PAPER

Using Trust of Social Ties for Recommendation

Liang CHEN[†a)], ***Student Member***, **Chengcheng SHAO**[†], **Peidong ZHU**[†], ***Nonmembers***, ***and*** **Haoyang ZHU**[†], ***Student Member***

SUMMARY Nowadays, with the development of online social networks (OSN), a mass of online social information has been generated in OSN, which has triggered research on social recommendation. Collaborative filtering, as one of the most popular techniques in social recommendation, faces several challenges, such as data sparsity, cold-start users and prediction quality. The motivation of our work is to deal with the above challenges by effectively combining collaborative filtering technology with social information. The trust relationship has been identified as a useful means of using social information to improve the quality of recommendation. In this paper, we propose a trust-based recommendation approach which uses GlobalTrust (GT) to represent the trust value among users as neighboring nodes. A matrix factorization based on singular value decomposition is used to get a trust network built on the GT value. The recommendation results are obtained through a modified random walk algorithm called GlobalTrustWalker. Through experiments on a real-world sparser dataset, we demonstrate that the proposed approach can better utilize users' social trust information and improve the recommendation accuracy on cold-start users.

key words: *social network, trust-based, collaborative filtering, random walk*

1. Introduction

Since the commercialization of the Internet in the 1990s, the global Internet has become the world's critical information infrastructure for global economic development and social progress. The Internet has quickly penetrated all areas of economic and social activities, and it promotes the global informatization process. Meanwhile, the development of information technology brings forth the "information overload" problem [1].

Human beings are integrating all the kinds of media content ever produced into the sea of bits through information technology, while continuing to produce content at an unprecedented rate. Moreover, information production and diffusion have mainly followed the principle of "filter-then-publish" [2] in the past. Mass media, churches, schools, and other authorities play a role as gatekeepers or filters. Today, the increasingly common phenomenon of content generation is "publish-then-filter".

Actually, people regard "eliminate uncertainty" [3] as a definition of information, but the multitude of information seems to have increased the uncertainty of human society. Therefore, recommendation methods are gradually attracting people's attention as an information filtering technology. Collaborative filtering (CF) is one of the most popular techniques in recommendation approaches. It can predict user interests directly by uncovering complex and unexpected patterns from a user's feedback data such as user-items ratings without any domain knowledge [4]. However, traditional collaborative filtering faces several challenges, such as data sparsity, cold-start users and prediction quality.

Given the huge amount of items, the user-items rating data become extremely large and users only rate a dozen or fewer items. As a result, the sparsity of available user feedback data is often greater than 99% [5]. Due to the data sparsity, collaborative filtering faces many difficulties in using the cosine similarity method or Pearson Correlation Coefficient (PCC) [24] similarity method to identify similar users or similar items, naturally causing the drop of recommendation performance.

With the rapid development of Web 2.0 technology in recent years, OSN has become an important platform for people to communicate and share information. In the Web2.0 era, network content is mainly generated by the users, and each user can generate his or her own content. People in the OSNs are not only consumers but also producers of information. All kinds of OSN activities produce a multitude of social interaction data. According to Statistic Brain [6], every 20 minutes the number of links shared on Facebook is more than 1 million, and more than 3 million messages are sent. Similar statistics are shown for Twitter.

Many empirical studies [7]–[11] found that utilizing OSN information can improve the quality of recommendation. Jamali and Ester [12] combined the trust propagation mechanism in the social network with the matrix decomposition model to improve the recommendation quality. The trust relationship from social information has been identified as a useful means of using social information to improve the quality of recommendation. Walter et al. [13] proposed a trust-based model in which agents use their social network to disseminate information, while using the trust relationship to filter information. However, these methods used only simple binary trust relationship information, whereby the user either trusts a user or not. Such trust information has the characteristics of coarse granularity, which is unable to fully tap the potential of the trust relationship. Moreover, due to the data sparsity, some users have little information either on trust or on ratings, which greatly reduces the

Manuscript received May 25, 2015.
Manuscript revised September 25, 2015.
Manuscript publicized October 30, 2015.

[†]The authors are with College of Computer, National University of Defense Technology, Changsha, China.
a) E-mail: thomaschen160@gmail.com

DOI: 10.1587/transinf.2015EDP7199

recommendation performance.

To address the above challenges, in this paper we propose a novel trust-based recommendation approach which uses the trust information in social networks. In this approach, we propose a term called GlobalTrust (GT). It represents the concept that the trust value between social network users is affected by the neighboring nodes. The GT value between users reflects the users' trust from the fine-grained perspective, which definitely improves the accuracy of the recommendation. A matrix factorization based on singular value decomposition is used to get a trust network built on the GT value. The recommendation results are obtained through a modified random walk algorithm called GlobalTrustWalker. We conduct several experiments on the Epinions [14] dataset to evaluate the accuracy and efficiency of the proposed approach.

The rest of the paper is organized as follows. Related works are discussed in Sect. 2. Section 3 defines the problem. We discuss the details of our proposed approach in Sect. 4. The experimental results and comparison with existing methods are discussed in Sect. 5. Finally, we conclude the paper in Sect. 6 and introduce some directions for future research.

2. Related Work

Traditional recommendation approaches use user feedback data (such as rating, purchase and click, etc.) as a basis to recommend content [4]. These methods can be generally classified into two categories: content-based methods and collaborative filtering-based methods.

