Group Recommendation Based on the PageRank

Jing Wang  
School of Computer Science and Technology, Xidian University, Xi’an, China  
Email: wangjing@mail.xidian.edu.cn

Zhijing Liu  
School of Computer Science and Technology, Xidian University, Xi’an, China  
Email: liuprofessor@163.com

Hui Zhao  
School of Electronic and Information Engineering, Xi’an Jiaotong University, Xi’an, China  
Email: huizhao@stu.xjtu.edu.cn

Abstract—Social network greatly improve the social recommendation application, especially the study of group recommendation. The group recommendation, analyze the social factors of the group, such as social and trust relationship between users, as the factors for the prediction model established. In this paper, PageRank algorithm is introduced in the recommendation method to calculate the member’s importance in the group respectively, and to amend the aggregate function of individual preferences. The aggregate function consider the relationship between various users in the group, and optimize the aggregate function according to users different influence on the group, which can better reflect the social characteristics of group. In short, the study on group recommended model and algorithm can take the initiative to find the user's needs. Extensive experiments demonstrate the effectiveness and efficiency of the methods, which improve the prediction accuracy of the group recommended algorithms.

Index Terms—recommendation system; collaborative filtering; social network

I. INTRODUCTION

With the rapid development of Internet, Web has become the world's largest source of information. However, it also brings the problem of information overload, which the user cannot obtain the useful information in too much information efficiently. The recommendation system is an information filter tool, which take the initiative to find resources and provide the appropriate services. So that, we use recommendation system to solve the problem of information overload.

In the fields of academia and industry, recommendation system has made great development. On one hand, ACM began hosting a special recommendation system international conference from 2008, such as, ACM Conference on Recommender System. And many companies open their data sets for study, one of the famous competitions of recommendation system organized by Netflix. On the other hand, the recommendation system is also successfully applied in many online stores, such as Amazon.com, CDNow.com, Barnes & Noble.com and MovieFinder.com, etc.

According to the different recommendation method, the recommendation system is usually divided into two categories: content filtering-based recommendation systems and collaborative filtering recommender systems. These recommendation systems are used for individual user, without considering the social relationship of the user. In recent years, with the rise of social network, Facebook and Twitter, social filtering is becoming the research focus of collaborative filtering techniques [1,2]. Social filtering methods use a common preference of the user and his friends to analyze the preferences of friends, in order to predict a user’s preferences[3,4]. The simplest social filtering algorithm is the neighborhood-based algorithm[5]. In addition to the simple neighborhood model, there are other social filtering algorithms. Graph model modeling method represent the user's social network and their preferences into a diagram, and then use the random walk algorithm to do the social recommendation for the user[6,7]. The matrix decomposition algorithm decompose the user's preference matrix and social network matrix to calculate a feature vector of the user and objects, and use of the vector point multiplication to measure the user’s preferences[8]. For the entity in social tagging system: users, items and tags, they built unified model to find their semantic in users, projects, and label latent[9]. For team composition and roles of relevant factors, they proposed the team's social recommendation[10]. In addition, social recommendation is use of data mining methods[11].

In real life, when you make your decision, obviously, you will be impacted by surrounding social environment. This is usually referred to emotional infection, which is usually proportional to the trust between the people, the people more you trust; the greater the impact will be on you. For example, your best friend say to you “*** movie
looks better, you may very likely to see the movie” or “xx is a very good book, will also affect you to buy this book”.

To take into account social factors, including an individual's personality, the density relationship between people and each other emotional infection, can improve the prediction accuracy of the group recommendation system. The group recommended generate a polymerization model based on the individual's preferences model to get the preferences of all users in the group[12,13].

There is also some research for the group recommendation system. For instance, MusicFX system nicely addresses everyday needs of groups of members who want to select music to be played while they exercise in a fitness center[14]. LET'S BROWSE recommends web pages to a group of two or more persons who are browsing the web together[15]. POLYLENS is a generalization of the MOVIELENS system that recommends movies to groups of users[16]. GroupRecoPF is an innovative group recommendation in a distributed platform[17]. INTRIGUE recommends tourist attractions for heterogeneous groups of tourists[18]. Happy Movie is a Facebook application where we provide a movie recommendation for a group of people planning to go to the cinema together[19].

Although the above method is to consider the relationship of trust between users, but did not consider the transfer of trust.

In our approach, we propose to study how people interact depending on their personality or their closeness in order to improve group recommendations. We introduce the PageRank algorithm to calculate the importance of each group to amend the aggregate function of individual preferences, which is more suitable for the group's social characteristics, and also improve the prediction accuracy of the group recommendation algorithm.

