A reputation-enhanced model for trust-based collaborative filtering recommender system

1st Linshan Shen  
College of Computer Science and Technology  
Harbin Engineering University  
Harbin, China  
shenlinshan@hrbeu.edu.cn

2nd Shaobin Huang  
College of Computer Science and Technology  
Harbin Engineering University  
Harbin, China  
huangshaobin@hrbeu.edu.cn

3rd Xiangke Mao  
College of Computer Science and Technology  
Harbin Engineering University  
Harbin, China  
maotiamo@hrbeu.edu.cn

Abstract—As information overload becomes increasingly serious, people face the problems of discovering meaningful and relevant information, products and services. Recommender system can leverage these problems, which can help satisfy users need of personalization over the increasing amount of information on the Internet. Most used techniques are collaborative filtering, but which suffers from the data sparsity and cold start problems. Trust-based collaborative filtering systems address these issues. Generally, there exist some unreachable users by propagating in the trust network, which has a higher reputation. In this study, we will propose a reputation-enhanced model for trust-based collaborative filtering recommended system, which combines the initial implicit reputation with explicit users feedback information to enhance trust metrics, a reputation-enhanced algorithm and a recommendation algorithm based on the proposed model are proposed, which can further alleviate the sparsity of trust network and improve the accuracy of recommendation. Experiments will be made to demonstrate the effect of our proposed model by using film trust dataset.

Index Terms—collaborative filtering, recommender system, reputation-enhanced model, user feedback

I. INTRODUCTION

With the development of Internet, the problem of information overload is becoming more and more serious. Web users cannot find the useful or accurate information that they want to get in time due to this problem [1]. Moreover, some users surf the Internet without knowing what they want. Recommender systems are introduced to deal with these problems, which can mine users requirements and provide a personal recommendation result based on each users or his/her friends situations [2], and help users make right purchasing decisions in an information overload environment [3].

Collaborative filtering(CF) is the most popular technique of recommendation. Collaborative filtering systems collect the opinions from the users in the form of rating on items and recommend the items on the basis of these opinions. In [4], they developed the recommender system which identified similar users and suggested the items that they liked in the past. Many researchers have applied collaborative filtering on many items such as, papers [5], news [6], websites [7], movies [8], books [9], and so on.

Generally, a CF algorithm includes three steps. Firstly, it computes similarity value between each pair of users using Cosine similarity or the Pearson correlation coefficients based on the rating data that they have rates on items in common, where 1 means completely similar and -1 completely dissimilar. Secondly, when the most similar users of the current user are found, to some items not rated by the current user will be predicted based on the rating of their most similar users. At the third step, the items with highest predicted rating are suggested to the current user [10].

Even though collaborative filtering recommender systems can deal with large number of users and items, there exist some problems such as cold start, sparsity and malicious attack, which affect the quality of recommendation.

Trust-based collaborative filtering recommendation helps a lot in improving the quality of recommendation and addressing the problems of sparsity, cold-start and malicious attack. Since only the few users rated the items explicitly, user similarity was computed only on the data sparsity, and the nearest neighbors cannot be found out accurately. Some potential neighbors can be mined by using a trust network, to some extent, which relieves the problem of data sparsity. Many researchers proposed methods to solve the problem. In [11], they proposed a novel method to alleviate the problem of data sparsity by merging trust metrics in collaborative filtering. In this process, Ratings of users trusted neighbors were merged to complement and represent the preferences of the user and to find other users with similar preferences. Meanwhile, they used this novel method to alleviate the cold-start problem. Cold-start problem is also an unresolved problem that recommender systems suffer. Zou Haitao et al. proposed a cold-start tolerant algorithm by pre-processing propagation of trust network in [12]. Then they utilized the algorithm to TrustRank, a novel recommender system. Another aspect, some fake profiles can be filtered by using trust information, and the problem of malicious attack can be alleviated by combining trust network and collaborating filtering. A hybrid two-phases procedure for shilling attack detection was proposed in [13]. First, a multidimensional scaling approach was adopted to identify distinct behaviors that help to detect and secure the recommendation activities, subsequently, clustering-based methods were proposed to discriminate attack users. Two recommendation algorithms by combining CF techniques with
trust information were proposed in [14] and [15]. One of the common weaknesses of these algorithms are that they only consider the distance as the key factor to compute trust values between two users and neglect other impacts among those nodes. That is to say, trust metrics are local concept in one community.

