

A Hybrid Recommendation Algorithm Based on Social and Collaborative Filtering

Guo Li^{1,a}, Yang Yijun^{2,b} and Huang Rong^{2,c}

¹ Department of Information Engineering, HBEPC Technology Training center(Wuhan Electric Power Technical College), Wuhan 430079, China

²School of Information Engineering Wuhan University of Technology, Wuhan 430070, China

^a email: 3203483@qq.com, ^b 443166189@qq.com, ^c1297487564@qq.com

Keywords: item-based collaborative filtering; social relationships; hybrid recommendation

Abstract. Social-based recommendation and collaborative filtering-based recommendation have their own characteristics. Considering that traditional collaborative filtering only makes use of users' behavior data but ignores users' social relationships, a recommendation algorithm combined with social and collaborative filtering was proposed in this paper. Traditional item-based collaborative filtering algorithm was improved first, and then the hybrid recommendation algorithm was constructed by considering the complementarity of users' behavior data and social relationships, which can relieve the existing problems of collaborative filtering such as data sparse and cold start and is proved to improve the accuracy of the recommendation through experiment.

Introduction

Traditional personalized recommendation technology is mainly calculating the users' preferences for items by mining users' behavior data and using user's existing preferences for items to predict items which user may like, then the items is recommended to target user. The current mainstream recommendation algorithm is collaborative filtering algorithm. Although collaborative filtering algorithm has been widely used, but there are still two major issues of data sparse and cold start. In order to solve these problems, many scholars have put forward solutions. Yu-fang Zhang et al [1] proposed a two-step filling matrix method combined with conditional probability algorithm, which can partly relieve data sparse and improve the effectiveness of the recommendation. Yang Yang et al [2] proposed a recommendation algorithm based on matrix factorization and user nearest neighbor model, which can effectively improve the accuracy of the prediction score. Shu-chao Ma [3] proposed a graph search algorithm based on two layer graph model, which can solve the problem of data sparsity and scalability. Sotirios P.Chatzis [4] proposed a dynamic Bayesian probability matrix decomposition model, which can reflect that users' behaviors change with time and is proved to have some certain superiority through experiment.

The above methods all can effectively relieve the data sparsity and partly improve recommender efficiency of traditional collaborative filtering. With the rise of the social E-commerce model, personalized recommendation combined with social network becomes a new research direction of the recommendation system. Previous studies have shown that recommendation combined with social information can effectively improve the effect of personalized recommendation [5-6]. Collaborative filtering-based recommendation reflect the continuation of the users' own interests or the preferences of the people with similar interests, while social-based recommendation can simulate the real society, which reflect the trust of friends. Therefore, combining collaborative filtering-based recommendation with social-based recommendation becomes a hot spot in current research. Ma et al [7] proposed a factor model based on matrix factorization to alleviate the data sparsity and poor prediction accuracy problems by incorporating social information. Yu et al [8] proposed a matrix factorization model with social information as the regular term, which use social information as a regular term to constrain the matrix decomposition model and is proved to improve the effect of Micro-blogs recommendation through experiment.

On the basis of existing researches, a hybrid algorithm was proposed, which is based on improved item-based collaborative filtering recommendation and social-based recommendation. The algorithm considers the users' behavior data and social relationships and is proved to improve the accuracy of the recommendation effectively through experiment.

Basic Algorithm

A. Improved item-based collaborative filtering algorithm

The basic idea of item-based collaborative filtering algorithm is recommending items which is similar to the items that user have enjoyed to the target user. Similarity between items is calculated by users' historical data. The main steps of collaborative filtering algorithm are as follows: users' preferences mining, generating nearest neighbors and prediction and recommendation.