II. THE DEFINITION OF GROUP

In recommender system, we have a set of users $U = \{u_1, u_2, K, u_n\}$ and a set of items $I = \{i_1, i_2, K, i_n\}$. Each user $u$ rates a set of items $R_{iu} = \{i_1, i_2, K, i_n\}$. The rating of user $u$ on item $i$ is denoted by $r_{ui}$. $r_{ui}$ can be any real number, but often ratings are integers in the range [1-5]. In a trust-based system, we also have a trust network among users. If $u$ trusts $v$, then $t_{uv}$ denotes the value of this trust as a real number in [0; 1]. Zero means no trust and one means full trust.

In general, group recommended method aggregates the personal preferences of individual users with some functions, which providing a recommended model for the entire group. The impacts of different users in a group are different on the group, so the aggregate function should be emphasis on the high-impact users. Our group recommended method is based on this idea. Firstly, analyze social factors in a group, and then calculate the group influence. Finally, adjust the aggregate function to improve the prediction accuracy of recommendation algorithm, based on group's social characteristics.

A. The Description of Group

In this paper, the group (G) is a social group, which can be described as a directed graph. In directed graph, each node represents a user, and the line (link) between the nodes stand for trust relationship between users. Link, point to the node, is in-link, and leave the node, out-link. In this article, the trust relationship is unequal, that is user a trust user b does not mean that the user b trust user a, which means you trust a person and not the man to trust you. For instance, in recent years, twitter or microblogging, who you are concerned about does not mean that he / she is also concerned about you.

There are four users is a group, and each other's trust relationship as shown in Fig. 1. User 1 trust user 2, $a_{21} = 1$, vice versa $a_{12} = 0$. In Fig. 1, user 1 trust user 2, $a_{12} = 1$, and user 2 does not trust the user 1, $a_{21} = 0$.

![Figure 1. The 4 users group](image)

B. The Analysis of Social Relationship

In the Fig. 1, we can see that the more in-links of user, the more users trust, thus the user has higher influence on group. The user1 and user2 only have one in-link, user3 and user4 both have two in-link, so user3 and user4 has higher influence than user1 and user2. Therefore, according to the number of in-links of user, we can calculate the influence of the user on group. The influence of the group G is as following:

$$f_{u,G} = \frac{\sum_{i=1}^{m} a_{ik}}{\sum_{j=1}^{m} a_{ij}}$$

As the above analysis, (1) can calculate the influence of the user on group. However, when two users have a same number of in-links, we will not be able to distinguish their influence with this method. For example, in Fig. 1, user3 and user4 has two in-links, the result of (1) is the same. In this case, we can further analyze their in-links.

The two in-links of user3 come from user1 and user2, and user4 from the user3 and user2. In addition to their common in-links user2, user3 and user4 have a different in-links, the user1 and user3. It can be seen that user3 is more important than user1, so we can infer to user4 is more important than user3. According to these analyzes, the influence of the user on group is not only dependent...
on the number of in-link, also depends on the user of the
in-link.

C. The Influence of User on Group

Google's PageRank algorithm determine the
importance of all the pages based on regression
relationship, which means pages link many high quality
pages, must be high-quality pages. A hyperlink to a page
counts as a vote of support, so Google determine the
importance of web pages based on the number of votes.
However, Google considers not only the number of votes,
but also importance of web pages of vote. Some of the
important pages of the vote are considered to have a
higher value, so that it links to pages will be able to
obtain a higher value.

PageRank definition: for a page A, if there are
pages \( T_i, T_j, \ldots, T_n \) pointing to the page A, PageRank
value of A and \( i \) are \( \text{PR}(A) \) and \( \text{PR}(T_i) \) respectively,
the number of out-link of page \( T_i \) is writed as \( \text{C}(T_i) \),
where \( i = 1, 2, K, n \).

\[
f_{u,G} = \frac{\sum_{i=1}^{m} a_{ik}}{\sum_{i,j=1}^{m} a_{ij}} \quad (2)
\]

And then, Google add damping factor (damping, factor)
to further amended (2), namely:

\[
PR(A) = \frac{1-d}{N} + d \sum_{i=1}^{m} \frac{PR(T_i)}{C(T_i)} \quad (3)
\]

where \( N \) is total number of pages, the damping
factor \( d = 0.85 \).

PageRank algorithm can calculate the influence of user
\( u \) on group \( G \). If user \( u \) has \( k \) trust users in the group \( G \),
which mean the user \( u \) has \( k \) in-links, so the influence
of user \( u \) on the group \( G \) is:

\[
f_{u,G} = \sum_{i=1}^{k} \frac{f_{u,G} * \text{C}(\text{in}_i)}{\text{C}(\text{in}_i)} \quad (4)
\]
or

\[
f_{u,G} = \frac{1-d}{|G|} + d \sum_{i=1}^{k} \frac{f_{u,G}}{\text{C}(\text{in}_i)} \quad (5)
\]

where \( |G| \) is the number of users in group G. When \( d = 1 \),
(5) is equal to (4).

