A Score Management System Design for Sports using Data Mining: Analysis of a New Study

Hao LIU[1, a]

1 Department of Physical Education
Changan University, Xi'an 710064, China
[a]liuhao@163.com

Abstract — In view of the current situation in physical education, there is urgent need to have duplications to allow teachers to use complex statistical and analytical work to solve practical problems currently facing colleges and universities. This should make their jobs easier, but also makes teachers more focused on performance improvement and more disciplined in their approach. How to better improve the physical quality of university students? To address the problem currently in common debate we use: i) scientific and rational analysis and data mining, and ii) feasibility and scientific programs for the sports section of teachers, schools, colleges and universities. To achieve this we combine their own work with current computer technology, and apply it to college students by designing a sports achievement mining systems to better provide accurate physical education to teachers for training and reference. We then present a general sports score overall framework for scientifically managing the process. We use data mining technology to establish a unified regulatory framework for scientific sports scores, and extend the management capabilities of performance to achieve enhanced decision support performance management.

Keywords -- Score Management System; ADO.NET; Sports; Data Warehouse; Data Mining

I. INTRODUCTION

With the reform of the education system and the education of widening social increasingly demanding high-quality and comprehensive physical talents, students are not just good grades, it also requires good physical fitness, so as to competent future work. Therefore, more and more attention in Physical Education, and the Ministry of Education for the education curriculum also developed a series of training standards and performance standards. Physical examination also included the student's academic performance in the past, as part of the test overall quality of students. During sports fitness tests, including students' body shape and body are two aspects [1-3]. For different physical exercise, there are different training methods and test methods, such as running, usually the length of time or location, minutes and seconds or meters. The use of these results expressed as a fraction, will be included in the final score, therefore, usually students completing these courses, educational achievement is responsible for converting these sports in accordance with relevant state standards. However, due to many students, while sports program includes a lot of content, which is very much related to the data, teacher data to be converted, then then enter the data into the system to the Senate, this process is very large workload, and this manual process requires teachers to complete, thus making inefficiency. This has to a certain extent, restricted the development of physical education [4].

Therefore, the actual situation of Physical Education, the urgent need to have a lot of duplication that allows teachers from complex statistical and analytical work of the system, to solve practical problems currently facing colleges and universities, make these jobs easier, but also makes teachers more focused summary and performance improvement in discipline. Thus play a guiding role of teachers in students to improve the quality and effectiveness of teaching. For questions and present situation of Physical Education exists for related information technology, full use of modern data mining technology to enable automated processing and analysis of large amounts of sports data and statistical analysis to facilitate the students' queries and teachers. Application of this system not only for various grades of inquiry and subtotals work quickly and accurately, including the achievement of sorting and filtering, etc., and can make the appropriate exercise program to develop the student's different physique and health of students at any time tracking and evaluation, so that students in their bodies understand and grasp the situation, in order to improve students' motivation to exercise, in order to achieve the combination of quality education and physical education.

In view of this situation, the use of data mining technology for application herein teaching management system, and remarkable achievements, the school management has been greatly improved. So we question against the existing universities proposed scheme Sports performance management system, designed to solve various problems facing the process of physical education in colleges and universities currently encountered by the application of the system to help students develop good exercise habits, enhance physical fitness of young people, provide a guarantee for good teaching quality and effectiveness.

II. THE BASIC TECHNOLOGY OF DATA MINING

Data mining is a lot of, incomplete, noisy, fuzzy, random data is extracted from implicit in them, people do not know in advance, but is potentially useful in the process of

