4] Yu Z, Guindani M, Grieco S F, et al. Beyond t test and ANOVA: applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. Neuron, 2022, 110(1): 21-35.

[5] Sohail A. Genetic algorithms in the fields of artificial intelligence and data sciences. Annals of Data Science, 2023, 10(4): 1007-1018.

[6] Garud K S, Jayaraj S, Lee M Y. A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models. International Journal of Energy Research, 2021, 45(1): 6-35.

[7] Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2021, 2(01): 20-28.

[8] Kaikkonen L, Parviainen T, Rahikainen M, et al. Bayesian networks in environmental risk assessment: A review. Integrated environmental assessment and management, 2021, 17(1): 62-78.

[9] Akramovich A A, Nazirjonovich X Z. Technologies for developing of future physical education teachers through media education tools. Academicia: an international multidisciplinary research journal, 2021, 11(2): 885-890.

[10] Mukhidinov A, Tillaev S. Ways to Apply Physical Education in the Family. International Journal of Research in Commerce, It, Engineering and Social Sciences. 2022, 16(3): 78-82.

[11] Vilchez J A, Kruse J, Puffer M, et al. Teachers and school health leaders' perspectives on distance learning physical education during the COVID-19 pandemic. Journal of School Health, 2021, 91(7): 541-549.

[12] Neville R D, Makopoulou K. Effect of a six-week dance-based physical education intervention on primary school children’s creativity: A pilot study. European Physical Education Review, 2021, 27(1): 203-220.

[13] Pesce C, Lakes K D, Stodden D F, et al. Fostering self-control development with a designed intervention in physical education: A two-year class-randomized trial. Child Development, 2021, 92(3): 937-958.

[14] Welch R, Taylor N, Gard M. Environmental attunement in health, sport and physical education. Sport, Education and Society, 2021, 26(4): 339-348.

[15] Cruickshank V, Pill S, Mainsbridge C. 'Just do some physical activity': Exploring experiences of teaching physical education online during Covid-19. Issues in Educational Research, 2021, 31(1): 76-93.

[16] Dupri D, Nazirun N, Candra O. Creative thinking learning of physical education: Can be enhanced using discovery learning model?. Journal Sport Area, 2021, 6(1): 29-36.

[17] Sevimli-Celik S. Moving between theory and practice: preparing early childhood pre-service teachers for teaching physical education. Journal of Early Childhood Teacher Education, 2021, 42(3): 281-298.

[18] Estevan I, Bardid F, Utesch T, et al. Examining early adolescents’ motivation for physical education: Associations with actual and perceived motor competence. Physical Education and Sport Pedagogy, 2021, 26(4): 359-374.

[19] Dyson B, Howley D, Wright P M. A scoping review critically examining research connecting social and emotional learning with three model-based practices in physical education: Have we been doing this all along? European Physical Education Review, 2021, 27(1): 76-95.

[20] Baena-Morales S, González-Víllora S. Physical education for sustainable development goals: Reflections and comments for contribution in the educational framework. Sport, Education and Society, 2023, 28(6): 697-713.