For example, as Fig. 1, after iterative computation of
(5), we can get its impact vector \( F = (0.2972, 0.1638, 0.2334, 0.3055) \). User4 has the largest influence on group,
followed by user1. Although the user1 has an only one in-
link, user4, user4 has the largest influence and only one
out-link, so the influence of user4 on the group \( G \)
passed to the user1. The number of in-links of user1
is less than the user3. However, the influence of user1 is
more than the user3.

III. GROUP RECOMMENDATION ALGORITHM

According to predicted user rating, the group
recommendation algorithm calculates predicted group
rating, which is usually achieved by a certain aggregate
functions. In this paper, our proposed group
recommendation algorithm is based on this approach. The
innovation of this paper is that our aggregate function
consider the relationship between various users in the
group, and optimize the aggregate function according to
users different influence on the group, which can better
reflect the social characteristics of group.

Given the predictive value of user \( u \) for item \( i \) is
denoted by \( \text{pred}(u,i) \), and the predicted value of group
\( G \) for item \( i \) is denoted by \( \text{pred}(G,i) \), we can get
aggregate functions which calculate \( \text{pred}(G,i) \) based on
\( \text{pred}(u,i) \). In this article, we use the following equation
to calculate \( \text{pred}(G,i) \):

\[
\text{pred}(G,i) = \sum_{u \in G} (f_{u,G} * \text{pred}(u,i)) \quad (6)
\]

where, \( f_{u,G} \) stand for influence of user \( u \) on the group \( G \),
which can get from (4) or (5), and \( \text{pred}(u,i) \) can get
from the recommendation system.

IV. EXPERIMENTAL TESTING RESULTS

A. Testing Data

We have tested our algorithm with groups of real users
and movies. We have conducted a real user group study
consisting of 7 members as in Fig. 2. There are 5 males
and 2 females in the group. Each member of group trusts
some members in the group but not all of them, thus
forming a connected social graph between certain
members. To get better result, each member was also
asked to pay close attention to other members who his/her
trusted in the group.

Figure 2. The group topology diagram

In the first stage, we obtain each user’s personal
preferences. Firstly, we get 50 movies from MovieLens
dataset, which including the name of the movie and genre
label, a total of 18 kinds of film genres label, and each film at least have one genre label. And then, we require each user to rate at least 30 films, ratings are integers in the range [1-5], and each user rates at least one film. After get rating data for each user, we can build the model of the user’s personal preferences. For user $u$’s rating data, we can generate the corresponding vector of each rating, a total of five vectors. Preference model of the user $u$ is expressed as:

$$P_u = (v_{u1}, v_{u2}, v_{u3}, v_{u4}, v_{u5})$$  \hspace{1cm} (7)$$

where, $v_{ui} [1 \leq t \leq 5]$ is vector of movie listings, whose rating is $t$. The vector gives movies description, including the name of the movie, director or actor list. In this paper, we use film genre label vector to represent movies. And when the genre label vector of two films has a high degree of overlap, we can say they have high similarity. So the predictive value of user $u$ to the film $i$ equal to rate of the most similar vector, namely:

$$\text{pred}(u, i) = t \quad \text{if} \quad \max(\text{rel}(v_{ji}, v_{ui}))$$  \hspace{1cm} (8)$$

where $v_{ji}$ is the vector of the film $i$, rel($v_j, v_{ui}$) is a similarity of vector $v_j$ and $v_{ui}$.

The second phase, all users rate on the other 100 films. The rules: do not require the user immediate to feedback rating for movie $m$, he/she can see the rating of trusted user, and then rate film $m$ eventually. After the second phase is completed, we can compute the actual ratings $r(G, m)$ of movie $m$. The calculation method is at following:

1. **average value**

$$r(G, m) = \frac{\sum_{u \in G} r_{u,m}}{|G|}$$  \hspace{1cm} (9)$$

where $r_{u,m}$ is the final rating of user $u$ for the film $m$.

2. **weighted average value**

$$r(G, m) = \sum_{u \in G} w_u \ast r_{u,m}$$  \hspace{1cm} (10)$$

where $w_u = f_{u,G}$ and $w_u$ is the weight of user $u$.

B. The Testing Results

In recommendation system, recommendation quality metrics commonly use mean absolute error (MAE). The MAE calculates the error between the predicted rating and actual rating of the user, to measure the accuracy of the forecasts. The smaller the MAE is, the better the recommend system is.

