

DISCUSSION CONCERNING THE APPLICATION OF DATA MINING TECHNOLOGY IN SPORTS PERFORMANCE MANAGEMENT



ORIGINAL ARTICLE
ARTIGO ORIGINAL
ARTÍCULO ORIGINAL

DISCUSSÃO SOBRE A APLICAÇÃO DA TECNOLOGIA DE MINERAÇÃO DE DADOS NA GESTÃO DO DESEMPENHO ESPORTIVO

DEBATE SOBRE LA APLICACIÓN DE LA TECNOLOGÍA DE MINERÍA DE DATOS EN LA GESTIÓN DEL RENDIMIENTO DEPORTIVO

Xiaojing Song¹ 
(Sports Education professional)

1. University of Management,
Jinan, China.

Correspondence:
Shandong Province Shandong
Jinan, China, 250357.
282712267@qq.com

ABSTRACT

Introduction: Finding the factors that contribute to success in student performance or failure is necessary for every teacher. **Data mining**, which is already used in companies for management processes, can be essential in this research. **Objective:** Discuss the data mining algorithms application in sports performance management. **Method:** A database was developed considering seasonal factors, health benefit index, and sports behavior characteristics. The data were entered under fuzzy logic, processed, and analyzed in IBM SPSS Modeler Software. Decision-making efficiency was improved with the target base interpolation analysis and the C spatial noise reduction methods. The fidelity of sports behavior was consolidated under Gauss time series analysis. **Results:** The relationship between the mining algorithm to find the existing problems and the association results in the mining rules provided valuable information for improving health guidelines to the physical activity students. **Conclusion:** The original data from the educational system can be transformed into useful information through the association rules algorithm, and the relationship between the performance can be obtained, providing the improvement in the decision making for the benefit of the physical level of the students. **Evidence Level II; Therapeutic Studies – Investigating the results.**

Keywords: Athletic Performance; Association Rules; Data Mining.

RESUMO

Introdução: Encontrar os fatores que contribuam para o sucesso no desempenho do aluno ou o seu fracasso é uma necessidade de todo professor. A mineração de dados, que já é utilizada em empresas para processos de gestão, pode ser uma importante aliada dessa pesquisa. **Objetivo:** Discutir a aplicação de algoritmos da mineração de dados na gestão do desempenho esportivo. **Método:** Um banco de dados foi desenvolvido considerando fatores sazonais, índice de benefício de saúde e características do comportamento esportivo. Os dados foram inseridos sob lógica Fuzzy, processados e analisados no Software IBM SPSS Modeler. A eficiência da tomada de decisão foi aprimorada com o método de análise de interpolação da base de alvo e o método de redução de ruído espacial C. A fidelidade do comportamento esportivo foi consolidada sob a análise de séries atemporais de Gauss. **Resultados:** A relação entre o algoritmo de mineração para encontrar os problemas existentes e os resultados da associação nas regras de mineração forneceram informações valiosas para o aprimoramento da orientação à saúde dos alunos praticantes de atividades físicas. **Conclusão:** Os dados originais do sistema educacional podem ser transformados em informações úteis por meio do algoritmo de regras de associação e a relação entre o desempenho pode ser obtida proporcionando o aperfeiçoamento na tomada de decisão para o benefício do nível físico dos alunos. **Nível de evidência II; Estudos Terapêuticos - Investigação de Resultados.**

Descritores: Desempenho Atlético; Associação de regras; Mineração de dados.

RESUMEN

Introducción: Encontrar los factores que contribuyen al éxito en el rendimiento de los alumnos o a su fracaso es una necesidad de todo profesor. La minería de datos, que ya se utiliza en las empresas para los procesos de gestión, puede ser un importante aliado en esta investigación. **Objetivo:** Discutir la aplicación de los algoritmos de minería de datos en la gestión del rendimiento deportivo. **Método:** Se elaboró una base de datos teniendo en cuenta los factores estacionales, el índice de beneficios para la salud y las características del comportamiento deportivo. Los datos se introdujeron bajo lógica difusa, se procesaron y analizaron en el software IBM SPSS Modeler. La eficacia de la toma de decisiones se mejoró con el método de análisis de interpolación de la base del objetivo y el método de reducción del ruido espacial C. La fidelidad del comportamiento deportivo se consolidó bajo el análisis de series temporales de Gauss. **Resultados:** La relación entre el algoritmo de minería para encontrar los problemas existentes y los resultados de la asociación en las reglas de minería proporcionaron información valiosa para la mejora de la orientación sanitaria de los estudiantes que practican actividades físicas. **Conclusión:** Los datos originales del sistema educativo se



