

A Big Data-Based Data Mining Tool for Physical Education and Technical and Tactical Analysis

<https://doi.org/10.3991/ijet.v14i22.11345>

Lili Pan

Baoji University of Arts and Sciences, Baoji, China

47587495@qq.com

Abstract—This paper attempts to develop a data mining tool to guide sports training, promote physical education and facilitate technical and tactical analysis. For this purpose, information techniques like mathematical statistics and big data analysis were employed to collect and analyse the information on competitive sports. Based on database and computer algorithm, the author designed a data mining tool applicable to the information of various competitive sports. The proposed tool can mine out valuable information from the big data, enabling trainers to realize targeted and efficient physical education. The mined information also helps improve the analysis of techniques and tactics of competitive sports. The research findings promote the application of information technology in physical education and competitive sports.

Keywords—Physical education, technical and tactical analysis, data mining, big data analysis

1 Introduction

The development of information technology and network technology brings great convenience to people's lives and work. In the era of information, professional skills and knowledge, people receive a lot of information on a daily basis [1-3]. While information enriches people's lives and enhances their personal skills, excessive information breaks the dynamic balance of human reception and digestion, making it difficult to extract more effective knowledge from the information base, and resulting in the useless work and waste of time [4-5]. The emergence of a new generation of information technology processing tools, namely data mining technology, has largely solved the problem of information flooding [6-7]. Such tools can help to screen out valuable information, and then perform processing, analysis and statistics, which in turn serves to improve professional skills, summarize rules and master high-use information [8-10].

In recent years, the function and importance of data mining have been gradually realized. In the field of competitive sports etc., data mining technology has started to be applied into professional information statistical analysis [11-13]. For instance, the US NBA system well applies the data mining tool to basketball teaching guidance and technical and tactical analysis as an auxiliary tool. Such tool can provide useful in-

formation for tactical optimization, player skill training and promotion of each ball team [14-15]. The information-based career of China's sports is also constantly developing, and has accumulated a large amount of information and competitive data, which has not yet been fully utilized effectively. Therefore, it has become an important task in the field of competitive sports on how to make full use of existing data and explore important information that contributes to the improvement of athletes' technique and tactics.

In view of the problems above, this paper aims to design a data mining tool for physical education and technical and tactical analysis. Information techniques like mathematical statistics and big data analysis were employed to collect and analyse the information on techniques and tactics of competitive sports. This shall be conducive to the in-depth analysis and summary of competitive sports information, and mining out valuable information, which lays a theoretical basis for the application of information technology in physical education and competitive sports.

2 Related Theory of Data Mining Technology

2.1 Big data analysis

Big data analysis generally refers to a method or technique for quickly acquiring valuable information from various forms of data [16-17]. This method involves not only the size of the data (video, audio, image, text, numbers, etc.) itself, but also the various elements of the collected or statistical information. Unlike conventional statistics, big data analysis typically includes the collection, importing or pre-processing of information, analytical statistics, and data mining [18-21]. Data statistics is only part of big data analysis, while big data analysis is the use of various algorithms to statistically calculate the data, and then extract high-level analytical data for meeting the actual needs based on the existing large amount of data processing. The big data is characterized by massiveness (multiple information, and huge data volume), high speed (fast information processing), diversity (various types of data), and utilization value (use value of collected data information).

In the field of competitive sports, the big data analysis is helpful. Firstly, it can properly predict the damage or injury that athletes may have during the competition. Secondly, based on the big data analysis of past trainings and games, it can help coaches and athletes avoid the empiricism in techniques and tactics; through the statistical analysis of the data in competition, athletes shall develop their strengths and avoid weaknesses, so as to maximize their potential; thirdly, during the sports competitions, it can help the interpreters to more vividly and accurately interpret the competition based on historical data. Finally, the big data analysis of the competitive sports can also enhance the interest and understanding of the audience or fans.

2.2 Data mining

Data mining is a kind of information processing method that integrates various technologies and related theories such as information technology, artificial intelligence, pattern recognition, mathematical statistics, machine intelligence and data visualization. It can help to extract and link the information that is potentially valuable, has a large guiding effect or application significance in a certain field or industry database, to facilitate the efficient utilization of potential information. In addition, data mining technology is considered to be a new method that can improve the efficiency and quality of commercial or scientific decision-making. Different from traditional data analysis, the information collected by data mining has the characteristics of practicability, validity and unknownness.

