A FOAF-based Framework for E-Commerce Recommender Service System

Yang Zhao
Wuhan University, Center for Studies of Information Resources
430072, Wuhan, P. R.China , zyaxjlgg_0813@hotmail.com
WWW home page: http://www.csir.whu.edu.cn/

Abstract. Recommender service systems have been widely and successfully applied in e-commerce to provide personalized recommendations to customers nowadays. The tremendous growth in the amount of available information and the number of visitors to websites poses some challenges for recommender service systems such as poor prediction accuracy, scalability, and dynamic changes of users. To address these issues and increase the performance of the systems, an e-commerce recommender service system framework based on FOAF (Friend of A Friend) is proposed in this paper. FOAF provides a RDF/XML vocabulary to describe individual users and their relationships with other users. A FOAF profile could allow a system to better understand users' personalized needs. Once the system extracts each user's preferences that are represented by FOAF document format, it can classify users with respect to their own preferences on real time, and also, it can recommend items in which some users are interested to a target user who has the highest similarity with them in the same group. Experimental results show that the proposed framework helps to reduce the recommendation time, while improving accuracy.

1 Introduction

With the fast development of internet infrastructure, recommender service systems have been successfully applied to improve the quality of service for users. A recommender service system on e-commerce has brought reduction of cost for searching by extracting information in which users are interested and listing the results that are likely to fit them. In order to provide more accurate recommendation results within an acceptable response time, a recommender service system is required to have the capacity to handle a large amount of information on real time.

Recommender service system framework generally applies the information filtering technique called collaborative filtering (CF)[1]to solve the problems in information processing. The typical CF approach employs statistical techniques to find a set of customers who have a history of agreeing with the target user. It has been widely used in a number of different applications, and overcomes lots of shortcomings of content-based recommender service systems. But with the rapid growth in the amount of users' information and products data, the calculation of the recommender algorithm becomes more and more complex, which leads to the scalability issue of CF. At the same time, the issue of dynamic changes of users' interests was not considered in CF, either. Recommender service systems are operated in the dynamic web environment in which at anytime a user may change his/her interests and decide to purchase some products. So the system should understand and handle users' interesting changes as quickly as it can. To solve these problems and increase the performance of e-commerce recommender service systems, a more effective system framework is needed.

In this paper, we proposed a Peer-to-Peer recommender service system framework based on FOAF (Friend of A of Friend) to improve the system scalability. FOAF[2]is a kind of semantic web technology that provides a RDF/XML vocabulary to describe a user's information, including name, mailbox, homepage URL, interests, friends, and so on. With the application of FOAF, it is easier to share and use information about users and their activities, to transfer information between web sites, and to automatically extend and re-use it on line[3]. User preference is asserted as a property of the FOAF document format by learning user's behaviors. Once the system learns user preference, it can classify users with respect to their preferences on real time, and also, it can recommend items in which users are interested to a target user who has the highest similarity with them in the same group on P2P network. Due to the reduced quantity of information processing, the system scalability will be greatly improved. At the same time, system can dynamically update and maintain the user profiles on real time by FOAF.

The paper is organized as follows: Section 2 introduces the approach for describing a user's information by FOAF profile. Section 3 presents the FOAF-based framework for implementing an e-commerce recommender service system and analyses the specific process of recommendation. Experiments with performance evaluation are given in Section 4. Section 5 concludes the paper.

2 Using FOAF Profiles for User Identification

We use the FOAF profile to identify a user. This type of profile contains information about social relations of the user next to traditional information that identifies the user. The profile could either be generated by hand, or more often, by copy, paste and edit of other people's FOAF, or by semi-automated tools such as FOAF-a-matic[4]. Part of an example profile is shown in Fig.1. This example illustrates several important things about FOAF. The *foaf:knows* property points to other people known by this person. Other FOAF profiles are linked through *rdfs:seeAlso*, allowing Semantic Web bots to crawl through FOAF space. At the same time, in this example we use the items that have been evaluated by this person to specify the interests of him/her. The `foaf:interest` property can be used to represent the web pages about these items, including the items name, user rating results and other information[5]. But the item in which a user is interested may not be his/her preference, because the user may rate a low score on this item. The RDF syntax of FOAF allows a system to better understand users' needs. When a user publishes a FOAF profile (just like the example below) on the e-commerce websites, the system is able to make use of that information, and then extract this user's preferences.

```
<rdf:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/" ... > ......
<foaf:Person>
<foaf:name>Zhao Yang</foaf:name>
<foaf:mbox rdf:resource="mailto:zymnesl@hotmail.com" />
<foaf:homepage rdf:resource="http://www.whu.edu.cn/zymnes" />
……
<foaf:interest dc:title="Gladiator"
rdf:resource="http://movielens.umn.edu/movieDetail?movieId=3578"/>
<foaf:interest dc:title="Harry Potter "
rdf:resource="http://movielens.umn.edu/movieDetail?movieId=40815"/>
<foaf:interest dc:title="Lord of the Rings"
......
</foaf:interest>
<foaf:knows><foaf:Person>
<foaf:name>Daisy Green</foaf:name>
<rdfs:seeAlso rdf:resource="http://martinmay.net/foaf.rdf"/>
</foaf:Person></foaf:knows>
……
</foaf:Person> </rdf:RDF>
```

