Physica A 461 (2016) 61–72

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

SoRS: Social recommendation using global rating reputation and local rating similarity

Fulan Qian[a,c], Shu Zhao[a,c], Jie Tang[b], Yanping Zhang[a,*]

[a] School of Computer Science and Technology, Anhui University, Hefei 230601, China
[b] Department of Computer Science and Technology, Tsing hua University, Beijing 100084, China
[c] Center of Information Support & Assurance Technology, Anhui University, Hefei, 230601, China

HIGHLIGHTS

- A recommender system combining global reputation and local similarity is proposed.
- This method of calculating reputation is completely independent of network.
- Predication performance of proposed algorithm has good understandability.
- The proposed framework also can do well work in small training sets.

ARTICLE INFO

Article history:
Received 18 December 2015
Received in revised form 29 March 2016
Available online 24 May 2016

Keywords:
Recommender system
User reputation
Social recommendation

ABSTRACT

Recommendation is an important and also challenging problem in online social networks. It needs to consider not only users' personalized interests, but also social relations between users. Indeed, in practice, users are often inclined to accept recommendations from friends or opinion leaders (users with high reputations). In this paper, we present a novel recommendation framework, social recommendation using global rating reputation and local rating similarity, which combine user reputation and social similarity based on ratings. User reputation can be obtained by iteratively calculating the correlation of historical ratings of user and intrinsic qualities of items. We view the user reputation as the user's global influence and the similarity based on rating of social relation as the user's local influence, introduce it in the basic social recommender model. Thus users with high reputation have a strong influence on the others, and on the other hand, the effect of a user with low reputation has been weakened. The recommendation accuracy of proposed framework can be improved by effectively removing nature noise because of less rigorous user ratings and strengthening the effect of user influence with high reputation. We also improve the similarity based on ratings by avoiding the high similarity with the less common ratings between friends. We evaluate our approach on three datasets including Movielens, Epinions and Douban. Empirical results demonstrate that proposed framework achieves significant improvements on recommendation accuracy. User reputation and local similarity which are both based on ratings have a lot of helpful in improvement of prediction accuracy. The reputation also can help to improve the recommendation precision with the small training sets.

* Corresponding author.
E-mail address: zhangyp2@gmail.com (Y. Zhang).

http://dx.doi.org/10.1016/j.physa.2016.05.025
0378-4371/

1. Introduction

In e-commerce, several service applications are driven by recommendation algorithms. With the development of Web 2.0, many social media allow users to build their social relations and provide recommendation of the service applications for users. Formed by users in the network, the social information provides an independent source of recommendation. Social recommendation [1–7] presents both opportunities and challenges to traditional recommender systems. As the common additional input, social relation is the most discussed problem. Social relation is usually described by a trust list or follower-and-followee relationship. A recommender system based on social relation falls into two broad categories. (1) Memory-based social recommender systems. How to find the appropriate user sets is the key problem in this category. These commonly used methods are information diffusion and random walk based on influence propagation theory [8–11] to find the appropriate user sets in the social relation network. (2) Model-based social recommender systems. Most existing social recommender systems in this class are based on matrix factorization. The social relation information is commonly introduced into the model-based social recommender system as regular parameter. Sometimes it is dealed directly with the linear combination of rating information and social information. The new models can improve the recommendation accuracy because of using the related social relation information [12–14].

In order to improve the recommendation accuracy, many factors including local and global influences are introduced into recommender system. Ma et al. [14] employed the similarity between users based on historical ratings as the local influence in recommender system. The framework they proposed can increase the mutual influences of users and improves the recommendation effect. This framework is also cited as SoReg which is a baseline algorithms in social recommendation. There are many methods to obtain the similarity between users, such as utilizing the social network topology [15–17] and social textual information [18]. User similarity reflects the local characteristic of social relation. As the factor with an important role in recommendation, user global influence has been considered by a few scholars in the social recommender system [19].

The reputation systems [20,21] calculate the reputation by collecting and gathering feedback information from previous behaviors of objects such as the previous supply information of supplier and historical ratings of user [22,23]. Hendrikx et al. [24] pointed that the reputation is "the estimation in which a person or thing is generally held". Thus the reputation can be used to represent the user global influence information.