In general, data mining generally relies on statistics and induction, decision trees, artificial neural networks, genetic algorithms, fuzzy technology, and visualization techniques, as shown in Figure 1. In the process of data mining, statistical analysis is used to summarize and extract valuable information; an established decision tree can use the tree structure to represent the set of decisions, classify and summarize the data; the artificial neural network can help to make data classification, clustering, finding feature rules, and then data mining; genetic algorithm and fuzzy technology optimize, identify and analyse big data through iterative process and fuzzy set theory respectively; finally, the visualization technology is used for the final representation of the information law in an objective and intuitive form.

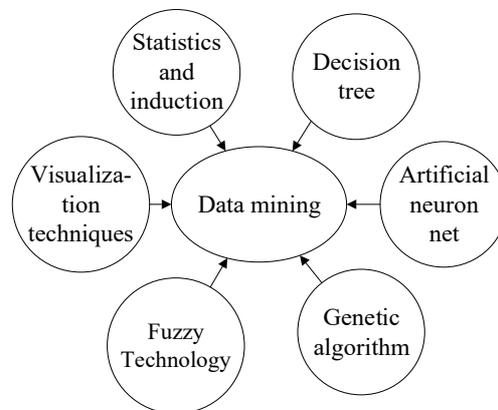


Fig. 1. Data mining technology

Data mining plays an important role in reform and decision-making optimization in various fields. In competitive sports, it is equally important to mine out valuable information from big data. Therefore, based on the big data analysis, this paper aims to explore a data mining tool that are beneficial to physical education and technical and tactical analysis in competitive sports.

The global data can be refined, that is, the sports technical and tactical database is subdivided into the basic situation of the athletes, the technical data, the technical statistics, and the technical video review (Figure 4). The team's database includes the basic situation of the team, tactics query, technical and tactical statistics, and tactical inquiry, and its flow chart is shown in Figure 5.

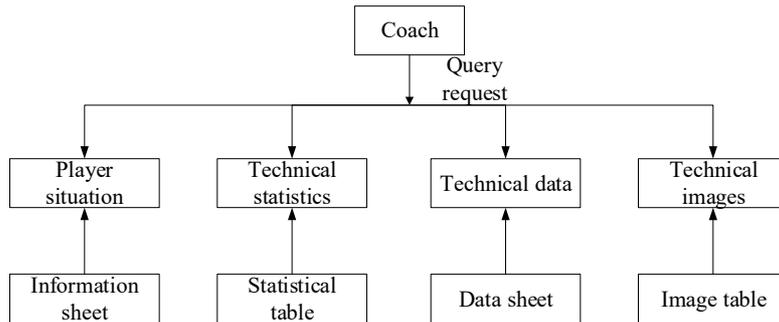


Fig. 4. Athlete refinement data flow

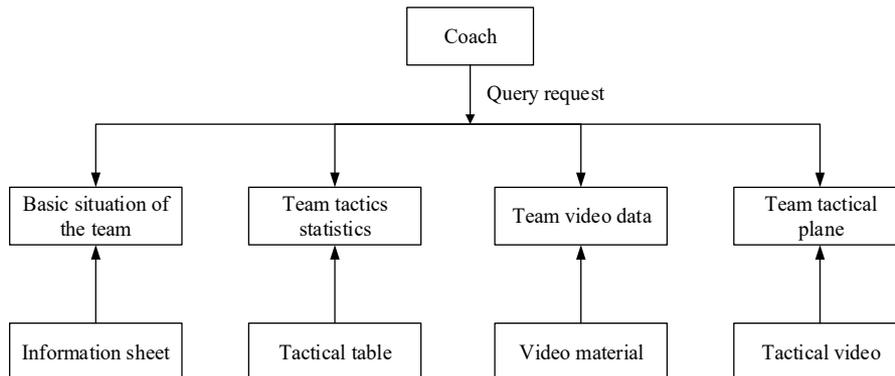


Fig. 5. Team refinement data stream

The basic situation of the team data such as the team name, logo, city, shirt colour, team history, list of active players, head coach, assistant coach, technical coach, defensive coach, team doctor, and honours needs to be scrolled through the screen; for the disciplinary data, the team's technical and tactical statistical data, their maximum value, average value, or limit value of any opponent are calculated through time partition; based on certain principle, an indicator was selected at a time for full statistics. Due to the huge amount of data involved in the database, large space is also needed. Thus, the video is called by storing the video path, and the original video is stored in the hard disk. The file names of the videos are sorted and stored, ensuring the one-to-one correspondence between the file name and the database table. Then, the database file can be retrieved and played in the physical location.