Fig.1. Example of a user's FOAF profile

3 The Proposed FOAF-based Recommender Service System Framework

Fig.2. illustrates the overall framework which consists of three major components: FOAF profiles, recommender component and user interface. The FOAF profiles of different users are generated automatically by semi-automated tools in this system. Users' information comes from user database of this e-commerce website. A user database stores and mangers all kinds of users' information, such as registration information, past buying and rating behaviors, page view logs, and etc. Users can also maintain and update their own FOAF profiles dynamically. By the FOAF profiles, system can find in which items different users are interested. The recommender component is used to generate recommendations for users based on FOAF profiles. It is composed of user grouping, similarity calculation and recommendation generating modules. The output of an e-commerce recommendation system is a list of the top-N recommender items, and the goal of the user interface is

to display these recommendation results in the manner such that the recommendation service user can absorb them easily and effectively. In this framework, in addition to FOAF method， some other recommendation algorithms have also been applied, such as collaborative filtering, associate rule based filtering, and etc. The specific process of the recommendation based on this FOAF framework will be analyzed below.

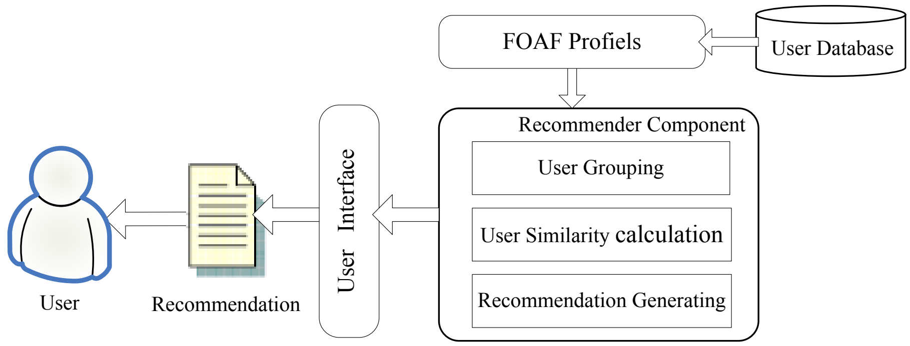

Fig. 2. The Proposed FOAF-based Recommender Service System Framework

3.1 Grouping Users According to Their Preferences

Grouping users makes it easy for the system to compute the similarity between users in the same group. Due to the extensibility of the FOAF-based distributed environment, it is possible to aggregate information from other users who have the same preference. The recommender service system extracts each user's preferences that are represented in FOAF profile. By the preferences of each user, recommender service system groups users. And then, when a user evaluates an item m_i, the system determines whether m_i is good enough to recommend to other users in the same group or not. In this paper, we take an online movie site as an example. We have distinguished movie genres into 15 categories [6]. Usually a user can have more than one preference about movie genre. Once a movie is rated by a user, the related page link of this movie will be added into the user's FOAF profile, and represented by the `foaf:interest` property. The page includes some basic information of this movie, such as movie name, movie genre, actors, user ratings, and etc.

In order to extract accurate preferences from each user, we employed a probabilistic method as shown in equation (1). Assume there are the set of users $U=\{u_1,u_2,\ldots,u_i\}$, the set of movies that is rated by a user u_i is $M=\{m_1,m_2,\ldots,m_j\}$ and the set of genre is $G=\{g_1,g_2,\ldots,g_k\}$. The function $f(P_{ik})$ extracts i-th user's preference about a genre g_k.

$$f(P_{ik}) = \sum_{j=1}^{n} \frac{g_k(m_j)}{\mu} R_{ij} \qquad \mu = \sum_{m=1}^{n} C_m(u_i, m_j) \qquad (1)$$

Where C_m is a set of genres included in movie m_j, and R_{ij} means a rating value of i-th user about movie m_j. Also μ means total count of genres that included in movies that are rated by user u_i. By comparing the value of $f(P_{ik})$, the system can classify users to different groups. Users in the same group have the same preference about movie genre. Sometimes, a user may have more than one preference, so he/she may be assigned to different groups at the same time.

