

Application of Data Mining Technology in the Subject Tactical Teaching of Badminton

<https://doi.org/10.3991/ijet.v13i07.8778>

Delin Yang

Henan Polytechnic, Henan Zhengzhou , China

delinyang18273@163.com

Abstract—In badminton field tactics teaching, there are problems of not timely, inaccurate and incomplete data collection of completion cases, not deep enough data and too few personalized training programs for athletes' tactics. Aiming at solving the above problems, the traditional Apriori association algorithm was optimized, and the algorithm was applied to the acquisition of badminton field tactics case data. Moreover, the multidimensional data mining was conducted, thus providing a scientific and personalized teaching plan for badminton field tactical teaching. The simulation experiments showed that data mining technology applied to badminton field tactics teaching could effectively put forward personalized teaching plan for data support from the tactical point of view, and improve the effect of badminton field tactical teaching.

Keywords—data mining technology; association algorithm; badminton; field tactical teaching

1 Introduction

Badminton is a very popular movement, and it is the focus gold medal project in the Olympic Games. China badminton team has a high level of skills and tactics, and has won many excellent results in major international competitions. By studying the factors that can affect the winning rate of the competition, including the athletes' physical function, psychological state, technical and tactical content, injury and rehabilitation, the researchers can provide many scientific supports for improving the athletes' performance and fully excavating the potential of the athletes.

Technology and tactics are an important factor affecting the performance of athletes. In modern sports, the importance of technology and tactics on various items is divided into three levels, which are decisive, important and basic, and the level that affects the network project is classified as decisive factor. This shows that, in the badminton, whether technique and tactics are applied properly are the decisive factors affecting the outcome of the game. At present, technical and tactical decision in badminton sport is mainly based on the personal experience of the coaches. Singly depending on the personal knowledge and experience and judgment ability cannot meet the scientific and accurate decision. Therefore, to improve the technical and tactical level of badminton sports, we are supposed to establish a scientific and tactical sup-

port system, and through the study of the tactical rules of badminton, to assist the coaches to do scientific decision-making.

Data mining technology provides an effective way for people to manage and analyze potential rules in massive data. At present, the national badminton team information platform has been put into use, and the data stored in the game is expanding. It is necessary to use these past statistical data to find new useful knowledge, and then provide scientific support for the coaches to make targeted training plans and to make decisions in the game. At present, the research of data mining in the field of physical education in China is mostly to meet the needs of sports management, and seldom improve the efficiency of system operation and the accuracy of decision-making from the perspective of computer science. In particular, using the traditional association rules algorithm in the process of mining the game data, the threshold introduced is limited and the constraint conditions for the user's interest are short. The algorithm runs long and many of the rules in which the users are not interested are redundant rules. Because badminton games are usually held in a short time, the long analysis process and the existence of a large number of redundant experimental results will have a negative impact on the efficiency of the system. Therefore, it is necessary to improve the traditional association algorithm to improve the efficiency of the system.

To sum up, through two aspects of sports information and data mining, the establishment of badminton tactical statistical analysis system is the requirement to adapt to the development of badminton sports. In the meanwhile, the establishment of this system can provide information management and tactical decision support services for badminton teams at all levels in China, and help to improve the coaches' level of subject tactical decision for training and competition, which is of great significance for further improving the level of badminton in China.

2 Literature review

Data mining technology (DMT) appears because it can discover hidden knowledge in large amounts of data. Wook pointed out that this technology was especially important for organizations to make wise decisions and make strategic plans [1]. Zhang analyzed the application of data mining technology in badminton court tactical analysis system [2]. Carpita and others studied the relationship between a football match (victory or defeat and draw) and a set of variables describing the action of a game through data mining analysis of the tournament data for a continuous period of 4 years [3]. The purpose of this research is to find out the factors that will result in a match success. Jelinek and other scholars made use of data mining technology to improve the accuracy of prediction of Australian football match results [4]. Islami and so on used data mining algorithm to study pregnant women's understanding of sports activities, and found that vector movement is beneficial to the health of fetus and pregnant women [5]. European sports clubs are often the object of European sports and social studies. Its size, age, and the number of movement times are all the factors that are studied. For instance, Wicker and others used data mining methods to analyze the main problems of the German sports clubs [6].

At present, the data mining research in China is mostly in the theoretical stage, and the number of practical applications is relatively small. In addition, most of the current research is simply to realize the requirements of sports management. There is little research on how to improve the efficiency and accuracy of data analysis from the perspective of computer science, which affects the practicability of the system to a certain extent. In this paper, through two aspects of sports information and data mining technology, the establishment of badminton tactical statistical analysis system is to meet the requirements of the development law of badminton sports, which has a certain guiding significance.