3.2 Application of data mining in technical and tactical analysis of badminton teaching

The sports technical and tactical analysis needs to carry out informationization, quantification and automation by relying on the tactics of the big data platform. For data statistics and classification in sports, the data mining based on multimedia database is of great significance and value. For example, in the actual badminton competition, the relevant information obtained by the coach is very large, and the real important information implied in the data cannot be acquired quickly, because the noise on the scene has a great interference with the coach's judgment, and the coach's conclusions through the experience and intuitive judgment is less reliable. Therefore, it is especially important to use data mining technology to discover the laws implied in the data.

The competition data was analysed through association rules, and combined with the characteristics of the ball to obtain the correlation between the players and between the players' own technical actions. Then, a corresponding data analysis system was established, and the statistical results were inputted into the quantity statistics sheets by selecting all the competition videos of the 2014 World Championships. After the data acquisition, the data was entered into the system by the mouse click input or script input. The data processor was responsible for statistical and comprehensive tactical analysis of the data; for routine data such as attack success rate, serve success rate, etc., the data mining technology was used for comprehensive analysis combined with the opponent's situation. The coach can query and analyse the data through the system. The system is built as shown in Figure 6.

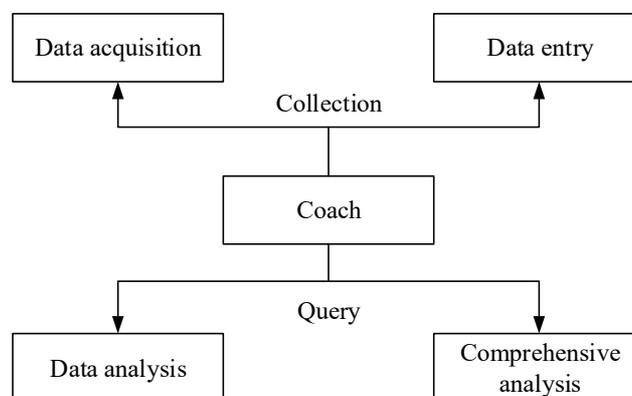


Fig. 6. System diagram

The system is divided into four modules, each of which can be sub-divided into several sub-modules, finally forming a system of badminton tactical analysis with compact relationship and simple structure. Figure 7 shows the general framework of system functional structure.

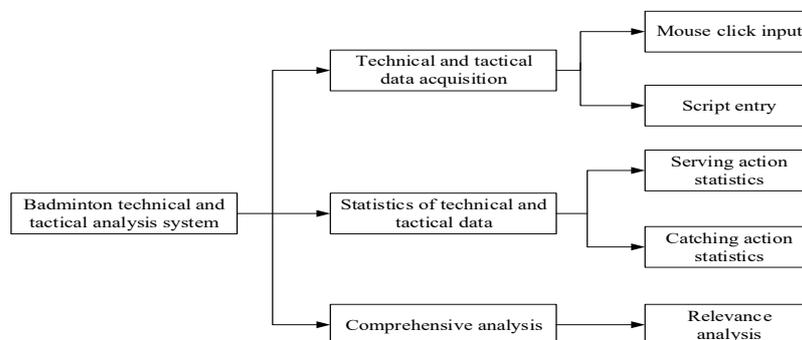


Fig. 7. General framework of system functional architecture

4 sheets were designed in the system design for building database, mainly including the sheet related to the competition information, the information sheet of the team, the player, the coach, the technical action type sheet, and the badminton trajectory. Table 1-4 show the designed sheets of the system.