Reputation is public trust derived from explicit or implicit information of users and items, which is the global concept. In [16], they gave this definition: Reputation can be considered as a collective measure oftrustworthiness (in the sense of reliability) based on the referrals or ratings from members in a community. Even though some previous papers [17]–[19] applied reputation in the trust-based and collaborative filtering recommender systems, algorithms or models, some unreachable users with high reputation to whom are neglected from the active user, which is the important factor to the quality and accuracy of recommendation. In this paper, we proposed reputation-enhanced model in trust-based collaborative filtering recommender system.

The contributions of this paper are illustrated as follows:

1) A novel model is proposed, which utilizes the users reputations dynamically updated to enhance the trust metrics by utilizing the serviced users feedback, comprehensively considering the situation of all users, from the local and global concepts to make recommendation, there will be good effect for collaborative filtering recommender systems.

2) In this paper, reputation-enhanced algorithm will be proposed to initialize and update the reputation of users dynamically, and recommendation algorithm will also be proposed to serve users.

3) Two datasets of experiments will be done to verify the effect of the proposed model and algorithms.

The remainder of this paper is organized as following. In section 2, the previous work on models about trust-based and collaborative filtering are reviewed, and some different approaches to modeling are analyzed. Following this, in section 3, the proposed model is described in detail. The reputation-enhanced algorithm and recommendation algorithm are proposed in section 4. Experiments will be demonstrated in section 5. The final section presents some concluding remarks and points to future work.

II. RELATED WORK

As above mentioned, collaborative filtering recommender systems existed the problems of data sparsity, cold-start and malicious attack [20], [21]. Researchers have started to study the trust-aware recommender systems to provide more personalized and accurate recommendation to users. Several trust-aware methods and models have been proposed to address the above problems in collaborative filtering recommender systems [22]–[25]. In [22], they proposed an effective trust-aware method. Two computational models of trust is computed and is incorporated into standard collaborative filtering frameworks. In [23], they proposed a novel personalized strategy to manage his own trust relationship with a mechanism, to make the recommendation more accurate by using part trust value as a complementary factor to user similarity. In [24], they proposed a trust-based architectural framework for collaborative filtering recommender system, in which trust metrics and rating matrix were taken as input and neighbors were generated using trust metrics and user similarity respectively and importance of trust over collaborative filtering was described. Hwang and Chen [25] developed an implicit trust filtering methods where the trust values were directly derived from the user rating data, then they proposed an improved technique to the Resnicks CF techniques [26] by integrating trust into CF recommendation process, making use of both trust propagation and local similarity neighborhoods.

Many researchers combined reputation with collaborative filtering to make recommendation. In a collaborative environment, reputation includes user reputation and item reputation, in this paper, we only consider the user reputation. Kevin et al. described a generic approach to modeling user and item reputation [27], WS model [28], PageRank [29] and HITS algorithm [30]. McNally et al. proposed the WS (Weighted Sum) model to compute the reputation of user by calculating the sum of its incoming edges, which simply calculated the reputation of a producer $n_i$ at some time $rp_i$ was shown as equation (1) [28]. Brin and Page proposed a well-known algorithm used by google to rank web search results [29]. Kleinberg proposed HITS algorithm to estimate user reputation [30]. HITS algorithm considered producers as authorities and consumers as hubs, which used repeated iterations to update the authority ($\text{author}(n_i)$) and hub values ($\text{hub}(n_i)$) for each user at each iteration.

$$rp_i = \sum_{e \in E_i} w_e$$

Where $E_i$ is the set of inlinks (from consumers) to (producer) $n_i$ and $w_e$ is the weight of inlink edge $e$. Some online communities offered the process and method of collecting user reputation. As the feedback-based mechanism was examined in [16] and [31], the method of utilizing feedback information to update user reputation got more attention. Hu Wei et al. proposed a contribution-based user reputation model in collaborative recommender system, a method of assigning reputations to nearest neighbors on basis of their contributions was proposed [31].