3 Method

The data used in the experiment are from the statistics database of the national badminton team. The data source is the technical and tactical statistics of the last beat of each round of a man's singles athlete in the last two years and the training in the daily team. The transaction database D' after the preprocessing is 9136, and item is 45. The item that each transaction contains includes the basic information of the athletes, the competition information, the number of rounds, the number of beats, the service area, the technology, the creation of direct scoring technology and so on.

Through the experiment, the ACARMI algorithm is compared with the Reorder algorithm before the improvement. At the same time, in order to verify the advantages of the implementation efficiency between the ACARMI algorithm and the other constraint based association algorithms, the most widely used Separate algorithm is selected to carry out the longitudinal ratio.

The subject technical and tactical statistical analysis system of badminton realizes the fast collection and data mining analysis of the athletes' game technology and tactics data. The system consists of four parts, which are data collection module, data conversion module, ARMCI algorithm realization module and user visual interface. Data collection module is divided into mobile terminal data mining module and PC terminal technology and tactics integration module. The role of the former is to realize the game statistics on the client side of the Android system, and transfer the game data to the comprehensive information database through the mobile network, so as to reduce the time consumption of the intermediate link of the data collection. The PC terminal technology and tactics integration module is achieved by extracting the data from the national badminton team in the excel table before it is transmitted to the database, and the data in the database can be downloaded to the excel form for printing. In addition, the system can also achieve preliminary data distribution statistics, and provide assistance for users to make constraints.

The preprocessing module is to convert the numerical statistical information into Boolean data, and to reduce the dimension of some merged items and delete some items that are not related to the interest of the users, so as to improve the scientific nature and accuracy of the mining results. The processed data will be transferred to the intermediate database to prepare for mining association rules.

The ARMCi algorithm realization module is mainly responsible for data mining in the intermediate database, and provides interface for the upper user visualization interface. The user visual interface can provide the user with the interface of data mining parameters such as constraints, minimum confidence, minimum support and other data mining parameters. At the same time, the mining association rules mined can be presented to the users intuitively, and the user can make the strategy according to the presented rules.

4 Result

4.1 ACARMI algorithm and Apriori algorithm

Apriori algorithm is one of the most widely used algorithms in mining many Boolean association rules. Since all the parent classes of infrequent itemsets are infrequent itemsets, the Apriori algorithm uses an iterative algorithm of layer by layer search to explore $(m+1)$ set of items through the m - item set. There are some shortcomings in the Apriori algorithm: there will be a large number of candidate item sets in combined growth; there are too many times to scan the transaction database; the adaptation range is limited; the algorithm lacks the interaction with the user. In the last shot of badminton, the technical statistics of the badminton contain the attributes of a variety of dimensions, and the number is huge. Using the traditional Apriori algorithm mining will produce a large number of candidate items. Not only the time is long, but also generates a large number of worthless association rules, which seriously affects the production of the decision making for the game. In order to solve this problem, we optimize the Apriori algorithm according to the characteristics of the badminton movement. At present, in the improved algorithm based on the Apriori algorithm, the mature one is the MultipleJoins, Reorder and Direct algorithm proposed by R.Srikant.

The ACARMI algorithm combines the advantages of the Reorder and Direct algorithms. In each stage, the number of candidate sets can be maintained at a lower level, which can effectively reduce the time and resources consumed in the generation of candidate items, thus improving the efficiency of the algorithm.

4.2 Comparison between ACARMI algorithm and Reorder algorithm

First of all, the constraint condition C is set to $\{\text{score}\}$, that is, the association rules obtained by data mining must contain the "score". The effect of the Reorder algorithm and the ACARMI algorithm running under the same minimum support S_{\min} is shown in Figure 1.

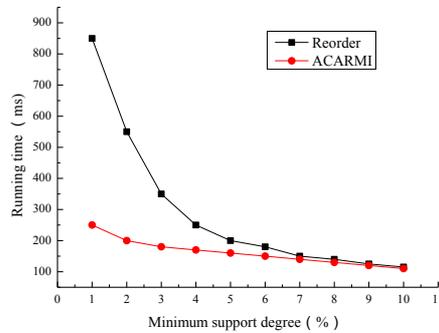


Fig. 1. Running time contrast diagram under different minimum supports

Then, under the same support degree ($S_{min}=2\%$), the running time of Reorder algorithm and ACARMI algorithm under different constraint numbers is shown in Figure 2.