INTRODUCTION

For school students' all-round development of body's attention, sports discipline by the school began to attention. Only if the students continue to strengthen their exercise and improve their physical qualities, they can learn better.¹ It can also be said that the discipline of physical education is the basis of other disciplines. The students' sports tests are also taking into account their total results, gradually changing the wrong idea that the students think sports are not so important.² For the calculation of sports performance, it is mainly converted to specific scores according to the relevant regulations of the school through testing students' physical education classes, such as high jump, long jump, and 300 meter dash. Because of the large number of students in the school, the statistics of students' test scores is also a very arduous task and a large amount of tasks. Therefore, we should apply new technologies to improve students' performance management.³ With the rapid development of computer technology, data mining technology is also improving and developing gradually. And more and more companies and enterprises are using data mining technology to improve their management system, so as to continuously promote their company's prosperity and development. For data mining, it is widely used in large companies, but it is relatively small in school applications. Because schools pay more attention to theoretical knowledge, and data mining technology is relatively strong in practice, so it is less used in university management system.⁴ But in fact, this is a loss to the university management system. Due to the lack of data mining technology application, students' related management work is heavy, resulting in low efficiency. Data mining technology has the advantages of good performance, simple operation and easy access. Because of these advantages, it is constantly promoting the rapid development of the technology.⁵ From the perspective of students' physical education curriculum, this paper further has analyzed the application of data mining technology in efficient management system, so as to continuously improve the management efficiency and teaching quality of school.

MATERIAL AND METHODS

The application of data mining in the analysis of sports data

In recent years, computer technology is developing rapidly, and data mining technology is rising gradually. How to find valuable knowledge from massive data and provide objective and correct information for decision makers has become a topic of concern recently.⁶ At present, data mining technology has achieved good application results in all walks of life in the society. Similarly, in the field of education, data mining technology has begun to be used by more and more researchers to analyze students' various campus behavior. With the problems in teaching, students physical health level has become the focus of attention, how to maximize the value of student achievement data, the educational system of school sports achievement data collation, analysis and mining, and provide useful information for the teaching of educational management and student physical health promotion work, into the important content in the teaching of Physical Education.⁷ When people are mining for data, the steps will vary with the application of different fields. Generally speaking, it can be roughly divided into six basic steps: Task Definition, Data Collecting, Data Preparation, Modeling, E-valuation,

and Envelopment. The data mining process of sports performance can also be roughly divided into six steps. (Figure 1) In the teaching of physical education, the methods of data mining are mainly in the following aspects. (Table 1)

