



Data Mining Algorithm in Volleyball Match Technical and Tactical Analysis

Yongfen Yang^(✉)

Kunming University, Kunming 650217, China

Abstract. This paper analyzes the difficulties in the application of data mining algorithm in the technical and tactical analysis of sports competition, and puts forward a scheme for mining the key factors of winning volleyball matches. By applying the data mining algorithm based on Markov process and calculating the system reliability difference, the problem of finding the key action conversion process in volleyball competition is solved. In order to solve the problem of data acquisition speed, this paper proposes a method to solve the problem of data acquisition in real-time by setting the threshold of data acquisition. The design experiment shows the correctness and feasibility of the scheme.

Keywords: Data mining · Markov process · Data preprocessing · Technical and tactical analysis

1 Introduction

In the volleyball match, the result of the analysis of on-the-spot technical and tactical data is an important basis for coaches to carry out strategic deployment. Therefore, the collection, statistics and analysis of on-the-spot technical and tactical data have become the key to affect the coach's decision-making. However, it is difficult to complete the statistical work of data by hand. Due to the speed restriction, it is impossible to record a large number of, comprehensive and accurate data, and it is impossible to conduct scientific and systematic analysis of the collected data in a limited time, which can not meet the data requirements for scientific guidance. Coaches can only rely on experience and extraordinary experience.

1) Data collection. Because volleyball is to change the score by turns, and in the process of each round, there are 12 players in two teams at the same time. Volleyball requires high speed, which makes the recorder have to record the situation on the field in a very short period of time, including a series of data such as player number, ball landing area, technology and tactics, which requires high real-time performance.

2) Data analysis. Because there are many tactical changes and contingency in volleyball, it is necessary to prevent the possibility of making wrong analysis results based on a small amount of data. There is a high demand for the ability of systematic analysis of data. It is more and more necessary to apply the data mining algorithm to the technical and tactical analysis of volleyball matches.

In this paper, we propose a method to analyze the key factors of winning volleyball match by using data mining algorithm based on Markov process. In view of the above two application difficulties, the solutions to meet the real-time requirements in the data acquisition process and the solutions to eliminate ambiguity in the data preprocessing process are given, and the correctness and feasibility of the solutions are demonstrated through experiments.

2 Principle of Data Mining Algorithm Based on Markov Process

The idea of data mining algorithm based on Markov process is: the mining object is regarded as a system composed of multiple states, and the transition process between states conforms to the semi Markov process. The state transition probability matrix is obtained by statistics [1]. The system reliability is calculated by using the matrix, and then the increment is set to calculate the system reliability difference. Thus, the sensitivity of reliability to transition rate is analyzed. The method to calculate the system reliability is: let Q_{ij} be the transition probability from state I to state state, and let $C1n$ be the system reliability from initial state to state state, then we can calculate it by the following equation.

$$\begin{bmatrix} C1n \\ C2n \\ \vdots \\ C3n \end{bmatrix} = \begin{bmatrix} Q11 & Q12 & \dots & Q1n \\ Q21 & Q22 & \dots & Q2n \\ \vdots & \vdots & \vdots & \vdots \\ Qn1 & QN & \dots & Qnn \end{bmatrix} \begin{bmatrix} C1n \\ C2n \\ \vdots \\ Cnn \end{bmatrix} \tag{1}$$

The system reliability from the initial state to the successful state is obtained. The calculation method of system reliability difference is: after calculating the system reliability, each item in the state transition probability matrix is added with a small increment, and then the system reliability is recalculated by using the added state transition probability matrix. The difference between the former and the latter is the system reliability difference, and the larger the difference is, It indicates that the value changed in the state transition probability has greater influence on the system reliability.

3 Application Research

When the algorithm is applied in volleyball match, the series of actions from serve to score and their conversion relationship can be regarded as a system, and each action is only related to the previous action. Each action is regarded as a state, and the transition between actions is regarded as a state transition. According to the principle of the algorithm, the change of the state transition rate can affect the change of the system reliability. Therefore, the sensitivity analysis of the system reliability to the state transition rate between each pair of actions can be used to determine whether the state transition rate of each group of actions has a significant impact on the final result, According to the data mining algorithm based on Markov process and the characteristics of volleyball match, the scheme of applying the algorithm in volleyball match is made.