3.2 Users Similarity Calculation

To an active visitor, users in the same group have different similarity. In order to improve the recommender accuracy for a particular user（active user）, we should find which ones have the highest similarity with this user. There are a number of different ways to compute the similarity between users. The most widely and successfully way is neighborhood-based collaborative filtering [7]. But the traditional neighborhood-based collaborative filtering suffers from the poor scalability problem, i.e., the time used to search for a group of neighbors is proportional to the number of all users in the system. In the FOAF-base framework proposed in this paper, we have already grouped users according to their preferences, it is not necessary for all users in the database to be weighted according to their similarity with the active user. So the neighbor searching time is reduced. On the basis of user grouping, we use person correlation coefficient algorithm[8]that is a major approach of collaborative filtering to compute the similarity between active a and another user i, as equation (2).

$$sim(a,i) = \frac{\sum_{j \in I_{a,j}}^{n} (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_{j \in I_{a,j}}^{n} (v_{a,j} - \bar{v}_a)^2} \sqrt{\sum_{i=1}^{n} (v_{i,j} - \bar{v}_i)^2}} \quad (2)$$

Where $I_{a,i}$ denotes the set of items rated by both of user a and user i. $v_{a,j}$ denotes a rating value of active user a about item j. $v_{i,j}$ denotes a rating value of user i about item j. $\bar{v}_a$ and $\bar{v}_i$ indicate the average rating of active user a and user i on items. As shown in equation (2), it can be seen that the value of $sim(a,i)$ is proportional to the similarity of user a and i's rating results. And our main goal is to find a number of k-nearest neighbors who have the highest similarity and this process is based on their preferences and some partial information of the active user.

3.3 Generation of Recommendation

Recommendation is a list of TOP-N items that the active user will like most. It is derived from the k-nearest neighbors of the active user. These items are recommended to the user in the form of "customers who liked/rented this movie also rented…".

When the user is an old one，the system has kept track of use's history data, such as his click-streams, his transaction history data, ratings, and etc. Using these data, system can group users and find neighbors for an active user. As neighbors and active user have the highest similarity weights, we can predict the active user's rating value about an item according to his neighbors' rating value about it [9]. The TOP-N highest prediction items will be added to the list and recommended to active user. The prediction of the active user *a* on the items *k* that he doesn't purchase could be calculated as equation (3).

$$P_{a,k} = \overline{R}_a + \frac{\sum_{u \in N} sim(a,u)(R_{u,k} - \overline{R}_u)}{\sum_{u \in N} |sim(a,u)|} \qquad (3)$$

sim(a,u) denotes the similarity between user *a* and user *u*, $R_{u,k}$ denotes a rating value of user *u* about the item *k*, $\overline{R}_a$ and $\overline{R}_u$ denote the average rating made by users *a* and *k* respectively. N is a set of nearest neighbor of active user a.

4 Experimental Evaluation

We carried out several experiments to evaluate the performance of the proposed FOAF-based framework, and to compare it with some related work.

4.1 Data Set

We used experimental data from an open dataset of MovieLens which contains 100,000 ratings of 3,900 movies from *6,040* users. Ratings are discrete values from 1 to 5 (1,2,3,4,5)[10]. We randomly selected 1,000 users who have rated 30 or more movies from the database. We divided the data into a training set (80%) and a test set (20%). We use data from the training set to compute predictions. Data from the test set is then used to evaluate efficiency and accuracy.

4.2 Evaluation Metric

Recommender service systems research has used several types of measures for evaluating the quality of a recommender service system. They can be mainly categorized into two classes: statistical accuracy metrics and decision support accuracy metrics [11]. Statistical accuracy metrics evaluate the accuracy of a system by comparing the numerical recommendation scores against the actual user ratings for the user-product pairs in the test dataset. Mean Absolute Error (MAE) between ratings and predictions is a widely used metric [12]. MAE is a measure of the deviation of recommendations from their true user-specified values. We choose MAE as our evaluation metric to report prediction experiments because it is most commonly used and easiest to interpret directly. MAE is defined as equation (4).

$$MAE = \frac{\sum_{i=1}^{N} |P_i - R_i|}{N} \qquad (4)$$

P_i denotes the predicted rating of product *i*, R_i denotes the true rating of product *i and* N denotes the number of items in the test set. Formally, the lower the MAE, the more accurately the recommendation engine predicts user ratings.

4.3 Experimental Results

After system made FOAF profiles of users, 1,000 users were divided into 15 groups according to their preferences. Therefore it was not necessary for all users in the database to be weighted according to their similarity with the active user. To confirm the effectiveness of the proposed improved recommendation method based on FOAF, we compare the traditional CF algorithm with the new recommendation algorithm. We carried out the experiment while changing the number of neighborhoods from 10 to 50, interval is 5.