knowledge. Data mining as an emerging multidisciplinary applications, decision support are all walks of life activities play an increasingly important role. Data mining include clustering, classification, prediction, Outlier analysis, association rules, and other aspects of visual description. Data mining association rule mining is an important research topic, but also one of the most widely used, in recent years has been the industry's research focus [5-6]. The main purpose of association rule mining is to find valuable items between large data sets relevant contact, as shown in figure 1.
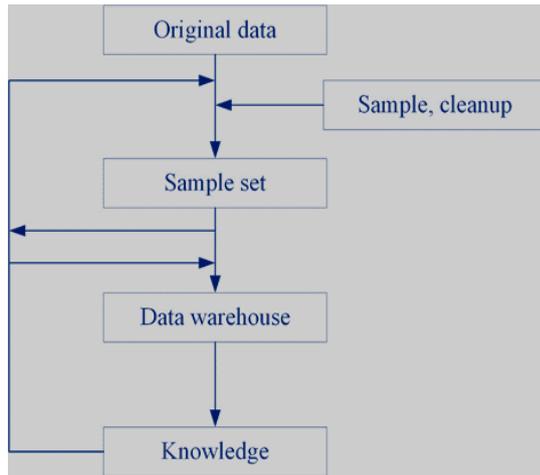

Figure 1.The basic process of data mining.

There are several methods of data mining to achieve, in terms of data mining tasks, introduces several common mining methods.

Data Summary. Refers to the summary data by a certain method to the existing data is concentrated, usually it refers to the use of the calculation method for some data, such as the process of summing the data, process the data to calculate the average value, which belong to a kinds of statistical methods, the results of these data calculated by a certain mode of expression chart, a common bar charts, pie charts and so on. Data mining focuses on deep data summary, it is a generalization from one point of view to summarize the data, usually in accordance with the order from the low to high variety of data abstraction, through different levels of analysis and abstraction. The data potentially draw some rules in order to predict the future trend of the data. There are two ways that can be achieved in the process of data generalization, and that is facing the inductive method and multidimensional data analysis method properties [7-9]. Multidimensional data analysis methods in data mining process common data warehouse technology, this method in actual use can also be called online analytical processing methods.

Classification Mining. Classification mining is the use of classification tools will need to handle data mapping, and mapping purpose is to be assigned to the same category. Often these data types are classified through a discrete type. When the classification is constructed, it can have a variety of methods to achieve, and usually several ways to get more

statistical method that contains the Bayesian method and the non-parametric method, its essence is by constructing discrimination It means a function to classify data. In addition to this construction method classifier, also it includes a number of other methods, including machine learning, decision tree method.

Clustering methods. Clustering refers to the same type of data classification, clustering of data to achieve such information or data gaps in the same category as far as possible a minimum, while the gap between the different categories will be the maximum possible, thus making this information and relationships between data become clear, for the latter part of data mining will play a very important role in guiding and necessary preconditions. In the implementation process of clustering methods, statistical methods and machine learning methods common method.

III. .NET AND DATABASE TECHNOLOGY

Comprehensive comparison of current popular development of various types of technology, in the framework of the selection system, preclude the use of the three-tier architecture, sports scores for student analysis, using data mining methods, decision tree analysis for data expansion.

ADO.NET data provider is a specific type of database processing class. Different types of data providers are a set of different classes, to realize their access to specific types of data storage functions, optimizing the various ADO. Network data storage type. Referred to as data provider must provide a set of basic functions, but special data providers can have a lot of additional properties and methods, they are just in order to access the memory data type [10-11]. Figure 2 shows a detail of ADO. Net object model, the system. The relationship between the core components and their data namespace. Including the six main objectives: connection object, command objects, data reader object data adapter object, the Data Set object, the data view objects. They provide a powerful data processing capabilities.

