Wireless Pers Commun (2018) 102:2797–2805
https://doi.org/10.1007/s11277-018-5304-6

Automatic Evaluation Model of Physical Education Based on Association Rules Algorithm

Tong Zhang[1]

Published online: 8 February 2018

Abstract As physical education rapidly develops in China currently, various technology, especially the computer technology is also used in the physical education. Automatic evaluation model of physical education based on association rules algorithm was studied in the paper. Association rule recommendation algorithm was firstly expounded in this paper. Then, the indexes of automatic evaluation model of physical education based on the collaborative filtering algorithms of association rule were established. Through the real teaching data of one university, the algorithm established in the paper was tested, and the process of physical education in colleges and universities was evaluated. Finally, it was concluded that the algorithm could effectively describe and evaluate the behavior of the physical education teaching and had the potential for popularization.

Keywords Association rules algorithm · Physical education · Automatic evaluation

1 Introduction

In recent years, with the rapid rise and the rapid development of the Internet industry, as well as the rapid upgrading of mobile terminal such as smart phones, mobile Internet has penetrated into our life in all aspects. Everyone has a smart phone every day and "closely contact" with mobile Internet from social networks to electronic commerce, from taxi-hailing app to take-out platform [1]. Thus, people begin to explore how to propose targeted evaluation method according to the data characteristics for physical education based on the existing commodity recommendation method of the traditional electronic platform [2]. Centering at the issue, discussion and the empirical analysis was conducted in the paper. This paper firstly introduced the present mainstream recommendation methods (mainly

✉ Tong Zhang
BurgessAaV@yahoo.com

1 Department of Physical Education, Zhongnan University of Economics and Law, Wuhan, China

including content-based evaluation, association rules evaluation and evaluation based on collaborative filtering) in the field of sports evaluation, and also introduced the core recommendation algorithm used by the evaluation methods. As for research methods, this paper explored the associated information between commodities and commodities, and commodity categories through the traditional method of association rule. Besides, with the classic Apriori algorithm in association rules algorithm as the foundation, the paper tried to make judgment by increasing the goods and users' location information, evaluated the physical education according to the real data of the mobile terminal behavior, and tested the effect.

2 State of the Art

In theory research, foreign scholars defined association rule for the first time in 1993, and put forward a kind of classic algorithm Apriori algorithm of mining frequent itemsets in 1994. Apriori algorithm laid a solid foundation for the studies of association rule mining algorithm; many scholars and experts improve or optimize it in their researches based on Apriori algorithm, and propose new methods such as DHP algorithm, close algorithm, Partition algorithm and sampling algorithm [3]. Later, some scholars put forward multi-level association rules in 1995. A great number lots of scholars carried on the thorough discussion of multi-level association rule mining problem, and put forward such as new algorithms such as ML_T2L1 [4]. System collected user ratings data through the browser page and displayed the evaluation results, and such way was more convenient and intuitive [5]. The system, such as GoogleNews was dedicated to personalized music evaluation, and specifically developed for large user groups and frequently updated news field [6]. Therefore, foreign personalized service and evaluation technology have been extensively used in various fields. These evaluation system not only provide convenience to the user experience, but also bring very considerable flow, profit and propaganda interests to merchants (Huang) [7]. Although China has got a certain development in the related researches of recommendation service and recommendation technology in recent years, China is still relatively backward in terms of the applied research on association rules in the commodity recommendation system. In addition, China needs to improve the existing recommendation system in depth, the recommended quality, scale and personalized recommendation mainly because China started late and lags behind in relevant theoretical research [8]. In recent years, especially after 2004, with the rapid development of domestic Internet industry, China has begun to undertake the corresponding theoretical researches and special applications.

3 Methodology

3.1 Association Rules Recommendation Algorithm

Association algorithm is important in data mining. Richard Armitage and others put forward the idea of mining the association rules between itemsets in customer transaction data itemsets for the first time in 1993, which core was recursion method based on the thought of two-stage frequent itemset [9]. The recommendation based on association rules has been successfully practiced in the real retail industry such as Wal-Mart, and widely used in online commodity recommendation. To be specific, association rules aim to explore the correlation between all kinds of goods.

As for the application scenarios of this recommended method, after a large number of users buy A, and then buy B and C, the system will recommend the users B and C once it finds that one user buys A. This application scenario is very appropriate for shopping, because shopping is often based on the current demand, that is to say, users' shopping points of interest is in change over time. Association rules can recommend users the goods that are not similar at all. For example, after users buy a digital camera, the system will recommend an SD card, digital camera batteries, etc. Professional term in the field of recommender systems is the diversity of the recommendation results. the most important of aspect of association rules is to find co-occurrence relation or rule (mining frequent itemsets), and the involved classic algorithms include Apriori algorithm, FP- Grmvth algorithm [10]. The association rules algorithm was further introduced in the following part.

