A Novel Hybrid Collaborative Filtering Algorithm Research for Music Recommendation

Yan Zhang
School of Electronic and Information Engineering
Harbin Institute of Technology(Shenzhen)
Shenzhen, China
ianzh@foxmail.com

Yu Li
School of Electronic and Information Engineering
Harbin Institute of Technology(Shenzhen)
Shenzhen, China
leeyuhit@163.com

***Abstract*—With convenient access to thousands of music pieces on the Internet, it is important to provide an effective recommendation system which could select possible preferable songs for users. In this paper, a novel hybrid algorithm is proposed for music recommendation. By importing a weighting factor concretely, item-based collaborative filtering algorithm is combined with user-based collaborative filtering algorithm to construct this new algorithm. The Mean Absolute Error (MAE) is employed to examine the performance of recommendation algorithms. It is clearly revealed that this new method could significantly improve the effectiveness of music recommendation system.**

Keywords—recommendation system; music; hybrid; collaborative filtering; MAE

I. INTRODUCTION AND RELATED WORKS

With the development of Internet, information overload[1] has become more and more serious in our life. The emergence of information overload has brought too much inconvenience[2] for production and daily life. For example, the increasing number of music pieces presents increased difficulties for searching preferable songs. Under usual conditions, we listen to music randomly that already exists in our music players. But when it comes to some special occasions, such as the situation that we prefer our favorite songs, the random model may not satisfy our needs. A special recommendation mechanism is needed to customize a special individual playing list for us.

User-based collaborative filtering and item-based collaborative filtering are two kinds of algorithms which is widely used in recommendation systems[3]. The user-based collaborative filtering[4] bring ratings of songs that are provided by the user to compare with ratings of other users on same songs. The system make predictions about the favorite songs of a user according to the preference of a user who shares a higher similarity with this user[5]. The item-based technique is similar with the user-based method. However, it make predictions according to the music similarities. It means the similarity between music pieces plays an indispensable role in this case. The user-item matrix are main original data resources in those two techniques. It is a serious sparse matrix when there is too many users or songs included at the same time with serious information shortage problem. What a pity is that these two techniques don't solve serious sparse problems perfectly in the predicting process.

Collaborative filtering (CF) is an effective method to predict automatically relevant values of the user by collecting information from other similar users or items. It has been widely used in e-commerce systems, such as Amazon web, Ebay web etc. It is proven that collaborative filtering algorithms have made great success in recommendation systems. There have been a lot of related works related to CF recommendation in both domestic and foreign countries. Zibin[6] combined user-based approach and item-based approach to predict Quality-of-Service (QoS) values for the current user by employing historical Web service QoS data from other similar users and similar Web services. Ye Zhang[7] presented a method to predict movie preferences by user-based collaborative filtering and item-based collaborative filtering method separately. SongJie Gong[8] used linear combination to both item rating similarity and item attribute similarity for predicting ratings to target user. But there weren't any experiments to support their assumptions. Sutheera[9] proposed an item-based collaborative filtering method which combines the attributes similarity of items with the rating matrix similarity. The user-music matrix in music recommendation is much sparser because information shortage problem is more serious in this case. Consequently, it is very important for us to propose a more effective method to recommend music songs for users.

In this paper, we propose a novel hybrid collaborative filtering algorithm. By importing a weighting factor, we combine item-based method with user-based method. In the process of implementation, we separately conduct several experiments to examine the efficiency of different predicting methods by varying with different influencing factors. The mean absolute error (MAE) method is employed to evaluate the experimental results. The lower the MAE, the more effective of the method. From the results of those experiments, we conclude that this novel hybrid method has an obvious improvement for recommendation efficiency compared with another algorithms since it has the lowest MAE.

II. COLLABORATIVE FILTERING ALGORITHM

Memory-based collaborative filtering algorithm is one of the most popular algorithm in recommendation system. It can be classified into two catalogues, which is user-based collaborative filtering[10] and item-based collaborative filtering[11] separately. The construction of these two collaborative methods is shown in the Fig. 1.