3.1 Data Acquisition

In order to collect the data needed to realize the data mining algorithm, it is necessary to record the execution process of each technical action in volleyball match. Because there are many changes in the athletes' technical and tactical movements in the competition, and all kinds of actions are completed in an instant, so it is challenging to record the process of the competition. In order to solve this problem, a script description language based on process is designed. This language uses mnemonic method to code the basic technical movements in volleyball match [2]. However, the design of this script description language requires the recorder to record a series of information including team member number, technical action, technical type, start area and end area, which makes the recorder work a lot.

In order to collect competition data more quickly and accurately, it is necessary to design a pattern based script description language. The so-called script description language based on pattern refers to: since most of the tactics used by both sides of the game have certain patterns in high-level volleyball matches, these patterns can be used as some formulas in the process of script description language design, and only a small amount of information about this mode is needed in the user's recording of the game process, Other information can be inferred from the tactical rules of volleyball matches.

In terms of acquisition mode, the main advantage of keyboard acquisition mode is faster acquisition speed, and the disadvantage is that the accuracy is slightly lower than that of mouse acquisition. Mouse acquisition has advantages in acquisition accuracy, but due to the limited collection interface, it can not cover all scripts, so it has certain limitations. In this scheme, the combination of mouse acquisition and keyboard acquisition is used to improve the collection efficiency.

- 1) Search the script data table to count the most frequently used scripts and their usage frequency.
- 2) Script sorting.
- 3) Output frequent scripts to acquisition interface.
- 4) For frequent scripts, the mouse collection method is used; for non frequent scripts, keyboard acquisition method is used.

3.2 Data Preprocessing

Data preprocessing can improve the quality of data and make the process of data mining more effective and easier. In this scheme, the necessary data preprocessing includes data classification and data ambiguity elimination.

4 Data Classification

Because the script description language based on pattern is designed for technical and tactical information, it is necessary to classify and integrate the collected data and integrate the data in different data tables to build a new small data warehouse. The separation of data warehouse and operational database is due to the different structure, content and

usage of data in these two systems [3]. Decision support needs historical data, but the operational database generally does not maintain historical data. In this case, although the data in the operation database is rich, it is often insufficient for decision-making. For different data types, such as offensive and defensive processes, different data warehouses will be generated. The data classification process is shown in Fig. 1.

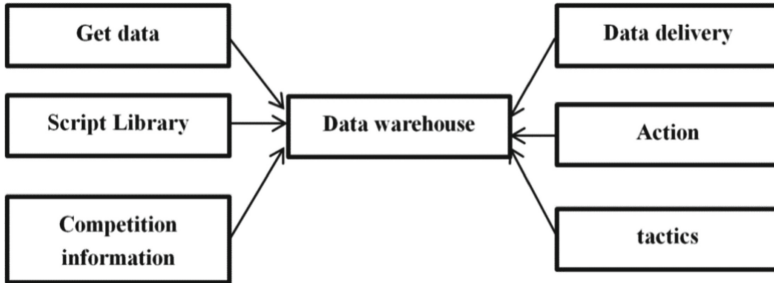


Fig. 1. Data classification process.

5 Simulation for Solutions

When applying the data mining algorithm based on Markov process to volleyball match, we must solve the problem of ambiguity in the process of action conversion. The origin of this problem is decided by the rules of volleyball match. In volleyball match, at most three consecutive actions can be carried out in each half court, which leads to the ambiguity in describing the process of action conversion.

In the process of preprocessing the collected data, the solution is as follows:

- 1) Set a threshold. The method of threshold setting is: due to the characteristics of volleyball match, the frequency of some action conversion process is very large, and some action conversion process only appears several times in the whole game. According to the collected state transition rate of action combination, take such a value, which can divide the numerical combination of state transition rate into two intervals [4]. The interval of state transition rate containing larger value is interval a, and the interval containing smaller value is interval B, which is close to the maximum critical value of B interval [5]. Figure 2 shows the solution simulated by threshold method.
- 2) For the action combination (a interval element) whose state transition rate is greater than the threshold value, the original practical significance is maintained and the reasonable action conversion is carried out. Figure 3 shows the use of different components to simulate the solution.
- 3) The action combination (b-interval element) whose state transition rate is less than or equal to the threshold value is ignored, or its value is incorporated into the action combination with similar actual meaning.