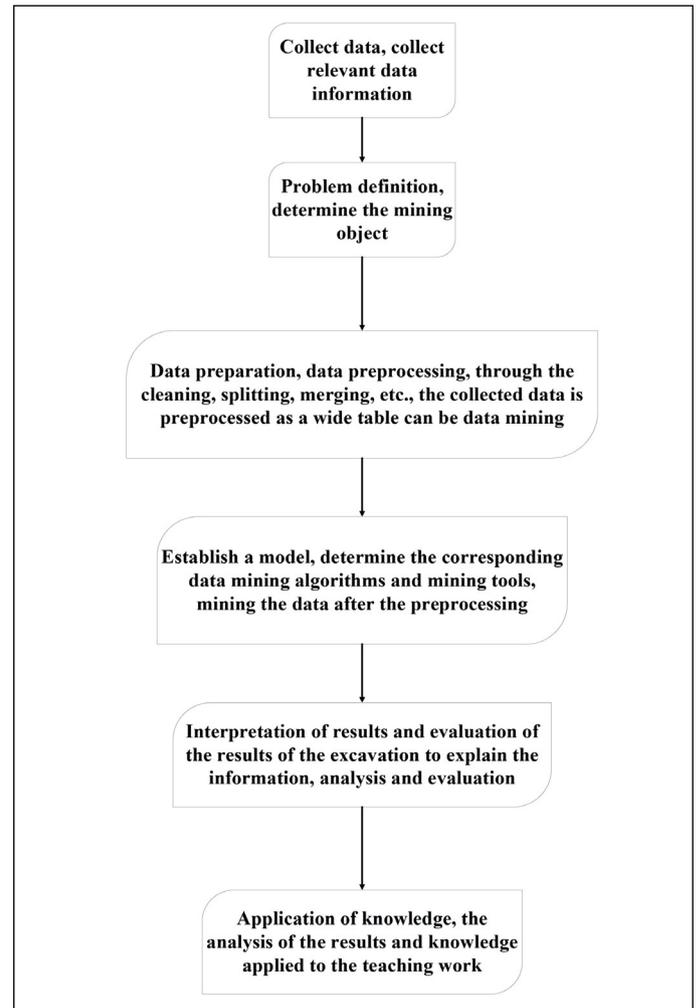


Figure 1. Data mining process of sports score.

Table 1. Data digging methods.

classification of methods	Specific method content
data classification	It is mainly classified according to the attribute and definition of the object, so as to establish the class group.
Data Regression	The next continuous value is continuously predicted by a certain number.
data prediction	The future values of the attribute are evaluated according to the past values of the related attributes of an object.
data clustering	A certain data is classified by difference and similarity. The main purpose is to classify data belonging to the same category into one class, so that different classes of data can be reduced as much as possible.
data association	It mainly refers to the analysis of an event, so as to find out similar events, for example, the factors that appear in the A event, and the proportion of B events in A events.

Database design

The database mainly stores the data in the target system and reads the request and operation of the target system. It is the key content of the system development. The design of the database needs to design the concept, physics, and logic of the database. The conceptual design of the database needs to make use of the E-R diagram to connect the objective things. The main entities in the target system have students' information, test types, test items, and grades, etc. Student information entities E-R map records the student's personal information and information types. Logical database design to design concept of the E-R to convert the database diagram, to follow the principle of conversion work, an entity corresponds to a relationship. A physical database design need to be designed according to the given data structure and method of DBMS storage.

Construction of entity model

1. Description of information fitting model. The data mining of sports behavior in large data environment and the construction of large data information resource network database for sports behavior. Edge sequence tree structure behavior of sports resources database for $\{e_1, E_2, e_r\}$, $e_i = \dots, (O_i, p_i + 1)$, $1 < i < r$, $\{p_1, P_2, p_i\}$. The seasonal factors of juvenile physical exercise meet the following factors:

$$\text{dist}(e_i) = \text{dist}(o_i, p_i + 1) = \text{dist}(\{p_1, p_2, \dots, p_i, p_i + 1\})$$

At this point $\{\text{dist}(e_1), \text{dist}(e_2), \dots, \text{dist}(e_r)\}$ is called the health benefit index of sports behavior. In the information storage model of health sports behavior benefit index, the index system of health benefit index is fitted by constructing characteristic data entity set, and the big data information fitting model is described as follows:

$$\begin{aligned} R_\beta X &= U \{E \in U / R | c(E, X) \leq \beta\} \\ R_\beta X_1 &= U \{E \in U / R | c(E, X_1) \leq 1 - \beta\} \\ \text{bnr}_\beta(X) &= R_\beta X - R_\beta X_1 \end{aligned}$$

The data of sports behavior characteristics are affected by factors such as fitness equipment, season, and school physical education curriculum and so on. We need to excavate and analyze these information data, and get static and dynamic query template set for data mining of juvenile physical exercise.

