2018 5th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC)

A Novel Algorithm for Group Recommendation Based on Combination of Recessive Characteristics

Yi Liu
Department of Information System
China Ship Development and Design Center
Wuhan, China
503864583@qq.com

Xian Zhong
Department of Computer Science and Technology
Wuhan University of Technology
Wuhan, China
23668229@qq.com

Lin Li
Department of Computer Science and Technology
Wuhan University of Technology
Wuhan, China
43598250@qq.com

Jingling Yuan
Department of Computer Science and Technology
Wuhan University of Technology
Wuhan, China
740080412@qq.com

***Abstract*—Recommendation system is an effective way to solve the problem of information overload, and it was widely used in the real life. In view of user groups have complex characteristics, many studies found that some important recessive characteristics could more truly reflect the interests of group users. In this paper, we propose a novel algorithm for group recommendation based on combination of recessive characteristics (GRBCRC). Firstly, on the basis of SVD++ recommendation algorithm, we combine the recessive characteristics and propose a novel personal recommendation (NPR) algorithm as the basis of group recommendation algorithm. By the NPR algorithm, we get the user's prediction scores, and then we propose GRBCRC by combining the group weight calculation based on recessive characteristics. The experimental results show the better performance of our proposed algorithm.**

Keywords-group weigh; recessive characteristics; group recommendation; fusion strategy

I. INTRODUCTION

As an extension of the personal recommendation system, the group recommendation system is no longer to serve only one person [1], but to make a suitable recommendation for a group. The complexity of group users' information determines that the algorithm of personal recommendation system can not adapt to this scenario [2].

To solve this problem, we propose a novel algorithm for group recommendation based on combination of recessive characteristics (GRBCRC). The remainders of this paper are structured as follows. Section 2 discusses the background knowledge of SVD++ algorithm, recessive characteristic, fusion strategy. The detail of GRBCRC is described in section 3. Section 4 firstly compares the performance of NPR algorithm, with SVD++ algorithm, and then compares the performance of GRBCRC with Average Strategy and Borda Count. Finally, we conclude the paper in Section 5.

II. PRELIMINARIES

A. Recessive characteristics

The recessive characteristics are mainly divided into two parts: the interaction in social map and the content information. The social map of group users can reflect the familiarity of the users, and the perception of mutual familiarity between users is different [3]. As shown in Fig. 1, A, B, C, D and E belong to a group, each single node represent a user, the familiar relationship between the two users is expressed by the arrow. For example, the arrow from A to C represents the familiarity of A to C, and the number on the arrow is degree of familiarity. We use the number of interactions between users to determine the degree of familiarity between users. The greater the number, the higher the degree. Interaction is a kind of recessive characteristic, which includes forwarding, @, comment, etc.

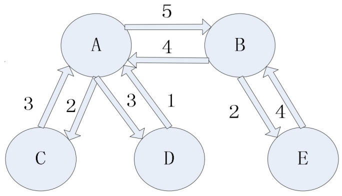

Figure 1. Social map.

Content information mainly includes labels and keywords. Labels summarize the main features and preferences of users. And the keyword is extracted from the contents of users' comment in social network. Each keyword will correspond to the corresponding TF-IDF weight.

B. SVD++ algorithm

SVD++ algorithm is the representative of the matrix decomposition algorithm. By reducing the dimension

978-1-7281-0207-8/18/$31.00 ©2018 IEEE
DOI 10.1109/BESC.2018.00011

method, SVD++ maps the sample data which set in high-dimensional space to the low-dimensional semantic space, and it also removes noise data and redundant data from sample data, reflecting the valuable part of the data.

The classic SVD++ model is as follows:

$$\hat{r}_{ui} = b_{ui} + q_i^T(p_u + \sum_{k \in N(u)} y_k \beta_k) \tag{1}$$

Here $\hat{r}_{ui}$ represents the evaluation of the user u to item i. In the differential formula $b_{ui} = \mu + b_u + b_i$, b_u represents the offset of the average score u, which is caused by personal reasons such as user's mental and physical condition, etc. b_i describes the deviation from the item itself. β_k is the weight coefficient corresponding to the character k, and y_k is the vector of character k. $q_i^T p_u$ represents the model of hidden factors, p_u is user's characteristic vector in n dimension, while q_i^T is the transition matrix of q_i.

