Ji, W., & Lv, T. (2022). Research on technical analysis of basketball match based on data mining. In Li, E.Y. *et al.* (Eds.) *Proceedings of The International Conference on Electronic Business, Volume 22* (pp. 674-679). ICEB'22, Bangkok, Thailand, October 13-17, 2022

Research on technical analysis of basketball match based on data mining

(Work-in-Progress)

Wen Ji [1]
Tao Lv [2,*]

*Corresponding author
[1] Lecturers, Hohai University, Nanjing, Jiangsu, China, jiwen@hhu.edu.cn
[2] Associate Professor, Hohai University, Nanjing, Jiangsu, China, lvtao@hhu.edu.cn

ABSTRACT

The aim of this paper is to preprocess basketball technology actions, to classify these actionswith data mining technology, to mine association rules among them. The main works are shown below:The common approaches of data mining are discussed, such as preprocessing technology, classification technology, clustering technology and mining rules technology. Both ID3 decision tree classification algorithm association and Apriori association rules algorithm are studied in detail.The paper discusses basketball technology actionsboth on a small scale and a large scale, J48 decision tree classification and Apriori association rules mining algorithm basketball are applied, all these research results should have useful instruction to team.

Keywords: Data mining; Apriori algorithm; Basketball Skill.

INTRODUCTION

Basketball game is the use of basketball basic skills, according to a certain form of tactical organization, a process of offensive and defensive changes (Robertson, Back & Bartlett, 2016). In the game, the athlete is in the technical, tactical or position mobility performance for the complex, open, random and non-linear competitive ability to organize and game system, to show their resilience and technical level.

But it can also provide the basis for the coaches to instruct the tactics training and the on-the-spot technical and tactical contingency by analyzing the changing characteristics of the data (Shi, 2015). In actual combat, the active factor is the coaches should pay special attention to grasp the presentation of the performance of the spot when the game features, the use of tactics, staffing, time allocation and other parties have reference and guidance. Basketball statistics there are many, but how deep-level analysis of relatively messy, relatively complex statistical data, how to find these data changes, how to use these laws to solve some problems, targeted training and before and after the combination of tactics, Improve the ability of the game, basketball data analysis is an urgent need to solve the problem (Lazer, et al, 2014) .

Data mining technology is undoubtedly the rise of technical and tactical analysis to provide a powerful help (Marmarinos, et al., 2016). Data mining technology is different from the traditional data statistical methods (Stein, et al., 2017). Statistics is a science of how to collect, organize, analyze and interpret digital information in data. Statistics can be divided into two broad categories: descriptive statistics and inferential statistics (Wang, 2014). Description Statistics involves organizing, accumulating, and delineating information in data; inferential statistics involves using sampled data to infer the population. Data mining is a kind of practical application algorithm (mostly machine learning algorithm), which can solve the problems related to various fields by using the data of each field (Zhang,2014). Combining the frontier of informatics with the analysis of basketball technique and tactics, it can provide effective scientific basis for coaches and athletes in training methods inside and outside the field (Li, 2014).