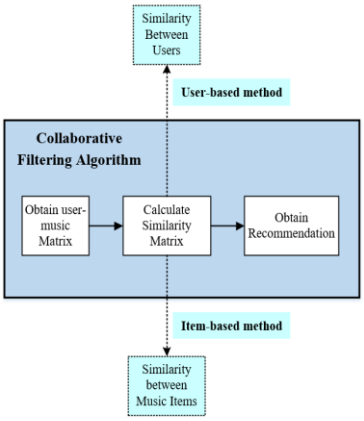

Fig. 1. Construction of general collaborative filtering algorithms for music recommendation

In the above Fig. 1, it shows that there are three main steps to carry out general collaborative filtering algorithm for music recommendation. First, we need to obtain user-music matrix from the original datasets. Second, we should calculate similarity matrix based on the user-music matrix. In user-based collaborative filtering algorithm, the calculation of similarity between users is needed. Similarly, we calculate similarity between music items in item-based filtering method. In the end, we need to recommend music items to users according to the predicting rating values. The calculation method of predicting rating is different in those two algorithms. But obtaining recommendation is the common goal in this step.

A. *Obtain User-Music Matrix*

From above fig 1, we can find out the first step is to obtain user-music matrix in collaborative filtering algorithms. The main dataset consists of user's information and corresponding music information which describes the preference of this user. A matrix denotes R is used to describe the user-music information. The value of music rated by the first user is stored in the first row of matrix R. When some users don't rate the music pieces, the value in the corresponding position in the matrix will be zero. The rating matrix is shown in the following Table I.

TABLE I. RATING MATRIX

Item User	item_1	item_2	…	item_n
user_1	R11	0	…	R1n
user_2	0	R22	…	0
…	…	…	…	…
user_m	0	Rm2	…	0

B. *Calculate Similarity Matrix*

The second step is to calculate similarity matrix. It means that we need to find out the neighboring users/music items of the target user/music. The difference between user-based method and item-based method begins in this step. The essence of this procedure is to calculate similarity between two vectors. Pearson correlation, cosine vector similarity and adjusted cosine vector similarity have been widely used to measure the similarity between two vectors. All of those calculating formulas are shown in the following,

1) *Pearson Correlation*

$$\mathrm{sim(i,j)} = \frac{\sum_{c\in I_{ij}}(R_{i,c}-\overline{R_i})(R_{j,c}-\overline{R_j})}{\sqrt{\sum_{c\in I_{ij}}(R_{i,c}-\overline{R_i})^2 \sum_{c\in I_{ij}}(R_{j,c}-\overline{R_j})^2}}$$

(1)

2) *Cosine Similarity*

$$\mathrm{sim(i,j)} = \frac{\sum_{c\in I_{ij}} R_{i,c} R_{j,c}}{\sqrt{\sum_{c\in I_{ij}} {R_{i,c}}^2 \sum_{c\in I_{ij}} {R_{j,c}}^2}}$$

(2)

3) *Adjusted Cosine Similarity*

$$\mathrm{sim(i,j)} = \frac{\sum_{c\in I_{ij}}(R_{i,c}-\overline{R_c})(R_{j,c}-\overline{R_c})}{\sqrt{\sum_{c\in I_{ij}}(R_{i,c}-\overline{R_c})^2 \sum_{c\in I_{ij}}(R_{j,c}-\overline{R_c})^2}}$$

(3)

C. *Obtain Recommendation*

The last but not least step is to obtain the recommendation for the active user. It means that we will get a list of predicting ratings to items and we recommend items which active user might rate higher values. Similarly, there are different techniques to predict rating values related to these two methods.

1) *User-based Method*

In user-based method, we predict the rating based on users' similarities.

$$\mathrm{R}^{pre}_{u,i} = \overline{R_u} + \frac{\sum_{a\in I_u} sim(u,a)*(R_{a,i}-\overline{R_a})}{\sum_{a\in I_u}|sim(u,a)|}$$

(4)

Where $\mathrm{R}^{pre}_{u,i}$ denotes the predicting rating of the active user u to the target music item i, $\overline{R_u}$ and $\overline{R_a}$ denotes the average rating of all music items rated by user u and user a separately, $sim(u,a)$ denotes the similarity between the user u and user a, I_u denotes the neighboring users set of the active user u, $R_{a,i}$ is the rating value of neighboring user a to the music item i.

Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 3 Issue 10, October - 2016

1) *Item-based Method*

In item-based method, prediction of rating is related to music items' similarities.

$$\mathrm{R}_{u,i}^{pre} = \overline{R_i} + \frac{\sum_{j \in I_i} sim(i,j) * (R_{u,j} - \overline{R_j})}{\sum_{j \in I_i} |sim(i,j)|} \quad (5)$$

Where $\mathrm{R}_{u,i}^{pre}$ denotes the predicting rating of the active user u to the target item i, $\overline{R_i}$ and $\overline{R_j}$ denotes average rating of music item i and music item j rated by the active user separately, $sim(i,j)$ denotes similarity between music item i and music item j, I_i denotes neighboring music items set of target item i, $R_{u,j}$ is rating value of neighboring music item j rated by active user u.

III. A Novel Hybrid Construction of CF Algorithm for Music Recommendation

We propose a novel hybrid construction of collaborative filtering algorithm for music recommendation in our research.

$$\mathrm{R}_{u,i}^{pre} = \lambda_u \left(\mathrm{R}_{u,i}^{pre}\right)_{user} + \lambda_i \left(\mathrm{R}_{u,i}^{pre}\right)_{item} \quad (6)$$

Where $\left(\mathrm{R}_{u,i}^{pre}\right)_{user}$ and $\left(\mathrm{R}_{u,i}^{pre}\right)_{item}$ denotes the ratings predicted by user-based collaborative filtering method and item-based collaborative filtering method respectively. λ_u and $\lambda_i$are the weighting value based on the similarity of users and music pieces respectively. What's more, it meets the conditions that

$$\lambda_u + \lambda_i = 1 \quad (7)$$

$$\lambda_u = \frac{\omega \times \sum_{x=1}^{N} sim(u,u_x) / N}{\omega \times \sum_{x=1}^{N} sim(u,u_x) / N + (1-\omega) \times \sum_{y=1}^{M} sim(i,i_y) / M} \quad (8)$$

$$\lambda_i = \frac{(1-\omega) \times \sum_{y=1}^{M} sim(i,i_y) / M}{\omega \times \sum_{x=1}^{N} sim(u,u_x) / N + (1-\omega) \times \sum_{y=1}^{M} sim(i,i_y) / M} \quad (9)$$

Where the weighting parameter $\omega \in (0,1)$. u_x denotes similar users of the active user u, N is the number of similar users, i_y denotes the similar music items of target music item i, M is the number of similar music items.

By importing a weighting factor, information in user similarity could be a kind of complement to the information in music similarity, vice versa. In a way, this method could provide more information in the processing of predicting ratings.

Theoretically, this novel method can significantly decrease the negative impact of sparse problem which usually occurs. The procedure for this new algorithm is shown in the following Fig. 2,

INPUT R_{user_music};

for i=1:Number_user
  do $sim(u,a)$;
end

for j=1:Number_songs
  do $sim(i,j)$;
end

for k=1:Chosed_number

$$\left(\mathrm{R}_{u,i}^{pre}\right)_{user} = \overline{R_u} + \frac{\sum_{a \in I_u} sim(u,a) * (R_{a,i} - \overline{R_a})}{\sum_{a \in I_u} |sim(u,a)|};$$

$$\left(\mathrm{R}_{u,i}^{pre}\right)_{item} = \overline{R_i} + \frac{\sum_{j \in I_i} sim(i,j) * (R_{u,j} - \overline{R_j})}{\sum_{j \in I_i} |sim(i,j)|};$$

$$\mathrm{R}_{u,i}^{pre} = \lambda_u \left(\mathrm{R}_{u,i}^{pre}\right)_{user} + \lambda_i \left(\mathrm{R}_{u,i}^{pre}\right)_{item};$$

end

OUTPUT $\mathrm{R}_{u,i}^{pre}$;

Fig. 2. A novel hybrid collaborative filtering algorithm

IV. Implementation and Results

A. *Dataset and Performance Measurement*

In order to prove the effectiveness of our novel method, we analyzed and compared our algorithm with the traditional user-based and item-based algorithm on MATLAB platform. We derived the data from Million Song Dataset website[12]. Million Song Dataset is a modern music collection set which can be used freely for research. The music dataset collection contains 1,019,318 unique users, 384,546 pieces of MSD tracks and 48,373,586 pieces of playing counts records. In our research, we chose 984 users and 816 pieces of songs and corresponding playing counts records as our dataset in order to obtain experimental results effectively. We assumed that the playing counts represented the users' preferences for music pieces. Obviously, it means that the user prefers certain music if he or she plays it more times than other songs. We assumed the rating value of songs which were not played by anyone as zero. The transformation rules are shown in the following table II.