Some methods are proposed by taking user reputation as the global influence in social recommender system. Tang et al. [25] described the user global influence by using user reputation obtained by calculating the PageRank value of each user node in the social network. The typical PageRank method is applied only to a directed graph and is not applied to an undirected graph. Costagliola et al. [26] viewed that the reputation has a global influence aspect and the trust is viewed as a local influence from users linked in trust-chain. Tang et al. [27] combined global and local trust for service recommendation, they set usual items as services, and computed the reputation of service as the global trust of the service. Essentially, unlike above papers, the user global influence was replaced with service global influence in this paper. DeMeo et al. [28] proposed the method to recommend users in social networks by integrating local and global reputation. Using the ego-network of each node, the model was built for a directed social network and unsuitable for an undirected social network.

The methods of user reputation in above mentioned papers are dependent on social network. User global influence or reputation based on social network rely heavily on network topology. We introduce the user reputation based on global ratings [29] as the user global influence into our social recommender model. Compared with user reputation based on social network structure, our proposed framework is used in a larger scale because of complete independence of social network. User reputation is obtained by iteratively calculating the correlation of historical ratings of user and intrinsic qualities of items. The greater the correlation, the higher the user reputation. The user reputation described with ratings employs more common sense and has good understandability. Therefore, users are more influential with their shopping or use experience (i.e., user historical ratings) and rigorous professional ratings (i.e., the correlation of historical ratings of users and intrinsic qualities of items is great), not only by simply having many friends.

In this study, we investigate the effect of recommendation accuracy by introducing the user reputation based on ratings into social recommender system and propose a novel framework, that is social recommendation using global rating reputation and local rating similarity (SoRS) framework. Our contributions are summarized as follows:

(1) We introduce the reputation system based on ratings into social recommendation, thereby provide an alternative approach to improve the recommender system by exploiting the reputation system.
(2) We verify the fact that SoRS framework (combined user reputation based on ratings and local rating similarity) has better recommendation effect than the framework used only one of the aforementioned information sources. This verification is conducted by choosing parameters.
(3) SoRs is evaluated on two real datasets, namely, Douban and Epinions. Empirical results demonstrate that SoRS achieves considerably improvement in prediction accuracy for both directed and undirected social networks. We also perform the experiments on the two Movielens benchmark datasets. The experimental results show that SoRS does work very well on general recommendation system without social network.
(4) We analyze the characters of reputation and show the effect of SoRS in small training sets. The experimental results show that using the reputation as complement can compensate the lack of training sets and improve the recommendation precision.

2. Method

2.1. Notation

We first introduce notations used in this paper. We use Latin letters to represent users and Greek letters to represent objects to distinguish them. Let U present user latent factor matrix and V present item latent factor matrix. We use $R \in \mathbb{R}^{n\times m}$ as the user-item rating matrix. $r_{i\alpha}$ presents the user i ratings α on item. Let $U_i \in \mathbb{R}^K$ and $V_\alpha \in \mathbb{R}^K$ be the user preference vector for u_i and item characteristic vector for v_α, respectively, where K is the number of latent factors.

2.2. User reputation based on historical ratings

We employ the method described as user reputation in Ref. [29]. User reputation is determined by the correlation coefficient between the rating vector of user and corresponding average rating vector of items. Each item has a certain true intrinsic quality. User reputation can be obtained by iteratively calculating the correlation of historical ratings of user and intrinsic qualities of items. User reputation can be depicted as follows.

The reputation of user i s initially assigned as Eq. (1).

$$C_i = |r(i)|/n \tag{1}$$

where $r(i)$ denotes the item set of items that have ratings given by user i.

The intrinsic quality of item α can be calculated by Eq. (2).

$$Q_\alpha = \frac{\sum\limits_{i\in U(\alpha)} C_i r_{i\alpha}}{\sum\limits_{i\in U(\alpha)} C_i} \tag{2}$$

where $U(\alpha)$ denotes the user set of users who rated item α.

The correlation of user i and intrinsic qualities of items can be described as Eq. (3).

$$Corr_i = \frac{\sum\limits_{\alpha\in r(i)} (r_{i\alpha} - \overline{r_i})(Q_\alpha - \overline{Q_i})}{\sqrt{\sum\limits_{\alpha\in r(i)} (r_{i\alpha} - \overline{r_i})^2} \cdot \sqrt{\sum\limits_{\alpha\in r(i)} (Q_\alpha - \overline{Q_i})^2}} \tag{3}$$

where $\overline{r_i}$ denotes the average value of rating that user i has provided. $\overline{Q_i}$ denotes the average value of intrinsic qualities of all items rated by user i.