In our group recommendation algorithm, we used the group mean absolute error (GMAE) to represent:

$$\text{GMAE} = \frac{\sum_{m \in M} |\text{pred}(G, m) - r(G, m)|}{|M|}$$  \hspace{1cm} (11)$$

where $|M|$ is the size of the test set.

We take (1) and (5) to calculate the influence of the user. And then calculate the predicted group rating for film $m$ according to the (6). Here, we use the (1) to calculate predicted group rating, defined as the predicted group rating method1 (PGRM1), and the (5) as the predicted group rating method2 (PGRM2). The predicted group rating method2 contains two case, $d=1$ and $d=0.85$.

According to rating of each user for film $m$, we use (9) and (10) to calculate the actual group rating respectively. Here, we use (9) to calculate the actual group rating, defined as actual group rating method1 (AGRM1), and (10) as actual group rating method2 (AGRM2). The actual group rating method2 also contains two case, $d=1$ and $d=0.85$.

According to (11), calculate the various circumstances GMAE, the results as Table 1:

<table>
<thead>
<tr>
<th>Method</th>
<th>AGRM1</th>
<th>AGRM2 ($d=1$)</th>
<th>AGRM2 ($d=0.85$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGRM1</td>
<td>1.2622</td>
<td>1.1356</td>
<td>1.0733</td>
</tr>
<tr>
<td>PGRM2</td>
<td>0.9013</td>
<td>0.8922</td>
<td>0.8578</td>
</tr>
<tr>
<td>PGRM2</td>
<td>0.8734</td>
<td>0.8633</td>
<td>0.7734</td>
</tr>
</tbody>
</table>

In the table, the value of GMAE of (AGRM2) and (PGRM2) is lower than that of (AGRM1) and (PGRM1) obviously, which indicate that (5) can better describe group’s real social relationship. So that, for the group recommendation the method2 is more effectively. In (5), different values of $d$ have little effect on GMAE, which can be seen from the Table 1.

Fig. 3 is a predicted group rating and actual user rating of 100 test films when $d=0.85$. The curve of predicted group rating is closest with the user1’s rating. Fig. 4 is MAE between predicted group rating and actual user rating. MAE value of user1 is the minimum in the group, and user6 and user7’s MAE values are the largest, which indicate impact of user1 on group is highest, and user6 and user7 are lowest. This is based on the topology of Fig. 2 the relationship between the influence of each user group is basically corresponds to.

V. CONCLUSION

In this paper, according to social characteristics of the group and the topological relationship between the various users in a group, we introduce the PageRank algorithm to calculate the impact of each user on a group, which are use their respective influence on the group to adjust the aggregate function to better reflect social impact of a group. In short, the study on group recommended model and algorithm can take the initiative to find the user’s needs, and provide fast and effective recommended method, which has theoretical significance and application value.
However, the proposed method has some limitations. On the one hand, we only given trust or mistrust to describe social relations of user are not suitable. For real-life relationship of trust between people is not only trust or mistrust, there is a certain degree of trust. In addition, the relationship of trust between people is not always static, and it may change because of certain behaviors. On the other hand, the experimental group and test sets we constructed are not big enough, so the more data need to be collected for verification.

ACKNOWLEDGMENT

This work was supported by Project 61173091 of the National Science Foundation of China.

REFERENCES


Jing Wang, born in Xi'an Shaanxi Province, on September 1981, graduated from School of Computer Science and Technology, Xidian University in 2011, received her Ph.D., M.Sc. and B.Sc. in Computer Science from Xidian University in 2011, 2008 and 2003, respectively. And research interests on data mining and machine learning.

She is a teacher at the School of Computer Science and Technology, Xidian University, China. Now she has an interest in the field of collaborative filtering recommendation.

Zhijing Liu, born in Xi'an Shaanxi Province, on April 1957, graduated from the Department of Computer Engineering of Northwestern Telecommunications Engineering Institute in 1982. And research interests on data mining, vision computing.

He is professor and advisor for doctoral students at the school of computer science and technology, Xidian University, China. And he currently serves as head of the Research Center of
Computer Information Research Application, and is director of the China and America Associated Laboratory of key Technologies of Mobile Electronic Commerce. His research works focus on the fields of data mining and vision computing.

Hui Zhao, born in Henan Province on July 1983, graduated from School of Computer Science and Technology, Xidian University in 2008. He got his Bachelor and Master degree from Xidian University 2008. And his research works focused on data mining, information retrieval.

He is a Ph.D candidate at the school of electronic and information engineering, Xi'an Jiaotong University. And his research works focus on the fields of data mining and mobile learning.