Data Warehouse. Data Warehouse has become the industrial economic management departments to support industrial and economic management decisions is an important platform. Established based on industry and economic operation monitoring data warehouse information system can implement a variety of industrial and economic information system integration of heterogeneous data sources, optimizing the economic analysis in the process industry, to improve and strengthen the industrial and economic monitoring and forecasting techniques to improve the efficiency of the industrial economic operation monitoring daily economic business analysis of industrial economic management departments to provide decision support industrial economic operation data acquisition, data extraction, data analysis, data modeling, OLAP and data mining and sharing for the industry to provide a platform to query the data and economic analysis and decision support and protection [12]. Data warehouse structure shown in Figure 2.
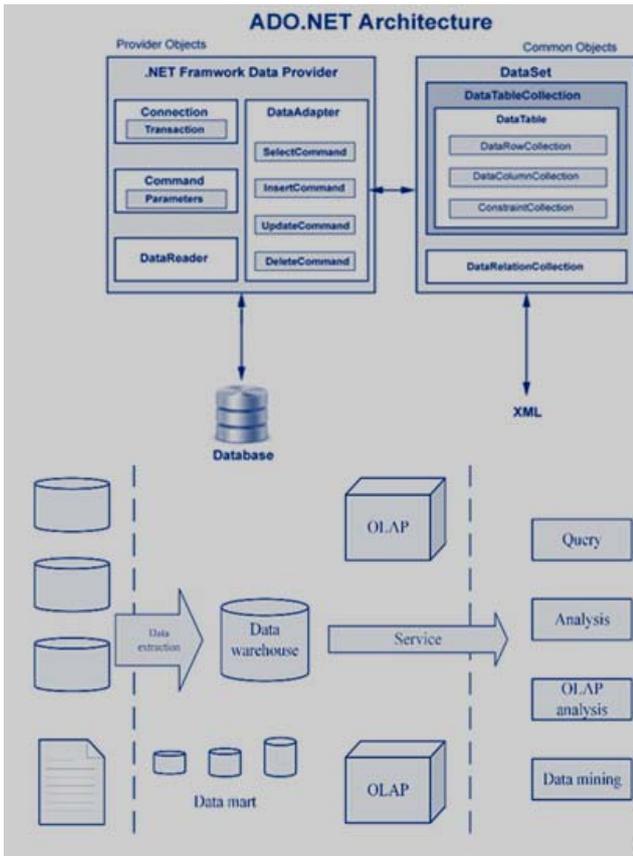
Figure 2. Structure of the ADO.NET technology and the data warehouse.

IV. REALIZATION OF THE DATA MINING

Among data mining applied to sports and physical analysis of performance management system, the first step is to define the main issues to be solved, and then in order to achieve effective information selected according to the issue, in order to achieve build the database. After the establishment of a database, through analysis of the data, build a data model. In the data model are built on the realization of the data evaluation and analysis, if there is an error, the need for effective adjustment. Finally, based on the interpretation of the corresponding model, as a basis to make a scientific decision. System mainly uses a decision tree data mining. A decision tree is a choice from a large number of heterogeneous data, identify data mining methods similar data through a series of simple decision tree rules. There are a variety of decision tree algorithms, such as ID3, CART, C4.5, SEES, SLIQ like. Among the most widely used and most influential algorithm C4.5 algorithm ID3 algorithm and improved, according to the characteristics of sports achievement data should be chosen to establish a decision tree ID3 algorithm. ID3 algorithm from the root node of the tree, select an attribute to distinguish different data samples. For each value of the property to produce a branch, the branch corresponding subset of the attribute value is moved to the newly created child nodes, the nodes until all samples

are partitioned into a class. Such top-down decision tree generation algorithm, can be divided into distinct subset of records according to different attributes of information provided, and affect the speed and quality of decision trees grown tree structure, which leads to the result of judgment rule the merits of relationship information.

In the decision tree generation process, comprising the steps are usually two steps, namely the generation of trees and tree pruning. Tree generation means generates a first initial tree, in the tree, all the data are first placed in the position of the root node, then a top-up of a recursive manner in which these data are scattered, placed in the appropriate nodes on the tree in such a way to generate a complete decision tree; and then the tree pruning, some weed out useless data, to form a reasonable tree science, thus, Decision Tree. Split in the decision tree process generally need to follow two conditions, namely, all the data on one node belong to the same category; no property can be re-used for the data partition. Only two conditions are met, the division can be stopped. When recursive manner established decision tree, usually require several database scanning and traverse the work can be completed, which also explains the use of this method can easily build decision trees, consuming fewer resources, at the same time It can be applied to a number of large databases to establish a decision tree. Figure 3 describes the steps simple decision tree generated.
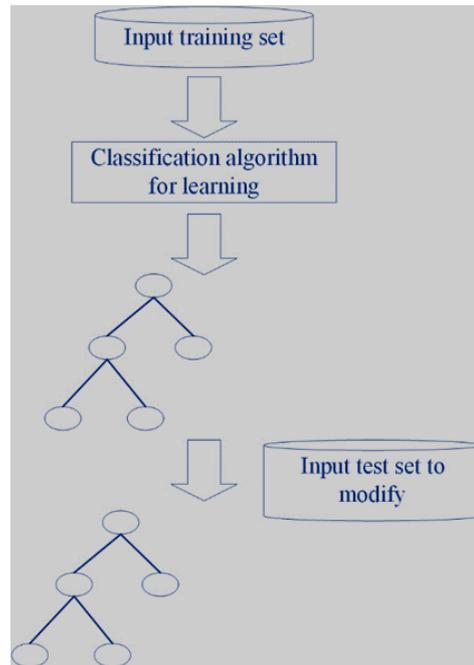