Table 1. Match information table

Serial number	Name	Type	Field size	NULLS
1	Competition number	Text	60	No
2	Name of competition	Text	60	No
3	Match time	Date	/	No
4	Place of competition	Text	30	No
5	Competition type	Text	20	No
6	Home team number	Number	10	No
7	Visiting team number	Number	10	No
8	home and away	Text	/	No
9	Home team score	Number	10	No
10	Visiting team score	Number	10	No

Table 2. Information sheet for teams, players and coaches

Serial number	Name	Type	Field size	NULLS
1	Team number	Text	20	No
2	player's number	Text	20	No
3	Coach number	Text	20	No
4	Full name	Text	30	No
5	Gender	Text	10	No
6	Date of birth	Date	/	No
7	Height	Number	20	No
8	Weight	Number	20	No
9	Date of registration	Date	/	No
10	Technical number	Number	10	No

Table 3. Technical action type table

Serial number	Name	Type	Field size	NULLS
1	Technical type number	Text	20	No
2	Technical type name	Text	50	No
3	Skill number	Text	20	No
4	Skill name	Text	50	No
5	Remarks	Remarks	50	No

Table 4. Badminton trajectory table

Serial number	Name	Type	Field size	NULLS
1	Trajectory number	Text	20	No
2	Trajectory name	Text	50	No
3	Trajectory alias	Text	50	No
4	Remarks	Remarks	50	No

After the data collection, all the records of one athlete at specific competition were analysed. The data was saved in the script, and each batting was segmented and stored. Then, the program automatically analysed the player's movements and the ball's landing point. Through analysis for competition data, it can be found that the keywords are the court, net, bound, and score. Storing the data to the string, a comparative analysis could be conducted. Ball impact analysis can easily find the athlete's technical movements and habitual hitting position. The statistical results are shown in Figure 8.

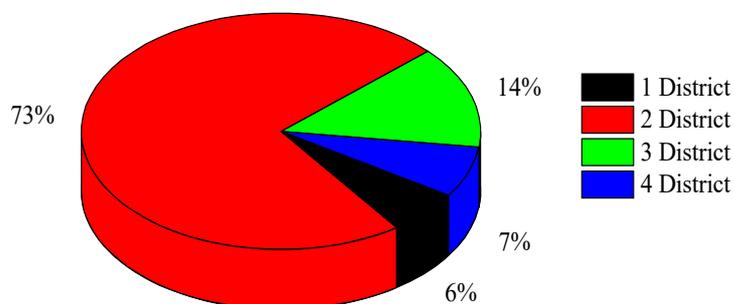


Fig. 8. Ball impact analysis

Figure 8 shows the drop point of the ball when the player uses a movement "back-hand serve". From the statistical probability point of view, when the opponent makes the backhand serve in the competition, the drop position of the ball can be roughly determined, so that the player can calmly respond on the spot. It is also possible to understand the opponent by analysing the technical movements of the opponent. Figure 9 shows the statistical data of technical movement used by the opponent in a competition. It can be seen in the figure, when the hitting action is the backhand serve, the opponent's most action is the forehand push, with the probability up to 40%, so that the opponent's action can be prejudged, and the next action is prepared in advance.

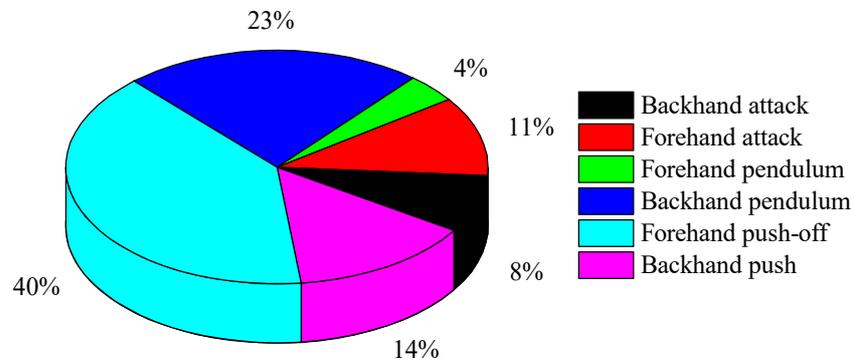


Fig. 9. Technical statistics of opponents

Figure 10 shows the scores of players' gains and losses in the forehand serve. The technical action of the forehand serve was used 50 times in the competition, of which there were 5 service aces, 0 hitting error, and 45 successful batting, indicating that the forehand shot success rate is 100%, and the probability of service ace is 10%. This action is very effective for the player to hit the ball.

