Application of Data Mining Technology in Sports Training Index Analysis

Meng Xiangwu, Wang Shun, Liu Chao

School of Physical Education, Huaibei Normaol University, Huaibei 235000, China

Keywords: Data Mining Technology; Index Parameters; Training Analysis Mechanism; Sports Training; Index Analysis; Analysis Model

Abstract: The conventional analysis method can provide a general analysis of sports training index, but its ability is relatively low when analyzing niche data. To solve this problem, this paper proposes data mining technology. First, the indicator parameter classification is determined, then the data mining technology is imported, the sports training analysis mechanism is established through this technology, and the construction of the index analysis model is completed. The model is used to analyze the process of niche data mining, and effective data of training indicators are obtained. Through the coverage test, accuracy test and immunity test, the variable parameters of the comprehensive analysis capability are determined. Further calculation of this parameter shows that the comprehensive ability of the data mining application analysis method is improved by 37.14% compared with the conventional method, which is suitable for analysis of niche sports training indicators of different data types.

1. Introduction

The conventional index analysis method adopts statistical method to make a general analysis of sports training indicators. When analyzing niche data, due to the small amount of statistical data, conventional methods have the disadvantage of low comprehensive analysis capabilities[1-2]. Therefore, the application of data mining technology in the analysis of sports training index is proposed in this paper. According to the characteristics of the data set, the classification of index parameters was determined, then the sports training analysis mechanism was established by importing data mining technology, and the analysis model was constructed. By analyzing the three processes of data preparation, data mining and result interpretation, the data mining results of training indicators are obtained and data analysis is completed. Finally, the data mining technology is applied in the analysis of sports training indicators. In the simulation test environment, two different methods were used for coverage test, accuracy test and immunity test to obtain variable parameters for comprehensive analysis. Through the calculation and comparison of this parameter, it can be seen that the analysis method proposed in this paper has a very high effectiveness.

2. Construction of the analysis model for sports training indicators

2.1 Import of Data Mining Technology

Determining the index parameters is based on the characteristics of the data set to find the concept description of the category which represents the overall information of such data, that is, the intension description of the category. The purpose of the classification is to analyze the input data. Through the characteristics represented by the data, the accurate description is found for each type of data, such description is often expressed as a predicate, and it is used to classify the subsequent data. Although the class labels for these data are unknown, their categories can still be predicted[3].

The classification can be described as follows: Given a set T of training data, where the element record is described by several attributes, there is only one attribute as a class attribute in all the attributes. This set is represented by a vector X=(X1, X2,..., Xn), where Xi ($1 \le i \le n$) represents a non-category attribute and may have different ranges. When the value range of the attribute is continuous, the attribute is called a Continuous Attribute; otherwise it is called the Discrete

  DOI: 10.25236/isaicn.2019.035

Attribute. C={Cl,C2,...,Ck} represents the data set with k different categories of attributes. Then the T determines a mapping function from vector X to C, that is, $H: f(X) \to C$. The purpose of the classification is to use data mining techniques to express the implicit function H. The expression of the function H is as follows [4]:

$$H = pH_0 + \log(a^n + ef) \quad (1)$$

Where the H represents the implicit function, the H_0 represents the initial state of the function, the p represents the defining attribute of the function, the a represents the range of the element record, the n represents the range of the condition, the e represents the range of the sports index and the f represents the discrete index of the sports index. Index parameter classification is generally divided into two steps:

The first step is to import data mining technology through known data sets.

The second step is to use the obtained model for the classification operation.

First, the accuracy of model classification is estimated. If the accuracy of the model is acceptable, the model can be used for classification. The first step, as shown in Figure 1, is to use the training data set for learning. Training sets are analyzed by classification algorithms to generate classification rules.

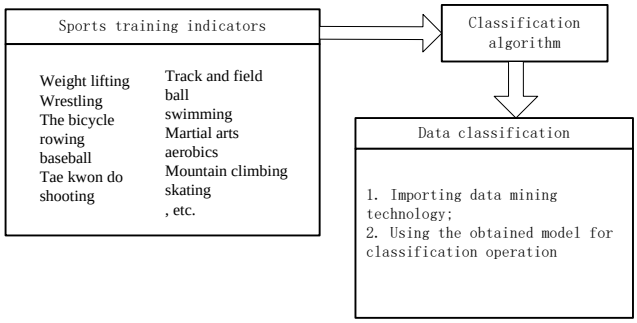

Figure 1 Learn using training set

Second, as shown in Figure 2, the test data is used to evaluate the model. If the accuracy rate is acceptable, the classification rules will be used to classify the new data.

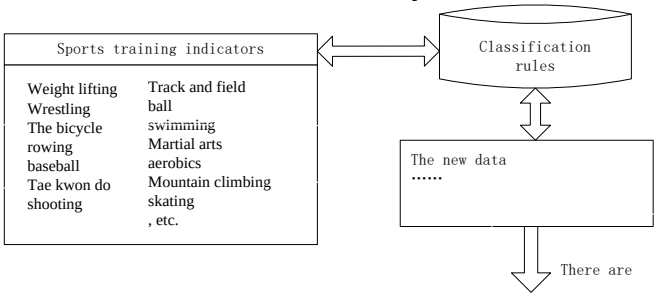

Figure 2 Evaluate the model with test data

Data mining methods are developed from artificial intelligence and machine learning methods, which combine statistical analysis methods, fuzzy mathematical methods and scientific computing visualization techniques. Data mining technology can be divided into the following six categories:

