

# Application of Data Mining in the Guidance of Sports Training

Fanhui Meng<sup>1</sup> Qingli Li<sup>2</sup>

Department of Physical Education Heilongjiang Bayi Agricultural University

Daqing, China 163319

<sup>1</sup>Mfhmfh2008@163.com <sup>2</sup>Liqingli@163.com

**Keywords:** sports training, data mining

**Abstract:** Data mining is the techniques of finding the potential law from the data by machine learning and statistical learning .This paper focuses on a number of problems existed in the currents ports training, discusses the application principle of the data mining technology in sports training, and applies the critical neural networks for forecasting the performances of the athletes .Experimental data show that prediction of athletic performance by the use of neural network has very good approximation ability. It shows a broad application space of the use of data mining technology.

## Introduction

The sport is a highly integrated disciplines, it contains human sports science, sports humanistic social science and so on, before 2000, the decision-making supporting system in sports field is limited to sports training, sports management evaluation. With a wide range of research areas in sports science, DSS in the field of sports applications still need to further expand. The research shows that, in recent years scholars on the movement selection field intelligent management system, the national fitness information analysis and decision support system are studied, it shows more and more scholars have already been aware of the powerful function of DSS support, and attempts the application to sports fields. With the rapid development of computer technology and in all aspects of development application, especially artificial intelligence theory and data mining technology, for students to realize the organic integration of different aspects of the theory, scientific training, advanced training methods applied to the students of sports training management become a possibility, the application of data mining technology in college students' Sport Training Assistant Decision Support System the theory is feasible. Based on the existing college students' sports, training and other aspects of the analysis, and combined with the students' specific physical quality and health standards, find out a set of scientific and reasonable training programmes, and through the auxiliary program in physical classroom use, to test the rationality, thus the existing decision support system to improve.

Data mining is the nontrivial extraction of implicit, previously unknown and potentially useful information from data, similar terms have the KDD ( knowledge discovery in database ), data fusion, decision support, data mining is defined from the stored in database, data warehouse or other repository of large amounts of data to obtain valid, novel, potentially useful, ultimately understandable patterns in data processing. The Ministry of education in 2002 promulgated the " students' physical health standard ". At present, all colleges and universities have accumulated years of students' physique health data, these data imply a large number of potential rules and knowledge. Around these data to establish a database, data mining, can find out the students' physical health changes, as well as with other kinds of relations between factors, for physical teaching reform to provide valuable information for decision making. is the nontrivial extraction of implicit, previously unknown and potentially useful information from data. [1].

Data mining can also be described as an attempt to create a database to describe the complex world of simple model, so we can also say that data mining is a method of processing a large amount of information, and it helps to more quickly find useful information. Many people think of data mining as a common terminology, knowledge discovery in databases or KDD, synonyms, and others are just the data mining as knowledge discovery in database is an essential step in the process of. The

knowledge discovery process has the following steps: data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, knowledge representation.

### **The Trouble Existing In Sports Training**

For competitive sports, in order to improve the athletes in the competition's result is usually in training through a variety of training methods to improve the athletes' competitive level and physical fitness. Have to go through the various training methods to improve athletes' basic quality, technical level and special quality and technical level, and through these indicators to evaluation of athletes' physical condition or predict athletes may athletic performance. Because in the process of sports training, different age groups of male, female athletes have different standards for quantitative detection. On the training of coaches through the test data analysis, can find out the gap between athletes, and so the development of training programs. But with the accumulation of test data, the traditional data processing, analysis methods cannot meet the actual demand, the following two prominent:

1. In the current sports training, many of the training items and contents are coaches based on practical experience of many years to develop, in the long-term training, the accumulation of a large number of training data, but these index exists between what kind of relationship, that is a result which can drive the other one or several grades improve improve? How, many coaches and not scientific analysis. Based on the above reasons, how the training data and index for further optimization, according to what criteria optimization, how to optimize? We must first clear the relationship between the physical fitness test items, in order to further optimize the physical indicators provide scientific basis.