Extract characteristics of fuzzy decision. A fuzzy decision method was applied to build a physical behavior characteristic entity model. The range of characteristic distribution of sports behavior big data was N discrete characteristic information points $A = \{a_1, A_2, \dots, A_N\}$, the mean time domain distribution of the sports behavioral benefit index is calculated as follows:

$$t_m = \frac{1}{E_x} \int_{-\infty}^{\infty} t |x(t)|^2 dt$$

The mean of frequency domain distribution is:

$$v_m = \frac{1}{E_x} \int_{-\infty}^{\infty} v |X(v)|^2 dv$$

In the form: P is the strength of physical exercise. The flexibility order $\alpha = p\pi / 2$ is the U domain. It is similar to the balance factor in the

sample of physical exercise database information and satisfies a $1 < a_2 < \dots < a_N$. The large data clustering technology is used to classify the feature information, and the X is divided into C classes, and the large data information sub sets are obtained.

$$\begin{aligned} V_1 &= \{> a_1, > a_2, \dots, > a_{N-1}\} \\ V_2 &= \{\geq a_1, \geq a_2, \dots, \geq a_{N-1}\} \\ V_3 &= \{< a_2, < \dots, < a_N\} \\ V_4 &= \{\leq a_1, \leq a_2, \dots, \leq a_N\} \\ V_5 &= \{= a_1, = a_2, \dots, = a_N\} \end{aligned}$$

The construction of sports behavior history log can fit the information distribution characteristics of many influential factors, and build a fuzzy decision and constraint model of data mining considering the healthy sports behavior information parameter sub grid sparse density. The Constraint Decomposition of 3 parameters of the kind of physical exercise, sports achievement and health promotion degree. C_distance (P) is defined as follows: $c_dist \tan ce(p) = \sum_{i=1}^n \frac{dist(e_i)}{i^2}$

To improve the utilization efficiency of sports behavior information by constructing the entity model of fuzzy decision making of sports behavior characteristic data. When the sports behavior data mining rule, the target base class Cbase Target weighted sum is $T_i = \sum_{j=1}^m \omega_{ij} * T(P_{ij}) \geq \tau_i$, the inverse distance interpolation analysis method is used to divide the sports behavior data XN (n) of length N into L segments, each length is M. The C spatial noise reduction method is used to carry out the anti-interference processing of the large data information characteristics of sports behavior, and the fidelity of the sports behavior feature data mining is obtained.

A Gauss time series y (n) is generated for the characteristic data of sports behavior

$$T(P_i) = CF * \sum_{j=1}^m \omega_{ij} * T(P_{ij})$$

A Gauss time series y (n) is generated for the characteristic data of sports behavior. Set $X = (x_1, X_2, x_{D_i})$ is a point in space, Q source data distribution characteristics of health behaviors of juvenile sports, this is the premise of rule established τ_i true rules.

Association rules mining algorithm

Algorithm is the core of data mining technology. The typical algorithm is Apriori algorithm, which focuses on finding out the occurrence of certain events in the database, so as to find those credible and representative rules. The basic idea of this algorithm is to first mine all the frequent item sets through iterations, and then use frequent item sets to construct the minimum user confidence rules. The Apriori algorithm is described below.

Input: the entire database transaction D

Output: the rule that satisfies the minimum confidence of the user

Step:

- scan the database, technology for each item, and find 1 candidate sets;
- Finding out 1 sets of frequent sets according to the support degree;
- Starting from 2 sets of sets, frequent k-1 itemsets produce frequent K item sets;
- The current production frequency K term concentrates only one item set at the end of the cycle.

There are two key steps in the Apriori algorithm connection and pruning. The following example illustrates the implementation of the Apriori algorithm. The minimum support degree is 2 and the minimum

confidence level is 70%. The original item set is shown in Table 2. The database is scanned, and each item is counted. The candidate 1 sets C1, I1:4, I2:3, I3:4, I4:1 and I5:2 are obtained. According to the minimum support, the number of items less than 2 is deleted, and the frequent 1 item set L1 is obtained. (Table 3)

L1 generates C2, {I1, I2}: 2, {I1, I3}: 3, {I1, I5}: 1, {I2, I3}: 2, {I2, I5}: 2, 1, delete the number less than 2, and get the frequent 2 sets. (Table 4) C3 is generated by L2, including {I1, I2, I3}, {I1, I2, I5}, {I2, I3, I5}, {I5}, {I2 and I1, because it doesn't belong to "Chi", so it is not a frequent itemset.