III. Our Algorithm

Our algorithm is divided into two parts:

A. NPR algorithm

Personal recommendation system is the foundation of group recommendation system. The NPR algorithm is based on the SVD++ algorithm, and we combine interaction in social map and content information as recessive characteristics to SVD++ algorithm.

Interaction in social map: If user u has interactions with user w, we can extend vectors of user characteristics p_u in SVD++:

$$f_u = \frac{1}{\|\varepsilon_n\|_2} \sum_{x \in act(u)} \varepsilon_{u,x} s_x + p_u \tag{2}$$

$act(u)$ in (2) represents the entire project set which u had interacted with. $\varepsilon_{u,x}$ represents the number of interactions, S_x is the No. x recessive characteristic, and f_u is an extension of vectors of user characteristics.

$$f_u^1 = W(u, y) \sum_{y \in K(u)} s_{kw(y)} + \frac{1}{\sqrt{|L(u)|}} \sum_{n \in L(u)} s_{tag(n)} + f_u \tag{3}$$

$$f_u^1 = W(u, y) \sum_{y \in K(u)} s_{kw(y)} + \frac{1}{\sqrt{|L(u)|}} \sum_{n \in L(u)} s_{tag(n)} + f_u \tag{4}$$

Content information: In (3) and (4), $L(\cdot)$ is the user's label set, $K(\cdot)$ represents the keywords of the user and y is one of them. $W(\cdot, y)$ is the weight of y. The higher the weight is, the stronger the connection would be between y and the user.

The final model of the combination of recessive characteristics can be obtained by:

$$\hat{r}_{ui} = b_{ui} + h_i^T(f_u^1 + \sum_{k \in N(u)} y_k \beta_k) \tag{5}$$

In NPR algorithm, $|D|$ is the number of samples of the training set. $\overrightarrow{N_u}$ is designed to express the average number of elements in the group f of the user's feedback information. The parameters are set as follows: η is learning efficiency, λ is penalty factor, and m is training times and d is dimension. The time complexity is $O(|D|\overrightarrow{N_u})$.

Algorithm: NPR

```
Import: training data set D in the file;
Output: the personal predictive scoring list list。
 while the number of iterations ≤ m do
  for u, I, r_ui in D do
    p_u^im ← 0⃗ ∈ R^{1×d}
     for j in F(u) do
// F(u)=R(u)∪N(u), F(u) is the intersection of explicit
// feedback and implicit feedback
       for f in [0,d) do
         p_{u,f}^im ← p_{u,f}^im + β_{j,f} y_{j,f}
       end for
     end for
     z ← 0⃗ ∈ R^{1xd}
     for j in F(u) do
      e_ui ← r_ui − r̂_ui
// r̂_ui is the score predicted by the parameter generated
// by last iteration
       b_u ← b_u + η(e_ui − λ_1 b_u)
        for f in [0,d) do
         z_f ← z_f + e_ui β_{j,f} q_{i,f}
         p_{u,f} ← p_{u,f} + η(e_ui q_{i,f} − λ_2 p_{u,f})
         q_{i,f} ← q_{i,f} + η(e_m (p_{u,f}^im + p_{u,f}) − λ_2 q_{i,f})
        end for
       end for
       for j in F(u) do
// update the characteristic vectors of feedback
        for f in [0,d) do
         y_{i,f} ← y_{i,f} + η(z_f − λ_3 y_{i,f})
        end for
       end for
       List = top(y_{i,f}, η)
     end for
  end while
```

B. GRBCRC algorithm

1) Group weight calculation based on recessive characteristics.

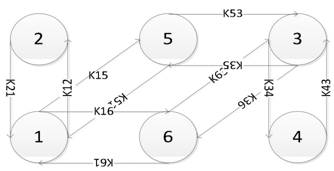

Figure 2. Group users' relationship based on recessive characteristics.

In Figure 2, there are 6 users marked 1 to 6. K_{ab} is the degree of familiarity from a to b, and the value of K_{ab} comes from the recessive characteristics. The Group Weight Calculation Based on Recessive Characteristics can be described by follows:

Step 1: Traverse all paths between user 1 and user 2 in the group, n is the number of all paths.

Step 2: If there is a direct path from user 1 to user 2 in the collection, which means no other intermediate nodes. Then, the degree of familiarity from user 1 to user 2 is $R_{12}= K_{12}$.

