

Design and Application of Basketball Tactics Analysis Based on the Database and Data Mining Technology

Tian Li-gang
 Sports department
 Hebei Agricultural University
 HeBei Baoding, China
 wangfushun@yahoo.cn

Wang Fu-shun
 College of Information Science & Technology
 Hebei Agricultural University
 HeBei Baoding, China
 wangfushun@yahoo.cn

Abstract—Database and data mining technology was applied in basketball tactics analysis , after pretreatment to the database, used classification and association rules of data mining technology, applied Apriori algorithm , solved the less stored data, statistical method crude, the single analysis results and so on. In the training and competition, digging out key technology information of each player's movement in basketball match, and providing important guidance information for players and coaches.

Keywords- basketball tactics, database, data mining

I. PREFACE

In the era of rapid development of computer technology, the basketball game informatization level gradually improved, the basketball game technical movement data also appeared a rapid growth, and in these large amounts of data set, we only care about and need a small part. In order to timely control the game, and take effective management, it is imperative to analyze motion with advanced computer technology. So, in such a data source to dig, extract interest and potential useful patterns and hidden information in basketball game action of technology, data mining technology is a new application project^[1].

The basketball tactics application used database and data mining technology, solved the less stored data, statistical method crude, the single analysis results and so on. Realized the basketball technology action for mining, finding useful information to provide coaches and players, helping the team for targeted training, improving techniques action and the competitive level, and digging out technical movement of strong relevance in the data, the analytical results have a high guidance, and contribute to the development of basketball.

II. ESTABLISHMENT OF DATABASE SYSTEM

The analysis system of basketball tactics based on the system of basketball game information acquisition, analyzing the collected basketball game data^[2]. The collection system of basketball game information and analysis system use the same database. Each one of these systems include all kinds of information table .The basketball game information acquisition system including the team information table, the information table, match type information sheet, competition information sheet, game script information sheet, and technical movement

information table, the action will happen area table, action effect table. Basketball tactics analysis system table including were script affairs table, frequent item sets table, frequent item sets separation table and association rules table.

The database layer provide the data source for this system; Data extraction and conversion layer mainly extract basketball game in the database data, and preprocess data, to provide mining data for mining algorithm layer; Algorithm encapsulation layer on the pretreatment of data mining, produce useful information; The user calls layer according to the needs of users, call mining tools in the specific mining algorithm, produce the information of user needs; The layer of visual display displayed useful information after mining through the visual graphics .Figure 1 is a basketball tactics analysis system structure.

III. BASKETBALL TECHNOLOGY ANALYSIS OF DATA MINING IN DESIGN AND APPLICATION

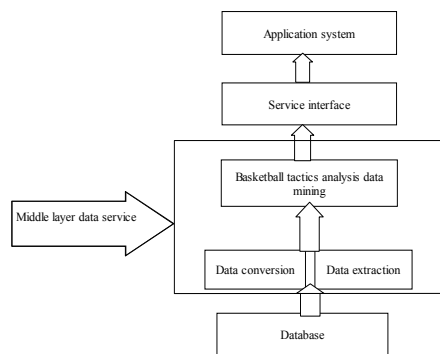


Figure1. Basketball tactics analysis system structure.

Data Mining is the process of find new useful association, pattern and trend from a large number of stored data, another words , it is a process to extract information of implicit and people didn't know, but it is potentially useful information and knowledge, from a large number of incomplete, noisy, and fuzzy, random data. The information is great value through the data mining extracted from a large number of data, this information hidden in the data can help people make a scientific and reasonable decision. Data mining is the process of multiple steps connected each other and repeated the process of human-computer interaction. It used of the data pretreatment technology, data

mining algorithm and its application.

A. Data mining technology of pretreatment

Some of the data mining algorithms compared to the processing of the data set usually have certain requirements, such as the better data integrity, less data redundancy, the small correlation among attribute. However, the actual system of data generally rarely can meet the requirements of the data mining algorithm. Besides, a large number of data more meaningless composition, seriously affected the efficiency of the mining algorithm, and because the noise will cause invalid induction. In data mining system data preprocessing has become one of the key issues in the implementation process.

Data mining in data preprocessing mainly to accept and understand the user's found requirements, make sure find the task, extraction and find the source of task related knowledge, according to the background knowledge of the mandatory rules for data inspection, through the clear ,induction and so on, generate target data for mining the core algorithm using, namely knowledge base. Realized extract script data meet basketball script design from the data.

B. Some of the commonly technology used in data mining

Some of the commonly technology used in data mining including: classification used for extract model of describing important data, the model also called classifier. The model can give the analysis of the data to one of the given category;cluster,the physical or abstract collections of objects group become the similar object composed of more than one set, clustering is generally operate a few data item at the same time; Mining of association rules found interesting correlation or relationship in the data between item sets.