THE VALUE OF DATA MINING IN BASKETBALL TECHNIQUE AND TACTICS ANALYSIS

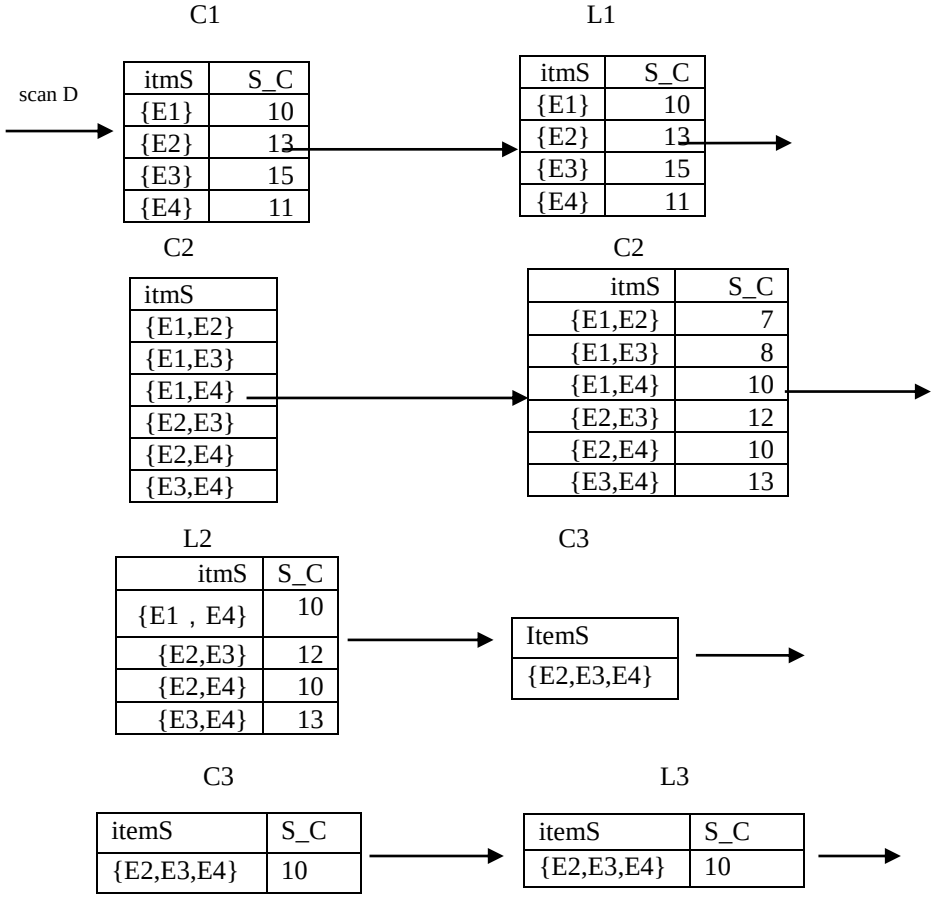
Data mining, as a discipline, is aimed at solving problems in various industries (Tashakkori et al, 2014). Different technologies and practices in different research fields are needed in the process of data mining. Data mining technology in has such a definition: from a large number of incomplete, noisy, fuzzy, random application of actual data, extract hidden in it, people do not know in advance, but it is Potentially useful information and knowledge on the process side (Martinez &Walton, 2014). It takes a global solution and applies it to the development of technical and tactical combinations to solve the same underlying problem.

For example, in a basketball game, there are breakthroughs in basketball players, shooting, passing, assists, break points, shooting, shooting, errors, but also including its shooting rate, free throw percentage, Ball hits, and so deep technical statistical analysis, as well as without the ball cover athletes with mobile, rebounds, mistakes, steals, blocks, effective defense, fouls and their nature, as long as the technical and tactical action can be defined, are The object of data mining. Through the computer software recording or video analysis, access to technical and tactical statistics, find the average and limit values, resulting in frequent itemsets and a number of association rules(Monroe, et al, 2013).

Basketball match between the players in the group and the pairing relationship between the pairing are always very complex and changeable. In the past five years, there are more and more interdisciplinary researches on the combination of physical education, sports training and data mining, such as the application of data mining algorithms in technical and tactical analysis of volleyball, table tennis and cricket (Shimizu, Louzada & Suzuki, 2014). The correlations between variables are determined by finding the correlation coefficient among the variables, and frequent itemset pattern knowledge is found from the given data set. Data mining is to analyze each data and find the regularity from a large amount of data. There are three steps (z): data preparation, regular search and regular expression. After obtaining the game data, how to do a good job of data mining depends on whether you can find the appropriate law.