RESULTS

Data sources and preparation

Obtains: Height and weight fraction is greater than or equal to 90 is divided into good, vital capacity, 800 meters or 1000 meters fifty meters run score, score, score of standing long jump, sit and reach scores were greater than or equal to 75 for the best sit ups or chin up scores greater than or equal to 65 is excellent. An excellent 1, or 0, gets a wide table that is suitable for data mining. (Table 4)

Data sources and preparation

Data source or collection. According to the analysis and excavation of the physical performance in a college undergraduate educational system, the 2011 grade -2014 level sports results are selected as the mining samples. (2) Data preparation. The educational system in grade 2011 -014 sports scores in the form of Excel raw data are stored, including properties are basically the same, are the basic information of students (grade school code number, class number, class, student number, name,

Table 2. Original item set.

Grade	Original item set
1	I1, I3
2	I1, I3, I4
3	I1, I2, I3
4	I2, I5
5	I1, I2, I3, I5

Table 3. Frequent 1 sets of sets L1.

Grade	Support
I1	4
I2	3
I3	4
I5	2

Table 4. Frequent 2 sets of sets L2.

Grade	Support
I1, I2	2
I1, I3	3
I2, I3	2
I2, I5	2

Table 5. Sports achievement wide table.

Grade number	class	Student Id	Height and body weight score	Vital capacity fraction	Fifty meter run score	Standing jump score	Anterior flexion fraction of the sitting body	Eight hundred meter or one thousand meter run score	Sit ups or pull up the score
41	1402	01143048	1	1	1	1	1	1	0
41	1403	01143081	1	1	1	1	1	1	0
41	1403	01143065	1	1	1	1	1	1	0
41	1402	01147062	1	1	1	1	1	1	1
41	1402	01147043	1	1	1	1	1	1	1
41	1303	06134073	1	1	1	1	1	1	0
41	1402	01147057	1	1	1	1	1	1	1

gender, nationality code, date of birth, the source of students, home address, ID number). Whether the waiver, waiver, score, standard score, additional points, running additional points, height (CM) and weight (kg) fraction, vital capacity (ML) and 800m scores, and the scores of 1000m and run, run score, 50 run (s) and fractional, standing jump Far (CM) and score, body antexion and scores, sit ups and pull ups and score points. First of all, in SPSSstatistics, we use merge function to merge 2011-2014 grades of sports achievements, supplement missing values, delete unnecessary columns, and modify attributes by computing function functions of the calculated variables. First of all, in SPSSstatistics, we use merge function to merge 2011-2014 grades of sports achievements, supplement missing values, delete unnecessary columns, and modify attributes by computing function functions of the calculated variables. Then the 0,1 process is carried out. IN the more than 4000 note, the number of the former 1/4-/3 is excellent for each achievement from high to low. IN order to get the height and weight fraction is greater than or equal to 90 is divided into good, vital capacity, 800 meters or 1000 meters fifty meters run score, score, score of standing long jump, sit and reach scores were greater than or equal to 75 for the best sit ups or chin up scores greater than or equal to 65 is excellent. An excellent 1, or 0, gets a wide table that is suitable for data mining. (Table 5)

Model establishment

Model building is an important stage in the process of data mining. (Figure 2) First, the new "sports association scores 11-14" flow in SPSSModeler, select the Statis-tics file source, and import the data, read data; secondly to define modeling field variables, choose "type" node is added. Because the concern is the relationship between performances, in the "role" in the definition of the "two", selected as the mean in the modeling of each variable in both conditions is the result. "Network node end graphics, operation can be associated network diagram of the line thickness represents the correlation degree. IN the final graph, the "network" node is selected, and the related network graphs can be run, and the degree of association of the line is expressed. IN the modeling, the Apriori node in the association is selected, the minimum

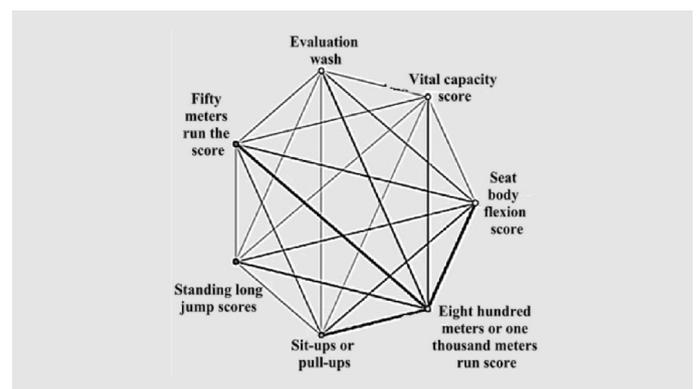


Figure 2. Sports Results Network Diagram.