The basic idea of content-based methods is to recommend items that are similar to the user's preferred ones. Most of the content-based approaches focus on items which contain textual information such as news, books and other documents [15], [16]. Mooney et al. [17] developed a book recommending system that utilizes semi-structured information about items gathered from the web using simple information extraction techniques. Balabanovic and Shoham [18] proposed a recommendation approach which combines content-based filtering and collaborative methods.

Collaborative filtering-based recommendation methods calculate the similarity of users or items to predict the interests of users. Additional information such as the user-item ratings and past behaviors of users are used to enhance the prediction accuracy. Many researchers use CF-based methods to make some good jobs. Koren et al. [19] built a combined CF-based model that analyzes similarities between products and users. Herlocker et al. [20] proposed an algorithmic framework for performing collaborative filtering and new algorithmic elements that increase the accuracy of collaborative prediction algorithms. Yildirim and Krishnamoorthy [21] proposed an item-oriented algorithm that infers transition probabilities between items based on their similarities.

The rise of OSN brings the richness of online social activities. And these online activities generate an ocean of social information, such as the following relations in Weibo, friendship information in Facebook and trust data in Epinions. Therefore, many recommendation methods have begun to focus on how to use social information to improve the quality of recommendation in recent years. Some representative work such as SoRec [22] and SocialMF [23] combined the matrix factorization method with social information to improve recommendation accuracy. TidalTrust [24] and MoleTrust [25] incorporated users' trust information into social network traversal-based approaches to get positive recommendation performances. Guo et al. [26] merged a user's trusted neighbors' ratings to represent the preferences of the users. Moradi and Ahmadian [27] proposed a reliability-based recommendation method to improve trust-aware recommender systems. Jin et al. [28] designed a trust-based top-k recommender system, which computed the trust values between users from users' interest similarity. Fazeli et al. [29] proposed a trust-based recommender algorithm with social data obtained from monitoring the Open Discovery Space (ODS), which is a platform for sharing eLearning resources online.

Different from the above work, we put the users' latent trust value as an important factor to enhance the trust-based recommender system. In order to explore the users' latent trust value, we think that calculation of the trust value between users should consider information on indirect trust users. Compared with the existing related work, the main contributions of our work are summarized as follows:

(a) We propose the concept of GlobalTrust, which denotes an improvement on simple binary 0/1 trust networks. It transforms the coarse-grained trust relationship to a fine-grained one.
(b) We design a random walk algorithm based on the proposed GlobalTrust. Compared with other random walk algorithms, our algorithm selects the walking nodes based on GT values. So the algorithm can accelerate the convergence rate and improve the accuracy of recommendation.
(c) We conduct several experiments on a real-world dataset to evaluate the accuracy of the proposed approaches. The real-world dataset, which is from Epinions.com [14], has the features of data sparsity and cold-start users. The experimental results show that the proposed approach can provide high recommendation accuracy and improve the recommendation quality on cold-start users.

3. Problem Definition

Calculations of GT value are derived from a 0/1 trust network. We consider that the trust value between each pair of users cannot simply be marked by 0/1, which has coarse granularity. Each pair of users' trust value in the GT network is calculated by the degree of nodes. Therefore, the calculation integrates the network structural information into a trust value and predicts the users' ratings on items by

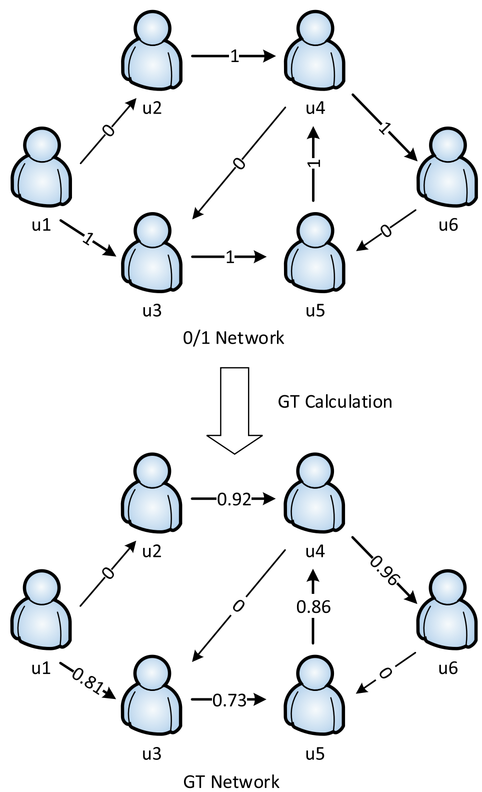

Fig. 1 Schematic diagram of the GT value calculation.

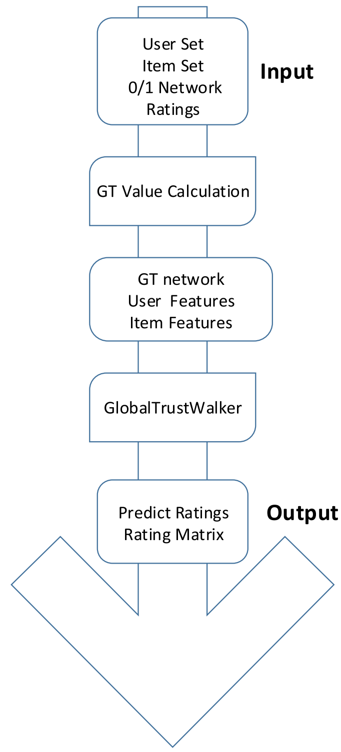

Fig. 2 Summary process of the approach.

analyzing the GT network and user-rating records. Figure 1 shows a schematic diagram of the GT value calculation.

