

Distance Education Recommended Sequence Key Technology Research and Implementation*

Z.F.QIAO, J.X.GUO, J.C.ZHAO

¹*Institute of Information on Science and Technology of Agriculture, Beijing Academy of agriculture and forestry Sciences, Beijing 100097, China*

²*The Research Center of Beijing Engineering technology for Rural Remote Information Services, Beijing 100097, China*

ABSTRACT: This paper analyzes user learned historical data's business characteristics of distance education platform. After comparison and analysis using PrefixSpan algorithm, proposed a model with the ability to track the sequence of interest is recommended changes. Research on key technology of sequence databases, sequence mining, sequence recommendation. Design a distance education platform sequence recommendation system, and achieved good results in practical applications.

KEYWORD: Sequence Recommendation; Sequence; Frequent Sequences; PrefixSpan

1 GENERAL INSTRUCTIONS

As a new intelligent tools, recommender systems have been more and more used in the web site in recent years. It can identify the user's preferences and recommend the real things of interest to the user, so as to improve the website service quality, reduce the processing burden on users in the choice of products, information. The collaborative filtering algorithm is the most widely used in the recommendation system, it is through the analysis of historical data of customers, generate interest in the most similar neighbor set, recommend their most things of interest to the user. Although when given enough clear preference information, this arithmetic can show good performance, but because of the interest of the user is constantly changing, with the increase of time, because of the static characteristics of simple integration of all the historical data, it will inevitably lead to recommend a low performance^[1]. This paper will consider the change of user interest degree, study and implementation of a recommendation algorithm based on user behavior sequence, for each user to provide video sequence with features of personality interest recommendation, which can solve the problem of changing customer interest in a certain extent.

2 RECOMMENDATION ALGORITHM BASED ON USER INTEREST DEGREE

In this paper, the proposed algorithm is based on such a consideration: Generally time of user selected something of interest is recorded. Use user learning record data in L time long for user analysis, L is divided into N segments, N-1 segment of anterior time target user data is known, find the N time segment recommend products for the target user. The realization of the algorithm has four main stages: data pretreatment, rule mining, recommended, feedback.

In the preprocessing stage, the user data are segmented by learning history, concept hierarchy and association rules to reduce the data sparseness, establish customer profile;

Rule mining stage, based on user preference files for clustering of customers, and get the user behavior sequence according to the model of each period, so as to obtain the behavior sequence association rule;

Recommendation phase, found the behavior sequence of target user, compare Similarity the behavior sequence of target user and rules, and recommend the selected rule;

The feedback phase, tracking whether the user click recommended resources, give feedback to the system, as a reference to recommend the next time.

*Foundation Items: Construction and application of provincial rural information service platform in developed area (2014BAD10B02); Development and Application of the Series Products of the NongKeTong for Beijing National Modern Agricultural City for Science and Technology (D141100004914002).

3 RECOMMENDED SEQUENCE DESIGN

3.1 Data Pretreatment

To make the sequence pattern algorithm can mine valuable information, must provide rules, accurate, clean, concise source data---view transaction set. Because watch video records is the number of incomplete rules, so the need for video viewing record data preprocessing.

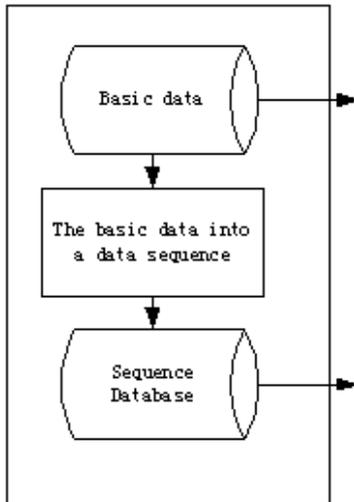


Figure 1. Data Pretreatment figure.

3.2 Sequence Rule Mining

Personalized video recommendation is for users to recommend, and the last learning video is the most worthy of recommendation according as the priority, so the last learning video. However, if only for the one last learning video to watch as the basis of accuracy is not high, we use a video sequence recently learn to recommend, it can better improve the recommendation accuracy. Because the user learning sequence is too long in the sequence algorithm will increase the number of calculations and will have a greater impact on the recommended amount, and the amount and time of the user viewing the video of uncertainty.