Step 3: If there is no direct path from user 1 to user 2, but there are other paths that contain intermediate nodes, traversing each indirect path. K_i is the sum of degree of familiarity calculated by path I, node_i is the number of intermediate nodes in path I, then the degree of familiarity from user 1 to user 2 is:

$$R_{12} = max\left\{\frac{K_i}{1+node\_i}\right\},\ i\in[1,n]. \quad (6)$$

Step 4: m is the number of users who have paths (both direct and indirect paths) with user 1. According to the methods from Step 1 to Step 3, calculated the sum of degree of familiarity from user 1 to other user s, and the value can be expressed by $R_1= \sum_{1\le j\le 5} R_{1j}$, the weight of user 1 is $W_1=R_1/m$.

2) Group recommendation algorithm based on recessive characteristics.

We get each user's prediction scores through the NPR algorithm, and then gather all the users' prediction scores to form a group predictive scoring matrix $\tilde{R}$, the matrix can be obtained by:

$$\tilde{R} = \begin{bmatrix} \tilde{r}_{1,1} & \tilde{r}_{1,2} & \cdots & \tilde{r}_{1,m} \\ \tilde{r}_{2,1} & \tilde{r}_{2,2} & \cdots & \tilde{r}_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{r}_{n,1} & \tilde{r}_{n,2} & \cdots & \tilde{r}_{n,m} \end{bmatrix} \quad (7)$$

$\hat{r}_{a,b}$ represents the user a's prediction score on item b. Since each user has a different weight in the group, we also use group weight calculation based on recessive characteristics to get the weight of each user. Then the weight matrix can be described by:

$$W=[w_1, w_2, w_3... w_n] \quad (8)$$

We use $\tilde{R}_{group}$ represents the prediction scores of group users, and $\tilde{R}_{group}$ can be obtained by:

$$\tilde{R}_{group} = \left|W^T \cdot \tilde{R}\right| = \left|\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}^T \cdot \begin{bmatrix} \tilde{r}_{1,1} & \tilde{r}_{1,2} & \cdots & \tilde{r}_{1,m} \\ \tilde{r}_{2,1} & \tilde{r}_{2,2} & \cdots & \tilde{r}_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{r}_{n,1} & \tilde{r}_{n,2} & \cdots & \tilde{r}_{n,m} \end{bmatrix}\right| = \left|\left[\tilde{R}_{group,1}\tilde{R}_{group,2}\cdots\tilde{R}_{group,m}\right]\right| \quad (9)$$

By multiplying the transition matrix of W and $\tilde{R}$, the weights of group users can be integrated into the prediction scores of group users. Finally, we get the final item recommendation list by arranging the prediction scores of group users in descending order.

IV. EXPERIMENTAL RESULT

To evaluate the performance of our algorithm, we choose the KDD Cup 2012 Track 1 as the data set, this data set records the data within a specific time interval of Tencent's micro-blog data set. And then we implement our algorithms on this data set.

A. Performance comparison among different personal recommendation algorithms

Firstly, we compare NPR algorithm's performance with SVD++ algorithm, randomly guessing and preference model. In the experiment we divide the data set to training data set and test data set. The algorithms are running on the training data set and the personal recommendation models are established. We use Mean Average Precision (MAP) as a test standard.

TABLE I. EXPERIMENTAL RESULT OF PERSONAL RECOMMENDATION

Model	Value of MAP	Training times
Randomly guessing	0.2085	---
Preference model	0.2335	30
SVD++ algorithm	0.2879	15
NPR algorithm	0.3357	15

The experimental results are shown that the MAP of NPR algorithm is larger than SVD++ algorithm, randomly guessing and preference model in all scenarios, which means NPR algorithm has a better performance than the other two algorithms.

B. Performance comparison among different group recommendation algorithms

We have obtained the users' prediction scores by personal recommendation algorithm, then we integrate the prediction scores into group scores by group algorithms. To evaluate the ability of GRBCRC algorithm, we compare the performance of GRBCRC algorithm with Average strategy and Borda Count.