Inductive learning is currently the focus of research, and it is mainly divided into two categories: information theory and set theory. The information theory approach uses information theory to establish the decision tree. In the field of sports training analysis, the decision tree is a simple index representation method. It gradually classifies cases into different categories. This kind of method has good practical effect and great influence. Since the method finally obtains a decision tree, it is generally called a decision tree method. The more distinctive methods in information theory methods are: ID3 and IBLE method[5]. In recent years, due to the development of the rough set theory, the set theory method has been rapidly developed. This includes the coverage of positive

exclusion exceptions (AQ method), the concept tree method and the rough set method. Their three relationships are shown in equation 2 [6].

$$A=\delta\sum_{i=1}^{j}[\psi\lambda]^2 \quad (2)$$

Typical biomimetic technology methods are neural network methods and genetic algorithms. These two methods have formed an independent research system and they have also played a huge role in data mining. The neural network method is based on the IP model and the Hebb learning rule, and the three types of neural network models are established. The neural network sports training index is a distributed matrix structure. Neural network learning is reflected in the gradual calculation of neural network weights. Using neural network techniques is particularly effective when it is difficult to obtain concepts from complex or inaccurate data. The trained neural network is like an "expert" with some kind of specialized sports training indicators, so it can learn from experience like people [7].

Genetic algorithm is the algorithm that simulates the process of biological evolution. It consists of three basic processes of breeding, crossing and mutating. The algorithm has played a significant role in optimizing calculations and classifying machine learning.

From this it can be seen that certain mathematic operations on several data variables can get the corresponding mathematical formulas.

The statistical analysis method uses statistical principles to analyze the data. It includes: common statistics, correlation analysis, regression analysis, variance analysis, cluster analysis and discriminant analysis[8].

The generation of fuzzy mathematics is due to the objective existence of ambiguity. And the higher the complexity of the system is, the stronger its ambiguity is. This is the principle of mutual grammar summarized by Zadeh. The fuzzy set theory can be used to make fuzzy judgments, fuzzy decisions, fuzzy pattern recognition, fuzzy association rules and fuzzy cluster analysis on practical problems. The expression of fuzzy mathematics method is shown in formula 3.

$$Pi(m)=\frac{P(Z_1+Z_2)}{x_1x_2}+\lambda \quad (3)$$

Where the Pi(m)means fuzzy mathematics, P is the representation of the complexity of the system. Z1 represents the fuzzy judgment, Z2 is the fuzzy decision, x1 represents the fuzzy pattern recognition, x2 represents the fuzzy cluster, and λ represents the fuzzy association rule.

The visual data analysis technology broadened the traditional charting function and enabled users to analyze the data more clearly. For example, turning multi-dimensional data into a variety of graphics, which play a very strong role in revealing the inherent nature and regularity of the data. The purpose is to enable users to browse data and mining process alternately, and improve the effect of mining. This technology plays an important role in all stages of data mining. In the preparation phase, the source data is displayed through scatter plots and histograms, which will lay the foundation for better data selection. In the mining phase, various mining processes are described in the visual form, and the user can see from which database the data is extracted, how to extract, how to preprocess and how to mine. In the presentation phase, the technique makes the training indicators easier to understand.

2.2 Establishment of training analysis mechanism

Data mining classification techniques include decision trees, Bayesian, neural networks and rough sets. The condition that the decision tree stops splitting is that the data on one node belongs to the same category and no attribute can be reused for segmentation. Building the decision tree can be done by scanning the database several times. This means that fewer resources are required and that it is easy to handle situations where there are many predictors. Therefore, the model of the decision tree can be built very quickly and is suiTable for applying to a large amount of data. Through the determination of index parameter classification, the data mining technology is imported and the

analysis mechanism is established. Finally the model is built.

3. The analysis of index

3.1 The analysis of data mining process

Data mining is a multi-stage process, which mainly includes data preparation, data mining and result interpretation. The data mining process of sports training index is the iterative process of these three phases.

3.2 Effective data analysis

If the sample belongs to the same class, this node is a leaf node and is marked with this class. Otherwise, the measure of information gain is used as heuristic information to select the attribute of the sample classification, which is the “test” or “decision” attribute of the node. Assumed that all attributes are classified, that is, take discrete values. The branch is created for each known value of the test attribute, and the sample is divided.

4. Case analysis

In order to ensure the effectiveness of the technology proposed in this paper, simulation test analysis is performed. The test uses different types of sports training index as objects for analysis. In order to ensure the validity of the test, the conventional index analysis method is used as comparison object. The test data is presented in the same chart and conclusions are reached through the calculation of comprehensive analysis capabilities.

5. Conclusion

This paper puts forward the application of data mining technology in the analysis of sports training indicators. It relies on the construction of index analysis model, through the analysis of mining process and data, the analysis of sports training index is completed. The experimental data shows that the method designed in this paper has extremely high effectiveness. It is hoped that the research in this paper can provide a theoretical basis for the analysis methods of sports training indicators.

Acknowledgement

Fund Project: 2019 Ministry of Education Humanities and Social Sciences Research Youth Project Funding: Research on the Practice, Problems and Commissioning Strategies of Youth Physical Health Policy Implementation, No. 19YJC890045.

References

[1] WangChong, WANGCong. Simulation of 3d visual action amplitude tracking method in sports [J]. Computer Simulation, 2017(1):245-248.

[2] ZhouPeng, WangShiwei, ZhangWuping. Simulation of high accuracy control of volley hit point on volleyball front [J]. Computer Simulation, 2017(12):246-249.

[3] Zhang L, Yang X, Sang C. Cloud computing and data mining application in enterprise profitability analysis based on the perspective of cash flow[J]. RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao, 2016, 2016:161-172.

[4] X. Ruan, G. Tao, H. Liu, et al. Application of data mining for investigating the cognition of how square dance promote community sports culture construction[J]. Boletin Tecnico/technical Bulletin, 2017, 55(13):594-600.