2. Exercise training aim is to hope that through training to improve the athletes' competitive state, so that the player during the game to achieve the best competitive state, give full play to the best level. The traditional method is to rely on the trainer's experience, in preparation for the games, the athletes by adjusting the intensive training of physical state to adjust to the best. This requires that coaches have rich training experience, and fully understand the athletes' physical condition. Obviously, this increases the coaches training decision training decision error and the possibility of pressure. Through the training data analysis of athlete's competitive state changes, to evaluate their competitive condition. Each athlete after long time of practice, accumulated a wealth of training data, these data are each athlete's competitive state changes in the important historical information. Whether through these historical data analysis of athletes' competitive state of development, changes in a certain period, predict athletes' competitive state, to provide a scientific basis for trainers to make training plan. The current domestic and foreign to the sports training data for scientific analysis, is widely used in the method of statistical analysis. For example, by means of the principal component analysis, cluster analysis identified in some sports every target important degree, through correlation analysis to determine the correlation among different indexes, by means of regression method for the prediction of athletic performance. Should say statistical analysis technology use in athletic training, guidance for coaches to athletes' training is valuable, because it broke the tradition of only rely on personal experience to guide training of coaches. But as a result of athletes, especially athletes, years of training, the accumulation of a large number of data, using training data were analyzed, using only traditional statistical analysis technique is a little low, because the training data is large, it is difficult to find a suitable model to describe the relationship between data. Therefore, the use of more advanced data analysis techniques is necessary. Data mining technology, to how to face a large number of complex motor training data for scientific laws and correlation, with optimization method. In the data mining technology but also by a large number of statistical tools, but is not limited to statistical analysis technique, which combines many statistics, artificial intelligence, neural network, data warehouse, information theory and other disciplines technology.

### **Design For Decision Support System**

Data mining and problem solving system includes two modules, respectively is the problem solving module and data mining module. The data mining module according to the model library

system model, the data warehouse multidimensional data sets for data mining, in order to obtain the required

Knowledge. Data mining need some specialized algorithms, such as association rules method, clustering algorithm, decision tree, these algorithms in the model library has a corresponding mining model. The result of data mining as the new knowledge and the model can be used to enrich the knowledge base and model base, new generation knowledge rules can provide decision support for the reform of PE teaching. Problem solving module is server of a man-machine interaction module user input in the form of data processing function module, the function module corresponds to the reasoning in expert system for action, it is in the form of Javabean package on the server. In the process of JavaBean will be based on the user input and the rules in the knowledge base to get fit before a match, students participate in a training program, which gives the personal training program and plan, with the user input, all users will be training project to be grouped, and finally in the input end of all results be aggregated to generate a set of collective training programme is returned to the user.

The following is the realization of system data mining methods of association rules mining:: in this system, by using Apriori algorithm for data warehouse data mining, the algorithm by calling the model library association rules mining model to realize the given by the expert, examine the effect of various factors of university students' physical quality between the relevance, it is to find out the students' physical quality is decided jointly by what factors and their weights, meanwhile, will produce about different sports new influence factor, and regards it as a new rule is added to the knowledge base in. Decision tree classification: sports according to its effects on the physical quality of different properties can be divided into: Endurance skills, accurate, strength and speed, phenotype, fighting against type. The system uses a decision tree classification algorithm, this algorithm by calling the model base of decision tree mining model to achieve, in which, the test attribute selection and how to divide the sample set is the key link of constructing decision tree, test attribute begins by sports experts experience income. Through the analysis of training data samples, eventually resulting in a decision tree. The results of the decision tree classification rule extraction, from the root to the leaves of each path to create a rule, form a set of rules. These rules into the knowledge base, in the problem solving module, can use these classification rules are given for faster users to participate in sports.

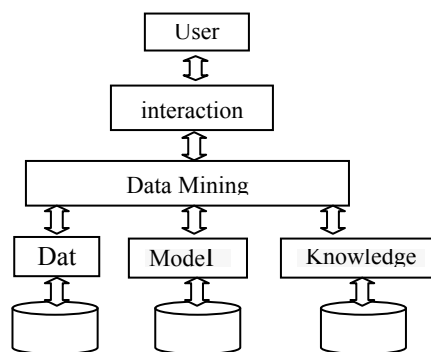


Figure 1. rchitecture Diagram for Agricultural Topic Detection.