TABLE II. Transformation of Playing Counts

Playing Counts(T)	Transferred Values
NONE	0
T=1	1.0
1<T≤2	1.5
2<T≤4	2.0
4<T≤7	2.5
7<T≤11	3.0

11<T≤16	3.5
16<T≤22	4.0
22<T≤29	4.5
T>29	5.0

After we got the rating matrix of user-music information, we randomly got train set and test set from an original dataset. Then we can carry out our experiments by employing this train sets and test sets.

In our research, we have implemented all of three methods to predict the rating of non-rating elements. The adjusted cosine vector similarity has been employed during the process of computing the similarity between users and music pieces. The metrics for evaluating the accuracy of a prediction algorithm can be either statistical accuracy metrics or decision-support metrics. In the following of our experiments, we used Mean Absolute Error (MAE), the statistical method, to evaluate the accuracy of prediction. MAE evaluates the accuracy of prediction by comparing its values. The smaller the value of MAE is, the more accurate the prediction algorithm is. The calculation of MAE is defined as follows,

$$MAE = \frac{\sum_{j=1}^{N}\left|R_j - R_j^{pre}\right|}{N} \qquad (10)$$

B. Implementation and Results

There are several factors which influences the accuracy of prediction. Such as the different predicting methods, the number of neighbors during the process of prediction and the percentage of testing data in the dataset. For the hybrid method, the parameter ω in the weighting factor is also an influence factor for the accuracy. In our research, we implemented several experiments to examine the effects of those factors.

1) Influence of the Number of Neighbors

In order to find the effects of number of neighbors in the experiment, we set the other variables as constants. Varying the number of neighbors for all those predicting methods, we got the results shown in the following figure 2.

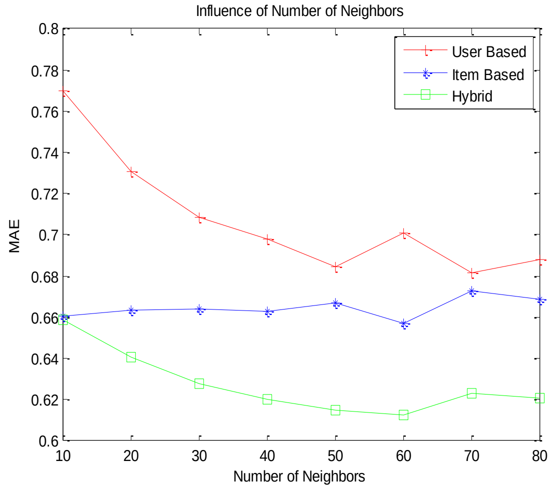

Fig. 3. MAE of three methods varying with increasing neighbors

2) Influence of Percentage of Test Data

In this experiment, we set the number of neighbors as constant. We chose the percentage of test data as variables and then we got the results for all of the three methods shown as following,

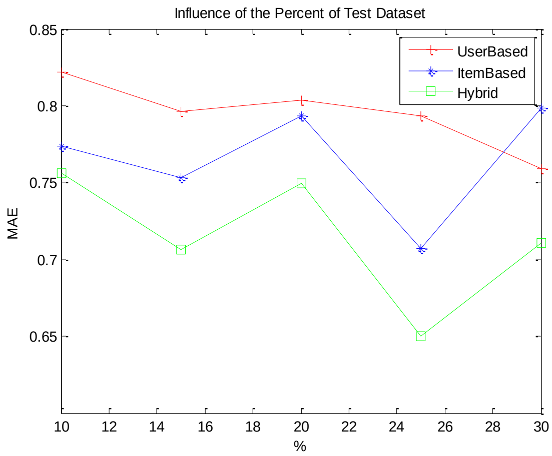

Fig. 4. MAE of three methods varying with increasing of percentage of test data

3) Influence of Weighting factor

For the hybrid method, we found the weighting parameter also influenced the experimental results. Varied the parameter ω in the weighting factor from 0.1 to 0.9, we got the results below.