The number of constraints is 1: {score}

The number of constraints is 2: {score, reboundin straight edge}

The constraint number is 3: {score, rebounding straight side, service area =1}

The number of constraints is 4: {score, rebounding straight side, service area =1, receiving area =10}.

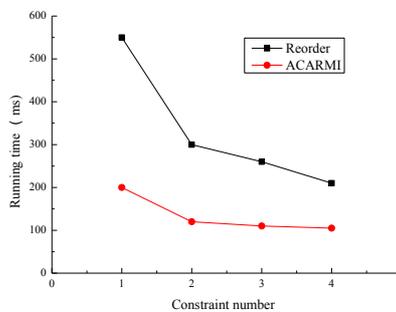


Fig. 2. Running time contrast diagram under different constraint numbers

From the analysis of Figure 1 and Figure 2, we know that the time consumption of ACARMI and Reorder algorithms decreases with the increase of n , which is because the higher the minimum support is, the more items that are filtered, and the time of calculation will be reduced accordingly. It is the same reason why the consumption time is reduced when the number of constraints increases. At the same time, we can see that under the same conditions, the efficiency of ACARMI algorithm is higher than that of Reorder algorithm. In the initial stage, the Reorder algorithm uses the incoming item set S to generate the candidate option set. The more the number of

items in the S is, the more time the algorithm consumes, so when the number of items is reduced, the consumption time of the Reorder algorithm is declining. Since the ACARMI algorithm uses the same method as the Direct algorithm in the initial stage to generate the candidate item set, the time consumed is greatly reduced compared to the Reorder algorithm. However, when the number of items is reduced, the decreasing trend of the ACARMI algorithm consumption time is less obvious than that of the Reorder algorithm.

4.3 Comparison between ACARMI algorithm and Separate algorithm

Constraint condition $C=\{\text{score}\}$. The effect of Separate algorithm and ACARMI algorithm under the same minimum support S_{\min} is shown in Figure 3 and Figure 4.

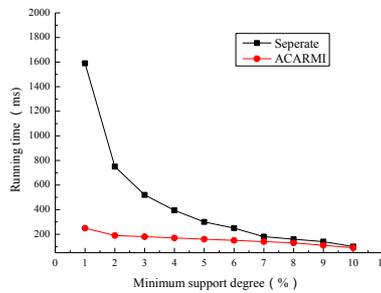


Fig. 3. Running time contrast diagram under different minimum supports

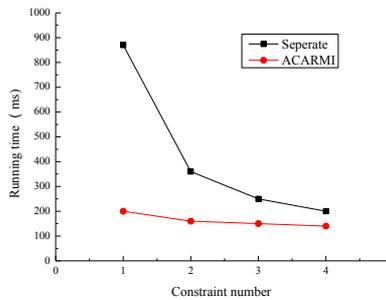


Fig. 4. Running time contrast diagram under different constraint numbers

From the analysis of figures 3 and 4, we can see that under the same support or constraint conditions, the calculation efficiency of ACARMI algorithm is much better than that of Separate algorithm. This is because the Separate algorithm optimizes the Apriori algorithm by introducing the constraint conditions. Because the method of generating candidate items like the Apriori algorithm is used, and each time one candidate item set generates, the database is scanned once, when the number of confidence and constraints is small, the time consuming is very large; when the number of

confidence and constraints is large, the efficiency is improved. The main time consumption of ACARMI algorithm occurs when generating $k+1$ candidate item sets, and only needs to scan database three times during execution. Therefore, the ACARMI algorithm is less affected by the minimum confidence and the number of constraints relative to the former, and the efficiency of the algorithm is relatively stable.

4.4 Design and implementation of statistical data collection module for technology and tactics

The data statistics module of the mobile terminal can provide the function of data statistics as shown in the figure above, and through the data upload module, upload to the server automatically by the wireless LAN whenever the capacity of the database is more than 500M, and the upload function can be operated manually. At the same time, the module also provides the data query function of the server side, and it can display the querying data graphically and intuitively. Because the data mining function of this system is mainly aimed at the technical and tactical statistics of the last shot, the design and implementation of this module will be emphatically introduced, and its software workflow is shown in Figure 5.