The first factor is support degree, and it is for itemsets. A itemset support degree is the number of the itemset occurrences divided by the total record number (transaction number). Support degree can measure the frequency of itemsets in the entire transaction set. When we found the rules, we focused on high frequency of itemsets.

$$\text{Support}(\{a,b\}) = \mathit{frequency}(\{a,b\})/\mathit{frequency}(\mathit{all\ item\ sets}) \tag{1}$$

The second factor is the degree of confidence, and it is for association rules. Formula of the confidence of association rules $\{a,b\} - > \{c\}$; rule confidence measures the proportion of the number of occurrences of a set{a, b, c} of the occurrence of {a, b}, that is the probability of occurrence of {c} under the condition of {a, b}.

$$\mathit{Confidence}(\{a,b\} \to \{c\}) = \text{Support}(\{a,b,c\})/\text{Support}(\{a,b\}) \tag{2}$$

For lift degree, we need to pay attention to a problem when looking for possible lift rules; when RH} support degree has been quite significant, this rule may be invalid even if the rules have high degree of confidence. For example, among the analyzed 10,000 transactions, 6000 transactions contained x; 7500 contained y; 4000 transactions contained both at the same time. Support degree of lift rules is 0.4, seemingly quite high, but the lift rules is a mistake. After users have purchased x, they will buy Y with the probability of $4000 = \text{b000}) - \text{D.bb7}$; without any preconditions, users will buy Y with the probability of $(7500 = 1\ \text{DDDD}) - \text{D}.\ 75$. In other words, setting up such conditions that users buy X can reduce the users' probability of buying Y, so the x and Y are mutually exclusive. Therefore, it is necessary to introduce this concept of lift. It can measure the independence of the item set {a, b} and {c}. to be specific, if $\mathit{Lift}(\{a,b\} - > \{c\}) = 1$, it shows that the two are mutually independent. If the value < 1, it suggests A condition (or the occurrence of event A) and B are mutually exclusive; we don't admit that mining association rules is valuable unless the lift degree in the usual data mining is more than 3.

$$\mathit{lift}(\{a,b\} \to \{c\}) = \text{Support}(\{a,b,c\})/\text{Support}(\{a,b\}) \times \text{Support}\{c\} \tag{3}$$

The generation of association rules is generally divided into two steps. The first step is to find frequent itemsets. N items can produce $2^\wedge(n-1)$ itemset. So, we needed to specify the minimum support so as to filter out the frequent itemsets. The second step is to find out the centralized rule in the first step. Thus, we needed to specify the minimum degree of confidence to filter out the weak rules.

3.2 Evaluation Index Based on Collaborative Filtering Algorithm

Collaborative filtering recommendation is one of the association rule algorithm, and its recommendation method in the recommended system of physical education curriculum is one of the earliest and most widely used method. It is based on the following assumption: to help A to find the type of physical education curriculum he will choose or tend to choose, the first step is to find the user B who has the similar interests with user A, and then to recommend the physical education curriculum that user B has chosen to user A. Usually, the distance between the users is calculated based on user class information record, and then how much user A likes physical education curriculum a is evaluated through the weighted value of user B's evaluation of physical education curriculum who is closest to user A, and finally recommendation is conducted for user A according to the degree of preferences to the curriculum. It is very easy to understand because we tend to refer to the opinions of the good friends or the previous physical education curriculum in daily life when choosing curriculum, and friends are the user B with similar interest to us. Collaborative filtering methods are divided into items-based collaborative filtering and user-based collaborative filtering; the former calculates the distance between the users and the latter calculates the distance between the physical education curriculum. Collaborative filtering recommendation can explore users' potential needs and preferences, and evaluate rather than just according to the historical curriculum information of users, and it does not need information about the characteristics of physical education curriculum. Thus, collaborative filtering evaluation can evaluate the physical education curriculum that is difficult to be labeled and which characteristics are difficult to be described. In fact, collaborative filtering algorithm also has many defects, and there are a lot of problems to be solved. Collaborative filtering algorithm involves users' physical education teaching evaluation, so the accuracy of prediction score is mainly in view of the collaborative filtering, while the evaluation index is mainly used for evaluating the system that needs to show users rating.