1) Users' preferences mining

Users' preferences are dug out by uses' historical behavior data, and then the users' preferences is converted to user preference matrix. User preference matrix is as shown in the formula (1).

$$P_{m*n} = ((p_{11} p_{12} \dots p_{1n}), (p_{21} p_{22} \dots p_{2n}), \dots, (p_{m1} p_{m2} \dots p_{mn})). \quad (1)$$

2) Generating nearest neighbors

According to the user preference matrix, similarity between items can be calculated, then we can get the nearest neighbors of each item. There are two common methods to select the nearest neighbors. One method is selecting K items with the largest similarity as the nearest neighbors. Another method is selecting items of which the similarity is greater than a certain threshold as the nearest neighbors. The first method is selected in this paper. When calculating the similarity of items, the cosine similarity is chosen to calculate. Similarity between items calculation is shown in the formula (2).

$$Sim(i, j) = (\sum_{u \in N(ij)} p_{ui} p_{uj}) / (\sqrt{\sum_{u \in N(i)} p_{ui}^2} \sqrt{\sum_{u \in N(j)} p_{uj}^2}). \quad (2)$$

Where $N(i)$ denotes the users who like item i , $N(j)$ denotes the users who like item j and $N(ij)$ denotes the users who like both item i and item j .

In the item-based collaborative filtering algorithm, similarity between items is determined by the number of common users who both like them, therefore, similarities based on a greater user support are more reliable. However, when similarity between items is calculated by cosine similarity, one situation may occur that some items which share only a handful of common users have greater similarity than items which share more common users. So a correction term CT can be added to improve the cosine similarity. Besides, the similarity between items is reflected by the users who like them, but not every user's contribution to the similarity between items is the same. Some users are rather active, they produce behaviors on many items, but those behaviors are not out of their own interests, therefore, they have no contribution to the similarity between items. Generally, the contribution of active users to similarity between items is less than that of inactive users. Therefore, a penalty term for active users PT can be added to improve the cosine similarity. Based on the analysis above, an improved similarity calculation method is as shown in the formula (3).

$$Sim(i, j) = CT \cdot PT \cdot \cos(i, j). \quad (3)$$

$$CT = N_{ij} / (N_{ij} + I). \quad (4)$$

$$PT = \sum_{u \in N(i) \cap N(j)} (1 / \log(1 + |I(u)|)). \quad (5)$$

Where N_{ij} denotes the number of users who like both item i and item j , I denotes correction factor which can be set according to the actual environment. $|I(u)|$ denotes the number of items which user u likes, $N(i) \cap N(j)$ denotes the users who like both item i and item j .

3) Prediction and recommendation

According to the user's preferences and the nearest neighbors of items, the value of the user preference for items can be calculated, then items is sorted by preference value and the N items with

the largest value of preference is recommended to user. The value of user preference for items can be calculated as shown in the formula (6).

$$p(u, j) = \sum_{i \in N(u) \cap S(j, K)} sim(i, j) * r_{ui}. \quad (6)$$

Where $N(u)$ denotes the items which user u likes, $S(j, K)$ denotes the nearest neighbors of item j .

B.Social-based recommendation algorithm

According to the data sets of user preference and social relations, a simple recommendation algorithm is that items which user's friends like is recommended to the user. In other words, the friends' preference for items can come to the target user's preference for items. Formula is as shown in the formula (7).

$$P_{ui} = \sum_{v \in F(u)} w_{uv} P_{vi}. \quad (7)$$

Where $F(u)$ denotes the friends collection followed by user u , P_{vi} denotes how the user's friend v likes the item i , w_{uv} denotes the weight between friends, generally consisting of two parts: one is the degree of familiarity between friends, the other is the similarity of interest between friends.

1) Degree of familiarity between friends

Degree of familiarity between friends reflects the degree of trust between friends, of which the value can be calculated by a series of interactive data between friends, common interactive operations are sharing content, sending private message, comment and praise and so on. Generally speaking, the lower frequency, the higher the weight, the weight can be sorted: {sharing content}> {sending private message}> {comment}> {praise}. We define degree of familiarity between friends as *familiarity*(u, v). An appropriate formula is as follows:

$$familiarity(u, v) = \left(\sum_{i \in types(u, v)} \log(1 + count(i)) * weight(i) \right) / \left(\max_{k \in F(u)} familiarity(u, k) \right). \quad (8)$$

Where $types(u, v)$ denotes the all types of operations, $count(i)$ denotes the occurrence times of the operation type i , $weight(i)$ denotes the weight of the operation type i .