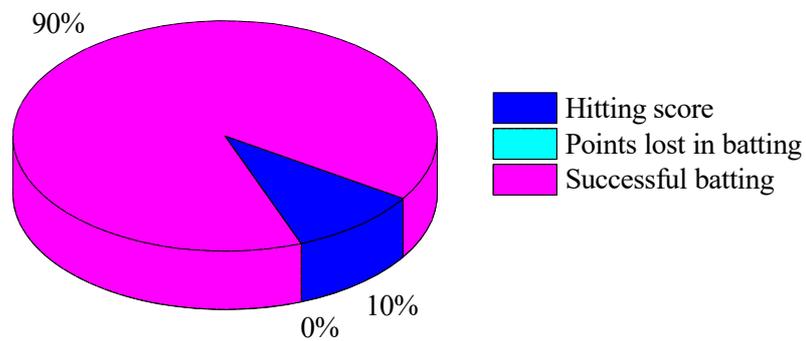


Fig. 10. Statistics of gains and losses

Through the use of data mining for statistical analysis of the techniques in competitive sports, it can be seen that the data mining technology has important significance to improving the athlete's techniques and PE. And the association rules help to systematically analyse the connection between the various technical actions of the athletes, and acquire the scientific and effective data, which provides great support for the technical and tactical analysis in competitive sports.

4 Conclusion

Considering the huge sports-related information and various types of data in competitive sports, it is the primary task of sports researchers to mine out valuable information and summarize the rules using the modern technology. For this, this paper attempts to explore a data mining tool that can be used to guide physical education and technical and tactical analysis. The main conclusions are as follows:

- The mathematical statistics, big data analysis, mathematical algorithms and various modern information technologies were combined together with the hierarchical management mode to establish the information database of competitive sports techniques and tactics.
- Based on the data mining theory, the badminton sport was taken as an example, and data mining tools were used to quickly identify and extract the valuable information from the massive sports competition data.
- Using the data mining tool to extract valuable information from big data, such fast and efficient method can not only assist the coaches to conduct physical education and guidance in a targeted manner, and help the athletes to develop strengths and avoid weaknesses, but also contributes to the improvement of techniques and tactics in competitive sports.

5 References

- [1] Li, Q.H. (2014). Strategy and analysis of sport events based on data mining technology. *Applied Mechanics and Materials*, 687-691: 1137-1140. <https://doi.org/10.10428/www.scientific.net/amm.687-691.1137>. <https://doi.org/10.4028/www.scientific.net/amm.687691.1137>
- [2] Zakamulin, V. (2013). Fooled by data-mining: the real-life performance of market timing with moving average and time-series momentum rules. *Journal of Asset Management*, 15(4): 261-278. <https://doi.org/10.10.2139/ssrn.2242795>.
- [3] Mariscal, G., Óscar, M., Covadonga, F. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2): 137-166. <https://doi.org/10.1017/s0269888910000032>.
- [4] Zhu, S.L. (2018). Research on data mining of education technical ability training for physical education students based on apriori algorithm. *Cluster Computing*, 1-8. <https://doi.org/10.1007/s10586-018-2420-8>.
- [5] Zhang, C. (2014). Research on construction of universities sports management information system based on data mining. *Applied Mechanics and Materials*, 3634(1378): 2702-2705. <https://doi.org/10.10428/www.scientific.net/amm.687-691.2702>.
- [6] Meng, F.H., Li, Q.L. (2013). Application of data mining in the guidance of sports training. *Advanced Materials Research*, 765-767: 1518-1523. <https://doi.org/10.10.2991/icsem.2013.102>. <https://doi.org/10.4028/www.scientific.net/amr.765-767.1518>
- [7] Liang, C.J., Yu, L.J., Wang J.H. (2012). Research on data mining of tennis techniques and tactics based on theory of association. *Computer Engineering & Applications*, 48(7): 246-248. <https://doi.org/10.10.1109/wcse.2010.53>