At the beginning of the approach, there is a set of users $U = \{u_1, u_2, \ldots, u_m\}$ and a set of items $I = \{i_1, i_2, \ldots, i_n\}$. Matrix $R = [R_{u,i}]_{m\times n}$ records the users' real ratings on items. The entry $r_{u,i}$ in matrix R denotes the rating expressed by user u on item i. $r_{u,i}$ can be any real number. In this paper, we set the ratings range as $[0, 5]$. In the GT network, each user u has a set G_u of direct trust friends, and $t_{u,v}$ denotes the value of GT between user u and v in the range $[-1, 1]$. A negative value means distrust and a positive value means trust. Matrix $T = [T_{u,v}]_{m\times m}$ is composed of entry $t_{u,v}$. A non-zero entry $t_{u,v}$ in matrix T denotes that there exists a trust relation between the corresponding users u and v. Obviously, the matrix T is non-symmetric.

Therefore, the task of a recommendation approach is as follows: Given a user $u_0 \in U$ and an item $i \in I$ for which $r_{u_0,i}$ is unknown, predict the rating for u_0 on item i using the matrix R and T. $\hat{r}_{u_0,i}$ denotes the predict rating and $\hat{R}$ is the predict matrix. In our approach, u_0 is called the source user and i is the target item.

4. Trust-Based Recommendation Approach

The general idea of our approach is to combine the GT value into a user-item rating matrix. The GT network can accurately portray the trust relations between users. In this section, we describe our trust-based recommendation approach in detail. First of all, we introduce the source and basic concept of the GT value. Then, the details of the GlobalTrustWalker algorithm will be presented. Finally, the calculation method of the predictive value is explained. A summary process of our approach is shown in Fig. 2.

4.1 GlobalTrust Value

As we mentioned before, many of the trust relationships between social network users are simply 0/1 trust and some networks do not even have trust data. A fine-grained trust value can improve the accuracy of the recommendation. Therefore, we propose the concept of GlobalTrust, which combines the structural information of network nodes into user similarity.

Definition 1 (Global Trust). Given users u and v, the GT value from u to v is as follows:

$$gt(u, v) = simU(u, v) \cdot t^*_{u,v}$$

where $simU(u, v)$ is the similarity of users u and v, and $t^*_{u,v}$ represents the network structure of 0/1 network information of the corresponding users u and v. The calculation of $t^*_{u,v}$ is as follows:

$$t^*_{u,v} = t_{u,v}\sqrt{\frac{d^-_v}{d^+_u + d^-_v}} \tag{1}$$

where $t_{u,v}$ represents users' trust mark information in the 0/1 network. The indegree of node v is denoted as d^-_v and the outdegree of node u is denoted as d^+_u. The part of the root represents the users' network structure information. We can note that the GT value signifies the unidirectional relationship between users u and v. So the values of $gt(u, v)$ and $gt(v, u)$ are not equal in general circumstances. The Epinions dataset contains 0/1 trust information. In this dataset, the indegree and outdegree of a node can be measured accurately. So it is very suitable for testing our approach. By computing the GT value between all connected users in a social network, we can obtain a GT-based network, where the weight of each edge is the value of GT, as shown in Fig. 1.

There exist a number of methods to calculate the similarity of social network users. Commonly used methods include: Pearson correlation similarity, Euclidean distance similarity, Tanimoto coefficient similarity and cosine similarity [30]. These methods are based on the users' characteristic vectors. In fact, the basic idea of these methods is to calculate the distance between two vectors. The closer the distance, the greater the similarity. In the recommendation scenario, we can use the user preferences for items as a vector to compute the similarity between users, or use the preferences of all users for an item as a vector to compute the similarity between items with the two-dimensional user-rating matrix.

In this paper, we use the cosine similarity method to calculate the similarity between users. The similarity calculation between users u and v is as follows:

$$simU(u, v) = \cos(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}| \cdot |\mathbf{v}|} \tag{2}$$

where $\mathbf{u}$ and $\mathbf{v}$ represent the characteristic vectors of users u and v.

4.2 GlobalTrustWalker Algorithm

The main challenge of our trust-based recommendation approach is to determine the random walk rule that determines how to select the next node and how to return the predict rating on the target item. We consider that an item which is rated by the user with a higher GT value is more recommendable than one with a lower GT value. We propose the *GlobalTrustWalker* algorithm based on the above consideration.

GlobalTrustWalker is a random walk algorithm which considers not only the item similarity, but also the GT value. Generally, the proposed algorithm consists of two major components. One is the random walk in the GT network and the other is the calculation of the prediction rating. The random walk which performs the search in the GT network always selects the node with the highest GT value. The calculation of the predict rating considers the same item or similar items to obtain the results. Definitions of the variables are shown in Table 1.

GlobalTrustWalker recommends a rating for a source user u_0 on target item i through several iterations. For each iteration, the algorithm performs random walks on the GT network. The random walk starts at source user u_0. Of course, each u_0 is not the same user. The algorithm will sequentially traverse all the users in the GT network. In the kth step, the walk reaches a certain node u. If the user u has rated the target item i, then the rating expressed by user u on item i is returned as the result for this walk. If user u does not rate item i, the algorithm has the following options:

Table 1 Variables for GlobalTrustWalker.