Figure 3. Step of decision tree generated.

School sports majors sports test scores, for example, through data mining, research the relationship between sports performance and the employment situation of. After statistics, sports scores and graduation years the whereabouts of a class-related data as shown in TABLE I.

8.3

TABLE I. TRAINING DATA

ID	Volleyball	Basketball	Long run	Long jump	Shot put	Employment situation
Student 1	81	87	58	75	77	C
Student 2	76	80	69	63	69	C
…	…	…	…	…	…	…
Student n-1	69	58	49	74	83	D
Student n	79	81	66	59	72	D

Based on the above training set data, using ID3 algorithm, considered the decision tree structure, a schematic diagram of the decision tree shown in Figure 4.

As can be seen from the figure, the relationship between the various subjects between test scores and the employment situation, for example, in basketball tests, the higher the score, the better the employment situation, which shows sports majors looking for a job, the employer Units are most concerned about is the test case of the student basketball program, followed by long-distance student achievement, because these two subjects to better reflect a student's physical condition.
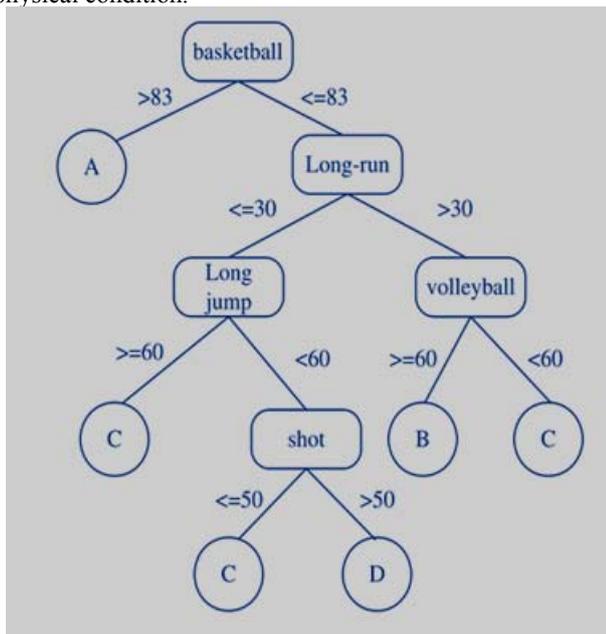

Figure 4. The resulting decision tree.

V. THE OVERALL ARCHITECTURE OF THE SYSTEM

Student sports performance management system developed using data mining techniques, in order to obtain a complete system functional requirements, needs analysis personnel to go to a Higher Colleges of Physical Education to conduct field research to learn more about the process of teaching assignment sports test type, project, etc. and student achievement records, statistics, analysis, etc., obtained by sorting system needs to implement functions. Aims to use students' sports performance management system to improve the efficiency of student sports teacher performance management, improve the level of information the school sports teaching and management, and provide a scientific basis for the construction of other information technology projects. The overall functional structure diagram of the system be represented, particularly as shown.

Development of the target system uses a three-tier .NET architecture, the entire application is divided into three levels, namely, the presentation layer, business logic and data access layers. The presentation layer is a direct user-oriented application layer, is responsible for the completion and interactive user interface is the use of this layer is typically accomplished through the design of some of the controls on the screen to complete the interaction. This layer is only responsible for receiving access requests issued by the user and the final result to user feedback, the specific method by which to achieve a user's request not to care; logic layer in the middle of the presentation layer and data access layer, it is the entire system of business core aspects of logic, is responsible for the user's request logical calculation and processing, while receiving data from the data access layer; data access layer is responsible for program execution and data access, and the results fed back to the user. Figure 5 shows the overall technical architecture of the system.