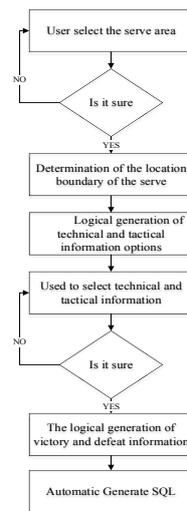


Fig. 5. The workflow of the last shot technical and tactical data statistics module

The PC terminal technology and tactics integration module is mainly divided into two sub modules: excel table data extraction and data statistics report generation. The former is to realize the rapid entry of the competition statistics in the form of excel form, and the latter is to achieve the display of data in the comprehensive database in the form of report according to the query requirement of the user. The workflow of excel form data extraction module is shown below in Figure 6.

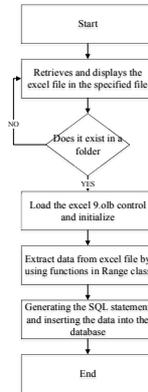


Fig. 6. The workflow of excel form data extraction module

4.5 Design and implementation of data preprocessing module

In view of the characteristics of badminton statistics, although the data has been normalized through software, most of the statistical methods before the application of the system come from the artificial statistics. In addition, there are many noises and redundancy in the original data and the format is not adapted to the Boolean data mining algorithm. As a result, before mining association algorithm, we must transform data into a standard form. The overall architecture of the data conversion model is shown in Figure 7.

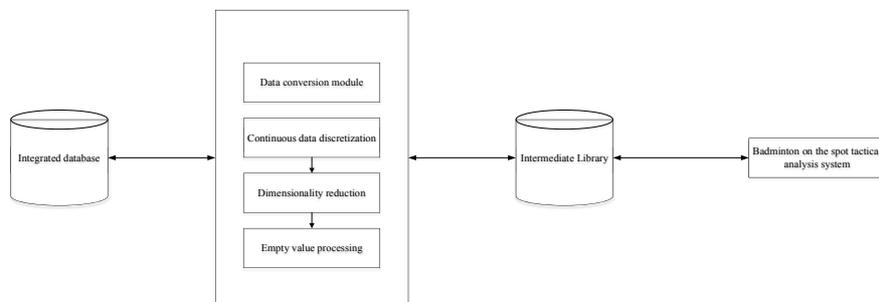


Fig. 7. The overall architecture of the data conversion model

As shown in the figure, data transformation is mainly divided into the following steps: firstly, continuous attribute discretization. In the data collected at present, there are many concrete continuous divided data. The data are large and not repeated, which will have a certain influence on the accuracy of the mining results. Therefore, the data should be discretized first of all. The so-called data discretization means that by setting several fixed value spaces, the data that belongs to a certain range of values is represented by a discrete value or symbol. The setting of value space will directly affect the accuracy of data mining results. If the range setting is too small, the effect

of data discretization will not be obvious, it cannot bring about the enhancement of the operation effect, and the data cannot be lost because the discrete data cannot reach the minimum support degree. If the range is set too large, it will make the accuracy of mining results reduced, and it will also be abandoned due to the fact that the data after discretization cannot reach the minimum support. The system uses the knowledge of badminton in the field of physical education to determine the value space. For example, taking the "time against time" as an example, there is a clear distinction standard in badminton field. Since more than 70% of the antagonism occurs below 12s, we can divide the property of "time against time" into two attributes, "normal" and "long", which are "1" and "0" with a Boolean variable. It is the data type matching with the association algorithm.

Secondly, dimension reduction. As a result of the complete related data statistics of a game, there are many kinds of attributes, and many of them have no relation to the gain or loss, mainly the objective factors such as the gender, nationality, and playing field of the athletes, which can be deleted to reduce the dimension of the whole database, so as to improve the pertinence and accuracy of the association rules. At the same time, we can combine to build a new attribute to reduce the dimension in terms of some attributes with similar influencing factors. For example, the two attributes of "last shot technology" and "creation of direct scoring technology" can be merged, and "serve point" and "drop point" can be directly merged into "Region - area" attribute.

Finally, the empty value processing. In the process of data collection, there are a large number of empty values and default items in the excel form files that have been manually recorded in the past, which is due to the fast pace of competition and the non-standard entry standard. This system has two ways to deal with the empty values in the collected data. In view of the inconsistency of two piece of data collection and the fact that each table in the same collection cannot be associated, it will sort out the time of the data collection and the number of the default values, delete the worst record of the statistical quality, and then merge the remaining records. Aiming at data duplication that may occur in database, the way to deal with it is to delete duplicate data. The way to deal with the default value is to judge the nature of the default first. If it is not a key item, fill in the pre - established normal value; if the default is the key item, such as the serve area, technical means, etc., the sample is deleted. At the same time, an empty value threshold is set for the whole sample. If the number of empty values is greater than the threshold value, it cannot judge the nature of the default, delete the sample directly, and then improve the efficiency of empty value processing.

