Joining User Clustering and Item Based Collaborative Filtering in Personalized Recommendation Services

SongJie GONG
Zhejiang Business Technology Institute,
Ningbo 315012, P. R. China
e-mail: gongsongjiezjbti@163.com

HongWu YE
Zhejiang Textile & Fashion College,
Ningbo 315211, P. R. China
e-mail: yehongwuzjbti@163.com

***Abstract*—Personalized recommender systems consist services that produce recommendations and are widely used in the electronic commerce. Many recommendation systems employ the collaborative filtering technology. With the gradual increase of customers and products in electronic commerce systems, the time consuming nearest neighbor collaborative filtering search of the target customer in the total customer space resulted in the failure of ensuring the real time requirement of recommender system. To solve the scalability problem in the collaborative filtering, this paper proposed a personalized recommendation approach joins the user clustering technology and item based collaborative filtering. Users are clustered based on users' ratings on items, and each cluster has a cluster center. Based on the similarity between target user and cluster centers, the nearest neighbors of target user can be found and pre-produce the prediction where necessary. Then, the proposed approach utilizes the item based collaborative filtering to produce the recommendations. The recommendation joining user clustering and item based collaborative filtering is more scalable than the traditional one.**

Keywords-personalized services; item based collaborative filtering; recommender system; user clustering

I. INTRODUCTION

As the development of the internet and electronic commerce systems, there are amounts of information arrived we can hardly deal with. Thus, personalized recommendation services exist to provide us the useful data employing some information filtering technologies [1]. Information filtering has two main methods. One is the content based filtering and the other is the collaborative filtering. Collaborative filtering has proved to be one of the most effective for its simplicity in both theory and implementation.

Collaborative filtering has been successfully used in various applications. The famous electronic commerce website Amazon and CD-Now have employed collaborative filtering technique to recommend products to customers and it has improved quality and efficiency of their services. It assumes that a good way to find a certain user's interesting items is to find other users who have similar interests with him [2,3]. Collaborative filtering methods operate upon user ratings on observed items making predictions concerning users' interest on unobserved items. With the adding of users and items in the user-item rating database, the scalability of ratings dataset is particularly important in domains. Different treatments are required and different prediction techniques must be employed depending on the scalability conditions, making the selection of an appropriate approach a cumbersome task.

With the gradual increase of customers and products in electronic commerce systems, the time consuming nearest neighbor collaborative filtering search of the target customer in the total customer space resulted in the failure of ensuring the real time requirement of recommender system. To solve the scalability problem in the collaborative filtering, in this paper, we proposed a personalized recommendation approach joins the user clustering technology and item based collaborative filtering. Users are clustered based on users' ratings on items, and each cluster has a cluster center. Based on the similarity between target user and cluster centers, the nearest neighbors of target user can be found and pre-produce the prediction where necessary. Then, the proposed approach utilizes the item based collaborative filtering to produce the recommendations. The recommendation joining user clustering and item based collaborative filtering is more scalable than the traditional memory based collaborative filtering algorithm.

II. EMPLOYING THE USER CLUSTERING TECHNOLOGY TO FILL THE VACANT VAULE

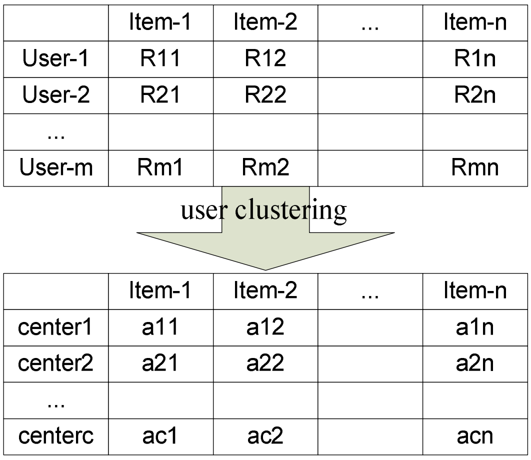
A. User clustering

User clustering techniques work by identifying groups of users who appear to have similar ratings. Once the clusters are created, predictions for a target user can be made by averaging the opinions of the other users in that cluster. Some clustering techniques represent each user with partial participation in several clusters. The prediction is then an average across the clusters, weighted by degree of participation. Once the user clustering is complete, however, performance can be very good, since the size of the group that must be analyzed is much smaller [4].

