Reconstructing Trust Matrix to Improve Prediction Accuracy and Solve Cold User Problem in Recommender Systems

Shunpan Liang*, Lin Ma, Fuyong Yuan
College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China
liangshunpan@ysu.edu.cn
*: corresponding author.

ABSTRACT

Recommender systems(RS) are a type of solution to the information overload problem suffered by users of websites that allow the rating of certain items. Collaborative filtering(CF) is one of the most widely used methods in personalized RS. The most critical part of collaborative filtering is to compute similarities among users using a user-item rating matrix based on which recommendations can be generated. However, CF suffers from several inherent issues, such as data sparsity and cold start, which affect the quality of recommendation seriously. To address these problems, we propose a reconstructing trust matrix measure in this paper, which combines user similarity and weighted trust propagation. Specifically, we first remove the trust relationship of those users whose similarity falls below a certain threshold. We then add the users that are not in the trust matrix into it when the similarity between them exceeds a certain threshold. Finally, weighted trust propagation is considered, aiming to distinguish trusted neighbors in a shorter distance with those in a longer distance and incorporate more trusted neighbors, especially useful for cold users. Experimental results on two real-world data sets show that our method achieves superior accuracy and it can solve cold user problem as well.

CCS Concepts

•Information systems → World Wide Web; •Web searching and information discovery → Social recommendation;

Keywords

Recommender systems; Collaborative filtering; Reconstructing trust matrix; Accuracy; Cold user

1. INTRODUCTION

The emergence of Web 2.0 applications has greatly changed usersąŕ styles of online activities from searching and browsing to interacting and sharing[1]. However, it is up against the challenge of information overload meanwhile, which is well-known as requiring to spend a lot of time to find useful information. Recommender systems (RS) are designed to cope with the problem and heavily used in e-commerce applications such as Amazon.com, Ebay.com, and Netflix.com etc.

Collaborative filtering(CF) has become the most well-known and commonly used techniques to generate recommendations in RS[2]. The intuition of CF is making predictions about a userąŕs preferences or tastes based on the preferences of a group of users that are considered similar to this user. However, CF suffers from several inherent issues, for example, data sparsity and cold start. Data sparsity refers to the difficulty in finding sufficient and reliable similar users due to the fact that users in general only rate a small portion of items, while cold start includes two main problems: (1) How to recommend to the new users who have not rated any items; (2) How to deal with the items never rated or purchased.To resolve the issue, additional information such as trust[3] is studied and incorporated into CF. But as cold users do not have a large number of trusted neighbors, we cannot use the trust information directly. Fortunately, trust can be propagated along with the web-of-trust. That is, if user A trusts B and B trusts C, it can be inferred that A trusts C in some degree. Therefore, it is necessary to propagate trust in order to find more trusted neighbors for cold users.

The majority of earlier approaches for prediction in trust-based systems make predictions utilizing all the trust statements present in the data. But user A trusting user B dose not signify that the similarity between A and B will also be high and the low similarity in trust relationship will impact prediction quality adversely. Thus in this paper we first remove the trust relationship of those users whose similarity falls below a certain threshold. We then add the users that are not in the trust matrix into it when the similarity between them exceeds a certain threshold. Besides, weighted trust propagation is considered, aiming to distinguish trusted neighbors in a shorter distance with those in a longer distance and to incorporate more trusted neighbors, which is especially useful for cold users.

The rest of the paper is organized as follows. In Section 2, we give a brief overview of related research on trust-based CF. The proposed approach is then elaborated in Section 3. Experiments on two real-world data sets are conducted in Section 4. Finally, Section 5 concludes our work.

2. RELATED WORK

To better model user preferences for the cold users who only rated a few items, additional information is often adopted and trust is of less ambiguity and more relevant to similarity. Till now many trust-based approaches have been proposed.

Jamali et al. designed the TrustWalker approach[4] to randomly select trusted neighbors in the trust networks, where trust information of the selected neighbors is combined with an item-based technique to predict item ratings. In contrast, our work focuses on generating predictions by combining trust information with a user-based technique.

Massa et al. proposed the MoleTrust algorithm[5], which performs depth-first search, to propagate and infer trust in the trust networks. Empirical results show that the coverage is significantly enlarged but the accuracy remains comparable when propagating trust.

Ray et al. presented another trusted method[6]. The trust links between two users will be removed if their similarity is lower than a threshold. But empirical results show that good performance is achieved at the cost of poor coverage, and it fails to function in cold conditions where user similarity may not be computable.

Recently, Deng et al. proposed a social network based service recommendation method with trust enhancement known as RelevantTrustWalker[7]. First, a matrix factorization method is utilized to assess the degree of trust between users in social network. Next, an extended random walk algorithm is proposed to obtain recommendation results.