Variables	Description
TU_u	The set of users that have a GT value with user u
$r_{u,i}$	The real rating expressed by user u on item i
$\hat{r}_{u,i}$	The predicted rating of u on item i
$\psi_{u,i,k}$	The probability of the random walk stopping at user u in the kth step
$M_u(v)$	The probability of selecting a user v from TU_u
$simI(I_i, I_j)$	The similarity of items I_i and I_j
$simU(u, v)$	The similarity of users u and v
RI_u	The set of items that user u has rated
$F_u(I_i)$	The probability of selecting item I_i from RI_u

(a) The random walk stays on the current node u and no longer continues. The probability of staying on node u is $\psi_{u,i,k}$. Then, we select the item j from RI_u that is most similar to the target item i. The rating $r_{u,j}$ expressed by user u on item j is returned as the result for this walk.

(b) The random walk continues with a probability of $1 - \psi_{u,i,k}$. Select a node v from among node u's direct trust neighbors. The node v has the maximum $gt(u, v)$ amongst the neighbors.

The algorithm has a probability $\psi_{u,i,k}$ of staying at user u in the kth step of the random walk. The probability is affected by the similarity between the target item i and the items that user u has rated. The greater the similarity between the rated item and the target item, the greater the chance of stopping at the current node.

Several methods can be used to calculate the similarity of items, but there is no attribute information of an item in the dataset. The only information about items is the users' ratings. Hence, the item-item similarity can be calculated by the users' ratings. We propose the cosine similarity method for calculating similarities between items. Values of cosine similarity are in the range $[-1, 1]$. The resulting similarity ranges from -1 meaning exactly opposite, to 1 meaning exactly the same, with 0 usually indicating independence, and in-between values indicating intermediate similarity or dissimilarity. We only consider items with positive similarity, because items with negative values are not useful in our algorithm. As mentioned in Sect. 4.1, the singular value decomposition of the user-rating matrix can be used to obtain the item characteristic vector. Each row of matrix Q can represent an item. So calculation of the similarity between given item I_i and I_j is as follows:

$$simI(I_i, I_j) = \cos(\mathbf{I_i}, \mathbf{I_j}) = \frac{\mathbf{I_i} \cdot \mathbf{I_j}}{|\mathbf{I_i}| \cdot |\mathbf{I_j}|} \tag{3}$$

Since we do not consider items with negative similarity, the values of $simI(I_i, I_j)$ are real numbers in $[0, 1]$. So we associate the maximum item similarity with the probability $\psi_{u,i,k}$. Furthermore, the farther the current node is away from the source user, the lower the accuracy of the recommendation. Therefore we cannot make the walk go too far and $\psi_{u,i,k}$ should increase when the walk steps increase. We introduce the step value k into the probability calculation through a sigmoid function. When the value of k increases, the sigmoid function makes the value close to 1. Due to the features of the GT network, the average shortest path of the network is short. Therefore, in order to accelerate the convergence of the algorithm, we use the sigmoid function for probability calculation. The training parameter of the sigmoid function is k/L. The calculation of $\psi_{u,i,k}$ is as follows:

$$\psi_{u,i,k} = \max_{I_j \in RI_u} simI(I_i, I_j) \times \frac{1}{1 + e^{-\frac{k}{L}}} \tag{4}$$

When the random walk stays on a user u with probability $\psi_{u,i,k}$, the algorithm selects one item from RI_u. The rating of the selected item is returned as the result of the walk. For each item $I_j \in RI_u$, we have a certain probability of selecting I_j from RI_u to obtain u's rating for the target item. The probability denoted by $F_u(I_i)$ is as follows:

$$F_u(I_i) = \frac{simI(I_i, I_j)}{\sum_{I_j \in RI_u} simI(I_i, I_j)} \tag{5}$$

If the random walk continues with the probability $1 - \psi_{u,i,k}$, the algorithm will choose the current user u's direct GT neighbors as the next step of the random walk. Many existing methods just randomly select a node for the next step, but we believe that the importance of neighbors for the current user is not the same. So we measure the neighbors' weights with a GT value. The higher the GT value, the greater the importance of the neighbors. The algorithm selects the neighbor with the largest GT value for the next step of the random walk. $M_u(v)$ denotes the probability of selecting a user v from TU_u with the largest GT value and is as follows:

$$M_u(v) = \max_{v \in TU_u} gt(u, v) \tag{6}$$

According to the above description of our algorithm, we can conclude that the random walk has three kinds of possibilities to stop:

(a) The current user has rated the target item. The rating expressed by the current user on the target item is returned as the result for this walk.
(b) The random walk reaches a certain user and stays on the node with a probability of $\psi_{u,i,k}$. The user does not rate the target item. The algorithm selects one of the items which is similar to the target item rated by the user. The rating expressed by the current user on the similar item is returned as the result for this walk.
(c) The random walk has a certain possibility of running forever. Therefore, we associate the maximum steps of the random walk with the average shortest path length of the GT network.

