An Improved Personalized Recommendation System Research

Xingyuan Li
Information Engineering College
Ningbo Dahongying University
Ningbo, China
19397334@qq.com

Qingshui Li[*]
Computer Science and Technology College
Zhejiang University of Technology
Hangzhou, China
mdusa@zjut.edu.cn
[*] corresponding author

Abstract—**In order to find information of interest and found valuable information resources in enrich Internet data. This paper describes a personalized recommendation system, personalized recommendation system is an intelligent recommendation system to help e-commerce site for customers to provide complete personalized shopping decision support and information services. for the User Rating data extreme sparseness, This paper presents nearest neighbor collaborative filtering algorithm based on project score predicted ,experiments show that this method can improve the quality of recommendation system.**

Keywords-Personalized recommendation; association rules ;Collaborative filtering recommendation; styling; Project Score

I. INTRODUCTION

With the growing popularity of Internet technology, the network was more widely used in people's lives, learning of all areas, and the Internet's vast information on complex data mining technology for a better application of the stage. As the internet information great wealth of online resources, personalized recommendations drawn increasing attention. It also demands access to information is no longer a full possession of information resources, but to get useful information resources, and thus the purpose of access to information gradually shift from the recall to precision, to meet specific needs of user-specific information become new environment target in the network of information services. Personalized recommendation system has become an important research content of e-commerce IT technology, and more and more get researchers' attention, At present, almost all large e-commerce systems, such as Amazon, CDNOW, eBay, Dangdang online bookstore, etc., are various forms using personalized recommendation system in varying degrees.

II. NTRODUCTION OF PERSONALIZED RECOMMENDATION SYSTEM

Personalized recommendations based on user characteristics and buying behavior of interest, recommended to the user of interest information and commodities. As e-commerce continues to expand, the number and types of rapid growth of goods, customers need to spend a lot of time to find their own to buy goods. This browser a lot of irrelevant information and products will undoubtedly submerged in the process of information overload in the continual loss of consumers. To solve these issues, personalized recommendation system came into being [1].

Personalized recommendation system is an advanced business intelligence platform which built on massive data mining basic, to help e-commerce site for their customers shopping to provide complete personalized decision support and information services.

III. PERSONALIZED RECOMMENDATION SYSTEM ALGORITHM

E-commerce recommendation system main algorithms are:

1) The recommendation algorithm based on association rules

2) The recommendation algorithm based content

The recommendation algorithm based content the advantage of simple and effective. The disadvantage is the limited capacity of feature extraction, over-refined, Content-based recommendation system can not find new customers interested in resources, and only find similar resources which customers have been interested. This method is usually limited to easily analyze the content of the products recommended, but for some of the more difficult to extract the contents of the goods, such as music CD, movie, etc. recommendations can not produce satisfactory results.

3) Collaborative Filtering Recommendation

Collaborative filtering quickly became a very popular technology in information filtering and information systems. With traditional content-based filtering is the different recommendations direct analysis of the contents, collaborative filtering analysis of user interest, specified in the user base to find similar users (interested) users, combination of these similar users of the evaluation of information to form a system for the specified user predict the preferences of this information [2].

IV. PROBLEMS TO BE SOLVED

Experiments show that collaborative filtering algorithms can provide more accurate recommendations, but also exposed some to be solved problems [3].

1) Data sparseness problem

E-commerce sites often have a large number of commodities, and each user to purchase or make an assessment of only a very small part, usually less than 1%, the data will lead to sparse algorithm accuracy rate.

The correlation coefficient need two users at least to evaluate the same product to more than two terms, actually two similar users, probably because of the lack of the same product but missed, the algorithm can be used to find any products recommended.

2) Scalability problem

E-commerce in a very large involved amount of data, and such as nearest neighbor algorithm efficiency with the increase in the number of users and products decline, How to improve the algorithm to make it more responsive to large-scale computing is one focus of the study.

3) Real-time problem

Real-time problem with require the online services a substantial increase in the customer becomes more and more prominent, effective way to solve this problem is to use distributed computing technology. In addition, project-based clustering algorithm for collaborative filtering [4] can significantly reduce the nearest neighbor of the query space, in order to effectively solve large-scale data processing system recommended facing real issues.