ASSOCIATION RULE MINING BASED ON APRIORI ALGORITHM

Association rule mining is a way to find out the meaningful relationship between many data. In this study, the purpose of mining association rules is to find the implicit association in the database of basketball technology action. Apriori algorithm is the most widely used method of association rule mining. It is an algorithm to find frequent itemsets by using candidate itemsets. The following is based on the design of basketball scripting language on the basis of the analysis of the Apriori algorithm in the excavation of basketball action association rules in the design and application (Rehman & Saba,2014). Apriori algorithm is a typical algorithm to find frequent itemsets in transaction database. Frequent itemsets are the itemsets with support ≥ minimum support (Cheng et al, 2016). Achieve this goal, anti-corruption scanning is needed for the database of things. This step wastes more time, restricting the operation of Apriori algorithm. Apriori algorithm can be recursive way, the transaction database to find out all the frequent itemsets. Specific operation is the first thing in the data table as a candidate for each item set, with Cm to represent (m value can be 1，2，3……), then the basketball technical action database The item sets with support ≥ minimum support are set as the set of frequent 1-itemsets, which can be expressed as Lm (the value of m can be 1，2，3……), and so on, until L is empty, the algorithm stops (Brooks, Kerr & Guttag, 2016).



APPLICATION AND ANALYSIS OF DATA ACQUISITION AND PRETREATMENT OF BASKETBALL TECHNIQUE

Data Collection

Basketball is a high-profile, participatory sport. With the development of communication technology, satellite broadcast, TV broadcast and broadband network, can be far away thousands of miles of basketball game screen instantly spread to millions of households. People through the camera, broadband network to record the game scene, and then burned into a CD-ROM burner. In this paper, the original data is to live through the broadband network broadcast NBA (National Basketball Assciation) basketball game scene saved to the hard disk, but also the various sites on the game data is also copied down; and then the game video Burner burned into VCD discs; the last by repeated viewing the game for manual statistics, combined with the major sites of data collection.

NBA League regulations, a total of 48 minutes a match, divided into 4 sections, each 12 minutes, each round of the attack to be completed within 24 seconds, a match 100 to 200 rounds, if more than 24 seconds, then team will lose the ball. If the players have a good grasp of technical action, good coordination between the players, then it can be completed within 24 seconds, on the contrary, it will lose time because of shooting opportunities. Therefore, in this statistical analysis, we have an offensive round for the time period, statistics of a basketball game 5, 20, 100 round of the general technical action (such as Table 1).

Table1: statistics of general technical action

Round	Basketball technical action
R5	Rebounds, Pick-and-roll, Dribbling,Dribbling, Passing, Pick-and-roll, 2-points
R500	Dribbling, Dribbling, Assists, Pick-and-roll, 3-points
R100	Fast break, long pass, Technical foul, Steals, penalty