The pseudo-code of the *GlobalTrustWalker* is as follows:

Algorithm 1. GlobalTrustWalker

Input: U represents the set of users $\{u_1, u_2, \dots, u_m\}$; I represents the set of items $\{i_1, i_2, \dots, i_n\}$; U is the rating matrix; u_0 denotes the source user; i denotes the target item.
Output: Predicted ratings are represented by r
steps $\leftarrow$ 0
targetUser $\leftarrow u_0$
max-depth $\leftarrow$ L //set the average shortest path as the max step
$r \leftarrow 0$
for steps $\leftarrow$ 0 **to** max-depth **do**
  $u \leftarrow$ selectUser(TU_u) // select u from TU_u as the target of the next step according to the probability $M_u(v)$
  if $r_{u,i} \neq 0$ **then** // user u has rated on target item i
    $r \leftarrow r_{u,i}$
    return r
  end
  elseif $\psi_{u,i,k} >$ random(0, 1) or steps == max-depth **then**
    $I_j \leftarrow$ selectItem(RI_u) //select I_j from RI_u according to the probability $F_u(I_j)$
    $r \leftarrow r_{u,I_j}$
    return r
  end
  steps $\leftarrow$ steps +1
end
return r

4.3 Recommendation

The recommended result is returned in two situations: a rating by a trusted neighbor on the target item and a rating by a trusted neighbor on an item that is similar to the target item. So the predicted rating on the target item should be accumulated by the result of multiple random walks. The predicted rating for source user u_0 on target item i is $\hat{r}_{u_0,i}$.

$$\hat{r}_{u_0,i} = \frac{1}{n} \sum_{j=1}^{n} r_j \tag{7}$$

In the above equation, r_j denotes the result of each random walk, n is the number of random walks. In order to obtain a reliable prediction, the algorithm needs to perform a sufficient number of random walks to make the predictions more accurate. We set the threshold value of the variance of the prediction result to control the condition for the termination of the algorithm.

The variance of all the prediction ratings is denoted by σ^2. The calculation of σ^2 is as follows:

$$\sigma^2 = \frac{1}{n-1} \sum_{j=1}^{n} (r_j - \bar{r})^2 \tag{8}$$

where r_j denotes the prediction rating of the jth random walk. $\bar{r}$ is the average result of all random walks. n denotes the total number of random walks. σ_j^2 is defined as the variance of the last j random walks. The termination of the algorithm is $|\sigma_{j+1}^2 - \sigma_j^2| \leq \varepsilon$. The value of ε can be set to control the termination of the algorithm.

5. Evaluation and Analysis

In this section, we conduct a series of experiments using the Epinions dataset. In the dataset, the users indicate which user is trusted. Then, we will describe our evaluation metrics and analyze the results obtained during the experiments. We make a comparison between our approach and other related recommendation methods at the end of this section.

5.1 Data Analysis

The Epinions.com is a consumer review site which provides comparative information on a variety of commodities. Members of the site can decide whether to trust each other. All the trust relationships interact and form the web of trust which is then combined with review ratings to determine which reviews are shown to the user. The trust information on the site builds a who-trusts-whom online social network. Several trust-based recommendation approaches [31]–[33] use the dataset from Epinions to verify their work. Therefore, the Epinions dataset is suitable for the evaluation of our approach.

Massa and Avesani [34] use the simple binary trust relations in the Epinions dataset to increase the coverage of recommendation. The dataset in this work contains binary 0/1 trust relation information. So, our proposed approach can easily build the GT network with the dataset. This dataset was collected by Paolo Massa [35] in a 5-week trawl from the Epinions.com website. It contains 49,290 users who rated a total of 139,738 different items at least once. The total number of issued trust statements is 487,181 and the total number of ratings is 664,824. Users and items are represented by anonymized numeric identifiers. Each user trusts 9.9 direct neighbors and rates 13.4 items on average. The sparsity of the user-item rating matrix is greater than 99%. From the above statistics, we can observe that the user-item rating matrix of this dataset is a large and sparse matrix.

5.2 Evaluation Metrics

To evaluate our approach, we adopt several widely used metrics in our experiments. The purpose is to quantify the performance of the approach. The first metric is the Mean Absolute Error (MAE). The MAE is a quantity used to measure how close predictions are to the actual ratings. The MAE is given by the following:

$$MAE = \frac{\sum_{(u,i)} |r_{u,i} - \hat{r}_{u,i}|}{N} \tag{9}$$

In the above equation, $r_{u,i}$ is the actual rating expressed by user u on item i. $\hat{r}_{u,i}$ denotes the corresponding prediction. (u, i) denotes the <user, item> pair. N is the number of all tested ratings.

Another important metric to measure the accuracy of recommendation results is the Root Mean Square Error (RMSE). The RMSE represents the standard deviation of the differences between predicted values and actual values. The RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{(u,i)}(r_{u,i} - \hat{r}_{u,i})^2}{N}} \tag{10}$$

These two metrics are both widely used in evaluating the prediction accuracy of the recommendation approach. The smaller the value of MAE/RMSE, the more precise the prediction.

The Epinions dataset we used is extremely sparse. Some recommendation methods have low coverage on sparse data. We need to test the coverage of our approach to compare it with other methods. So we use the coverage metric to measure the percentage of <user, item> pairs. The coverage can help to obtain recommendations for cold-start users. For a <user, item> pair, if a recommendation method does not obtain the predicted rating of the corresponding pair, then this method cannot cover this pair. The calculation of the coverage metric is as follows:

$$Coverage = \frac{\hat{R}}{N} \tag{11}$$

In the above equation, $\hat{R}$ denotes the number predicted, N is the total number of ratings. We use the F-Measure metric to get a harmonic mean between RMSE and coverage. In order to combine the RMSE and coverage into the F-Measure formula, we have to convert the RMSE into a precision metric in the range of $[0, 1]$. Since the user-item ratings of the dataset are in the range $[1, 5]$, we used a formula deformation on the RMSE to calculate the precision metric. So the precision is denoted as follows:

$$Precision = 1 - \frac{RMSE}{4} \tag{12}$$

According to the statistical analysis, the F-Measure is denoted as follows:

$$F - Measure = \frac{2 \times Precision \times Coverage}{Precision + Coverage} \tag{13}$$

5.3 Experimental Results

In this subsection, we present the experimental results and comparison results between the different methods. We first introduce the different methods for comparison, then present the results on cold-start users and all users.