- [8] Kamiya, K., Funahashi, H., Mano, Y. (2015). A study on the occurrence of sudden death during sports activities in schools, using text-mining analysis. *Journal of Japan Society of Sports Industry*, 25(2): 313-325. https://doi.org/10.10.5997/sposun.25.2_313.
- [9] Castro-Sánchez, A.M., Moreno-Lorenzo, C., Matarán-Pearrocha, G.A., Feriche-Fernández-Castanys, B., Granados-Gámez, G., Quesada-Rubio, J.M. (2011). Connective tissue reflex massage for type 2 diabetic patients with peripheral arterial disease: randomized controlled trial. *Evidence-Based Complementary and Alternative Medicine*, 804321. <https://doi.org/10.10.1093/ecam/nep171>.
- [10] Lupton, D. (2015). Data assemblages, sentient schools and digitised health and physical education (response to gard). *Sport Education & Society*, 20(1): 122-132. <https://doi.org/10.10.1080/13573322.2014.962496>.
- [11] Reddy, V.S., Rao, T.V., Govardhan, A. (2017). Data mining techniques for data streams mining, *Review of Computer Engineering Studies*, 4(1): 31-35. <https://doi.org/10.18280/rces.040106>
- [12] Chahbandarian, G., Bricon-Souf, N., Megdiche, I., Bastide, R., Steinbach, J.C. (2017). Predicting the encoding of secondary diagnoses. An experience based on decision trees, *Ingénierie des Systèmes d'Information*, 22(2): 69-94. <https://doi.org/10.3166/ISI.22.2.69-94>.
- [14] Ntoumanis, N. (2001). A self-determination approach to the understanding of motivation in physical education. *British Journal of Educational Psychology*, 71(P2): 225-242. <https://doi.org/10.10.1348/000709901158497>.
- [15] Standage, M., Duda, J.L., Ntoumanis, N. (2003). A model of contextual motivation in physical education: using constructs from self-determination and achievement goal theories to predict physical activity intentions. *Journal of Educational Psychology*, 95(1): 97-110. <https://doi.org/10.10.1037//0022-0663.95.1.97>.
- [16] Dimitrijević, R., Koropanovski, N., Dopsaj, M., Vučković, G., Janković, R. (2014). The influence of different physical education programs on police students' physical abilities. *Policing: An International Journal of Police Strategies & Management*, 37(4): 794-808. <https://doi.org/10.10.1108/pijpsm-05-2014-0060>.
- [17] Hagger, M., Culverhouse, T., Chatzisarantis, N., Biddle, S. (2003). The processes by which perceived autonomy support in physical education promotes leisure-time physical activity intentions and behavior: a trans-contextual model. *Journal of Educational Psychology*, 95(4): 784-795. <https://doi.org/10.10.1037/0022-0663.95.4.784>.
- [18] Mordal-Moen, K., Green, K. (2014). Physical education teacher education in norway: the perceptions of student teachers. *Sport Education & Society*, 19(6): 806-823. <https://doi.org/10.1080/13573322.2012.719867>
- [19] Chen, S., Kim, Y., Gao, Z. (2014). The contributing role of physical education in youth's daily physical activity and sedentary behavior. *BMC Public Health*, 14(110): 1-7. <https://doi.org/10.10.1186/1471-2458-14-110>.
- [20] Fogel, V.A., Miltenberger, R.G., Graves, R., Koehler, S., Zarcone, J. (2010). The effects of exergaming on physical activity among inactive children in a physical education classroom. *Journal of Applied Behavior Analysis*, 43(4): 591-600. <https://doi.org/10.10.1901/jaba.2010.43-591>.
- [21] Baert, Q., Caron, A.C., Morge, M., Routier, J.C. (2018). Fair task allocation for large data sets analysis, *Revue d'Intelligence Artificielle*, 31(4): 401-426. <https://doi.org/10.3166/RIA.31.401-426>.
- [22] Bikku, T. (2019). An indigenous tool (NoJavaCloud) to handle virtual nodes to simulate the cloud tasks, *Mathematical Modelling of Engineering Problems*, 6(1): 85-91. <https://doi.org/10.18280/mmep.060111>.

6 Author

Lili Pan (1985.03.19) female, born in Baoji, Shaanxi province. She is a postgraduate, and work in Baoji university of arts and sciences as a lecturer. She interests in physical education training and sports culture. In the undergraduate course teaching evaluation, she has made certain contribution to the teaching and research section, at the same time participated in the revision of "aerobics" teaching syllabus, often participated in teaching and research activities of the teaching and research section, teaching and research activities without being frequent. Actively participated in the judging work of aerobics and cheerleading competitions in various districts of Shaanxi province and Baoji city. In order to continuously improve my teaching skills, I published more than 10 papers, including 3 core journals, 3 utility model patents, and presided over and participated in 5 subjects.

Article submitted 2019-07-21. Resubmitted 2019-09-13. Final acceptance 2019-09-15. Final version published as submitted by the authors.