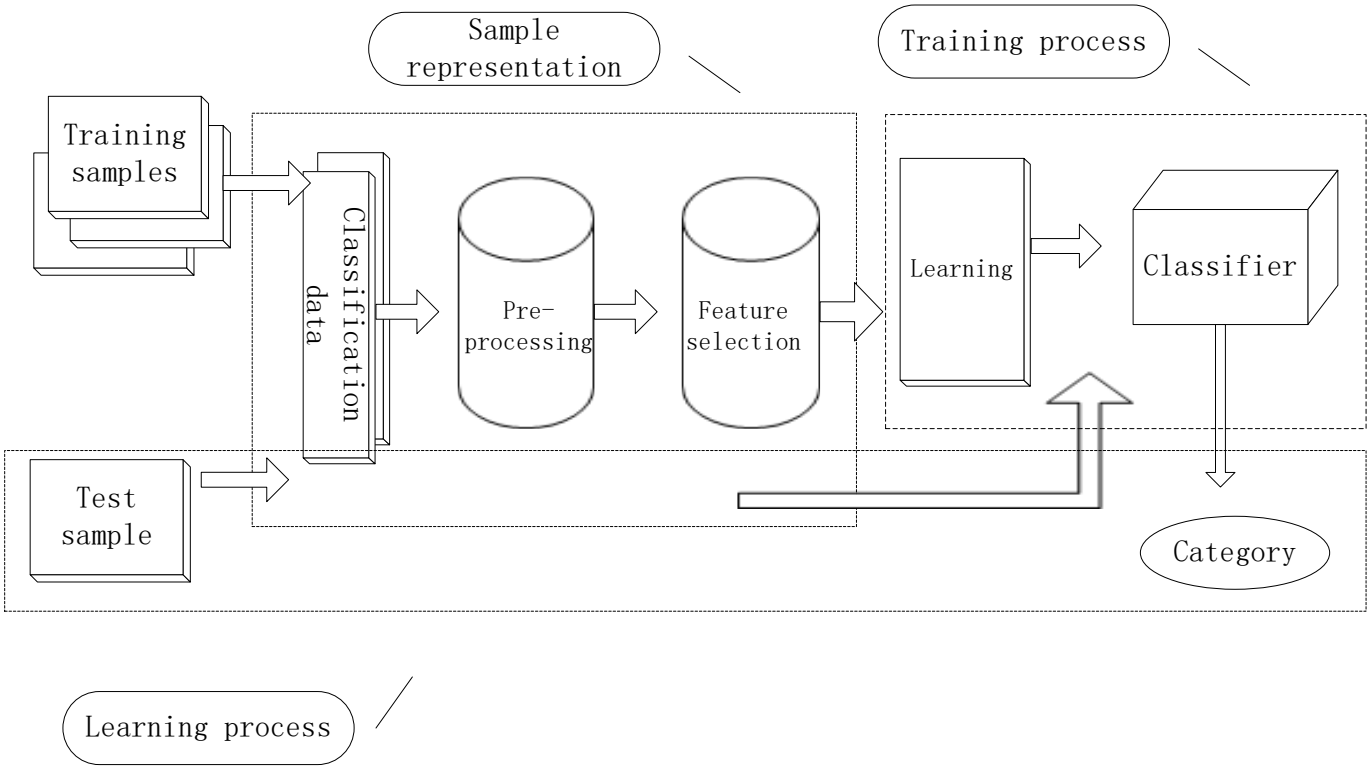


Figure2:The machine learning process of classification

Data Cleansing And Integration

Incomplete and inconsistent data can be used to fill the vacancy value, to correct data inconsistencies.

(1) Fill the vacancy value for no data items, according to the meaning of the data item, define a default value, and then use it to replace the missing vacancy value.

For example, T8 this round, "dribble" and "ball" is the same technical action, should be consistent, are corrected to "dribble", and so on.

(2) To correct inconsistencies in data collection Sometimes there are inconsistent data (what is the data inconsistency? Such as "dribbling" and "ball driving", both of them are the same technical action, they are inconsistent). There is a certain correlation between some data, A, B correlation can be measured by the following formula:

$$r_{A,B} = \frac{\sum\left(A-\bar{A}\right)\left(B-\bar{B}\right)}{(n-1(\delta_A\delta_B))}$$

where, $\bar{A} = \frac{\sum A}{n}$, $\bar{B} = \frac{\sum B}{n}$, $\delta_A = \sqrt{\frac{\sum\left(A-\bar{A}\right)^2}{n-1}}$, $\delta_B = \sqrt{\frac{\sum\left(B-\bar{B}\right)^2}{n-1}}$; r> 0, A and B are positively correlated; r = 0, A and B independent, irrelevant; r <0, A and B negative correlation. The correction method is:

If action = "ball forward"

Then action = "dribble"

If action = "2 points into the ball"
Then action - "2 points goal"

Data Integration

Multiple data sets are stored together in a consistent data store. There are data on the network and the data manually recorded by the video, now to be combined storage, storage before the merger with a technical action data. There are several "dribbling" actions, where they can be combined into one. The same "pick and roll" is also appeared several times, also merged into one. See Table 2.

Table2:NBA Technology Actions Statistics

Tid	Basketball technical action
T8	Rebounds, ?, Dribbling,?, Passing, ?, ?, 2-points
T15	?, Passing, ?, Assists, ?, ?, 3-points
T102	?, Steals, ?, Long Pass, ?, Fast break, ?

Data Reduction

Complex data analysis and mining on large amounts of data will take a long time to make this analysis impractical and unnecessary. What is the data reduction?