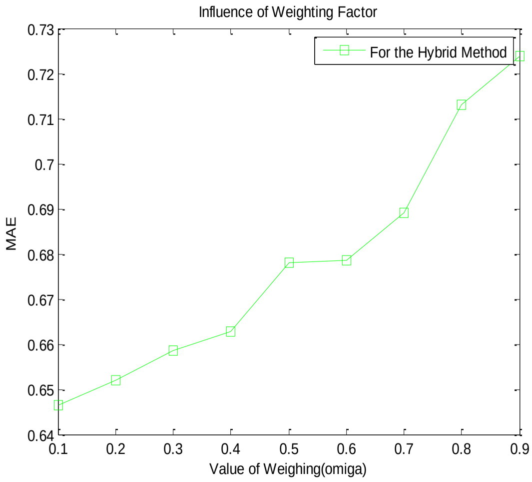

Fig. 5. MAE of hybrid CF algorithm varying with weighting parameter

V. Conclusion

In this paper, we propose a novel hybrid collaborative algorithm which combined user-based collaborative filtering method with item-based collaborative filtering method together. Theoretically, this new method can decrease the negative influence of sparse matrix by importing the weighting factor parameter. Experimentally, the recommendation effectiveness can be improved. As shown in the experimental results, the mean absolute error of this new method is lower than of user-based method and item-based method separately. It means the accuracy of predicting rating is increased. At the same time, the mean absolute error optimizes as the increasing of neighbor numbers. Varying with increasing the percentage of testing data, the MAE decreases non-monotonically. However, the prediction accuracy is the most optimal when the percentage is 25%. What's more, the weighting parameter ω in the novel hybrid method is also an influencing factor. The MAE optimizes gradually with the decreasing value of ω.

Acknowledgment

The work is supported by Microelectronic Technology Center in Harbin Institute of Technology. It provides us with excellent research environment and professional equipment.

References

[1] Y. Zhenhai, F. Yonghao, Z. Yikun and L. Shufen, "The Research of Modified Collaborative Filtering Recommendation Algorithm," *2015 7th International Conference on Information Technology in Medicine and Education (ITME)*, Huangshan, 2015, pp. 692-695.

[2] D. Margaris, P. Georgiadis and C. Vassilakis, "A collaborative filtering algorithm with clustering for personalized web service selection in business processes," *2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*, Athens, 2015, pp. 169-180.

[3] J. Herlocker, "Understanding and Improving Automated Collaborative Filtering Systems," Ph.D. T., University of Minnesota, pp 34-52, 2000.

[4] S. Puntheeranurak and T. Chaiwitooanukool, "An Item-based collaborative filtering method using Item-based hybrid similarity," *2011 IEEE 2nd International Conference on Software Engineering and Service Science*, Beijing, 2011, pp. 469-472.

[5] Y. Dai, H. Ye and S. Gong, "Personalized Recommendation Algorithm Using User Demography Information," *Knowledge Discovery and Data Mining, 2009. WKDD 2009. Second International Workshop on*, Moscow, 2009, pp. 100-103.

[6] Z. Zheng, H. Ma, M. R. Lyu and I. King, "QoS-Aware Web Service Recommendation by Collaborative Filtering," in *IEEE Transactions on Services Computing*, vol. 4, no. 2, pp. 140-152, April-June 2011.

[7] Y. Zhang ,W. Song, "A Collaborative Filtering Recommendation Algorithm Based on Item Genre and Rating Similarity," International Conference on Computational Intelligence and Natural Computing, pp 72-75, 2009.

[8] S. Gong ,H. Ye, X. Shi, "A Collaborative Recommender Combining Item Rating Similarity and Item Attribute Similarity," 2008 International Seminar on Business and Information Management, pp 58-60, 2008.

[9] S. Puntheeranurak and T. Chaiwitooanukool, "An Item-based collaborative filtering method using Item-based hybrid similarity," *2011 IEEE 2nd International Conference on Software Engineering and Service Science*, Beijing, 2011, pp. 469-472.

[10] K. Zhao and P. y. Lu, "Improved collaborative filtering approach based on user similarity combination," *2014 International Conference on Management Science & Engineering 21th Annual Conference Proceedings*, Helsinki, 2014, pp. 238-243.

[11] Y. Huang, "An Item Based Collaborative Filtering Using Item Clustering Prediction," ISECS International Colloquium on Computing, Communication, Control, and Management, pp54-56, 2009.

[12] labrosa.ee.columbia.edu/millionsong/lastfm.