III. PROPOSED MODEL

In this section, we start from the research motivation, then illustrate the proposed model. Data repository and modules in this model are described in detail. Many previous researchers have proved that trust can be propagated from one user to other users at a defined maximum propagation distance in a trust network, but there exist the unreachable users that cannot construct trust relation with the active user by propagating, which have higher reputation. The reasons of unreachable include the distance of the active user and objective user is beyond the maximum propagation distance, or the active user to objective user cannot construct trust relation by propagating because no
pathway between them. User reputation evaluates the ability of a user who gives reliable, fair, and trustworthy ratings, so that we cannot neglect the reputation. User reputation plays an important role in making recommendations because reputation is the collective opinions of a whole community, trust metrics is a kind of local concept for users, it is important to construct an implicit relations with the active user by considering the collective and local concepts. When we can find out those users with higher reputation, then based on the original trust relations to build the new trust relation, and a new trust matrix is generated. This paper proposed a reputation-enhanced model for trust-based collaborative filtering recommender system (RTCFRS), the proposed model is shown as Fig. 1.

The model includes data repository, data processing and recommendation three parts. In the latter two parts, there are five modules, namely user similarity metrics module, trust metrics module, and reputation enhancement module, recommendation generation module, and user feedback module.

A. Data Repository

Data repository is a database, which includes rating matrix of user-to-item, trust metrics between users, and feedback information from the served users.

Explicit user preferences are included in the data repository denoted by a rating matrix $R$, which has $M$ rows and $N$ columns, $M$ is the number of users and $N$ is the number of items. Here, $R$ is an input to the proposed model. The value in the matrix $R$ indicates the user’s rating of the item. Trust metrics between users is also included in the data repository denoted by a trust matrix $T$, which has $M$ rows and $M$ columns, $M$ is the number of users. The trust matrix is also as an input into the proposed model, the value of the matrix $T$ denotes the extent of trust that a trustor trusts a trustee.

There is also feedback information included in data repository. When a user gains his/her recommendation service, generally, he/she is asked to rate the service, 1 denotes satisfaction, 0 denotes no rating, and -1 denotes dissatisfaction, which will be fed back to and saved in the proposed model. When another user requires his/her recommendation service, the feedback information is used to update the reputations of users combined with original reputation in the data repository.

B. User Similarity Metrics Module

Above mentioned, the key step of collaborative filtering process is to compute the similarity between each user based on the rating matrix, produce a $M \times M$ user similarity matrix $S$, in which ith row contain the similarity value of $i$th user against every other user, a rating $r_{i,j}$ is in the range of $[1,5]$, denotes the rating of user $i$ have rated item $j$. The most used techniques of computing the similarity values between two variables with linear relation is Pearson correlation coefficient [31], its value is between -1 and 1, -1 denotes completely negative correlation, 1 denotes completely positive correlation, and 0 denotes no correlation. There are user $u$ and user $v$, their similarity value is computed by Equation 2.

$$sim_{u,v} = \frac{\sum_{i=1}^{I_{uv}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i=1}^{I_{uv}} (r_{u,i} - \bar{r}_u)^2 \sqrt{\sum_{i=1}^{I_{uv}} (r_{v,i} - \bar{r}_v)^2}}}$$

(2)

Where $sim_{u,v} \in [-1,1]$ represents the similarity between $u$ and $v$(a larger value indicates a higher similarity); $|I_{uv}|$ represents the number of items that have been rated by user $u$ and $v$ at the same time; $r_{u,i}$ and $r_{v,i}$ represent the ratings of item $i$ by user $u$ and $v$ respectively; $\bar{r}_u$ and $\bar{r}_v$ represent the average ratings of user $u$ and $v$ on $I_{uv}$.