Data warehouse management system. The first is the collection of college students sports scores and physical health questionnaire data, and carries on the record, and then the acquisition of the source data is a series of cleaning and transformation, to remove invalid and error data, and then built in the SQLServer data source based on the cube, is convenient for data mining.

The system design of data warehouse in the popular star mode. The star schema is based on relational data warehouse, is for online analysis of a multidimensional data organization form. Relational structure can better adapt the multidimensional data representation and storage. The body of students basic information as fact table, and selected the four related dimensions, each dimension table has its own attributes, dimension and fact tables by dimension keyword associated.

Data warehouse management system for decision support instead of transaction processing and design, so every time a user to perform data mining decision support, only the useful data from the

operation environment of extracted and input into the data warehouse. The input is completed, the data in the data warehouse becomes read-only, only in special conditions can be modified.

**Data Mining Application in Specific Sports Behavior**

Data mining application in specific sports behavior .Algorithm: first, using the relative positive region of consistent and inconsistent information system of core. Then, in order to attribute dependency defined attribute importance degree, according to the important degree of attributes as heuristic information, select the important degree of attributes in the attributes to add to the reduction set, until a termination condition is satisfied now, assuming that the dependence of threshold is  $\epsilon$ .

On physical education comprehensive quality index by, through data mining based on rough set attribute reduction algorithm. If the ratio of height and weight, vital capacity, fitness function to measure physical qualities, can use this method to optimize.

TABLE I. PHYSICAL TABLE

U	Condition attributes C			Decision attribute D
	Ratio of height and weight (c1)	Vital capacity (c2)	Fitness function (c3)	Comprehensive index
E1	Y	Y	Y	G
E2	Y	N	Y	G
E3	N	--	Y	G
E4	Y	Y	N	B
E5	N	N	N	B
E6	N	Y	Y	G
E7	N	Y	Y	BA
E8	N	Y	N	B

Calculation of the relative positive region, calculate the attribute core:

$$\begin{aligned}
 &POS_{(A-\{C1\})}(d) \neq POS_A(d), \\
 &POS_{(A-\{C2\})}(d) = POS_A(d), \\
 &POS_{(A-\{C3\})}(d) \neq POS_A(d), \\
 &R = \{C1, C3\} \\
 &\{ \text{Ratio of height and weight, Vital capacity, Fitness function} \} \text{ the attribute Physical} \\
 &U / R = \{ \{E1\}, \{E2\}, \{E3\}, \{E4\}, \{E5, E6\}, \{E7, E8\} \} \\
 &U / R \cup \{C2\} = \{ \{E1\}, \{E2\}, \{E3\}, \{E4\}, \{E5, E6\}, \{E7, E8\} \} \\
 &U / \{d\} = \{Y1, Y2\}, \\
 &Y1 = \{E1, E2, E3, E6\}, \\
 &Y2 = \{E2, E4, E5, E8\} \\
 &k_R(d) = 4/8, \\
 &k_{R \cup \{a2\}}(d) = 4/8, \\
 &SIG(a2, R, d) = 0.
 \end{aligned}$$

The C2 attribute significance degree is 0, so do not select C2 joined the reduction set, i.e. the last attribute reduction set is Red = { Ratio of height and weight, Fitness function } Insect. This incompatibility in the table, the algorithm to derive attribute core, through the core attributes to determine attribute reduction set, finally determined the main factors,.

Skowron based on discernibility matrix attribute reduction algorithm of information system

Poland A.Skowron scientists first proposed in 1991 by using discernibility matrix to represent knowledge, this approach to knowledge expression system of nuclear, reduction and other conceptual representation and computation has many advantages.