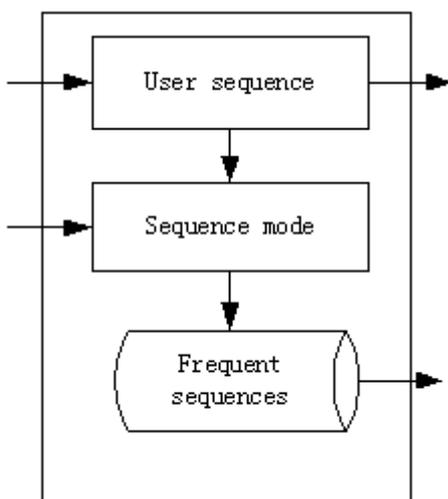


Figure 2. Sequence Rule Mining figure.

To this end, we use reverse generated user learning sequence method to construct user known sequence, Firstly, do all the user learning recording data processing, find the number of videos learned daily average as the length of the sequence to generate a user sequence. And then, from the user finally learning video recording sequence begins, select forward to watch the recorded data as a known sequence. The length of known sequence is users average daily number of videos watched.

After comparing sequences recommendation of GSP, SPADE, PrefixSpan. we chose to use PrefixSpan algorithm. Based on user learning historical data, the use of PrefixSpan algorithm calculates the number of average learned video every day, put this value as a user sequence length, using the user ID to obtain the last record, this video is user recommended sequence starting point, reverse find user recommended sequence. To user sequence as frequent sequences in the algorithm, the timing task processing good sequence database as a sequence database, the above two frequent sequence and the sequence database as PrefixSpan algorithm parameters, using the algorithm to generate frequent sequence is saved to the frequent sequence database.

PrefixSpan uses the idea of divide and conquer, projection database for multiple smaller constantly generated sequence database, then in each projection sequence pattern mining. PrefixSpan algorithm is a kind of pattern sequences based on projection growth algorithm.

PrefixSpan algorithm is a depth first search algorithm[2], the basic idea is to use the frequent prefix partition the search space and the projection sequence database, and search for related sequences. First check the prefix sub sequence, only the corresponding suffix subsequences projection to the database. The algorithm also uses the divide and conquer strategy, the projection database for multiple smaller constantly generated sequence database, then in each projection sequence pattern mining.

PrefixSpan algorithm implementation steps are described as follows:

- (1) The sequence database scanning once got all the frequent item n, that all the frequent sequences with 1 length set;
- (2) A subset of the complete set of frequent sequences of frequent sequence is divided into n with a different prefix;
- (3) By projecting the database structure and the corresponding in which recursively find a subset of frequent sequence mining.

The sequential database such as shown in Table 1, for a given minimum support threshold $min_sup=2$.

Table 1. sequential database

ID	sequential
1	<a(abc)(ac)d>
2	<(ad)c(bc)a>
3	<abdcb>
4	<acbc>

The projection database and sequence patterns such as shown in Table 2.

Table 2. Sequential patterns

prefix	projection database	sequential pattern
<a>	<(_bd) (bcd) (bcd)> <(_bc)><(_b)(bcd) >	<(ab)><ac><ad> <abc>< abb><abd>
	<(_d) (bcd) (bcd)> <(abc)><(bcd)>	<bb><bc><bcd> <bbcd><bd>
<c>	<(_d) (bcd)><(_d)>	<cd>
<d>	< bcd) (bcd)>	<d>

3.3 The personalized recommendation and analysis

Personalized analysis is to analyze the user history records and generate user learning sequence. Historical data for data mining to the customer, the maximal frequent sequence generated by PrefixSpan algorithm. Then compare similarity. Finally by choosing rules to select the appropriate video collection. Choose appropriate amount of video recommended to the user.

Step 1: use of credibility to select video. In the maximal frequent sequence (including the father sequences of projection sequence itself and all the subsequence projection sequence), According to the order of the number of individual and individual in a sequence of sequence in the parent and child generated similarity values, According to the similarity of the high and low to sort all frequent sequences, Recommended according to the similarity in the process of high and low priority selection father sequence projection sequence, then a recursive select various subsequence projection sequence, respectively. In the sequence of selected support is further used to select video.