2) Similarity of interests between friends

In addition to considering degree of familiarity between friends, the similarity of interests between friends need to be considered, for example, we are very familiar to our parents, but our interests does not resemble. The similarity of interests between friends can be calculated by the items they like, the formula is as follows:

$$similarity(u, v) = \left(\sum_{i \in I_{uv}} P_{ui} P_{vi} \right) / \left(\sqrt{\sum_{i \in I(u)} P_{ui}^2} \sqrt{\sum_{i \in I(v)} P_{vi}^2} \right). \quad (9)$$

Where I_u and I_v respectively denote the items which the user u and v like, I_{uv} denotes the items which user u and user v both like.

Normalizing *familiarity*(u, v) and *similarity*(u, v), then the weight between friends is calculated by linear weighted sum of *familiarity*(u, v) and *similarity*(u, v), the formula is as follows:

$$w_{uv} = a \cdot familiarity(u, v) + (1-a) \cdot similarity(u, v). \quad (10)$$

By changing the value a in the formula above to adjust the weight of the user's familiarity and similarity of interests.

Hybrid Recommendation Algorithm

Item-based collaborative filtering algorithm is simple and effective, which has been the mainstream personalized recommendation algorithm, while it still has the problem of data sparse and cold start.

Yet, there are a lot of good friendships based on trust and common interests in social networks, and Social-based recommendation can increase the user’s confidence in the results of the recommendation, as well as ease the problem of sparse data and cold start. Combined with the characteristics of two algorithms above, a hybrid recommendation algorithm is proposed in the paper, which considered both users’ own preferences and friend’s preferences and can improve the accuracy of the recommendation. Besides, when the user doesn’t have behavior data, items can be recommended to the target user through the user’s social relationships, thus relieving the cold-start problems of collaborative filtering.

Definition: let L denotes the length of final lists of recommendation. Let G_{cf} denotes recommendation results of item-based collaborative filtering, whose length is $K*L$. Let G_{sn} denotes recommendation results of social-based recommendation, whose length is $K*L$. And then the value of user preference for items in the two recommendation results is normalized and linear weighted. The linear weighted sum method is shown in the formula (11).

$$p(u,i) = a \cdot p_{cf}(u,i) + (1-a) \cdot p_{sn}(u,i). \tag{11}$$

In the above equation, $i \in G_{cf} \cup G_{sn}$, if $i \notin G_{cf}$, then $p_{cf}(u,i) = 0$; if $i \notin G_{sn}$, then $p_{sn}(u,i) = 0$. The union of the two recommendation results is sorted according to the final value of preference, then L items with largest value of preference can be recommended to target user. The implementation process of the algorithm mentioned above is shown in Figure 1.

As shown above, the hybrid algorithm actually is constructing a recommendation algorithm with both the feature of social and the characteristic of item-based collaborative filtering, thus obtaining the stronger ability to describe features of user and improving the effect of the recommendation.

In the algorithm above, the weight selection is rather important. If $a > 0.5$, the recommendation results bias towards item-based collaborative filtering; if $a < 0.5$, the recommendation results bias towards social-based recommendation. According to the following experimental analysis, you can find that accuracy of item-based collaborative filtering recommendation is higher than that of social-based recommendation, therefore, in order to improve the accuracy, generally choosing $a > 0.5$.