A similarity matrix $S$ can be formed by computing the similarity of every two users based on the (2).

C. Trust Metrics Module

In this module, trust matrix $T$ includes $M \times M$ trust values $t_{i,j}$, which represents the trust extent of a trustor $i$ trusts a trustee $j$. We also set the value of $t_{i,j}$ as the range of $[0,1]$. The trust matrix $T$ can be used to assign a predicted trust to every other user by propagating from a given source user. Assume the maximum propagation distance is $d$, a user at distance $n$ from source user will have a predicted trust value of $\frac{n}{(d-\frac{n+1}{4})}$. For example, in Fig. 2, nodes are users and edges are trust statements. Values on edge are one of the undefined and predictable trust statements.user B has issued trust statement in user C with value 0.8 and user C has issued trust statement in user E with value 0.9. We assume the maximum propagation distance is 4 in this trust network. The predicted trust value from user B to user E is $\frac{4-\frac{3+1}{4}}{4} = 0.75$.

Users that are unreachable within the maximum propagation distance have no predicted trust value [22]. At the same time, there is a linear decay in propagating trust, users closer in the trust network to the source user have higher predicted trust values.

D. Reputation Enhancement Module

Trust network only considers local trust metric, which considers the very personal and subjective views of the users, and finally predicts the trust values through trust propagation. Reputations of users or items belong to the global concept, which considers users or items from a community. Many researchers [16]–[19], [31] tried to combine trust metrics and reputation metrics to make recommendation, but they neglect isolated nodes, which are unreachable to the active nodes by propagating but have high reputations. In this study, how high will be consider? We consider those nodes that their reputation values are greater or equal to the average reputation value of all the nodes in the trust network. Then we construct an implicit trust relation with the weight value which is the average trust value of all from the active user to the trustee users. Generally, the trust matrix is sparse, although some predicted trust values are computed to update the trust matrix in the trust network.

Firstly, we need assign every user an initial value, since users reputation in ranking need not be considered, so we use McNally’s WS algorithm [28] to compute the initial reputation for every user, the equation (1) is shown in the section 2. We
can show an example to demonstrate our idea, a trust network is shown in Fig. 3, user 2 is the active user, it is not reachable to user 6, but the user 6 has higher reputation when we use Equation 1 to compute reputation value for every user, which is shown in the Table I. In order to improve the accuracy of recommendation to the active user 2, the opinion of user 6 cannot be neglected, we can construct an implicit trust relation from the user 2 to the user 6.

In the proposed model, when several users with higher reputation are selected, their reputations will be used to enhance the trust metric, the trust matrix $T$ will be updated. It is an important step to improve the quality of recommendation by considering local and global opinions for the active user in common.

**TABLE I**

<table>
<thead>
<tr>
<th>Node</th>
<th>n1</th>
<th>n2</th>
<th>n3</th>
<th>n4</th>
<th>n5</th>
<th>n6</th>
<th>n7</th>
<th>n8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reputation</td>
<td>1</td>
<td>1</td>
<td>2.1</td>
<td>2.4</td>
<td>0</td>
<td>2.3</td>
<td>0</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**E. Recommendation Generation Module**