Step 2: select video by support. The user sequences or user subsequence has an a to multiple projection sequences in the frequent sequences, In the projection sequence has a multiple video sequences behind user last learning sequence. Calculated maximal frequent sequence for each video in the algorithm, count the number of sequence of length 1 occurrences. preserve these videos occurrence number after the user last learned video . The occurrence number is support value. In the case of the sequence of the same length, in recommending the video to the user, to support high video is preferred, when the number does not meet

the recommended system requirements, you can select the lower support degree video , and turn recommended sequence until meets the recommended requirements.

Step 3: the candidate is too small solutions. In the initial stages of system operation, the number of users to learn too little, or not the amount of registered users, viewing fewer, candidate sequences situation appears inadequate. shorten the length of the sequence by the user, reducing the relatively constrained frequent sequences, increasing the range to meet the sequence that contains the user.

Step 4: Recommended video. Through the above analysis and screening process for special cases, the system will be able to filter out high to low sequence similarity with the user's frequent sequences, with the confidence and support system filter out the candidates of videos, then, select the video recommended to the user to complete the video testimonials.

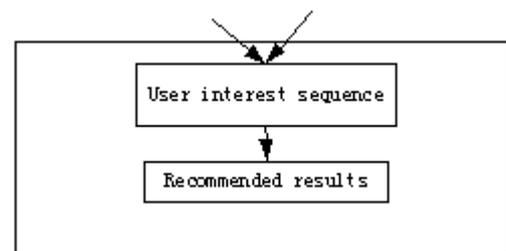


Figure 3. Sequence recommendation.

4 CONCLUSIONS

On the basis of the sequential pattern algorithm of Data Mining, full use of historical data user learning in distance education platform. From the application point of view, presents a series of personalized recommendation model. Using this method to extract user preference model views and recommendations. Has the accuracy, timeliness and operability, Distance education can make faster, more accurate and better information services to meet the needs of individual users. Core aspects of this model is sequential pattern mining frequent sequences and user preferences sequence generation.

REFERENCES

- [1] Che Gaoying, Zhang Lei, Zhang Luxu. Users browsing behavior patterns based on the sequence of extraction and analysis. Computer Technology and Development, sep.2012, Vol.22 No.9.
- [2] Wang Zheng, Li Jinhong. Application of improved PrefixSpan algorithm in counting passenger flow. Computer Era No. 6 2013.
- [3] Chen Ling, Bhowmick S S, Nejd W. COWES: Web user clustering based on evolutionary web sessions. Data & Knowledge Engineering, 2009, 68(10): 867-885.

- [4] Yang Zhenyu, WangWeiping. User access sequences based on the real-time web recommendation. *Computer system & applications*, 2008, (5):50- 54.
- [5] A.M.Ferman and A.M.Tekalp, Two-stage hierarchical video summary extraction to match low level user browsing preferences. *Multimedia, IEEE Transactions on*, v01.5, PP. 244-256, 2003.
- [6] Song Leyi, Xiong Hui, Zhang Rong. Towards the next generation of mobile recommender systems. *Journal of East China Normal University (Natural Science)*. May 2013, No.3.
- [7] Guo Fuliang, Zhou Gang, Cui Liangzhong. Research on model of user behavior habit authentication in cloud computing. *J.Huazhong Univ. of Sci. & Tech. (Natural Science Edition)*, Dec. 2013, Vol.41 Sup. II .
- [8] Xia Ying, Sun Chongwu. POI recommendation method based on pattern matching of spatio-temporal sequence. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, Jun. 2011, Vol23 NO.3.
- [9] He Jinjing. Mining Users Potential Interested in Personalized Information Recommendation Service. *Journal of Modern Information*, Apr.2013, Vol.33 NO.4.
- [10] Chen Yifei, LiZhiJun, Jiang Shouxu. Towards Sequential Recommendation of Passenger-finding Locations in An Urban Taxi Service. *Intelligent Computer and Applications*. Dec. 2013, Vol.3 NO.6.