In short, Prediction score, the evaluation index the accuracy of prediction score is to calculate difference between the system prediction score and the actual scores of users. Many indexes can be used to calculate the difference, such as the mean absolute error, the standard mean absolute error, mean square error (MSE) and root mean square error. Assume T is the test set; rAa is the user A's rating of curriculum A; rA is the user A's rating of curriculum A predicted by the system, rMAX and rMIN are the highest score and lowest rating of the user. So, the calculation formula of above indexes is as follows:

$$MAE = \frac{1}{|T|}\sum_{A,a \in T} |r_{Aa} - r_{Aa}| \tag{4}$$

$$NMAE = \frac{MAE}{r_{MAX} - r_{MAX}} \tag{5}$$

$$MSE = \frac{1}{|T|}\sum_{A,a \in T} (r_{Aa} - r_{Aa})^2 \tag{6}$$

$$RMSE = \sqrt{\frac{1}{|T|}\sum_{A,a \in T} (r_{Aa} - r_{Aa})^2} \tag{7}$$

However, subjective factors play major part in user's rating. Even if the system can do excellently in recommending product to the user, there is usually is always larger deflection in score prediction for the users, so most of the websites provide the recommendation of commodity rather than commodity rating predicts.

3.3 Data Source

In this paper, we adopted the true teaching data of one university, with the time span of a month (November 18 ~ December 18), simulated and recorded users' behavior, and evaluated the whole curriculum. All the data were divided into two parts. The first part was the users' mobile terminal behavior data in the curriculum information system, and table was named with tiatichi_mobile_recommend_train_user. The second part was curriculum subset and the table was named wih tianchi_mobile_recommend_train_item. First, different userid field corresponded to different users. Because the same user could register several accounts in the system for curriculum selection, userjd and real person didn't follow one-to-one correspondence. In addition, to protect the privacy of users, the field has been was desensitized, and it is not real user ID. Similarly, itemjd and teaching mode followed one-to-one correspondence, and item_category was coding of teaching curriculum's teaching mode, and it was desensitized. Behaviorjype corresponded to the users' different operation types of behavior.

4 Result Analysis and Discussion

In this paper, the algorithm was tested. First, the data information of sports teaching was explored through the algorithm so as to know the users' behaviors, including click and collection. Users select teaching mode by click, and add their favorite teaching mode in favorite. Favorites itself have capacity limits (50 teaching modes). User can directly select the teaching modes, or delete the contents in the favorites. The data didn't provide the behavior that users deleted the curriculums in the favorites; favorites itself has no capacity limits, so users can select curriculums directly in the favorites, or delete curriculum data in the favorites that do not provide users curriculum deletion behavior in the favorites.

Table 1 user_geohaSh is the hash value of user space location. The hash value of the location is generated through a space hash algorithm. In order to protect users' privacy, the specific calculation process is confidential. The hash algorithm refers GeoHash algorithm, and the range of accuracy represented the hash value is about 150 m × 150 m of rectangle. The prefix of hash value represents a bigger particle size of rectangular grid; for example, the prefix of length 6 corresponds to UOOmMOOm rectangle; 99suwas corresponds to a 150 m × 150 m rectangular, and 99suwa corresponds to a 1200 m × 600 m rectangular that contains 99suwas. Time field corresponds to the specific time of the user operation, and its precision reaches hourly level. For example, 2014112517 shows that users operate the curriculum from PM 17 to PM 18 November 25th, 2014. In the table of tianchi_mobile_recommend_train_item, item_id and item_category has the exactly same meanings with tianchi_mobile_recommend_train_user; item_geohash is the hash value of corresponded curriculum spatial location (Tables 2, 3).

Users' operation behavior was preliminarily explored. There were a total of 12256906 pieces of data in tianchi mobile _ _ how _ train_user table, involving 2876947 curriculum information, which could be divided into 8916 classes. (With item_id and item_category as

Table 1 tianchi mobilerecommendtrainuser

Field	Field description	Extraction description
userid	User identity	Sampling & field desensitization
itemid	Commodity identification	Field desensitization
behaviortype	User to commodity	
Type of behavior	Including browsing, collecting, adding shopping cart and buying, the corresponding values are 1, 2, 3 and 4 respectively.	
usergeohash	The spatial identity of the user location can be empty	Generated by the algorithm of latitude and longitude by secrecy
itemcategory	Commodity classification label	Field desensitization
Time	Behavior time	Precision to hourly level

Table 2 tianchi_mobile_recommend_train_item

Field	Field description	Extraction description
itemid	Commodity identification	Sampling & field desensitization
item_geohash	The spatial identification of the location of a commodity can be empty	Generated by a secrecy algorithm
itemcategory	Commodity classification label	Field desensitization