The another important process of the proposed model is to find the nearest neighbors which have the similar interests with the active user, but collaborative filtering technology exits the problem of data sparsity [11], it is not enough to find the optimal neighbors to recommend only using the user similarity. Trust metrics address this problem, and which was enhanced
by the user reputation, from the global and local opinions to make recommendation. How to combine user similarity and enhanced trust metrics? In [23], they proposed a complex computing method, which combined the similarity matrix $S$ with enhanced trust matrix $ET$ to compute a weight and get the neighbor matrix $NE$ using the equation (3). Where $i$ and $j$ represent the $i^{th}$ and $j^{th}$ users, $K$ neighbors with highest weight are selected to make prediction for the active users to rate his/her unrated items from the neighbor set $NE$ of the active user.

$$\text{weight}_{i,j} = \begin{cases} 
\frac{2 \text{sim}_{i,j} \cdot t_{i,j} + t_{i,j}}{\text{sim}_{i,j} + t_{i,j}}, & \text{if } \text{sim}_{i,j} > 0 \text{ and } t_{i,j} > 0; \\
\text{t}_{i,j}, & \text{if } \text{sim}_{i,j} = 0 \text{ and } t_{i,j} > 0; \\
0, & \text{if } t_{i,j} = 0;
\end{cases}$$ (3)

The next step is to generate predictions of the active user to the unrated item referred to $K$ neighbors, the active user predicts the $i^{th}$ item using the equation (4), and gets the predicted rating value $p_{a,i}$, $r_{\text{a}}$ represents the average rated values on the items by the active user $a$, $r_{u,i}$ represents the average rating on the items by the active user $u$, $r_{a,u}$ represents the rating value on the item $i$ by the user $u$.

$$p_{a,i} = r_{\text{a}} + \frac{\sum_{u=1}^{K} (r_{u,i} - r_{\text{a}}) \text{weight}_{a,u}}{\sum_{u=1}^{K} \text{weight}_{a,u}}$$ (4)

Some unrated values are predicted by the $K$ nearest neighbors to the unrated items which the active user has not been rated before. So a list of ranked items from rated values is recommended to the active user.

F. User Feedback Module

After one transaction, the active user is required to give his/her rate, 1 denotes satisfactory score, -1 denotes unsatisfactory, if he/she does not give any his/her rate, we assign the feedback value is 0. The rate value will be fed back to recommender system and stored in data repository, and the reputations of every user will be updated by combining the original reputation values with the feedback rate when a new user comes to require for his/her service. Herein, we utilize the weight value computed by the equation (3) to cast the feedback value. Combined the original values $r_{p_i}$ of users reputation to update current reputation value $r_{p_i}$, the method is shown in equation (5).

$$r_{p_i} = r_{p_i} + \text{weight}_{a,i} FR_u$$ (5)

Where $r_{p_i}$ represents reputation of user $i$ before transaction, $FR_u$ represents the rate that user $u$ gives after one transaction.

When a new active user requires recommendation service, the updated reputation of every user will be used to enhance the trust metrics, if the trust value is greater than 1, the trust value is set as 1.

Next, we will introduce the reputation-enhanced and recommendation algorithms in a formal way.

IV. ALGORITHMS

In this section, the reputation-enhanced and recommendation algorithms will be proposed.

A. Reputation-enhanced Algorithm

In the reputation-enhanced algorithm, there are three steps to describe the reputation-enhanced process:

Algorithm 1 Reputation-enhanced Algorithm

1: when an active user requires the service, if he/she is the first customer of model, then initialize the reputations of users based matrix $T$ using equation (1); else update the reputations of users using equation (5).
2: check the unreachable users from the active user by propagating, if their reputations are greater or equal to the average reputation value of all users, then construct implicit trust relations from the active user to these users with weight which is the average trust weight value of the active users trust values to his/her trustees.
3: update the matrix $T$ according to the current explicit and implicit trust relation.

B. Recommendation Algorithm Based on the Proposed Model

In this algorithm, we have eight steps to describe the recommendation process of the proposed model.