Heap reduction redundancy deleted the value. 9 technical moves, including: assists, 2 points, rebound, steal, and other steals in the game, such as "dribbling", "passing" 3 points, 3 free throws, cloak, break and block o Data compression using nine letters A, B, C, D, E, F , G, H, I will be the 9 kinds of technical actions to encode, compress the data set. A represents assists, B represents 2 points, C represents rebounds, D stands for steals, E for 3 points, F for free throws, , G for pick and roll (Clk), H for break (Brk) and I for cap (Blk).

In this study, we use WEKA mining software, in order to more convenient basketball data mining, the need for the relevant format conversion, based on the previous step, one after the corresponding if there is action, then use y to represent, if No action, then use n to represent, see table 3.

Table3: Basketball technology action Judgment

Round	Basketball technology action
R5	y, n, n, n, n, y, y, n
R20	n, n, y, n, n, y, n, n, y
R100	n, n, n, n, n, n, y, n, y

Finally, the data storage file format to be converted to WEKA mining software identified (Attribute Format File) format to save. This only needs to paste the statistical table to WordPad, and then add arff file required attribute header @ attribute, and the file header with @ relation, data with @ data. As shown below:

```
@relation Basketball technology
@attribute Tid{ T8, T15, T102, ...}
@attribute A{ yes, no}
@attribute B{ yes, no}
@attribute C{ yes, no}
@attribute D{ yes, no}
@attribute E{ yes, no}
@attribute F{ yes, no}
@attribute G{ yes, no}
@attribute H{ yes, no}
@attribute I{ yes, no}
@data
TB, no, yes, yes, no, no, no, no, yes, no
T 15, ye s, no, no, no, yes, no, no, yes, no
T 102, no, no, no, yes, no, no, no, no, no
```

Figure3:Arff format of basketball technology actions statistics

EMPIRICAL ANALYSIS

Frequently frequent 1 -phase set L1: B, C, F, G, H; behind the number of support count or call frequency;
Frequent 2-phase sets L2: AB, AC, BC, BF, CH;
Frequent 3-itemsets L3: ABC;
Experiment Parameters: Selects the default setting except the minimum support and minimum confidence and the number of output rules. Also change the output of frequent itemsets from false to true.

Experiment parameters: delta = 0.05, minimum support 0.6, minimum confidence 0.4, the number of output rules 30, significant degree -1.0 If changes in parameters, such as change Minimum support, minimum confidence, the number of output rules, the degree of salience can be other experimental results.

When the data volume is large, the mining algorithm based on the analysis of Chapter 4 can also mine many association rules.Such as the NBA (American Basketball Association) League 2010-2011 season part of the game's basketball action using Apriori algorithm mining results are as follows (Figure 4).

Scheme: Weka assoiciation Apriori
Instances:261
Attributes:10
Tid
A
B
C
D
E
F
G
H
I
Associator model (full training set)
Minmum support:0.6(157 instances)
Minmum metric(confidence):0.4
Mumber of circles performs:8
Generated sets of large itemsets:
Size of set of large itemsets L(1):7
Large itemsets L(1)
A=no 203
D=no 202
E=no 230
F=no 210
C=no 159
H=no 161
I =no 242
Size of set of large itermes L(2):9
Large iterms L(2)
A=no D=no 166
A=no E=no 192
A=no I=no 188
D=no E=no 173
D=no F=no 163
D=no I=no 185
E=no F=no 179
E=no I=no 215
F=no I=no 191
Size of set of large itermes L3:3
Large iterms L3:
A=no E=no I=no 177
D=no E=no I=no 160
E=no F=no I=no 164

Figure4:The result of Apriori arithmetic algorithm association rules

Experimental parameters: In addition to the Minimum Support and Minimum confidence and the number of output rules other than the default settings. In addition, the output of frequent itemsets is changed from false to true. delata, Minimum support, Minmetric 0.4, numrules of output rules 30, singnificance level -1.0 . If you change the parameters, such as changing the minimum support, minimum confidence, the number of output rules, salience, etc. can be other experimental results.