The generation of data mining table is divided into the following steps: first, the "last beat statistics table", "technology list", "line list", "round list", "competition information table" and "athlete information table" are transformed into the middle table in the source database. Then, the steps of continuous attribute discretization, dimension reduction and space value processing are carried out on the generated middle table. Finally, the last technique mining table is generated. All data mining operations related to the final technology can be operated on this table.

4.6 Design and implementation of ACARMI mining module

Data mining module is the core of this system. The module is based on open source spring+struct+hibernate framework, and the implementation language is JAVA. The core idea of ACARMI algorithm is to reduce the number of candidate item sets by introducing constraint conditions, and reduce the number of scanned database by combining the advantages of Reorder algorithm and Direct algorithm. The concrete steps are as follows:

Step 1: first of all, define a plastic array, and through this array, record the number of frequent 1 set of statistics in the first traversal of the database process. Judge the frequent 1 set based on the ratio of the number of times and the minimum support ratio. At the same time, we define minLen and maxLen to record the minimum length and maximum length of intermediate frequent item sets in e_i .

Step 2: scan the database for the second times and define an array to record the entry option set S satisfying the constraint condition C . Because the frequent -2 itemsets generate the set S , the number of frequent 2 item sets is less than frequent 1 itemsets, and there is no memory overflow in the process of generation. At the same time, the Direct algorithm is used to generate frequent itemsets and determine whether it is a frequent itemset based on whether it contains the items in the set S , and this process has been iterated for the maxLen times.

Step 3: when the length of frequent k - itemsets satisfying the constraint condition C exceeds maxLen, use Reorder algorithm to generate frequent itemsets and determines whether they are frequent itemsets according to whether they contain items in set S . In this process, the database will be traversed for the third times.

The module calls the data mining table in the intermediate library through the interface provided by the intermediate library, and saves the generated association rules in the rule base. In addition, the module also provides an interface that interacts with the visual interface of the upper user. The user can set data mining parameters according to the requirements and display the mining rules in the visual interface.

Before data mining starts, users, according to badminton technology mining indexes, fill in the minimum support and minimum confidence, and select constraints. For example, the minimum support of this experiment is 4%, the minimum confidence is 10%, and the constraint condition is "lost", that is, the "lost" property must appear on the right side of the rule. In addition, the system provides constraint conditions editing. Complex constraints can be formulated by the input of Boolean expressions, and the constraints of the new editing will appear on the left side of the rule. After the user selects the "start mining" button, the system will use the ACARMI algorithm to process the data in the data mining table, and the results will be presented in the interface in the form of the table, and the results will be sorted from high to low according to the support value. In addition, the system also provides the export report function, and can import the generated rules into the badminton information library. The user can click the "query" button to query the analysis player ID.

4.7 Analysis of the effect of system implementation

In the practical application, the relevant data information of the last shot of the athletes in the "national badminton team comprehensive information processing platform" will be excavated. Therefore, the data used in this experiment are related data of a men's single athlete in the past two years and the statistics from team training. All the data are saved in the form of Excel table, so the experimental database is introduced into the experimental database through the PC end technology and tactics extraction integration system, and the data preprocessing module is standardized. There are 9136 transactions after the completion of the process. Because there are many uncertainties in the badminton movement itself, in order to prevent the loss of valuable knowledge, we set the minimum support (S_{min}) at 4%, and the minimum confidence (C_{min}) is set to 10%. Table 1 is the experimental result derived from the report generation function.

Table 1. ACARMI algorithm mining results

Rule	Support degree	Confidence degree
6—D Anti killing →Points	15.41%	11.11%
6—A Anti high pair→Points	11.15%	13.26%
5—A In the height→Points	9.05%	16.17%
6—J Head block→Points	8.87%	23.47%
11—J Go far and straight→Points	8.64%	31.08%
12—J Reverse alignment→Points	8.37%	32.03%
7—L Anti blocking→Points	7.96%	14.07%
9—K Anti blocking→Points	7.03%	12.06%
6—G Reverse the distance to the distance→Points	6.79%	34.06%
4—F Positive killing→Points	6.33%	13.09%
4—L Reverse block pair→Points	5.32%	36.38%
6—B Anti high school→Points	4.06%	17.06%