In our experiments, we compare our methods with different methods. A description of the different methods follows:

(a) Item-based CF: The item-based collaborative filtering method in our comparison uses the Pearson Correlation as the item similarity metric.
(b) User-based CF: The user-based collaborative filtering method in our comparison uses the cosine similarity method to calculate the user and item similarity.
(c) MoleTrust: This is the approach used in Massa and Avesani's work [25]. MoleTrust is a social network traversal based approaches which traverse the user's neighborhood in maximum-depth search and query the rating of the target item from the users within maximum-depth. Maximum-depth in MoleTrust is independent of any user and item. Backward exploration

Table 2 Comparison results for cold-start users.

Algorithms	MAE	RMSE	Coverage	Precision	F-measure
Item-based CF	1.1895	1.5006	22.59%	0.6248	0.3318
User-based CF	1.1491	1.4498	18.48%	0.6375	0.2865
MoleTrust	1.0811	1.3639	56.90%	0.659	0.6107
TidalTrust	0.9702	1.2262	59.31%	0.6934	0.6393
TrustWalker	0.9453	1.1932	72.59%	0.7017	0.7135
GlobalTrustWalker	0.8925	1.1552	77.75%	0.7112	0.7428

is performed to compute the trust value between indirectly connected users. The trust value between users is the aggregation of trust values from one user to users that directly trust another user.

(d) TidalTrust: This is a trust-based method proposed by Golbeck [24]. TidalTrust works in a breadth-first way which searches for raters that directly connect to the source-user. The trust value of source-user u to target-user v is aggregated from u's direct neighbors' to v. TidalTrust searches for raters that the source-user knows directly. If there is no direct connection, it searches two hops to find paths that connect the source-user to raters. The trust value is calculated for each rater on the path.

(e) TrustWalker: This method is a random walk approach based on trust and item similarity [32]. TrustWalker employs random walks in social networks based on users' binary 0/1 trust value. It visits a user's direct and indirect friends during the random walk. It improves the prediction accuracy and coverage by considering ratings for similar items. TrustWalker considers ratings from similar users and ratings of similar items. It is a combination of the item similarity based approach and trust-based approach. Different from TrustWalker, our approach uses the fine-grained trust value to form the trust network and performs the random walk based on the fine-grained trust value.

Table 2 shows the MAE, RMSE, coverage and F-measure for all the compared methods on cold-start users. We can observe from Table 2, that *GlobalTrustWalker* has lower error than all the other methods on cold-start users. Two classic recommendation algorithms, item-based CF and user-based CF, have the worst performance on the experimental results. The main reason for these results is that the cold-start users expressed few ratings on items, which means the two algorithms cannot accurately calculate the similarity on users and items.

Figure 3 shows the chart comparing the three metrics of coverage, precision and F-Measure on cold-start users. Since the introduction of the trust information between users, TidalTrust and MoleTrust perform better than the previous two algorithms, mainly because they can make recommendations for users by utilizing trust relations. However, there was no obvious accuracy improvement for the two algorithms. The TrustWalker algorithm shows a effective improvement on coverage and F-measure. This is because the algorithm combines the user trust relationships and item similarity to improve the recommendation

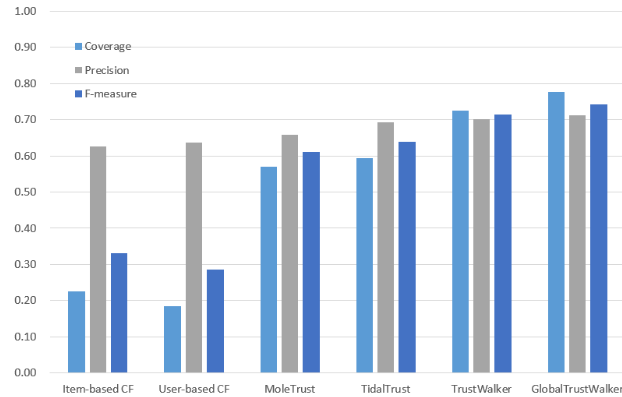

Fig. 3 Comparison of F-measure together with coverage and precision for cold-start users.

Table 3 Comparison results for all users.

Algorithms	MAE	RMSE	Coverage	Precision	F-measure
Item-based CF	1.1042	1.3934	68.39%	0.65165	0.6673
User-based CF	0.9556	1.2021	71.28%	0.699475	0.706
MoleTrust	0.9368	1.1859	87.51%	0.703525	0.7799
TidalTrust	0.9252	1.1675	85.16%	0.708125	0.7732
TrustWalker	0.8941	1.1282	96.27%	0.71795	0.8225
GlobalTrustWalker	0.8473	1.0692	98.69%	0.7327	0.841

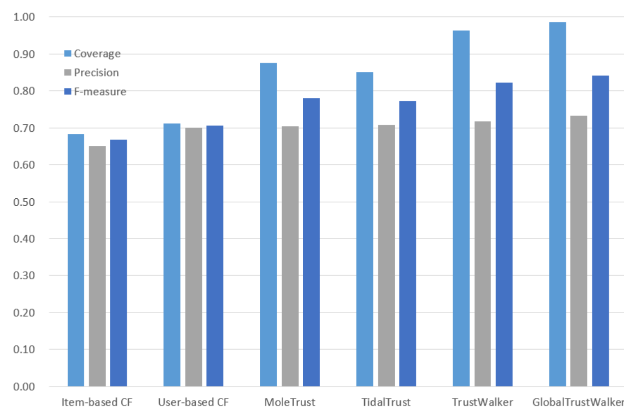

Fig. 4 Comparison of the three metrics for all users.

performance. Compared to TrustWalker, our algorithm selects the next walking nodes based on the GT value, rather than a random selection. Therefore our algorithm has greater recommendation accuracy than TrustWalker.