Given by the above figure shows, the target system will be a three-tier entire application layer partition, the drawing for the top-level control page, the user object it points to sports teachers, students and system administrators; in the middle of the application layer is business logic layer, the main job is responsible for logical judgment and execution, and data layers interact simultaneously; data access layer is mainly related to the completion of the database interaction, perform a variety of data requests. The reason for the use of the three-tier model, because through the three-tier division, can make these applications layers separated from each other, only focus on the task of processing within its own competence, not only reduces the coupling between modules, but also improve the cohesion . And the use of such a development framework, programmers can make the workload is effectively reduced, when business needs change, you can easily just need to realize these demands by certain adjustments business logic, without rewriting code or conduct a comprehensive revision, therefore, makes the program development efficiency has been greatly improved, but also development costs to get better control, cost reduction has a very important role.
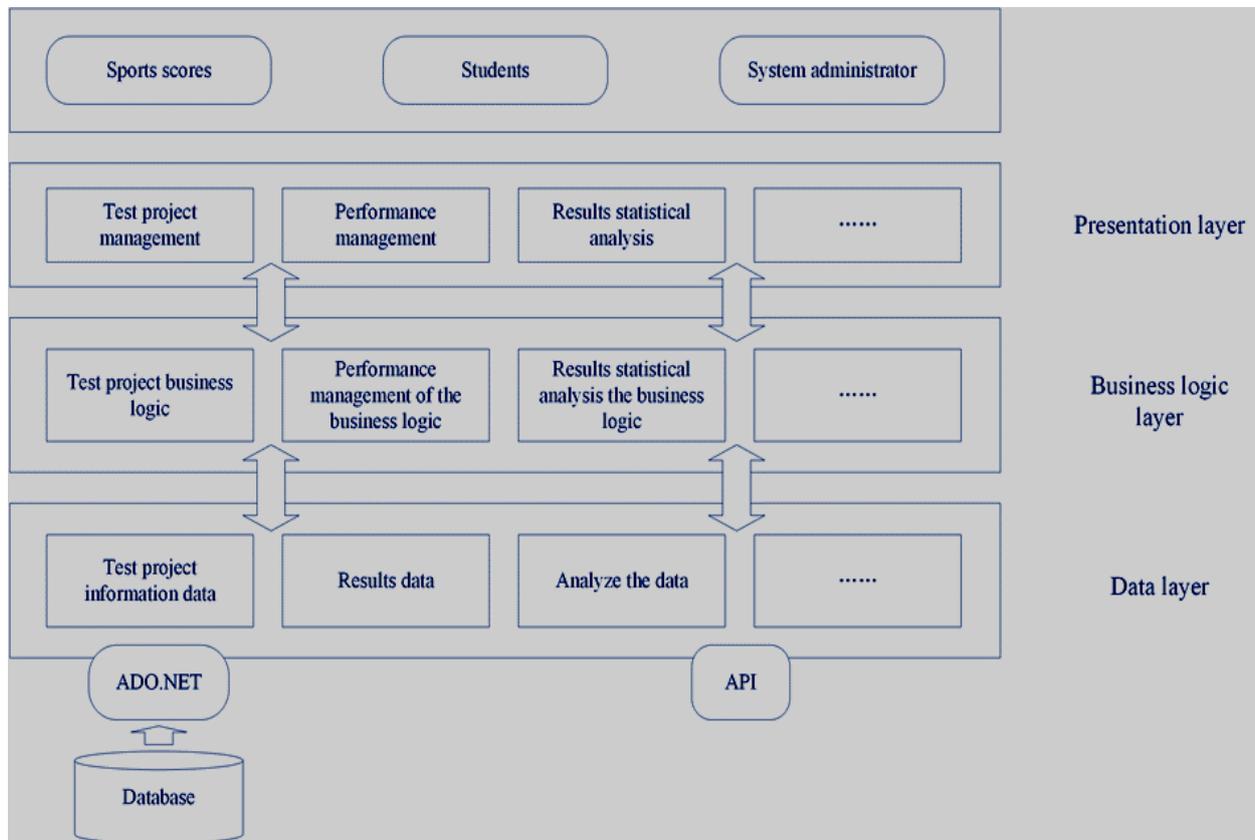
Figure 5. Architecture of the score management system based on data mining.