Guo et al. presented a merged method called Mergex[8], which merged the ratings of trusted neighbors in order to form a new and more complete rating pro?le for the active users based on which recommendations can be generated by integrating similarity and trust into CF.

Our work focuses on generating predictions by combining weighted trust propagation with a user-based technique. In order to achieve a better result, we first remove the trust statements between two users if their similarity is lower than a threshold considering low similarity in trust will affect the prediction accuracy. Since high similarity in trust statements can improve the recommendation results, the trust links between two users will be added into the original trust matrix if their similarity overtops a certain threshold. Finally, weighted trust propagation will be considered aiming to distinguish trusted neighbors in a shorter distance with those in a longer distance and find more trusted neighbors, which is especially useful for cold users.

3. OUR METHOD

In this section, we will present the specific method. We will introduce how to incorporate similarity and trust to reconstruct trust matrix. Then the weighted trust propagation will be explained.

3.1 Reconstruct trust matrix algorithm(RTMA)

The majority of earlier approaches for prediction in trust-based systems make predictions utilizing all the trust statements present in the data. But as we all know, user A trusting user B does not mean that the correlation between A and B will be also high. So we present the method of combining similarity and trust to reconstruct trust matrix. Pearson Correlation Coefficient(PCC) is a preferable method[9] and we adopt PCC to compute similarity in recommender systems, it is defined as follow:

$$S_{a,b} = \frac{\sum_i (r_{a,i} - \overline{r_a})(r_{b,i} - \overline{r_b})}{\sqrt{\sum_i (r_{a,i} - \overline{r_a})^2}\sqrt{\sum_i (r_{a,i} - \overline{r_a})^2}} \tag{1}$$

Where i represents the set of commonly rated items by users a and b. $r_{a,i}$ and $r_{b,i}$ denote the rating of item by users a and b respectively. $\overline{r_a}$ and $\overline{r_b}$ are the average rating value of users a and b .

In a recommender system, $Trust$ presents the traditional trust matrix, and $t_{a,b}$ is the trust value between a and b. Generally, the trust value is binary, namely 0 and 1, where 0 stands for distrust and 1 indicates absolutely trust. We set two similarity threshold α and β when reconstructing trust matrix. First, the low similarity in trust relationship will reduce the quality of rating prediction, thus, for the two trust users a and b, namely $t_{a,b} = 1$, we will remove their trust statement if $S_{a,b} < \alpha$ and reserve their trust relationship if $S_{a,b} > \alpha$. It is defined as (2):

$$t_{a,b} = \begin{cases} 0, & t_{a,b} = 1 \text{ and } S_{a,b} < \alpha \\ 1, & t_{a,b} = 1 \text{ and } S_{a,b} > \alpha \end{cases} \tag{2}$$

The high similarity in trust matrix will improve recommendation results, accordingly, for the two users a and b which are not in the trust matrix, namely $t_{a,b} = 0$, if $S_{a,b} > \beta$ we will add the trust relationship between a and b. However, if $S_{a,b} < \beta$, the trust relationship cannot be added for a and b. It is defined as (3):

$$t_{a,b} = \begin{cases} 1, & t_{a,b} = 0 \text{ and } S_{a,b} > \beta \\ 0, & t_{a,b} = 0 \text{ and } S_{a,b} < \beta \end{cases} \tag{3}$$

Through the above steps, the trust matrix is reconstructed, which is defined as $UpTrust$.

3.2 Weighted trust propagation algorithm(WTPA)

The cold users are generally defined as the users who have rated less than five items[10]. Since cold users are usually less active in the systems, they may not have a large number of trusted neighbors. Fortunately, trust can be propagated along with the web-of-trust. That is, if user A trusts B and B trusts C, it can be inferred that A trusts C in some degree. Therefore, it is necessary to propagate trust in order to find more trusted neighbors for cold user problem. MoleTrust[11] is the method using the above trust propagation to infer the trust value of indirectly connected users. Note that the trust value in the reconstructed trust matrix $UpTrust$ is binary, i.e., 0 or 1. As a result, the inferred trust value by the MoleTrust will be also binary, and thus we cannot distinguish trusted neighbors in a shorter distance with those in a longer distance. Hence, we adopt a weighting factor to devalue the inferred trust in a long distance:

$$t'_{a,b} = \frac{1}{d} \times t_{a,b} \tag{4}$$

Where $t_{a,b}$ denotes the trust value in the trust matrix, $t'_{a,b}$ is the weighted trust value and d is the shortest distance between users a and b determined by a breath first search algorithm. Note that the greater d is, the more trusted neighbors will be inferred. However, the more cost will be taken and more noise is likely to be incorporated. In this work, we restrict $d \leq 3$ to prevent meaningless searching and save computational cost for large-scale data sets. In fact, as we will show later, our method works well enough when d is small. The trust matrix is $UpTrust\_d$ after using weighted trust propagation.