Table 3 The hash value and its spatial location range

Length	Lattice size
1	5009.4 km × 4992.6 km
2	1252.3 km × 624.1 km
3	156.5 km × 156 km
4	39.1 km × 19.5 km
5	4.9 km × 4.9 km
6	1.2 km × 0.61 km
7	152.9 m × 152.4 m

the standard), the table included 10,000 user ID. Within a month, all users browsed the curriculum; 6730 users add the curriculum to favorites; 8886 users carried out on the physical curriculum. However, it was important to note that most of the users only operated a curriculum within a month, and that a lot of types of curriculums were only operated by a user within a month. For example, if behavior_type 4 (the case of class) was only considered, 94.99% of the curriculum was operated by some users in a month. The problems of the data themselves set up certain obstacles for subsequent association rules mining.

In addition, as shown in Fig. 1, the user behavior in class had distinct time characteristics. User operation time was concentrated on 7 o'clock in the morning time to 22 o'clock, showing that the user behavior on the mobile end mainly happens at the school learning time. In addition to A.M. 1 ~ 9 to P.M. 4 ~ 6, the customer s had similar behavior activity in class. The explanation for this phenomenon was as follows: the former time range was the time for rest and going to school for most people, and the latter time range was after school time. Within the two periods, because of the user's work and rest time and the environment, most users won't involve sports teaching. To sum up, users often browsed, selected curriculum in their spare time, and the rest time was the main sports teaching time.

Figure 2 describes the users' overall operating time distribution respectively, and the distribution of browsing behavior and collection time. As seen from it, the users were more obviously different during different time periods, suggesting browsing, collection, etc. in the users' various types of behaviors did not finally lead to the development of physical education curriculum.

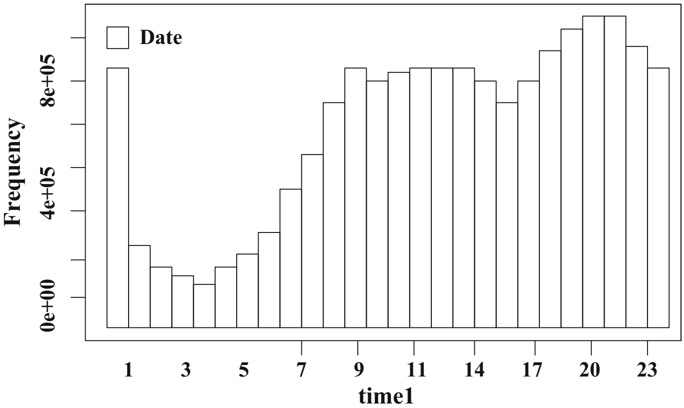

Fig. 1 User's operating time distribution

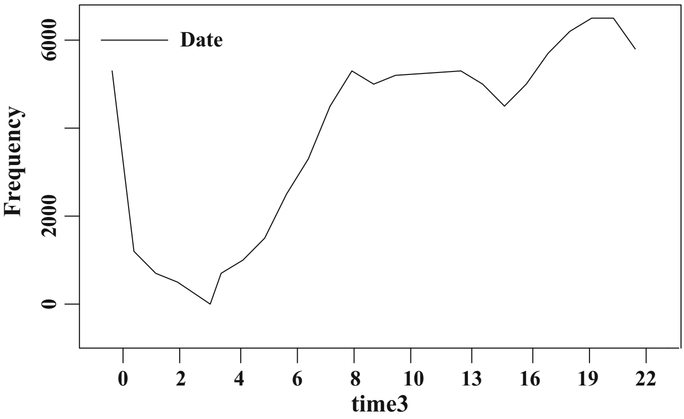

Fig. 2 Distribution of users' purchase time

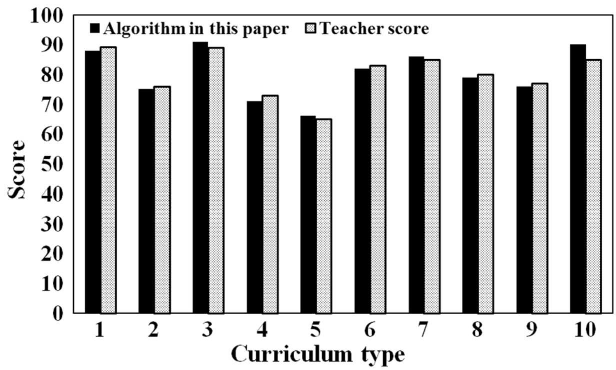

Fig. 3 Comparison between the scores of teachers and the results of the algorithm in this paper