Algorithm 2 Recommendation Algorithm

1: computing similarity matrix $S$ of users by using equation (2).
2: predicting trust value according to the maximum propagation distance, updating trust matrix $T$.
3: updating trust matrix $T$ using the reputation-enhanced algorithm.
4: computing the weights of every pair of users by using equation (3).
5: predicting the rating values that the active user has not rated by using equation (4).
6: recommending to the active user.
7: the served user gives his/her feedback value.
8: saving the feedback value to data repository.

V. EXPERIMENTS

In this section, we do two experiments to provide evidence supporting our claim that the reputation-enhanced algorithm enhances the performances of trust-based CF system.

The data set we used in the experiments is filmtrust dataset proposed by [32] in 2011, which included two files, ratings.txt and trust.txt. ratings.txt includes 35497 item ratings in the format: userid, moviedid, movieRating, which comes from 1508 users rating 2071 items with the sparseness of 98.86%. The trust.txt includes 1853 directed trust ratings in the format: trustorid, trusteeid, trustRating. By analysing the trust.txt, in this trust network, there are 732 users who attend to rate items, so we can know that the matrix $T$ is also very sparse.
In the first experiment, we consider feedback as a reputation enhancement to affect the trust relation when the maximum propagation distance is different. When the maximum propagation distance \( d \) is 2, 3, 4, the predicted trust values are computed by the \( \frac{(d-n+1)}{d} \) from the source user, when the trust matrix is updated by predicting trust values, then the numbers of directed trust rating are counted respectively. As expected, increasing the maximum propagation distance implies that the number of potential neighbors can be increased for every single user on the average.

In this experiment, we only consider the feedback value is 1, because there is no improvement to trust metrics when the feedback value is 0 or -1, which denotes no feedback or satisfaction from the served user. 100 users were selected randomly as the served users to require the services of the proposed model, when they gained their services, assume they gave the feedback value 1 to the model, the average numbers of the trust ratings were computed when the maximum propagation distance is 2, 3, 4. The comparison of original values and enhanced values of the above experiments is shown as Fig. 4. From the experiment, we can see that the higher the propagation distance, the less sparse the resulting predicted trust matrix, which directly affects the performance of recommendation.

Next, we demonstrate the recommendation performance of the proposed model, we introduced the Mean Absolute Error (MAE) [33] as one metric, since it is the most appropriate and useful for evaluating prediction accuracy of recommendation systems or recommendation model. This measure is similar with the clarification accuracy in the traditional learning issues [34]–[37].

In this experiment, we compare the process of the trust-based CF system with the recommendation algorithm of the proposed model in section 4. For the trust metrics module, the maximum propagation distance is set as 1, 2, 3, 4, we also select randomly a user to require services, assume he gives the feedback value 1 when he has finished one transaction, according to the reputation-enhanced algorithm, the trust matrix \( T \) is updated, and according to the recommendation algorithm of the proposed model, the neighbors are generated combined with the user similarity matrix. At the current state, we make the real ratings hide and try to predict them. The predicted rating is then compared with the real rating and the difference (in absolute value) is the prediction error. Averaging this error over every prediction gives the overall MAE, by this process, we compare the proposed model with the trust-based CF system. Through computing average MAE values, the comparison of two models is shown as Fig. 5. From the graphical representation in Fig. 5, the average values of MAE are smaller than that of the trust-based collaborative filtering system, the enhanced model performs better.

VI. CONCLUSION

In order to improve the accuracy of recommendation, in this study, we proposed a reputation-enhanced model for trust-based collaborative filtering recommender system. The proposed model effectively combines the initial implicit reputation with explicit users feedback information to enhance trust metrics, and makes the predicted rating value for the active user more accurate. Reputation-enhanced and recommendation algorithms for the proposed model were proposed, from the global and local concepts, we utilized users’ reputation to enhance the trust metrics by using a novel combination method. In our experiments, with a public dataset, we consider the enhanced results at deferent maximum propagation distance, look from the results of experiments, the proposed model performs better.

In the future work, we will develop the recommender system based on our model, and utilize it to some other areas, select some public data set to verify further the proposed model. At
REFERENCES