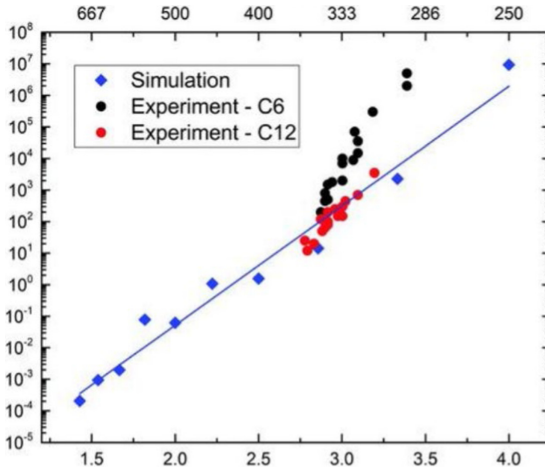


Fig. 2. Using threshold to simulate the solutions.

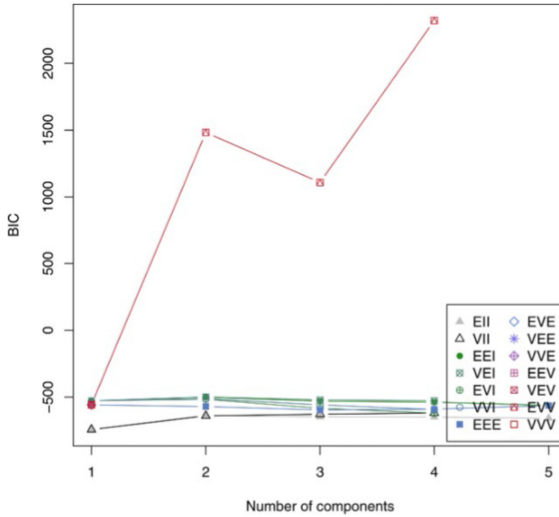


Fig. 3. Using the different components to simulate solutions.

6 The Design and Implementation of the Statistical Analysis System for Volleyball Match

6.1 System Design Objectives

The goal of the system design is to realize a volleyball match on-the-spot technical and tactical statistical analysis system which integrates the functions of data acquisition, data statistics, data analysis, data management and video management. Through this software, coaches can record and analyze the game data by computer, so as to provide technical support for coaches' strategy making [6].

The detailed design objectives are as follows: to build a basic platform with the functions of collection, statistics, analysis and management of on-the-spot technical and tactical information of volleyball matches. It can realize the collection, statistics and analysis of technical and tactical data, as well as the management of team, player and game information. At the same time, video detection and video location technology are applied to realize the function of playing video according to script. Based on the in-depth analysis of the technical and tactical composition of volleyball match, the script description language suitable for describing the technical and tactical process of volleyball match is designed, and the parser of the language is written. Applying data mining algorithm to deeply mine a large number of data, this paper proposes a data mining algorithm which is more suitable for sports computing. The main algorithms to be applied are: applying Markov process principle to data mining, looking for the state transition process which has the greatest impact on the system reliability in the state from serving to scoring. The association rules algorithm is used to find the association information between two or more actions to help coaches make strategic deployment. The application of classification and prediction technology to athletes in the game of different technical movements and technical and tactical combination of classification, while establishing a prediction model, given a player or the whole team in the past game for a specific tactic of technical and tactical information, predict their future face a certain tactic of countermeasures [7, 8].

6.2 System Function Structure Design

The function of the system mainly includes four parts: data management, data acquisition, data analysis and processing, system management. In addition, it also involves the application of video positioning technology. Data management is mainly the management of team information and player information, data collection is mainly through the mouse collection and keyboard collection of two collection schemes to collect data, data analysis and processing is through the massive data mining, so as to get some coaches interested in information, including the function of data statistics and analysis. Video management mainly involves some content of video positioning [9].