3.3 The description of our method

According to the above description, the algorithm of RT-MA and WTPA are expressed as follows:

Step1: According to user-item rating matrix and PCC algorithm, compute similarities between every two users.

Step2: Predefine threshold α and β.

Step3: For two users u and v in user-item rating matrix, reset $t_{u,v}$ according to definition (2) and (3).

Step4: Get the $UpTrust$.

Step5: For two users u and v in UpTrust_d, reset $t^{'}_{u,v}$ according to definition (4).

Step6: Get the $UpTrust\_d$.

4. EXPERIMENTS

To verify our method, we conduct experiments on two real-world data sets using the 5-fold cross validation method. The data set is split into five disjoint sets; for each iteration, four folds are used as training set and one as testing set. We apply the K-Nearest Neighbor(KNN) approach to select a group of similar users whose ranking is in the top K according to similarity; we vary K from 5 to 50 with step 5. The ratings of selected similar users are aggregated to predict itemsąŕ ratings by a mean-centering approach[12].

4.1 Data sets

Two real-world data sets are used in our experiments, namely FilmTrust and Epinions. They are available data sets that contain both the trust matrix and user-item rating matrix.

FilmTrust is a trust-based social site in which users can rate and review movies. It includes 1986 users, 2071 movies and 35497 ratings. The ratings take values from 0.5 to 4.0with step 0.5. In addition, 1853 trust ratings that are issued by 609 users are gathered. The sparsity is 98.86%.

Epinions is a website in which users can express their opinions about items (such as movies, books, and software) by assigning numerical ratings and writing text reviews. The data set consists of 49K users who issued 664K ratings over 139K different items and 478K trust statements. The ratings are integers rated from 1 to 5 and the sparsity is 99.95%. The trust values of both data sets are binary (either 1 or 0).

4.2 Evaluation metrics

The evaluation metrics are mean absolute error (MAE), root mean square error (RMSE) and rating coverage (RC) respectively. They are defined as follows:

$$MAE = \sum_{(u,i)\in T} |r_{ui} - \widehat{r}_{ui}|/|T| \tag{5}$$

$$RMSE = \sqrt{\sum_{(u,i)\in T} |r_{ui} - \widehat{r}_{ui}|/|T|} \tag{6}$$

Where T represents the set of prediction results and $|T|$ is the number of the set, and $\widehat{r}_{ui}$ is the prediction rating of user u to item i.

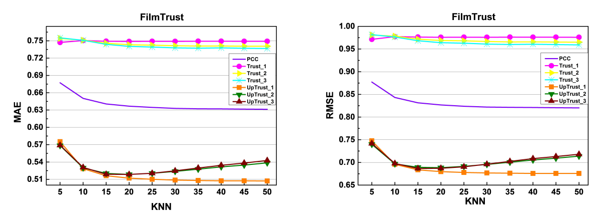
$$RC = \frac{M}{N} \tag{7}$$

Where M and N are the number of predictable and all the testing ratings, respectively.

4.3 Results and analysis

In this section, we will verify our method of reconstructing trust matrix.

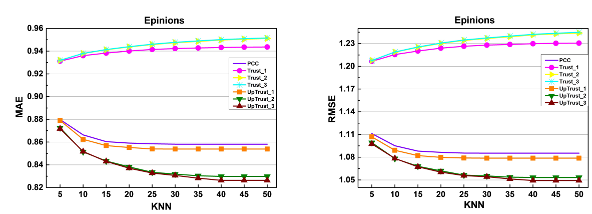
Fig. 1(a) and (b) are the performances of these approaches on FilmTrust in terms of MAE and RMSE respectively. The threshold α and β are set as 0.2 and 0.3 respectively. The results show that our method UpTrust_d is the best method. PCC is better than Trust_d and it illustrates that the low similarity in trust statement can deteriorate the prediction results. From the figure, we can observe that with the lengthening of trust propagation, Trust_3 is better than Trust_2 and Trust_2 is better than Trust_1, we may conclude that trust propagation is helpful to improve recommendation performance. However, in the method UpTrust_d, UpTrust_1 is the best method and it better than UpTrust_2 and UpTrust_3. This is because although more trusted neighbors can be identified via trust propagation, it does not guarantee that the rating profile will cover a lot more items and hence increase accuracy greatly. Rather, it is possibly that adding few trusted neighbors may result in some noisy ratings, and hence harm the predictive performance.