Table 3 shows the comparison results for all users. Figure 4 shows the chart comparing the three metrics of coverage, precision and F-Measure on all users. From the experimental results, we can observe that the recommendation accuracy improved on all the algorithms, including the item-based CF and user-based CF. This improvement is mainly because there are sufficient user-item ratings to calculate the user similarity and item similarity. But TidalTrust and UserTrust have no obvious advantage over the item-based and user-based algorithms. The two trust-based algorithms use trust relations to improve the coverage, but do not improve accuracy, which is because these methods do not mine

Table 4 Comparison results for cold-start users.

Algorithms	n	mean	SD	*p*-value	*df*
TrustWalker	30	0.7140	0.0056	-	-
GTWalker	30	0.7414	0.0035	-	-
Total	60	0.7277	0.0146	0.0254	58

Table 5 Comparison results for all users.

Algorithms	n	mean	SD	*p*-value	*df*
TrustWalker	30	0.8219	0.0039	-	-
GTWalker	30	0.8416	0.0045	-	-
Total	60	0.8317	0.0110	0.0373	58

the latent features of users and items. By contrast, our algorithm considers the user similarity and item similarity, and uses random walk based on GT value to select the most trusted user to improve the recommendation accuracy.

5.4 Significance Test

In this subsection, we present the significance test results on F-measure between our approach and TrustWalker. We use the *t*-test for the significance test. The *t*-test assesses whether the means of two groups are statistically different from each other. We test the results on cold-start users and all users respectively. We have carried out 30 times of experiments by using two approaches on two type of users.

The hypothesis of significance test is that there is statistically difference between the two approaches. The significance level of the test refer to the likelihood that the experimental results is support of the hypothesis. We set the significance level to 0.05. Table 4 and Table 5 shows the comparison of the two approaches on two types of users.

The above tables present the results of significance: times (n), mean value, standard deviation (SD), significance (*p*-value), degree freedom (*df*). Therefore, we accept the hypothesis that there is statistically difference from the two approaches.

6. Conclusions

Recommendation has become an important means of information filtering. Some traditional recommendation approaches, such as item-based CF and user-based CF methods, have been successfully applied in many applications. However, these traditional recommendation methods face many challenges, such as data sparsity, cold-start users and recommendation quality. Some recommendation methods use user trust relations to replace user similarity, which improves the recommendation quality and coverage on cold-start users. We proposed the concept of GlobalTrust to address these problems. GlobalTrust describes the trust relationship between users from the fine-grained perspective. We propose a random walk algorithm based on GlobalTrust called *GlobalTrustWalker*. Compared with other existing random walk methods, the proposed algorithm selects the target node for each step based on the GT value rather than randomly selecting the next node. We conducted several experiments on a real-world dataset, demonstrating that *GlobalTrustWalker* can provide recommendations with high accuracy and coverage.

In this paper, we focus on addressing the recommendation challenges of data sparsity, cold-start users and recommendation accuracy. The proposed trust-based approach effectively solves those problems. We believe that there is still much room for improvement. The GlobalTrust concept was set as context-independent. So we plan to expand the GlobalTrust into the contextual environment in our future work. Furthermore, in order to continuously improve the recommendation research, new recommendation algorithms will be needed to better mine various kinds of newly available social information.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (NSFC) under Grants No. 61170285.

References

[1] M.J. Eppler and J. Mengis, "The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines," The Information Society, vol.20, no.5, pp.325–344, 2004.
[2] J.D. Lasica, "Blogs and journalism need each other," Nieman Reports, vol.57, no.3, pp.70–74, 2003.
[3] Wikipedia entry on information, available at <http://en.wikipedia.org/wiki/Information>
[4] J. Tang, X. Hu, and H. Liu, "Social recommendation: a review," Social Network Analysis and Mining, vol.3, no.4, pp.1113–1133, 2013.
[5] M. Grčar, D. Mladenič, B. Fortuna, and M. Grobelnik, "Data sparsity issues in the collaborative filtering framework," in Advances in Web Mining and Web Usage Analysis, Lecture Notes in Computer Science, vol.4198, pp.58–76, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
[6] facebook statistics, available at <http://www.statisticbrain.com/facebook-statistics/>
[7] I. Guy and D. Carmel, "Social recommender systems," Proceedings of the 20th International Conference Companion on World wide web, pp.283–284, ACM, 2011.
[8] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy, "Make new friends, but keep the old: recommending people on social networking sites," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp.201–210, ACM, 2009.
[9] V. Agarwal and K.K. Bharadwaj, "A collaborative filtering framework for friends recommendation in social networks based on interaction intensity and adaptive user similarity," Social Network Analysis and Mining, vol.3, no.3, pp.359–379, 2013.
[10] W.-Y. Chen, J.-C. Chu, J. Luan, H. Bai, Y. Wang, and E.Y. Chang, "Collaborative filtering for Orkut communities: discovery of user latent behavior," Proceedings of the 18th International Conference on World wide web, pp.681–690, ACM, 2009.
[11] M. Jiang, P. Cui, R. Liu, Q. Yang, F. Wang, W. Zhu, and S. Yang, "Social contextual recommendation," Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, 2012.
[12] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation," Social Networks Proceedings of the 2010 ACM Conference on Recommender systems (RecSys),