The following conclusions can be derived from Eq. (3). If the rating of user i rated on items is close to their intrinsic qualities, then the correlation between user i and intrinsic qualities of items is high. In some extent, the high correlation indicates that the user is professional and rigorous; otherwise, the user is not serious.

For introducing the user reputation as regular parameter into the train model, we employ a mapping function, which is expressed as Eq. (4), to bound the range of $Corr_i$ into $[0, 1]$.

$$C_i = (Corr_i + 1)/2. \tag{4}$$

Iterating Eqs. (2)–(4) until the results satisfy Eq. (5), we obtain the reputation of each user.

$$\frac{1}{n}\sum_{\alpha=1}^{n} (Q_\alpha^{(n)} - Q_\alpha^{(n-1)}) \le 10^{-6}. \tag{5}$$

We discuss effect of recommendation with user reputation in Section 3.2.1. The relationship of the distribution of user reputation and accuracy improvement is investigated in Section 3.2.2. Under relatively small training set, the recommendation performance with reputation as additional information is showed in Section 3.2.3.

2.3. Local rating similarity function

Social correlation theories such as homophily theory and social influence theory pave the way for utilizing the social information in recommendations. Homophily means that the more similarity preference users has, the more connection users likely create in social network. Social influence suggests that users who are socially connected are more likely to share similar tastes. Considerable literature utilize the aforementioned theories to improve the recommendation effect by constructing related regular parameters. We use the homophily theory to build the individual-based social regularization term model proposed in Ref. [14] (that is SoReg) to impose constraints between the user and his/her friends. The social regularization term can be expressed as Eq. (6).

$$\frac{\lambda_1}{2} \sum_{f \in f(i)} sim(i,f) \left\| U_i - U_f \right\|_F^2, \tag{6}$$

where λ_1 is social parameter to controls the capability of social information, we will discuss this parameter in Section 3.2.4. The intimate level of friendship can be expressed by the distance of user latent factor vectors. If the distance of user latent factor vectors is long, then the difference between friends is large; otherwise, the difference between friends is small. We further use the similarity between friends, namely, $sim(i,f)$, as the regular parameter to strengthen the interplay between friends with high similarity. When the $sim(i,f)$ is large, the distance of U_i and U_f is long, and when $sim(i,f)$ the is small, the distance of U_i and U_f is short. We use the notation $f(i)$ to denote the friend set of user i in the undirected networks and outlink friend set of user i in the directed networks.

The evaluation of similarity $sim(i,f)$ between user i and user f can be calculated by many methods. We can define the similarity based on ratings which is notated as $sim(i,f)_{RB}$ by measuring the previous common rating of the two users. $sim(i,f)_{RB}$ can be calculated by Eq. (7).

$$corr(i,f) = \frac{\sum\limits_{\alpha \in r(i) \cap r(f)} (r_{i\alpha} - \overline{r_i})(r_{f\alpha} - r_f)}{\sqrt{\sum\limits_{\alpha \in r(i) \cap r(f)} (r_{i\alpha} - \overline{r_i})^2} \cdot \sqrt{\sum\limits_{\alpha \in r(i) \cap r(f)} (r_{f\alpha} - \overline{r_f})^2}} \quad f \in f(i), \tag{7}$$

where $\overline{r_i}$ and $\overline{r_f}$ denote the average ratings of all items given by users i and f, respectively. The value of $corr(i,f)$ ranges from -1 to 1. We also employ a mapping function to bound the range of $corr(i,f)$ into [0, 1] which is expressed as $sim(i,f)_{RB}$ can be defined as Eq. (8).

$$sim(i,f)_{RB} = \frac{corr(i,f) + 1}{2(1 + e^{\frac{-|r(i) \cap r(f)|}{2}})}. \tag{8}$$

In following experiments, we will explore this similarity based on ratings whether improve the recommendation accuracy by comparing the $sim(i,f) = sim(i,f)_{RB}$ the computed by Eq. (8). And we further use other similarity functions to replace the $sim(i,f)$. The details are discussed in Sections 3.2.1 and 3.2.2.

2.4. Framework of the algorithm

To utilize the user reputation and social relation, we propose a novel framework. The framework is intended to solve the optimization problem, which is expressed as Eq. (9).

$$\begin{aligned} \min_{U,V} L(R,U,V) &= \frac{1}{2} \sum_{i=1}^{n} \sum_{\alpha=1}^{m} I_{i\alpha} C_i (r_{i\alpha} - U_i^T V_\alpha)^2 \\ &+ \frac{\lambda_1}{2} \sum_{i=1}^{n} \sum_{f \in f(i)} sim(i,f) \left\| U_i - U_f \right\|_F^2 + \frac{\lambda}{2} (\|U\|_F^2 + \|V\|_F^2). \end{aligned} \tag{9}$$

User reputation C_i can be obtained by iteratively calculating using the method in Section 2.2.