The result of the experiment, as shown in Table 1, shows that the association rule expression in the experimental results is "12 - J counter - missing", which means "the opponent hits the ball from the area 12 to the area J and the technique is counter-hook". The probability of the incident at the same time of "losing the score" is 8.87%. When "the opponent hits the ball from the area 12 to the area", the probability of the loss is 32.03%, which indicates that the probability of losing the score is very high once the opponent is returning the ball from the area 12 to the regional J using the counter-hook technique in the region, and the support degree is 8.87% that the player has realized that the condition can lead to the loss of its own. We should try to avoid it during the competition, but we should also train this type of events in training. According to the rule "6 - D against 1", we can also see that the probability of "the oppo-

nent hit the ball from the area 6 to the area D and the technique is to kill the pair" and "lose the score" occurring at the same time is the highest. Whereas, when "the opponent hits the ball from the area 6 to the area D and the technique is counter kill", the probability is only 11.11%, maintaining at a lower level. This may indicate that the technical weakness of the athlete has been detected by other athletes, but it is clear that the athlete pays attention to the training for the weakness. Although the probability of the event is high, the loss rate remains at a normal level. In addition, according to the rule "4 - L reverse blocking 1", it can be seen that although the probability of the rule is not high, it is the most likely to lead to the loss of score. The athletes should pay attention to the training of the technical means in training. As we can see from the rules of the above chart, the frequency of area J is very high in the top ranking event, which indicates that there are some shortcomings in the player's pre-network technology, and the opponent has made a lot of attempts on this weakness. The player needs to strengthen his own pre-network technology to cope with this situation. In addition, from the top priority of the event, it can be seen that when the opponent uses backhand attack, the success rate is very high, which shows that the team has some shortcomings on the way to judge the counter-handball route, which can help the coaches to make a targeted training plan.

5 Conclusion

This paper introduces the application of data mining technology in badminton spot tactics teaching in detail. First of all, the ACARMI algorithm is introduced. It is proved that the efficiency of ACARMI algorithm is better than that of Reorder algorithm and Separate algorithm. Then, three functional modules, data collection, data preprocessing and data mining, are divided according to needs. In the data collection module, the application program based on Android tablet and the collection program based on the windows platform are designed and implemented, and the standardization and efficiency of data collection are improved. In data preprocessing module, a data preprocessing model suitable for badminton statistics data is designed and implemented. In the data mining module, the ARMCI algorithm is implemented using the JAVA language. Finally, the system is used to excavate the existing experimental data, and the feasibility of its auxiliary tactical decision is verified. In the future research process, we need to combine more badminton characteristics to make more logical judgment rules, so that it can be more conveniently used by coaches.

6 References

- [1] Wook, M., M. Yusof, Z., & Nazri, M. Z. A. (2014). Data mining technology adoption in institutions of higher learning: a conceptual framework incorporating technology readiness index model and technology acceptance model 3. *Journal of Applied Sciences*, 14(18), pp. 2129-2138. <https://doi.org/10.3923/jas.2014.2129.2138>
- [2] Zhang, M. (2017). Application of data mining technology in badminton spot tactical analysis system. *Agro Food Industry Hi Tech*, 28(1), pp. 3398-3401.

- [3] Carpita, M., Sandri, M., Simonetto, A., & Zuccolotto, P. (2015). Discovering the drivers of football match outcomes with data mining. *Quality Technology & Quantitative Management*, 12(4), pp .561-577. <https://doi.org/10.1080/16843703.2015.11673436>
- [4] Jelinek, H. F., Kelarev, A., Robinson, D. J., Stranieri, A., & Cornforth, D. J. (2014). Using meta-regression data mining to improve predictions of performance based on heart rate dynamics for australian football. *Applied Soft Computing*, 14(1), pp. 81-87. <https://doi.org/10.1016/j.asoc.2013.08.010>
- [5] Islami, F., Bagheri, F., & Mohammadi, F. (2016). Surveying the knowledge of pregnant women towards sport activities during pregnancy using data mining algorithms., 18(1), pp. 8. <https://doi.org/10.15314/tjse.30959>
- [6] Wicker, P., & Christoph, B. (2016). Analysis of problems using data mining techniques – findings from sports clubs in germany. *European Journal for Sport & Society*, 7(2), pp. 131-140. <https://doi.org/10.1080/16138171.2010.11687851>

7 Author

Delin Yang is with Henan Polytechnic, Henan Zhengzhou ,450046, China.

Article submitted 25 April 2018. Final acceptance 07 May 2018. Final version published as submitted by the author.