7 Research on Data Acquisition Solution

7.1 Design Idea of Script Description Language

The basic idea of script description language is: most of the techniques and tactics used by both sides in volleyball matches have the characteristics of fixed patterns, which are represented by non ambiguous codes. At the same time, in the process of recording the game, a series of description methods and reasoning rules are customized. You need to input as few codes as possible, and use the logic reasoning realized by computer to supplement the remaining codes and infer Hide the information, so that users can get as much information as possible [10].

According to the basic design idea of script description language, the basic design principles of script description language are summarized as follows: intelligent reasoning - through the analysis of volleyball skills and tactics, summed up as many patterns and rules as possible, users use script description language to describe the game data, input as little data as possible, get as much information as possible, Effectively shorten the length of a large number of code. Simplified memory - in order to make users with different professional backgrounds, different age groups and different cultural levels be proficient in the use of script description language as soon as possible, Pinyin abbreviations are used to record competitions. Pinyin, as one of the bases of Chinese language knowledge, is more suitable for Chinese memory habits and can meet the needs of users at different levels. Key record - in the process of script description language recording, due to the fast pace of volleyball match and the ever-changing situation, it is unrealistic to require users to record all data in real time and accurately in a short period of time, which must be recorded selectively. The script description language designed in this paper supports users to record the most interesting and final data first, Select the key process with the most reference value and analysis value to record [11].

7.2 Code Design of Script Description Language

The script description language uses Chinese Pinyin as the description carrier, which is convenient for users of different ages and cultural backgrounds [12]. In the process of designing the code, we mainly follow the following principles:

The first is convenient memory. For the choice of characters, the first letter of Pinyin is generally chosen, such as “F” for serve, “g” for attack and “1” for block. The advantage of this choice is that as long as people can say the word, they can immediately associate with the code that represents the phrase, which provides great convenience for memory. There is also a problem in doing so, that is, the ambiguity of letters. For example, the first letter of two words is the same character. In the process of code design, in this case, two different solutions will be adopted according to the specific needs: one is to use another character to represent one of the words, which is not the first letter of the word, but has other connections with the word, which can also facilitate the memory. The second method is that words with different meanings can be described by the same symbols if they are not in the same category. This ambiguity can be overcome in the process of script parsing through the design of script parser. In addition, the design of script parser supports the case discrimination of the same character, so it expands the availability of characters [13].

The second is convenient collection. Because the process of collecting data is mainly completed through the keyboard, so in the process of selecting characters, this topic tries to choose those keys that users can easily touch, or only need to press a key to record characters. The code design of script description language is shown in Fig. 4.