The idea is to divide the users of a collaborative filtering system using user clustering algorithm and use the divide as neighborhoods, as Figure 1 show. The clustering algorithm may generate fixed sized partitions, or based on some

978-0-7695-3618-7/09 $25.00 © 2009 IEEE
DOI 10.1109/IIS.2009.70

similarity threshold it may generate a requested number of partitions of varying size.



	Item-1	Item-2	...	Item-n
User-1	R11	R12		R1n
User-2	R21	R22		R2n
...				
User-m	Rm1	Rm2		Rmn

	Item-1	Item-2	...	Item-n
center1	a11	a12		a1n
center2	a21	a22		a2n
...				
centerc	ac1	ac2		acn

Figure 1. Collaborative filtering based on user clustering

Where Rij is the rating of the user i to the item i, aij the average rating of the user center i to the item j, m is the number of all users, n is the number of all items, and c is the number of user centers.

B. *User clustering to smoothing the rating matrix*

In this paper, we user the k means clustering algorithm to cluster the users into some groups as clustering centers.

Specific algorithm as follows:

```
Input: clustering number k, user-item rating matrix
Output: smoothing rating matrix
Begin
  Select user set U={U1, U2, …, Um};
  Select item set I={I1, I2, …, In};
  Choose the top k rating users as the clustering CU={CU1, CU2, …, CUk};
  The k clustering center is null as c={c1, c2, …, ck};
  do
    for each user Ui∈U
      for each cluster center CUj∈CU
        calculate the sim(Ui, CUj);
      end for
      sim(Ui, CUm)=max{sim(Ui, CU1), sim(Ui, CU2), …, sim(Ui, CUk);
      cm=cm∪Ui
    end for
    for each cluster ci∈c
      for each user Uj∈U
        CUi=average(ci, Uj);
      end for
    end for
  while (C is not change)
End
```

As we cluster the users in some groups, then we can smooth the user item rating matrix.

III. PRODUCING THE PREDICTION

A. *Measuring the item rating similarity*

There are several similarity algorithms that have been used [5,6]: Pearson correlation, cosine vector similarity, adjusted cosine vector similarity, mean-squared difference and Spearman correlation.

We use the cosine measure, as following formula, which looks at the angle between two vectors of ratings as the target item t and the remaining item r.

$$sim(t,r) = \frac{\sum_{i=1}^{m} R_{it} R_{ir}}{\sqrt{\sum_{i=1}^{m} R_{it}^{2} \sum_{i=1}^{m} R_{ir}^{2}}}$$

Where R_{it} is the rating of the target item t by user i, R_{ir} is the rating of the remaining item r by user i, and m is the number of all rating users to the item t and item r.

B. *Prediction using item-based CF*

Since we have got the membership of item, we can calculate the weighted average of neighbors' ratings, weighted by their similarity to the target item.

The rating of the target user u to the target item t is as following:

$$P_{ut} = \frac{\sum_{i=1}^{c} R_{ui} \times sim(t,i)}{\sum_{i=1}^{c} sim(t,i)}$$

Where Rui is the rating of the target user u to the neighbour item i, sim(t, i) is the similarity of the target item t and the neighbour item i, and c is the number of the neighbours.

IV. DATASET AND MEASUREMENT

A. *Data set*

We use MovieLens collaborative filtering data set to evaluate the performance of proposed algorithm. MovieLens data sets were collected by the GroupLens Research Project at the University of Minnesota and MovieLens is a web-based research recommender system that debuted in Fall 1997. Each week hundreds of users visit MovieLens to rate and receive recommendations for movies [1]. The site now has over 45000 users who have expressed opinions on 6600 different movies. We randomly selected enough users to obtain 100, 000 ratings from 1000 users on 1680 movies with every user having at least 20 ratings and simple demographic information for the users is included. The ratings are on a numeric five-point scale with 1 and 2 representing negative ratings, 4 and 5 representing positive ratings, and 3 indicating ambivalence.

B. *Performance measurement*

Several metrics have been proposed for assessing the accuracy of collaborative filtering methods. They are divided into two main categories: statistical accuracy metrics and decision-support accuracy metrics. In this paper, we use the statistical accuracy metrics [7].