1) Application of classification in data mining

Classification is the use of the training sample set through the relevant algorithm and get the result^[3]. Data classification operation usually have the following two steps: the first step, according to the given training set, find the appropriate mapping function $H: F(X) \rightarrow C$ to describe the model, this phase is often referred to as the stage of training model. The second step, using the model of previous step created to predicted data category, Or use the function model of data concentration of each category description, form classification rule.

A kind of algorithm in the process of classification is using decision tree to classify, decision tree is a similar to the flow chart of the number structure, each an internal node said in an attribute in test, each branch represents a test output, and each leaf node on behalf of the class or kind of distribution. The top of the tree is the root node. As shown in figure 2 decision tree, it says concept "score", it predicts what technical movement may lead to score, rectangular represents internal node, and elliptic represent leaf node^[4].

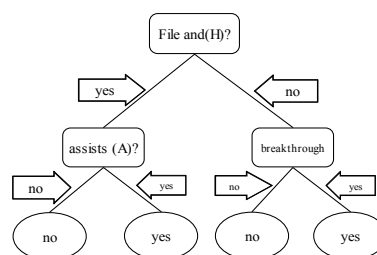


Figure2. Decision tree

The above points out that what kind of technical action brings 3 points score, the internal node (not leaf) of the decision tree saying an attribute in test, each leaf node representing a class (3 points score = yes or 3 points score = no). When decide the structure of the tree, many branches may reflect the training data concentration of noise or isolated point. Tree pruning tried to cut off the branch, in order to improve the accuracy of the unknown data classification. Decision tree has been widely used from medical treatment to bo different theory and business fields. The decision tree is the foundation of business rule induction system.

Decision tree induction of the basic algorithm is the greedy algorithm. It is recursive crush one by one to construct decision tree with top-down.

2) Association rules analysis technology and its algorithm

Basketball technique features analysis adopted the method of association rules analysis. This function mainly includes two parts, the association rules analysis and association rules back. The association rules analysis used the Apriori algorithm, then through the application class of CAssociationRulesDlg to reslize. The paper introduces the Apriori algorithm. It is a kind of algorithm using the candidate itemsets to find frequent itemsets, and it is one of the most influential algorithm. The mining process is mainly divided into two stages: the first stage is to find out all the technical movement to meet the value of minimum support, in other words, the data quantity of the basketball affairs database containing must be greater than or equal to the specified minimum support degree. Then, these meet the minimum support degree is called a frequent itemsets. If a term set contains k technology actions, called k – itemsets, and if a k - itemsets meet minimum support degree, then it is called frequently k - itemsets. The second stage is to calculate with all the association rules according to the former stage finding frequent itemsets and minimum confidence as the condition. For example if ABC is frequent itemsets 3 -, A,B,C including D, if the association rules AB to C meet minimum confidence, then the association rules is established.

In the realization of the system the algorithm will be realized and finish the appointed analysis function. The application including generate frequent item sets application and association rule generation application. In the process of generated association rules, it needs to

operate the database table, therefore it also includes the database algorithm provides the interface^[5].

VI. THE APPLICATION OF DARA PROCESSING AND ANALYSIS OF THE RESULTS

Based on the description of Apriori algorithm in data mining of association rules technology, in order to further understand, there is an example to illustrate how to use the association rules dig the most common action.

A. The example of association rules

Now there has a basketball game transaction data as is shown in table 1, $D = \{A, B, C, D, E, F, G, H, I\}$, A representative technical movement is assists, B is 2 points ball score, C is rebounds, D is steals, E is 3 points ball score, F is free throws, G is A breakthrough, H is pick-and-roll, I was blocked shots. $T = \{T1, T2, T3, T4, T5, T6, T7, T8, T9, T10\}$, assume that the minimum support degree is 30% and above (i.e. number of minimum support for three times the technical action), minimum confidence is 40%.

TABLE1. THE BASKETBALL GAME TRANSACTION DATA

Game material Numbers	The technical action
T1	A,B,G,H
T2	A,B,C,E
T3	A,B,C,D,F
T4	C,G,H
T5	A,B,C,G
T6	C,D,H
T7	B,E,F
T8	B,C,F,I
T9	C,H
T10	B

Mining frequent itemsets process is as follows:

- a) Each item is the set of candidate itemsets C1 members, scanning all affairs, Count the number of times each item appears;
- b) Based on minimum support $\min_sup = 3/10 = 30\%$ (that is, count 3), can determine the frequent 1 - itemsets L1;
- c) To find frequent 2 - itemsets, the use of connection $L1 \infty L1$ to generate candidate2 - itemsets C2, It taken by L1 in any combination of two items, total $C26 = 15$ 2 - item;
- d) Scanning the transactions in the D, calculation of C2 each candidate item support number;
- e) Determine the frequent 2 - itemsets L2, it consists of item meet the minimum support degree candidate 2 - a centralized;
- f) Through the L2 own connection to generate candidate itemsets 3 - C3, according to the \min_sup to frequent cut 3 - itemsets L3;
- g) Use $L3 \infty L3$ produce candidate 4 - itemsets, found that is empty. End the algorithm;
- h) Finally, find out the final validation of frequent itemsets meet 3 - not less than the minimum confidence $\min_conf = 40\%$.