The inspection standards are as follows:

- Mean Absolute Error(MAE)

$$\mathrm{MAE} = \frac{\sum_{u,i\in T}\left|r_{ui} - \hat{r}_{ui}\right|}{|T|} \tag{10}$$

- Precision

$$\mathrm{Precision} = \frac{\sum_{u\in U}\left|R(u)\cap T(u)\right|}{\sum_{u\in U}\left|R(u)\right|} \tag{11}$$

- Recall

Recall=The number of hit items / Total sample items (12)

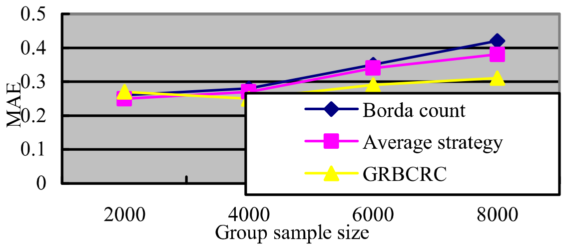

Figure 3. MAE variation with different group sample size.

The lower the value of MAE in the experimental results is, the smaller the gap would be between the predicted and the true values, which means that the algorithm is more accurate. Fig. 3 show the MAE of all algorithms increase with the size of group sample increasing. As a whole the MAE of GRBCRC is lower than Average strategy and Borda Count, and the MAE ascendant extent of GRBCRC is smaller than others. It proves GRBCRC is more accurate than Average strategy and Borda Count.

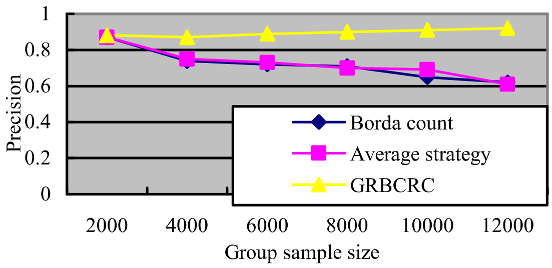

Figure 4. Precision variation with different group sample size.

Fig. 4 shows the variation of Precision with different group sample size. In all scenarios, the Precision of GRBCRC is higher than that of Average strategy and Borda Count. Further more, the Precision of GRBCRC ascends as the size of group sample increases, while that of two other algorithms descend.

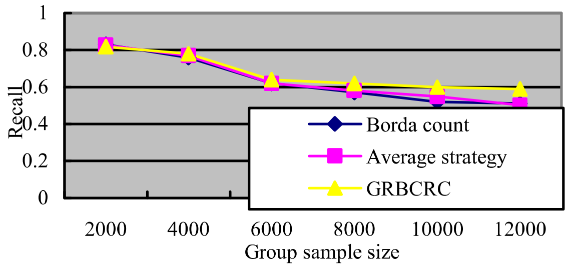

Figure 5. Recall variation with different group sample size.

We also compare the Recall of GRBCRC with Average strategy and Borda Count. In Fig. 5, the experiment results show that the Recall of all algorithms descend with the size of group sample increasing. And the Recall of GRBCRC is the highest in all scenarios. It reflects that the performance of GRBCRC is the best of all algorithms.

CONCLUSION AND DISCUSSION

The potential relationships and preferences of group users can be mined from the recessive characteristics. In this paper, firstly, we combine the recessive characteristics and propose a novel personal recommendation (NPR) algorithm as the basis of group recommendation algorithm. By the NPR algorithm, we get the user's prediction scores, and then we propose a novel algorithm for group recommendation based on combination of recessive characteristics (GRBCRC) by combining the group weight calculation based on recessive characteristics. To evaluate the performance of GRBCRC algorithm, we simulated it on the KDD Cup 2012 Track 1 data set, and compared its performance with Average strategy and Borda Count in several scenarios. The experimental results proved that GRBCRC can reach higher performance.

REFERENCES

[1] S. Seko, T. Yagi, M. Motegi, et al., "Group recommendation using feature space representing behavioral tendency and power balance among members", Proc. of the 5th ACM Conference on Recommender Systems, Chicago, IL, USA, pp.101-108, 2011.

[2] Meng Xiangwu, Ji Hong, "Recommendation system in big data environment", Journal of Beijing University of Posts and Telecommunications, Vol.38, No.2, pp.1-5, 2015.

[3] Zhang Yujie, Du Yulu, Meng Xiangwu, "Group recommendation system and its application", Journal of Computer Science, Vol.39, No.4, pp.745-764, 2016.