We use $L(R,U,V)$ to denote the object function in Eq. (9). The derivations of $L(R,U,V)$ with respect to U, V are represented as Eqs. (10) and (11).

$$\frac{\partial L}{\partial U_i} = \sum_{\alpha=1}^{m} I_{i\alpha} C_i (U_i^T V_\alpha - r_{i\alpha}) V_\alpha + \lambda U_i + \lambda_1 \sum_{f \in f(i)} sim(i,f)(U_i - U_f) \tag{10}$$

$$\frac{\partial L}{\partial V_\alpha} = \sum_{i=1}^{n} I_{i\alpha} C_i (U_i^T V_\alpha - r_{i\alpha}) U_i + \lambda V_\alpha. \tag{11}$$

A local minimum of object function $L(R,U,V)$ can be obtained through a gradient descent optimization method. Eqs. (12) and (13) are used to iteratively calculate until the values of U and V reach the convergence.

$$U_i = U_i - \Delta \left(\sum_{\alpha=1}^{m} I_{i\alpha} C_i (U_i^T V_\alpha - r_{i\alpha}) V_\alpha + \lambda U_i + \lambda_1 \sum_{f \in f(i)} sim(i,f)(U_i - U_f) \right) \tag{12}$$

$$V_\alpha = V_\alpha - \Delta \left(\sum_{i=1}^{n} I_{i\alpha} C_i (U_i^T V_\alpha - r_{i\alpha}) U_i + \lambda V_\alpha \right) \tag{13}$$

where Δ denotes the learning rate.

The detailed algorithm is shown in Algorithm 1.

Algorithm 1 SoRS

Input: The rating matrix R, the social information, a few parameters : λ_1, λ, learning rate Δ,and the number of latent factors K
Output: User latent factor matrix U and item latent factor matrix V
1. Calculate the user reputation.
1.1 Repeat
1.2 Initialize the user reputation according to Eq. (1).
1.3 Iteratively calculate the user reputation according to Eqs. (2), (3), and (4).
1.4 Until convergence
2. Calculate the user similarity.
3. Calculate the derivations of $L(R, U, V)$ with respect to U, V.
3.1 Repeat
3.2 Update U according to Eq. (13).
3.3 Update V according to Eq. (14).
3.4 Until convergence

2.5. Complexity analysis

The user reputation C_i is interactively computed by Eqs. (2)–(4). The computational complexities of Eqs. (2)–(4) are $O(m \cdot n)$, $O(m \cdot n)$ and $O(n)$. So the computational complexity of C_i is $l \cdot (O(m \cdot n) + O(m \cdot n) + O(n))$, that is $O(l \cdot m \cdot n)$, where l is the number of iteration. The value of l is related to dataset, but $l \prec m$, $l \prec n$. So we can regard l as a constant.

Computed by Eqs. (7) and (8), the computational complexity of $sim(i, f)_{RB}$ is $O(n^2)$.

In actually, the values of C_i and $sim(i, f)$ usually are computed in advance offline. The main computation of gradient methods is evaluating the object function L and its gradients against variables U and V. Because of the sparsity of matrices R, the computational complexity for gradients $\frac{\partial L}{\partial U}$, $\frac{\partial L}{\partial V}$ both are $O(\eta k)$, where η is the number of nonzero entries in matrices R, and k is the dimension of user/item latent factor vector. Adding the linear look-up time cost, the computational complexity of evaluating the object function of L is $O(n + \eta k)$ online.

3. Experiment

In this section, we present various experiments to compare our method with other methods in recommendation prediction accuracy. We analysis the effect of reputation and user similarity to improve recommendation in SoRS. The characters of reputation are specially analyzed. At last, we analyze how the change of social parameter λ_1 can affect the final recommendation accuracy.

3.1. Experiment setup

3.1.1. Data

We choose two real datasets to evaluate our proposed framework, namely, Douban [14] and Epinions [25]. These datasets are publicly available from the homepages of relevant references of the author. The social network is an undirected graph in Douban, whereas the social network is a directed graph in Epinions. We select the top 8000 users from the original Douban datasets, and use these users as seeds to further obtain their social network and movie ratings. We maintain the users who provide ratings at least 3 times and items that are rated at least 3 times in Epinions. We first prune the two datasets for our analysis. We also delete a few users and items with less rating (only one rating),which make no sense in statistics. A few statistics of these types of datasets are presented in Table 1.