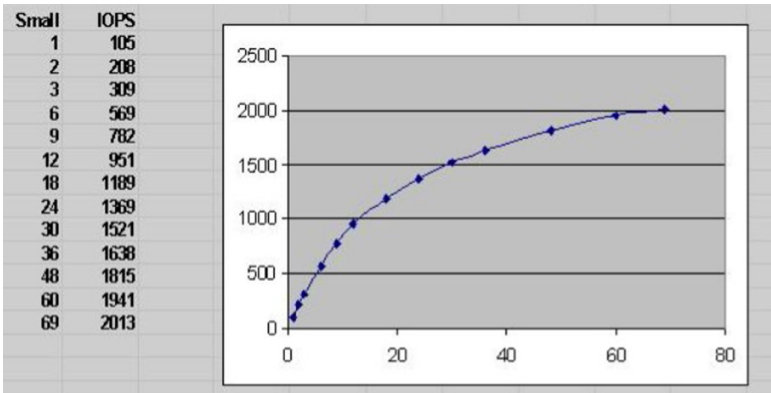


Fig. 4. Code design of script description language

7.3 Working Steps of Script Parser

- 1) Field split, split based on double spaces.
- 2) Judge whether it is valid one by one (single code, mode code or invalid code), add the valid code and invalid code into the code table on the right, and the invalid code is highlighted to prompt modification.
- 3) For the valid code, analyze it one by one. The correct code before parsing is entered into the code table, and the parsed code is entered into the parsing table.
- 4) If the field to be parsed is a single code (it is already a valid code at this time): judge whether there is an action. If the action is empty, the default action is serve. In the single code, only the action must exist, and the default area information is supplemented according to the action. Inference of default information. Data storage.
- 5) If the field to be parsed is a pattern code (it is already a valid code at this time), for the pattern field split into several segments, the analysis starts from the last segment. First, the mode number and action field are put into the back section of the analysis mode. If there is a mode number, the technology type is put into the database. If there is a mode effect, the action effect of several records is put into the database. In the front part of the analysis mode, if there is no action, the default is spike action. Analyze the non pattern fields from the back to the front in turn. If there are repeated actions, the first one analyzed (that is, the one close to the pattern field) shall prevail. That is, if the content storage unit to be entered is not empty, the current content to be entered will be abandoned [14].

Script translation: when the currently used scripts are no longer suitable for the needs of users and need to change the script format or the characters contained in the script, it is inconvenient for users to analyze the data described by the old scripts because a large number of previous scripts still exist in the database [15]. In order to unify the old and new scripts, the system provides a script editor, which can make the old scripts in the database be translated into new scripts. In the process of translation, the script structure design is involved. In order to meet the different needs of users, the script editor provides a new script syntax customization module. The grammar customization

module provides six fields for editing, namely: team member number, team member action, technical details, effect, start area and end area. In addition, at the end of the six fields, you can add any separator character. These elements can be freely defined by the user. After the new grammar rules are successfully customized, the script translation can be carried out. In the process of script translation, we should first find out the list of scripts to be translated, and then start the translation. The new script is written into the database instead of the old script.

8 Conclusions

This paper studies the difficulties of data mining algorithm applied to sports calculation, and puts forward a scheme of applying data mining algorithm based on Markov process to analyze volleyball match techniques and tactics. In the process of data collection, the scheme improves the collection efficiency by searching for frequent scripts. In the process of data preprocessing, a method of setting threshold is proposed to make the data meet the application conditions of the algorithm. The experimental results show that the scheme can dig out the data results that can not be obtained by visual observation and simple statistics, but have important significance for coaches to guide the game, and can mine the key action conversion process. The scheme has a certain role in mining the technical and tactical information of volleyball matches. The next step is to apply the algorithm to the analysis of key area transfer process and key tactics transfer process in volleyball match, as well as the application of association rules and classification prediction algorithm to the analysis of volleyball match technical and tactical information.

References

1. Yujia, Z., Zhao, H., Jiewei, W.: Research on the application of data mining algorithm in volleyball match technical and tactical analysis. *Comput. Appl.* **12**, 3027–3029 (2006)
2. Hanm, K.M.: *Data Mining Concepts and Technologies 1*. China Machine Press, Beijing (2005)
3. Editorial board of Encyclopedia of safety science and technology. *Encyclopedia of safety science and technology*, China Labor and social security press, Beijing (2003)
4. Rong, J., Ben, H.: Application of association rules in the analysis of volleyball skills and tactics. *Glamour China* **28**, 133–134 (2008)
5. dataproject[EB/OL]. <http://www.dataproject.com>
6. Zhang, J.: The application of data in sports news report. *Chin. Journalist* (2003)
7. Jiawei, H., Micheline, K.: *Data Mining Concept and Technology*. Mechanical Industry Press (2005)
8. Olivia Parr RUD, data mining practice, China Machine Press (2003)
9. Orlando, F.: Data mining and knowledge discovery: theory, tools, and technology. *Proc. SPIE.* **11**, 259-264 (2000)
10. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. *Proc of Very Large Data Bases* (1993)
11. ZPawlak *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Boston (1991)
12. Lianjie, S.: *Encyclopedia of Safety Science and Technology*. China Labor and Social Security Press, Beijing (2003)

13. Zhao, H., Wang, G., Gao, Y.: Abstract model of software architecture. *J. Comput. Sci.* (2002)
14. Zhao, H., Sun, J., Wang, G., Gao, Y.: Component based software reliability model. *Minicomputer Syst.* 730–736 (2002)
15. Langley, P.: *Elements of Machine Learning*. Morgan Kaufman, San Francisco (1996)