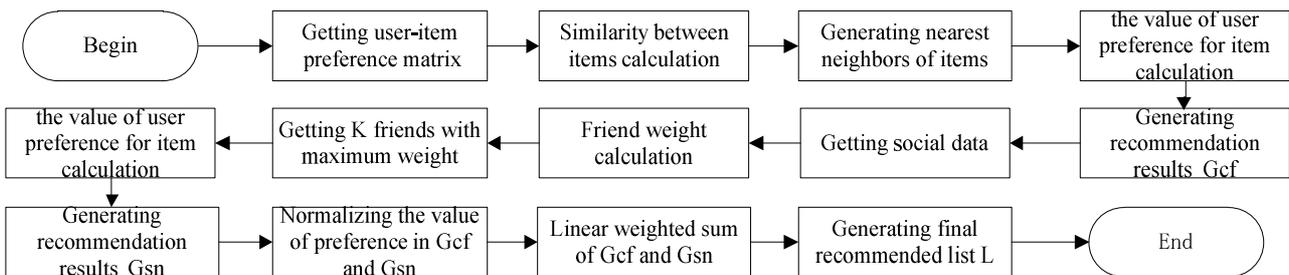


Fig.1. Implementation process of the algorithm

Experimental Analysis

A.Experimental data set

In order to verify the effectiveness of personalized recommendation algorithm proposed in this paper, Epinions Dataset was chosen as the data source for the experiments. Epinions Dataset comes from a well-known knowledge sharing and review site that was established in 1999, which contains two data sets, one is a user rating data set, and another is a trust relationship data set. The trust relation of trust data set is one-way, which can simulate the one-way social network like Twitter.

As the data set is relatively large, subset of data is extracted to experiment. The total data extracted in the paper contains 1607 users, 735 films, 16874 user ratings and 22492 trust relations.

The user rating data can be treated as extraction from users' behavior data, and rating value can indicate the value of user preference for items. According to the concept of sparsity, sparsity of the data set extracted in the paper is 1.42%.

B.Experimental design

User rating data set was randomly divided into 5 parts, one of which was selected as the testing set, others were merged into training sets. Building a recommendation model on the training set, then using the recommendation model to predict and recommend items for users in the testing set, finally, observing the precision and recall of recommendation. The main contents of the experiment are as follows:

- Comparing the precision and recall of traditional item-based collaborative filtering algorithm (ItemCF) and improved item-based collaborative filtering algorithm (ImproItemCF).
- Comparing the precision and recall of improved item-based collaborative filtering algorithm (ImproItemCF), social-based recommendation (SN) and the hybrid recommendation proposed in the paper (SNCF).

C.Experimental result

1) Experiment 1

Through experiment, comparing the precision and recall of traditional item-based collaborative filtering algorithm (ItemCF) and improved item-based collaborative filtering algorithm (ImproItemCF). The number of nearest neighbors of each item K in this experiment is set to 20, and the parameter l in the formula (4) is set to 5, the precision and recall curves are shown in Figure 2 and Figure 3.

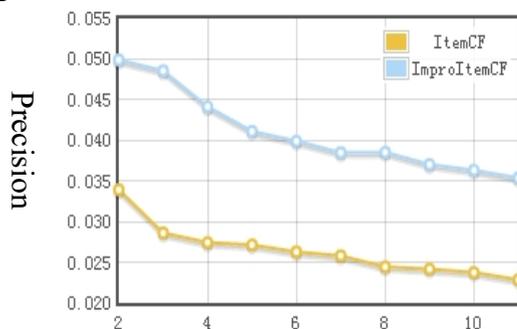


Fig.2 Precision Curve

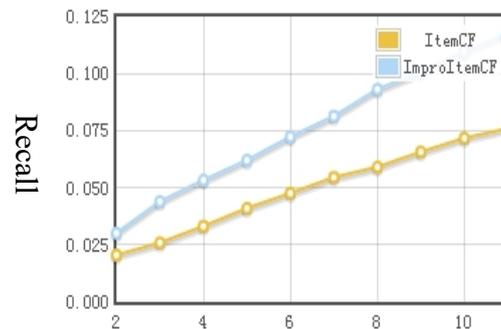


Fig.3.Recall Curve

As shown in Figure 2 and Figure 3, the precision and recall vary with the length of the recommendation lists. Through a comprehensive comparison of precision and recall, the effect of improved item-based collaborative filtering recommendation (ImproItemCF) was obviously better than that of traditional item-based collaborative filtering recommendation (ItemCF).

2) Experiment 2

Through experiment, comparing the precision and recall of improved item-based collaborative filtering algorithm (ImproItemCF), social-based recommendation (SN) and the hybrid recommendation proposed in the paper (SNCF).