Use the same method to deal with all possible formation of association rules, can guide the players and the coach to practice key technical action.

B. The analysis of experimental results

From the figure can get the following informations:(Set the minimum support degree is 30%, 40% minimum confidence. Note: from the graph 1, 2, 3 rule see, have assists and penalty must be points into account (100% confidence))

Associator output	Associator output	Associator output
run information	minimum support:0.3	B=yes F=yes 3
scheme:associations -I -N	minimum metric-confidence>=0.4	C=yes H=yes 3
relation:篮球	number of cycles performed:14	size of set of large itemsets:L(3):
instance:10	generated sets of large itemsets:	large itemsets L(3):
attributes:10	size of set of large itemsets:L(1):6	A=yes B=yes C=yes 3
7id	large itemsets L(1):	Best rules found:
A	A=yes 4	1.A=yes 4=>B=yes 4 conf(1)
B	B=yes 7	2.A=yes C=yes 3=>B=yes 3 conf(1)
C	C=yes 7	3.F=yes 3=>B=yes 3conf(1)
D	F=yes 3	4.A=yes 4=>B=yes C=yes 3 conf(0.75)
E	G=yes 3	5.A=yes B=yes 4=>C=yes 3 conf(0.75)
F	K=yes 4	6.B=yes C=yes 4=>C=yes 3 conf(0.75)
G	size of set of large itemsets:L(2):	7.H=yes 4=>C=yes 3 conf(0.75)
H	large itemsets L(2):	8.A=yes 4=>C=yes 3 conf(0.75)
I	A=yes B=yes 4	
Associator model	A=yes C=yes 3	

Figure3. The result of mining data

Frequent itemsets 1 - L1: A, B, C, F, G, H; the back of the digital to support the number or call frequency.

Frequent itemsets 2- L2: AB, AC, BC, BF, CH;

Frequent itemsets 3- L3: ABC;

The figure generated rules. Such as A to BC means A this technical movement contains action of B and C ,from figure 3 of article 4 association rules that support the number for the 3, meet the minimum support degree 30%, at the same time confidence for $3/4 = 75\%$, also meet minimum confidence 45%, this suggests that assists can bring rebounds and score of great master (confidence 75%).

Thus, from the mining process can see that in the transaction database the frequent technical movement is A (assist), B (2 points score), C (rebounds) and their combination, they will meet the threshold \min_sup and \min_conf . These combinations such as A to BC said assists to at least 30% of the support degree and 40% confidence to bring two or rebounds, by the same way can understand other combination.

For large amount of data, use the analysis of the mining algorithm, can also dig out many association rules. From the above rule 1, no assists will not have three goals, 95% confidence; Rule 2 showed no assists nor blocks will be 94% confidence without three goals. From it can see how important it is to assist, still can find other useful rules. The useful rules can be used to guide the team training, as in the breakthrough points of the ball to three points of the line no defender, let them throw three points.

VII. CONCLUSION

Along with the sports competition intense strengthen and increasing of computer technology application, how to carry out targeted technical movement training, become the players and the coach must face and solve the problem. Through the data mining and the application of database, to understand each player on the court action better. Dig out

the key technology information of each player's movement in basketball match, and discover of the advantages and disadvantages of players better. The coach can obtain players' information better, then distribute their position reasonably on the pitch as well as their own duty, at the same time can make the players understand themselves better.

REFERENCES

- [1] DataProject[EB/OL].[http:// www.DataProject.com](http://www.DataProject.com).
- [2] Savasere A, Omiecinski E, Nvaathe S. An efficient algorithm for mining association rules in large data bases[A]. Umeshwar D, Peter MDG, Shojior N. VLDB'95, Proceedings of 21st International Conference Very Large Data Bases[C]. Zurich, Switzerland: Morgan Kaufmann, 1995: 432-443.
- [3] B. Liu, R. Grossman, and Y. Zhai. Mining data records in web pages. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003.
- [4] Agrawal R., Imielinski T., Swami A., Database mining: a performance perspective. IEEE Transaction on Knowledge and Data Engineering, Dec. 1993, 5(6): 914~925.
- [5] Heikki Mannila. Data mining: machine learning, International conference on Scientific and Statistical Database Management, Stockholm, June 18-20 1991. 1~8.