Define ( discernibility matrix ) knowledge expression system  $KRS = ( U, A, V, K )$  on the domain objects is a non empty set  $U = \{X1, \dots, Xn\}$ , definition

$$M_{n \times n} = (c_{ij})_{n \times n} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix}$$

$$c_{ij} = \{\alpha \in A \mid f_{\alpha}(x_i) \neq f_{\alpha}(x_j)\}$$

By definition of  $c_{ij}=c_{ji}$ , so for the matrix, just consider the upper triangular element, or lower triangular elements.

Theorem: information system is equal to the difference in nuclear matrix elements all simple attribute

$$CORE(C) = \{\alpha \in C \mid \exists c_{ij} \in M_{n \times n}, \text{ s.t. } c_{ij} = \alpha\}$$

Theorem: any C in B, B is an attribute reduction of information system, if it satisfies the following conditions:

- 1) CIJ is not empty, there are  $B \cap CIJ$  non empty
- 2) B is independent of the

This gives 5 kinds of behavior of A1, A2, A3, A4, A5, 5 kinds of behavior to improve physical function. And the 8 sample, X1 - X8 is expressed as follows

TABLE II. IMPROVE PHYSICAL TABLE

U	a1	a2	a3	a4	a5
x1	1	0	2	2	0
x2	0	1	1	1	2
x3	2	0	0	1	1
x4	1	1	0	2	2
x5	1	0	2	0	1
x6	2	2	0	1	1
x7	2	1	1	1	2
x8	0	1	1	0	1

Structural difference matrix

$$M_{8 \times 8} = \begin{bmatrix} \emptyset & & & & & & & \\ C & \emptyset & & & & & & \\ 1345 & 1235 & \emptyset & & & & & \\ 235 & 134 & 1245 & \emptyset & & & & \\ 45 & C & 134 & 2345 & \emptyset & & & \\ C & 1235 & 2 & 1245 & 1234 & \emptyset & & \\ C & 1 & 235 & 134 & C & 235 & \emptyset & \\ C & 45 & 1234 & 1345 & 123 & 1234 & 145 & \emptyset \end{bmatrix}$$

We can see that  $c_{72}\{a1\}$ ,  $c_{77}=c_{63}\{a2\}$ , according to the above theorem  $CORE(C) = \{a1, a2\}$ . In the inclusion and combination of properties, satisfy the theorem 3.2 of only  $\{a1, A2, a4\}$ ,  $\{a1, A2, a5\}$ . So  $RED(C) = \{\{a1, A2, a4\}, \{a1, A2, a5\}\}$ . By this operation, can guide the sports training, thereby strengthening sports training science.

### Conclusions

This paper analyzes the physical training in some of the problems, to solve these problems to design a data mining data warehouse solution method, and the data mining algorithms to solve the practical problems, can the sports evaluation and physical guidance to make the correct guidance.

## References

- [1] AK Z. Rough sets[J]. Information and Computer Science. 1982 11(5): 341-356.
- [2] ERPAL BHANDARI, EDWARD COLET, JENNIFER PARKER, ZACHARY PINES, RAJIV PRATAP, KRISHNAKMAR RAMANUJAM. Advanced Scout: Data Mining and Knowledge Discovery in NBA Data. Data Mining and Knowledge Discovery, 1997, 1: 121~125.
- [3] tetsky-ShaPiro, "Datamining and knowledge discovery in business Databases[J], "ISMIS", 1996: PP. 56-67.
- [4] hai, A. Velivelli and B. Yu. A Cross-collection Mixture Model for Comparative Text Mining[A]. In Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. Seattle: Association for Computing Machinery, 2004: 743-748.
- [5] [://www.stcsm.gov.cn/learning/lesson/xinxi/20021125/lesson.asp](http://www.stcsm.gov.cn/learning/lesson/xinxi/20021125/lesson.asp)
- [6] A Gupta, V Harinarayan, D Quass. Aggregate-query Processing in Data Warehousing Environment. In Proc. 21th Int. Conf. Very Large Data Bases. Zurich, Switzerland. 1995 (9) : 358~369.
- [7] V Harinarayan, J D Ullman, A Rajaraman. Implementing Data Cubes Efficiently. In Proc. of the ACM SIGMOD Conference on Management of Data.