VI. CONCLUSION

This paper aims to study how to combine data mining technology and performance management to extract hidden data into useful information from large data. According to the specific circumstances of the current performance management we introduce data mining knowledge, detailed analysis and comparison of the data mining technology on the contents using the decision tree method of selection and rough set theory applied to student achievement analysis system. According to the specific circumstances of the current performance management, the paper describes the role of performance management, inadequate status quo and the existing performance management, decision tree algorithm and rough set theory introduced in detail, their role in the achievement of management carried out in detail. ID3 algorithm application for different types of courses of student achievement are analyzed to identify potential factors that affect student achievement, so that students can better maintain a good learning state to provide decision support information for the teaching sector, lead to better carry out teaching work to improve the quality of teaching; the application of rough set theory to a teaching classes in English achievement analysis to identify student achievement overall impact on the most important factor, in order to change teaching methods and methods of foreign language teachers, to improve the overall quality of teaching to provide evidence that the method can also be extended to other classes of different grades the test subjects were analyzed to provide technical support to further improve the performance management system.

REFERENCE

[1] Liao S H, Chen J L, Hsu T Y. "Ontology-based data mining approach implemented for sport marketing". Expert Systems with Applications, vol. 36, no. 8, pp.11045-11056, 2009.

[2] Schumaker R P, Solieman O K, Chen H. "Predictive modeling for sports and gaming". Springer US, 2010.

[3] Tsoumakas G, Katakis I, Vlahavas I. "Mining multi-label data" Data mining and knowledge discovery handbook. Springer US, pp.667-685, 2010.

[4] Tsai H H. "Global data mining: An empirical study of current trends, future forecasts and technology diffusions". Expert systems with applications, vol. 39, no. 9, pp.8172-8181, 2012.

[5] Liao S, Ho H, Yang F. "Ontology-based data mining approach implemented on exploring product and brand spectrum". Expert Systems with Applications, vol. 36, no. 9, pp.11730-11744, 2009.

[6] Bhatt C A, Kankanhalli M S. "Multimedia data mining: state of the art and challenges". Multimedia Tools and Applications, vol. 51, no. 1, pp. 35-76, 2011.

[7] Xiong H, Wu J, Chen J. "K-means clustering versus validation measures: a data-distribution perspective". Systems, Man, and

Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 39, no. 2, pp.318-331, 2009.

[8] Vaclavik L, Lacina O, Hajslova J, et al. "The use of high performance liquid chromatography–quadrupole time-of-flight mass spectrometry coupled to advanced data mining and chemometric tools for discrimination and classification of red wines according to their variety". Analytica chimica acta, vol. 685, no. 1, pp. 45-51, 2011.

[9] Huang S C, Chang E C, Wu H H. "A case study of applying data mining techniques in an outfitter's customer value analysis". Expert Systems with Applications, vol. 36, no. 3: pp. 5909-5915, 2009.

[10] Li N, Wu D D. "Using text mining and sentiment analysis for online forums hotspot detection and forecast". Decision Support Systems, vol. 48, no. 2, pp.354-368, 2010.

[11] Zheng V W, Zheng Y, Xie X, et al. "Collaborative location and activity recommendations with gps history data" Proceedings of the 19th international conference on World wide web. ACM, pp. 1029-1038, 2010.

[12] Hopkins W, Marshall S, Batterham A, et al. "Progressive statistics for studies in sports medicine and exercise science". Medicine+ Science in Sports+ Exercise, vol. 41, no. 1, pp. 3-10, 2009.

[13] Saravanan D, Srinivasan S. "Data mining framework for video data" Recent Advances in Space Technology Services and Climate Change (RSTSCC), 2010. IEEE, pp. 167-170, 2010.