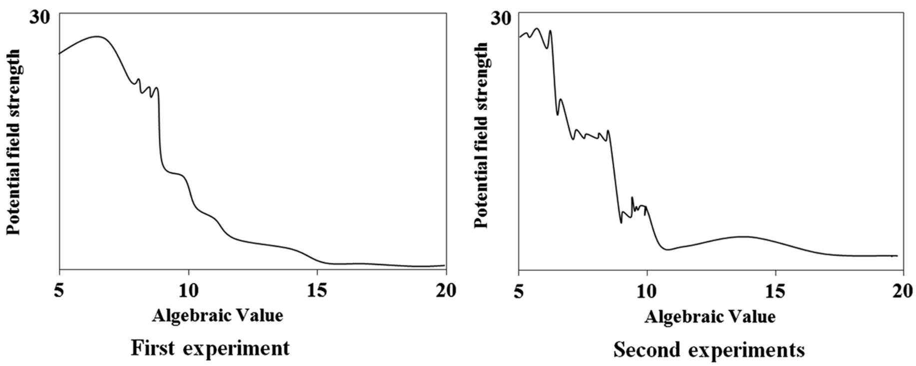

Fig. 4 Convergence test of the algorithm

Then, the curriculum was evaluated through the algorithm built in this paper. We selected 10 types of physical education, and 10 professional sports teachers, and asked the teachers to rate the effect of the classroom, and solved the average scores of 10 teachers (the full score was 100) (Fig. 3). Besides, we compared the scores of teachers and the results of the algorithm in this paper to examine the accuracy of the algorithm. The following results were obtained: the algorithm was similar to the artificial scoring results; the biggest difference occurred in the tenth section of class, which error reached 5%. Next, the convergence of the algorithm was investigated, and the Fig. 4 was obtained. The convergence result showed the algorithm had good convergence and strong stability.

5 Conclusion

At present, the undertakings of physical culture and sports are developing vigorously, and sports teaching have also made fairly rapid growth; sports teaching integrates a multi-disciplinary technology, especially computer and Internet technology; the fusion of technology promotes the sports teaching China to achieve a new development. In this background, automatic evaluation model of physical education based on association rules algorithm was studied in the paper. Firstly, this paper expounded association rules recommendation algorithm, collaborative filtering algorithm based on association rules, and established a series of automatic evaluation indexes of physical education curriculum on the basis. Secondly, through the real teaching data of one university, association rules algorithm built in this paper was used for data mining, and the users' a series of operation on curriculum was mainly analyzed. Besides, the curriculum was evaluated with the algorithm constructed in the paper. The comparison between the scores of teachers and the results of the algorithm in this paper showed that the algorithm had good precision. Furthermore, the convergence test suggested that the algorithm had strong stability and the potential for popularization.

References

1. Watanabe, K., Lopez-Colon, D., Shuster, J. J., & Philip, J. (2015). Abstract 15507: CPR education in middle school students during physical education class. *Circulation, 132*(Suppl_3 Su), 145.
2. Zhou, L. J., Li, S., & Liu, T. (2017). Application of association rule in students' results in University. *Advanced Materials Research, 1100*(181), 991–992.
3. Shu Juan Zhang, & Wang, Q. M. (2016). The research and application of association rules algorithm. *Applied Mechanics and Materials, 2418*(325), 1090.
4. Peng, N. X. (2016). An efficient weighted association rules mining algorithm. *Applied Mechanics and Materials, 2490*(333), 377–379.
5. Zhang, L., & Zhang, J. L. (2017). Association rules mining algorithm based on linked list. *Applied Mechanics and Materials, 2617*(380), 216–217.
6. Qi, W. Q., Li, Y. R., Ye, H. F., Duan, D. P., & Jiang, X. C. (2017). Research on classification of partial discharge of switchgear cabinets based on a novel association rule algorithm. *Applied Mechanics and Materials, 2808*(448), 667.
7. Hung, J. S. (2015). Analysis model of computer network security teaching founded on apriori algorithm. *Applied Mechanics and Materials, 3727*(707), 1117.
8. Song, A., Ding, X., Chen, J., Li, M., Cao, W., & Pu, K. (2016). Multi—objective association rule mining with binary bat algorithm. *Intelligent Data Analysis, 20*(1), 389.
9. Chithra, R., & Nickolas, S. (2017). A novel algorithm for mining hybrid-dimensional association rules. *International Journal of Computer Applications, 1*(16), 176–179.
10. Cao, X. (2016). An algorithm of mining association rules based on granular computing. *Physics Procedia., 33,* 765.

Tong Zhang (1986) male, Xi'an, Shaanxi Province, Department of physical education, Zhongnan University of Economics and Law. The lecturers, Master graduate students, mainly engaged in the study of physical education and training.