Statistical accuracy metrics evaluate the accuracy of a prediction algorithm by comparing the numerical deviation of the predicted ratings from the respective actual user ratings. Some of them frequently used are mean absolute error (MAE), root mean squared error (RMSE) and correlation between ratings and predictions. All of the above metrics were computed on result data and generally provided the same conclusions. As statistical accuracy measure, mean absolute error (MAE) is employed.

Formally, if n is the number of actual ratings in an item set, then MAE is defined as the average absolute difference between the n pairs. Assume that p1, p2, p3, ..., pn is the prediction of users' ratings, and the corresponding real ratings data set of users is q1, q2, q3, ..., qn. See the MAE definition as following:

$$MAE = \frac{\sum_{i=1}^{n} | p_i - q_i |}{n}$$

The lower the MAE, the more accurate the predictions would be, allowing for better recommendations to be formulated. MAE has been computed for different prediction algorithms and for different levels of sparsity.

C. Comparing with the traditional CF

Figure 2 illustrates the sensitivity of the algorithms in relation to the different numbers of neighbors, which compares the performance of two different CF algorithms. The results indicate that the accuracy of the proposed algorithm is better than the traditional CF algorithms.

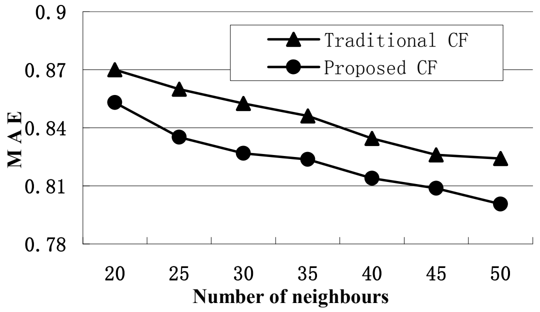


Figure 2. Comparing the proposed CF algorithm with the traditional CF algorithm

V. CONCLUSIONS

Personalized recommendation consist services that produce recommendations and are widely used in the electronic commerce. Many recommendation systems employ the collaborative filtering technology. With the gradual increase of customers and products in electronic commerce systems, the time consuming nearest neighbor collaborative filtering search of the target customer in the total customer space resulted in the failure of ensuring the real time requirement of recommender system. In this paper, we proposed a personalized recommendation approach joins the user clustering technology and item based collaborative filtering. Users are clustered based on users' ratings on items, and each cluster has a cluster center. Based on the similarity between target user and cluster centers, the nearest neighbors of target user can be found and pre-produce the prediction where necessary. Then, the proposed approach utilizes the item based collaborative filtering to produce the recommendations. The recommendation joining user clustering and item based collaborative filtering is more scalable than the traditional collaborative filtering.

REFERENCES

[1] Sarwar B, Karypis G, Konstan J, Riedl J. Item-Based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International World Wide Web Conference. 2001. 285-295.

[2] Chong-Ben Huang, Song-Jie Gong, Employing rough set theory to alleviate the sparsity issue in recommender system, In: Proceeding of the Seventh International Conference on Machine Learning and Cybernetics (ICMLC2008), IEEE Press, 2008, pp.1610-1614.

[3] Yu Li, Liu Lu, Li Xuefeng, A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce, Expert Systems with Applications 28 (2005) 67–77.

[4] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, Recommender systems for large-scale e-commerce: Scalableneighborhood formation using clustering, Proceedings of the Fifth International Conference on Computer andInformation Technology, 2002.

[5] Songjie Gong, Chongben Huang, Employing Fuzzy Clustering to Alleviate the Sparsity Issue in Collaborative Filtering Recommendation Algorithms, In: Proceeding of 2008 International Pre-Olympic Congress on Computer Science, World Academic Press, 2008, pp.449-454.

[6] SongJie Gong, The Collaborative Filtering Recommendation Based on Similar-Priority and Fuzzy Clustering, In: Proceeding of 2008 Workshop on Power Electronics and Intelligent Transportation System (PEITS2008), IEEE Computer Society Press, 2008, pp. 248-251.

[7] Huang qin-hua, Ouyang wei-min, Fuzzy collaborative filtering with multiple agents, Journal of Shanghai University (English Edition), 2007,11(3):290-295.