(a) The results of MAE (b) The results of RMSE

Figure 1: The performances on FilmTrust

Fig. 2(a) and (b) are the results of those methods on Epinions in terms of MAE and RMSE respectively and the threshold α and β are set as 0.1 and 0.3. Likewise, UpTrust_d is the best approach of all and PCC is better than Trust_d. However, with the lengthening of trust propagation, Trust_1 is the best method in Trust_d while the UpTrust_3 is the greatest method in UpTrust_d, which is just opposite with the results on FilmTrust. This also explains that trust propagation can not guarantee better results will be received, although more trusted neighbors can be found, it is likely to add noisy information and decrease the accuracy of recommendation..

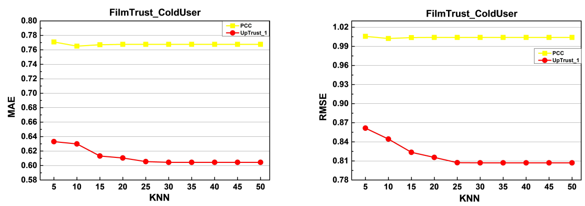


(a) The results of MAE (b) The results of RMSE

Figure 2: The performances on Epinions

In addition, to verify whether our method can solve cold start problem, we conduct experiments on FilmTrust in terms of MAE, RMSE and RC on cold users. Similarly, we compare UpTrust_1 with PCC and the performances are shown in Fig. 3. We can get that UpTrust_1 is much better than

PCC in terms of cold users which declares that our method can solve cold start problem to some extent. Table 1 is the performances of the two methods in terms of RC, which can further verify that our method can solve cold start problem effectively.



(a) The results of MAE (b) The results of RMSE

Figure 3: The performances on FilmTrust in terms of Cold Users

Table 1: The RC in the view of Cold Users on FilmTrust

Method	RC
PCC	43.57%
UpTrust_1	58.80%

5. CONCLUSIONS

This paper presents a reconstructing trust matrix method to improve the prediction accuracy and solve cold user problem of collaborative filtering recommender systems. Considering the similarity in trust matrix will affect recommendation results, we reconstruct traditional trust matrix. Besides, to recommend better for cold users and distinguish trusted neighbors in a shorter distance with those in a longer distance, weighted trust propagation is considered. The experimental results on two real data sets demonstrate the effectiveness of our methods in improving the prediction accuracy and solving cold user problem of recommender systems.

6. REFERENCES

[1] X. Zhou, Y. Xu, and Y. Li et al, The state-of-the-art in personalized recommender systems for social networking[J], Artif. Intell. Rev., 2012, 1-14.

[2] D. Jannach, M. Zanker, and M. Ge et al, Recommender systems in computer science and information systemsĺCa landscape of research[J], E-Commerce Web Technol, 2012,76-87.

[3] J. Audun and D.K. Walter Quattrochicchi, Taste and trust[J], Trust Management V, 2011, 312-322.

[4] M. Jamali, and M. Ester, Trustwalker: a random walk model for combining trustbased and item-based recommendation[C], in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2009, 397-406.

[5] P. Massa and P. Avesani, Trust-aware recommender systems[C], in: Proceedings of the 2007 ACM Conference on Recommender Systems, 2007, 17-24.

[6] S. Ray and A. Mahanti, Improving prediction accuracy in trust-aware recommender systems[C], in: Proceedings of the 43rd Hawaii International Conference on System Sciences, 2010, 1-9.

[7] S. Deng, L. Huang, and G. Xu, Social network-based service recommendation with trust enhancement[J], Expert Systems with Applications, 2014(41), 8075-8084.

[8] G. Guo, J. Zhang, and D. Thalmann, Merging trust in collaborative fltering to alleviate data sparsity and cold start[J], Knowledge-based Systems, 2014(57), 57-68.

[9] J. S. Breese, D. Heckerman, and C. Kadie et al, Empirical analysis of predictive algorithms for collaborative filtering[C], in Proceedings of the 14th Conference on University in Artificial Intelligence(UAI98), 1998, 43-52.

[10] P. Massa and P. Avesani, Trust-aware recommender systems[C], in: Proceedings of the 2007 ACM Conference on Recommender Systems, 2007, 17-24.

[11] I. Guy, I. Ronen, and E. Wilcox, Do you know recommending people to invite into your social network[C], in: Proceedings of the 14th International Conference on Intelligent User Interfaces, 2009, 77-86.

[12] C. Desrosiers and G. Karypis, A comprehensive survey of neighborhood-based recommendation methods[J], Recommender Systems Handbook, 2011, 107-144.