Rule 1 shows that there is no 3-point goal in the absence of assists, with a 95% confidence level. Rule 2 indicates that no assists and no blocks will result in a 94% confidence goal. From which to see how important assists. You can also find other useful rules. Of course, with some useless rules, such as Rule 3: No 3 goals without blocks, confidence 94%, this rule has no practical significance for players and coaches. Useful rules can be used to guide team training, such as multi-point break in the ball to the three-point unguarded players, so that they vote for three points.

CONCLUSION

Although there are studies on the physical effects of basketball on college students, the theoretical research on the sustainable development of basketball, the impact of mass media on the development of CBA league and the key of basketball video analysis Technology research, etc., but both atdomestic and foreign, the basketball technology movement mining research is almost zero, there is no ready-made results of the case.

In this paper, based on the analysis of application and analysis of data acquisition and preprocessing of basketball technical movement, this paper studies the application of data mining technology in basketball technical movement through the association rule mining method based on Apriori algorithm. The creative combination of movement and computer technology opens the way for the study of the law of basketball technology movement, and provides more accurate learning resources for coaches and athletes.

REFERENCES

Brooks, J., Kerr, M., & Guttag, J. (2016, August). Developing a data-driven player ranking in soccer using predictive model weights. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 49-55).

Cheng, G., Zhang, Z., Kyebambe, M. N., & Kimbugwe, N. (2016). Predicting the outcome of NBA playoffs based on the maximum entropy principle. *Entropy*, *18*(12), 450.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), 1203-1205.

Li, W. (2014). Analysis on Competitive Sports Based on the Data Mining Technology. In *Proceedings of the 2012 International Conference on Cybernetics and Informatics* (pp. 2359-2366). Springer, New York, NY.

Marmarinos, C., Apostolidis, N., Kostopoulos, N., & Apostolidis, A. (2016). Efficacy of the "pick and roll" offense in top level European basketball teams. *Journal of human kinetics*, *51*(1), 121-129.

Martinez, M. G., & Walton, B. (2014). The wisdom of crowds: The potential of online communities as a tool for data analysis. *Technovation*, *34*(4), 203-214.

Monroe, M., Lan, R., Lee, H., Plaisant, C., & Shneiderman, B. (2013). Temporal event sequence simplification. *IEEE transactions on visualization and computer graphics*, *19*(12), 2227-2236.

Rehman, A., & Saba, T. (2014). Features extraction for soccer video semantic analysis: current achievements and remaining issues. *Artificial Intelligence Review*, *41*(3), 451-461.

Robertson, S., Back, N., & Bartlett, J. D. (2016). Explaining match outcome in elite Australian Rules football using team performance indicators. *Journal of sports sciences*, *34*(7), 637-644.

Shi K J (2015). Research on the Construction and Improvement of the Content System of the Expanded Training Course in Colleges form the Perspectives of Obtain Employment and Ability Enhancement[J]. Executive Chairman.

Shimizu, T. K., Louzada, F., & Suzuki, A. K. (2014). Analyzing Volleyball Data on a Compositional Regression Model Approach: An Application to the Brazilian Men's Volleyball Super League 2011/2012 Data. *arXiv preprint arXiv:1412.5848*.

Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., ... & Grossniklaus, M. (2017). How to make sense of team sport data: From acquisition to data modeling and research aspects. *Data*, *2*(1), 2.

Tashakkori, R. M., Parry, R. M., Benoit, A., Cooper, R. A., Jenkins, J. L., & Westveer, N. T. (2014, March). Research experience for teachers: data analysis & mining, visualization, and image processing. In *Proceedings of the 45th ACM technical symposium on Computer science education* (pp. 193-198).

Wang, J. (2014). Research on Key Technology of Data Mining for Volleyball Game Based on Service System. In *Applied Mechanics and Materials* (Vol. 543, pp. 4698-4701). Trans Tech Publications Ltd.

Zhang, D. Y. (2014). Research on Data Mining Technique Based on Visualization of Parallel Coordinate Method. In *Applied Mechanics and Materials* (Vol. 513, pp. 2384-2388). Trans Tech Publications Ltd.