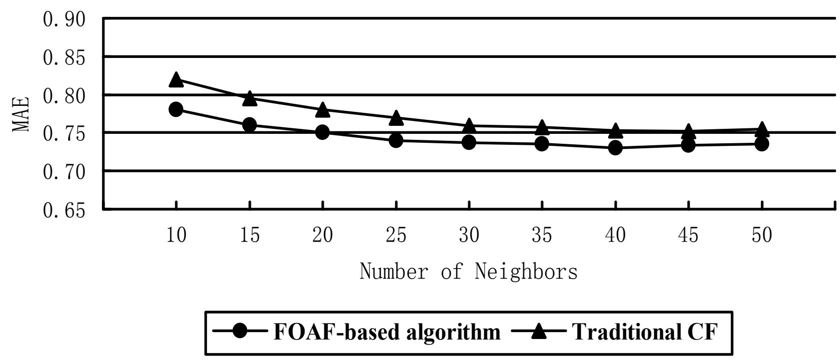

Fig. 3. Comparison of prediction quality of FOAF-based algorithms and traditional CF

In Fig.3, we could find that no matter how many neighbors we choose, the line of the traditional CF is higher than the line of our proposed one, which indicates that the FOAF-based algorithm can provide predictions of better quality than the traditional CF algorithm. Furthermore, the FOAF-based algorithm reduces the time used to search neighbors and generates the recommendations more quickly and effectively than traditional method.

5 Conclusions

Recommender service systems are a powerful new technology for extracting additional value for e-commerce websites from its users. These systems benefit users by enabling them to find items they like. Conversely, they help the e-commerce by generating more sales. Typical recommender service systems suffer from poor scalability and the lack of ability to handle dynamic changes of users' interests. In this paper, a new P2P recommender service system framework was proposed to address these issues. This framework is based on the FOAF. Since the recommendation system applies FOAF profiles for user describing and grouping, it reduces the averaged recommendation time and increases the recommending accuracy for users. The effectiveness comparison results also indicate that the proposed FOAF-based recommendation method performs better than the traditional CF algorithm in this recommender service system. The next research step is to validate and improve algorithm robustness of the proposed method.

Acknowledgment

The research reported in this paper is supported by the project Research on Digital Information Services in Network Environment under No. 06JJD870006, sponsored by Key Project of Key Institution in Humane and Social Science Research, Education of Ministry of P.R.China.

References

87. 1. A.K. Joseph, G.T. Loren and John, valuating Collaborative Filtering Recommender Systems, *ACM Transactions on Information Systems*, 22(1), 5-53(2004).
88. 2. E. Dumbill, Support online communities with foaf: How the friend-of-a-friend vocabulary addresses issues of accountability and privacy,
89. http://www-106.ibm.com/developerworks/xml/library/x-foaf.htm. [2006-08-13].
90. 3. S. Grzonkowski, A. Gzella, H. Krawczyk, S.R. Kruk, F.J.M.R. Moyano and T. Woroniecki, D-FOAF-Security Aspects in Distributed User Management System, *TEHOSS'2005*.
91. 4. S.R. Kruk and S. Decker, Semantic Social Collaborative Filtering with FOAF Realm, *Semantic Desktop Workshop, ISWC* 2005.
92. 5. FOAF Vocabulary Specification. http://xmlns.com/foaf/0.1/. [2007-01-20].
93. 6. MovieLens Project Web site.http://movielens.umn.edu. [2007-01-20].
94. 7. S.J. Ko, Prediction of Preferences through Optimizing Users and Reducing Dimension in Collaborative Filtering System, *IEA/AIE* 2004, 1259-1268(2004).
95. 8. C. Zeng, C.X. Xing and L.Z. Zhou, Similarity Measure and Instance Selection for Collaborative Filtering, In *12th international conference on World Wide Web*, Budapest, Hungary, 2003.
96. 9. J. Herlocker, J. Konstan, A. Borchers and J. Riedl,, An Algorithmic Framework for Performing Collaborative Filtering, In *Proceedings of ACM SIGIR'99*.ACM press.
97. 10. B. Sarwar, G. Karypis, J. Konstan and J. Riedl, Analysis of recommendation algorithms for E-commerce, In *Proc. of ACM'00 Conf. on Electronic Commerce,* 158–167 (2000) .
98. 11. Y.H. Guo and G.S. Deng , An improved Collaborative Filtering based E-Commerce Recommendation System with Case-based Reasoning, In *International Conference on Serveice System and Service Management,* 780-784(2004).
99. 12. W. Nejdl, M. Wolpers, W. Siberski, C. Schmitz, M. Schlosser, I. Brunkhorst and A. Loser, Super-Peer-Based Routing and Clustering Strategies for RDF-Based Peer-To-Peer Networks, In *Proc. of the Twelfth International World Wide Web Conference,*2003.