V. COLLABORATIVE FILTERING RECOMMENDATION ALGORITHMS BASED ON ASSOCIATION RULES ITEMS PREDICT

A. Algorithm Processes

the core of collaborative filtering recommendation algorithms based on association rules items predict is divided into two parts: The first is conduct the project score prediction by association rules algorithms , which will

effectively solve the data sparseness problem. The second part gives the prediction based on the results of the first part conduct the collaborative filtering recommendation. The algorithm uses association rules and collaborative filtering recommendation to achieve information recommend, its processing is divided into offline and online processing recommended two parts. Processing flow chart shown in Figure 1.
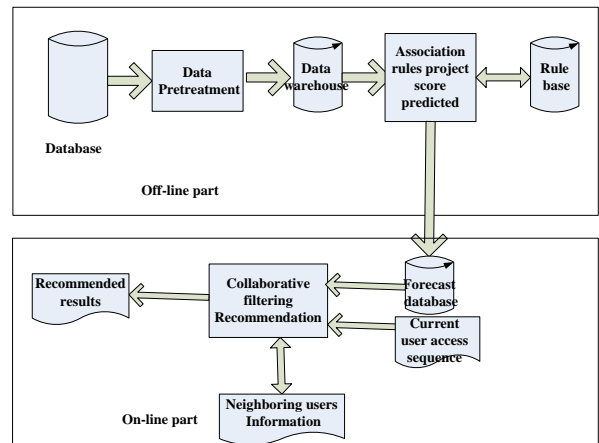

Figure 1. Flow chart of collaborative filtering recommendation algorithms based on association rules items predict

B. Data Preprocessing

Business data pre-processing is carried out off-line processing and update the basis, the main steps include: data cleaning, user identification, user access operation identification, path improvement and user transactions and pattern recognition. The data pre-processing detailed description step see Literature [5].

C. The Use of Association Rules for the Project Score Prediction

After completion business data of the data pre-processing, the next step will detail how to use association rules for item score predicted. For the association rules to achieve the two steps:

- The first step to identify for all with than the minimum support the support of the item set.

- The second step re-uses of large item sets to generate the required rules. In order to generate all frequent item sets, the algorithm uses the recursive method, the algorithm framework [6] shown in Figure 2, for simplicity, all of which support the project sets the number with its support express.
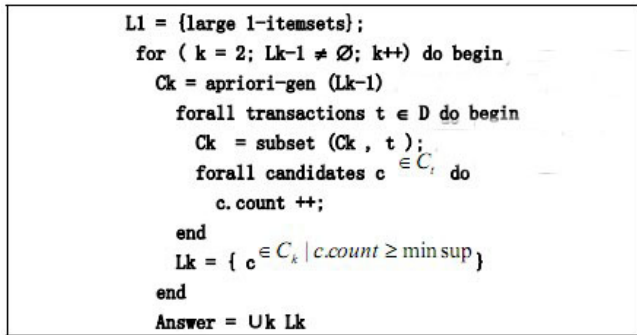
```
L1 = {large 1-itemsets};
for ( k = 2; Lk-1 ≠ ∅; k++) do begin
    Ck = apriori-gen (Lk-1)
        forall transactions t ∈ D do begin
            Ck = subset (Ck , t );
            forall candidates c ∈ Cₜ do
                c. count ++;
        end
        Lk = { c ∈ Cₖ | c.count ≥ min sup }
end
Answer = ∪k Lk
```

Figure 2.   Association rules framework diagram

Figure 3.   Effects of for the largest indirect of neighbor proportion for Algorithm

VI.   EXPERIMENTAL ANALYSES AND EVALUATION

A.   Selected Sample Data

In tests is used the Movie Lens as data set. Movie Lens dataset contains 943 users for the 1682 film about the 100,000 score data. Score is divided into five levels, namely 1,2,3,4,5 integer. . Select Movie Lens dataset reason, because it is the number of collaborative filtering algorithm used in the common data set, facilitate with other collaborative filtering algorithms to comparison. The data set available for free download on the Internet site at: http://www.movielens.org/.

B.   Testing Process

Parameters of effects for indirect of nearest neighbor, for the correlation indirect Pearson algorithm parameters effect of on the algorithm, we made the following experiment.. Take the test indicator to predict the average absolute difference, the maximum indirect neighbors of different proportion of X% (X = 0,10,20,30,40,50,60,70,80,90,100) for testing. This data set selected 40% of the randomly training set, In Given20 test case, the parameter indirectly from the largest to take 1. As show in Figure 3, from the average absolute deviation and average direction error can be seen, when the largest indirect neighbors proportion of to take 0% to 40%, the forecast can reach the best. When the largest indirect neighbors proportion more than 40%, the prediction error increased. After careful analysis we concluded that, because the indirect neighbors is through their own score estimated from the data, accuracy, and compared with the real gap between the neighbors, so when using the indirect neighbors too much time but will reduce the accuracy of the forecasts.
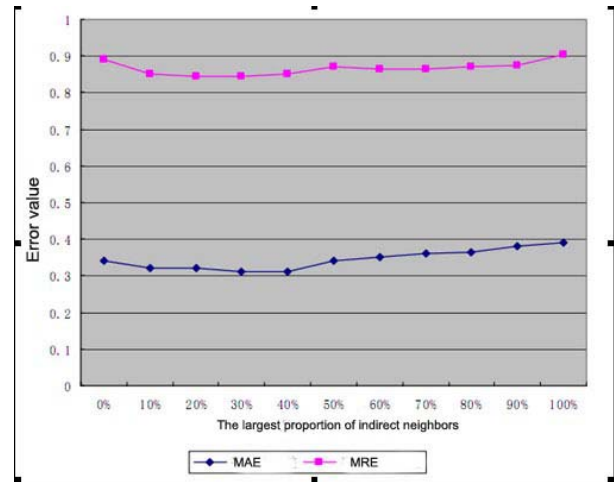
Consider the use of more indirect neighbor, predicted more time spent, we take the analysis that 30% of indirect neighbors is more appropriate.