condition support is 10, the minimum rule set confidence is 80%, and the association rules can be excavated with six rules: Sit ups or pull up outstanding achievement = height and weight, vital capacity; excellent achievement and sit and reach outstanding achievement of height and weight > = eight hundred meters; or one kilometer outstanding achievement of height and weight = >> =; standing long jump height and body weight of outstanding achievement; sit and reach top = > height weight excellence; sit and reach excellent, fifty meters height and weight = > outstanding excellence. It is found that there is a great relationship between the students' physical performance, and all the results are related to the height and weight.

DISCUSSIONS

Data mining technology is widely applied in the field of physical education, such as teaching evaluation, student induction learning, physical education curriculum setting, and teaching methods and so on. The set of points are as follows : (1) The data mining technology is applied to the evaluation of physical education. Through the application of data mining technology, we establish an effective evaluation system of physical education teaching, analyze teaching evaluation, find out the shortcomings in physical education, so as to change teaching plan and improve teaching quality. In the evaluation of physical education, students score the grade of the completion of the teaching task of the PE teachers, evaluate the teaching effect of the PE teachers and put forward the teaching opinions. (2) The application of data mining in students' inductive learning. Data mining technology is applied to the analysis and processing of students' knowledge points, which can

provide scientific guidance for students in the new course of physical education (3). The application of excavating in the setting of physical education curriculum. In the course of physical education, all the courses are arranged in a certain order, and the difficulty and difficulty of learning are in accordance with the difficulty of getting to the difficulties. (4) the application of data mining technology in sports teaching methods. By effectively recording the teaching methods of physical education teachers and combining the students' academic achievements, one can analyze the relationship between these two.

CONCLUSIONS

With the rapid development of Internet technology, data mining technology has been widely used in Colleges and Universities. At the same time, the rise of intelligent campus makes college students' various campus behaviors begin to store in database or data warehouse in the form of data. As the main force of the construction of the motherland, the physical quality of the college students will directly affect the process of China's modernization. In this paper, the application of data mining technology in the management of sports performance is studied and analyzed. Through association rule algorithm, the original data in education system can be transformed into useful information, and the relationship between achievements can be obtained, which provides decision-making and guidance for the reform of physical education and the improvement of students' physique level.

The author declare no potential conflict of interest related to this article

AUTHORS' CONTRIBUTIONS: This work execution, writing of the article and the critical review of its knowledge content are completed by the author. This paper can be used as the final draft of the manuscript. XS: writing and execution.

REFERENCES

1. Natek S, Zwilling M. Student data mining solution—knowledge management system related to higher education institutions. *Expert systems with applications*. 2014;41(14):6400-7.
2. Chiang WY. To mine association rules of customer values via a data mining procedure with improved model: An empirical case study. *Expert Systems with Applications*. 2011;38(3):1716-22.
3. Ramesh V, Parkavi P, Ramak K. Predicting student performance: a statistical and data mining approach. *International journal of computer applications*. 2013;63(8):35-9.
4. Schumaker RP, Solieman OK, Chen H. Sports knowledge management and data mining. *Annual review of information science and technology*. 2010;44(1):115-57.
5. Kabakchieva D. Student performance prediction by using data mining classification algorithms. *International Journal of Computer Science and Management Research*. 2012;1(4):686-90.
6. Zarsky TZ. Mine your own business: making the case for the implications of the data mining of personal information in the forum of public opinion. *Yale JL & Tech*. 2002;5:1-56. <http://hdl.handle.net/20.500.13051/7839>
7. Tang H Z, Peng J C. Research on Synthetic and Quantificated Appraisal Index of Power Quality Based on Fuzzy Theory. *Power System Technology*, 2003; 27 (12): 85–88.