2010.
[13] F. Walter, S. Battiston, and F. Schweitzer, “A model of a trust-based recommendation system on a social network,” Autonomous Agents and Multi-Agent Systems, vol.16, no.1, pp.57–74, 2008.
[14] Epinions, available at <http://www.epinions.com/>.
[15] M. Pazzani and D. Billsus, “Learning and revising user profiles: The identification of interesting web sites,” Machine Learning, vol.27, no.3, pp.313–331, 1997.
[16] H.C. Wu, R.W.P. Luk, K.F. Wong, and K.L. Kwok, “Interpreting tf-idf term weights as making relevance decisions,” ACM Trans. Information Systems (TOIS), vol.26, no.3, pp.13, 2008.
[17] R.J. Mooney, P.N. Bennett, and L. Roy, “Book recommending using text categorization with extracted information,” Proc. Recommender Systems Papers from 1998 Workshop, Technical Report, WS-98-08, 1998.
[18] M. Balabanović and Y. Shoham, “Fab: content-based, collaborative recommendation,” Communications of the ACM, vol.40, no.3, pp.66–72, 1997.
[19] Y. Koren, “Factorization meets the neighborhood: a multifaceted collaborative filtering model,” Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.426–434, ACM, 2008.
[20] J.L. Herlocker, J.A. Konstan, A. Borchers, and J. Riedl, “An algorithmic framework for performing collaborative filtering,” Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.230–237, ACM, 1999.
[21] H. Yildirim and M.S. Krishnamoorthy, “A random walk method for alleviating the sparsity problem in collaborative filtering,” Proceedings of the 2008 ACM Conference on Recommender Systems, pp.131–138, ACM, 2008.
[22] H. Ma, H. Yang, M.R. Lyu, and I. King, “SoRec: social recommendation using probabilistic matrix factorization,” International Conference on Information and Knowledge Management (CIKM), 2008.
[23] M. Jamali and M. Ester, “A matrix factorization technique with trust propagation for recommendation,” Social Networks, Proceedings of the 2010 ACM Conference on Recommender Systems (RecSys), 2010.
[24] J. Golbeck and J. Hendler, “FilmTrust: movie recommendations using trust in web-based social networks,” Proceedings of the IEEE Consumer Communications and Networking Conference, pp.282–286, Jan. 2006.
[25] P. Massa and P. Avesani, “Controversial users demand local trust metrics: an experimental study on epinions.com community,” Proceedings of the 25th American Association for Artificial Intelligence Conference, 2005.
[26] G. Guo, J. Zhang, and D. Thalmann, “Merging trust in collaborative filtering to alleviate data sparsity and cold start,” Knowledge-Based Systems, vol.57, pp.57–68, 2014.
[27] P. Moradi and S. Ahmadian, “A reliability-based recommendation method to improve trust-aware recommender systems,” Expert Systems with Applications, vol.42, no.21, pp.7386–7398, 2015.
[28] J. Jin and Q. Chen, “A trust-based Top-K recommender system using social tagging network,” Proceedings of the 9th IEEE International Conference on Fuzzy Systems and Knowledge Discovery, pp.1270–1274, 2012.
[29] S. Fazeli, H. Drachsler, and F. Brouns, “Towards a social trust-aware recommender for teachers,” Recommender Systems for Technology Enhanced Learning, pp.177–194, Springer, New York, 2014.
[30] S.H. Cha, “Comprehensive survey on distance/similarity measures between probability density functions,” International Journal of Mathematical Models and Methods in Applied Sciences, vol.1, no.2, 1, 2007.
[31] P. Massa and P. Avesani, “Trust-aware recommender systems,” ACM Conference on Recommender Systems, pp.17–24, 2007.
[32] M. Jamali and M. Ester, “TrustWalker: a random walk model for combining trust-based and item-based recommendation,” Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.397–406, ACM, 2009.
[33] S. Deng, L. Huang, and G. Xu, “Social network-based service recommendation with trust enhancement,” Expert Systems with Applications, vol.41, no.18, pp.8075–8084, 2014.
[34] P. Massa and P. Avesani, “Trust-aware bootstrapping of recommender systems,” ECAI Workshop on Recommender Systems, pp.29–33, 2006.
[35] P. Massa and B. Bhattacharjee, “Using trust in recommender systems: an experimental analysis,” Proceedings of iTrust2004 International Conference, pp.221–235, 2004.

Liang Chen is a Ph.D. candidate in School of Computer Science, National University of Defense Technology. His main research interests are Social Networks and Cybersecurity.

Chengcheng Shao is a Ph.D. candidate in School of Computer Science, National University of Defense Technology. His main research interests are Complex Networks and Network Science.

Peidong Zhu was born in 1971. Professor in the School of Computer Science, National University of Defense Technology. His research interests include network routing, network security and architecture design of the Internet and various wireless networks.

Haoyang Zhu is a Ph.D. candidate in School of Computer Science, National University of Defense Technology. His main research interests are Complex Networks and Cloud Computing.