After several experiments, the parameter a in the formula (11) is set to 0.7, and another parameter a in the formula (10) is set to 0.5, since there is no friend interactive data in the experimental data set, the familiarity between friends in social-based recommendation is set to 1. The precision and recall curves of three recommendation are shown in Figure4 and Figure 5, here, in the hybrid recommendation algorithm, the length of the single recommendation results is set to $5*N$.

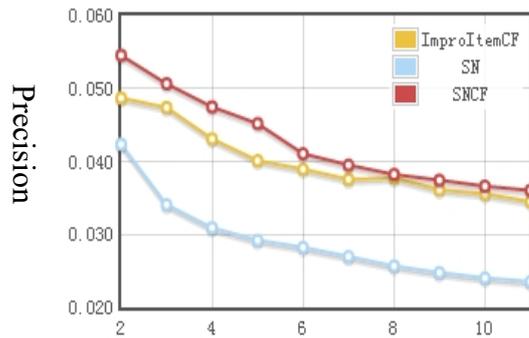


Fig.4.Pprecision Curve

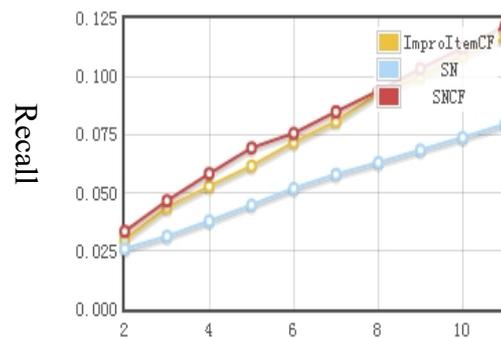


Fig.5. Recall Curve

As shown in Figure 4 and Figure 5, the effect of item-based collaborative filtering recommendation was better than that of social-based recommendation, the effect of the hybrid recommendation algorithm proposed in the paper was better than that of the other two separate recommendation algorithm, which was proved that the hybrid recommendation method proposed in this paper is effective.

Conclusions

Firstly, similarity between items calculation of item-based collaborative filtering was improved, a correction term for common users CT and a penalty term for active users PT were added to improve the cosine similarity. Then on the basis of item-based collaborative filtering and social-based recommendation, a hybrid recommendation algorithm was proposed in the paper, which was proved to improve the precision and recall of recommendation through experiment. The personalized recommendation algorithm fusion of social network is of great theoretical value and practical value. How to use social data to improve the recommendation effect of the existing recommendation system need further research.

References

- [1] Yu-fang Zhang, Jin-long Dai, Zhong-yang Xiong. Collaborative filtering algorithm based on two step filling for alleviating data sparsity [J]. Application Research of Computers, 2013, 30(9): 2602-2605.
- [2] YangYang, Yang Xiang, Lei Xiong. Collaborative filtering and recommendation algorithm based on matrix factorization and user nearest neighbor model [J]. Journal of Computer Applications, 2012, 32(2): 395-398.
- [3] Shu-chao Ma. Recommendation Algorithm based on Network Structure [D]. University of Science and Technology of China, 2013.
- [4] Sotirios P.Chatzis. Dynamic Bayesian Probabilistic Matrix Factorization[C]// Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014: 1731-1737.
- [5] De Meo P, Ferrara E, Fiumara G, et al. Improving recommendation quality by merging collaborative filtering and social relationships[C]// Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on IEEE, 2011:587-592.
- [6] Sedhain S, Sanner S, Braziunas D, et al. Social collaborative filtering for cold-start recommendations[C]// ACM Conference on Recommender Systems. ACM, 2014:345-348.
- [7] Ma H, Zhou T C, Lyu M R, et al. Improving Recommender Systems by Incorporating Social Contextual Information[J]. Acm Transactions on Information Systems, 2011, 29(2):219-230.
- [8] Yu Y, Qiu R G. Follower recommendation in microblog using matrix factorization model with structural regularization [J]. Scientific World Journal, 2014, 2014(3):435-444.