C.   The Results Comparison and Analysis

The following table 1 is still using 40% of the user data for the training set, the Given5, Given20, Given All But1 three cases when the test results, MAE on behalf of Mean Absolute Error, Pred Time, represents the predicted value calculated for each average time spent. Which correlation indirect neighbor Pearson algorithm in the parameters maximum neighbor to take 1 , the maximum proportion of 30% indirect neighbors.

TABLE I.   THE RESULT OF EXPERIMENTS

	Given	MAE	Pred.Time
association rules algorithm	5	0.87	9.42
	20	0.83	10.01
	All but 1	0.79	10.33
simple Pearson algorithm	5	1.02	8.21
	20	0.88	8.29
	All but 1	0.85	8.42
Association Indirect Pearson algorithm	5	0.90	8.47
	20	0.86	8.71
	All but 1	0.85	9.23

In order to better compare the prediction algorithm, the average absolute error of the table raised to a separate, plotted as a histogram,

Which SP is behalf of simple Pearson algorithm, AR is behalf of association rules, A.I.P is behalf of Association Indirect Pearson: from the MAE accuracy analysis, as shown in Figure 4
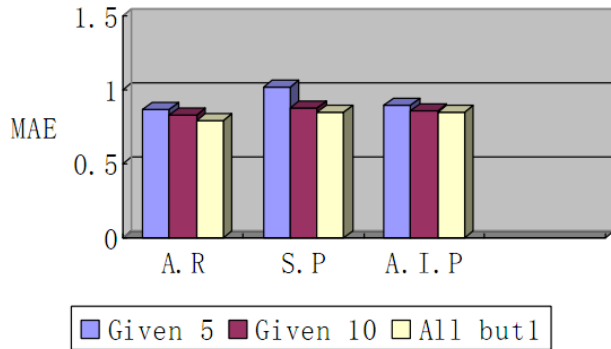
Figure 4.   From the MAE accuracy analysis

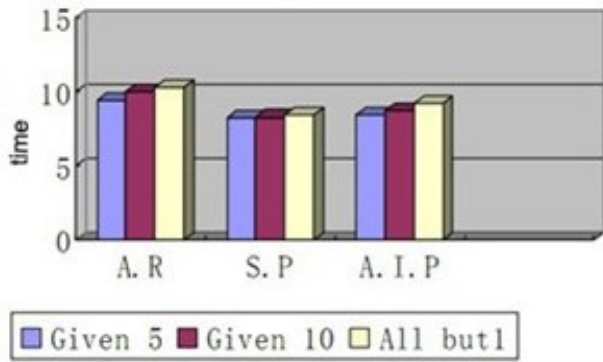From the Pred Time consuming analysis, as shown in Figure 5

Figure 5.   From the Pred Time consuming analysis

VII.   SUMMARY

This paper describes personalized recommendation of knowledge, and detailed analysis algorithm of the personalized recommendation system. For personalized Recommendation Classical algorithm- collaborative filtering algorithm that although the to provide more accurate recommendation, also exposed some of the need to solve problems, this paper presents an improved personalized recommendation algorithm.

REFERENCES

[1] Zhao Liang, Hu naiJing, Zhang Shouzhi, "Personalized recommendation algorithm design," Computer Research and Development, Vo1.39, No.8 Aug. 2002

[2] Breese J, Hecherman D, Kadie C, "Empirical analysis of predictive algorithms for collaborative filtering," In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI'98). 1998. 43~52.

[3] He An, "Collaborative filtering in e- E-commerce recommendation systems research," Zhejiang: ZhejiangSchool of Computer Science, 2007

[4] Deng Ailin Zhu Yang Yong, Shi Bole, " based on the project rating collaborative filtering prediction algorithm,".Software,2003,14(9) :1621-1628.

[5] Lu Lina, Yang Yiling, and Guan Xudong, "the study of Web log mining in data preprocessing ," Computer Engineering, 2000, 26 (4): pp. 66 - 72.

[6] Shi Jing, Gong Zhenyu, "personalized service system based on Web usage mining," University of Electronic Science and Technology, 2002.31 (4).