For each dataset, we partitioned the total ratings record dataset into tenfold cross-validation sets. The model is trained on 8 out of 10 sets, 7 out of 10 sets, and 6 out of 10 sets (80%, 70%, 60%) as the training sets and the remaining (20%, 30%, 40%) as the testing sets. The results presented are the averaged tenfold cross-validation performance. And the user reputation is obtained by Eq. (4) through different training sets, that is 60%, 70%, 80% of total ratings record dataset.

3.1.2. Evaluation protocol

We use two popular metrics, the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as the metrics to measure the prediction accuracy of our proposed method compared with other methods.

The metric RMSE is defined as Eq. (14).

$$RMSE = \sqrt{\frac{\sum\limits_{(i,\alpha)\in E^{Test}} (r_{i\alpha} - \widetilde{r_{i\alpha}})^2}{\left|E^{Test}\right|}}. \tag{14}$$

The metric MAE is defined as Eq. (15).

$$MAE = \frac{1}{\left|E^{Test}\right|} \sum_{(i,\alpha)\in E^{Test}} (r_{i\alpha} - \widetilde{r_{i\alpha}})^2, \tag{15}$$

Table 1
Statistics of the datasets.

Datasets	Douban	Epinions
User Number	7700	22168
Item Number	6497	41369
Ratings	851220	583968
Ratings Sparsity	0.017	0.000636
Relations	5874	342037
Relations Sparsity	0.00019817	0.000696
Clustering coefficient	0.063	0.176
Relations with common ratings	4119	16196

Table 2
Performance comparison of different recommender systems in Epinions datasets.

Training	Metric	PMF	SocialMF	SoRec	SoReg	SoRS
60%	RMSE	1.1721	1.1344	1.1420	1.1409	**1.1333**
	Improve	3.31%	0%	0.76%	0.67%	
	MAE	0.9188	0.8824	0.8849	0.8837	0.8806
	Improve	4.16%	0.20%	0.49%	0.35%	
70%	RMSE	1.1608	1.1263	1.1301	1.1292	**1.1213**
	Improve	**3.40%**	0.44%	0.78%	0.70%	
	MAE	0.9075	0.8723	0.8791	0.8742	0.871
	Improve	4.02%	0.15%	0.92%	0.37%	
80%	RMSE	1.1397	1.1237	1.1278	1.1251	**1.1099**
	Improve	2.61%	1.23%	**1.59**%	**1.35**%	
	MAE	0.8897	0.8749	0.8776	0.8763	0.8709
	Improve	2.11%	0.46%	0.76%	0.62%	

where $r_{i\alpha}$ denotes the rating of item α from user i and $\widetilde{r_{i\alpha}}$ denotes the predicted rating user i gives to item α by algorithms. E^{Test} denotes the set of (user, item) pairs that have actual ratings and $\left|E^{Test}\right|$ denotes the number of actual ratings in the testing sets. Small RMSE and MAE significantly indicate better performance of the recommendation algorithms.

3.2. Result

3.2.1. Comparisons

We conduct experiments to assess the performance of SoRS. In addition, we compare SoRS results with the following four methods. Three methods are social recommendation algorithms. If the similarity function is not specified, $sim(i, f)_{RB}$ is default.

(1) PMF [30]: this is the regular MF method and it is widely used in collaborative filtering recently. It only uses the user-item rating matrix for recommendations.
(2) SoRec [12]: this is the framework with social information.
(3) SoReg [14]: this is the basic framework in this paper. It is also the framework with social information.
(4) SocialMF [11]: this model is incorporated the mechanism of trust propagation of social network in the matrix factorization model.

The comparison results are demonstrated in Tables 2 and 3. The proposed algorithm obtains better recommendation performance both in the Epinions datasets (with directed social network) and the Douban datasets (with undirected social network). SoRS obtains 3.40%, 1.59%, and 1.35% relative best improvement in terms of RMSE in Epinions compared with PMF, SoRec and SoReg. In addition, SoRS obtains 2.60%, 0.5%, and 0.53% relative best improvement in terms of RMSE in Douban compared with those methods.

3.2.2. Effect analysis of reputation and similarity

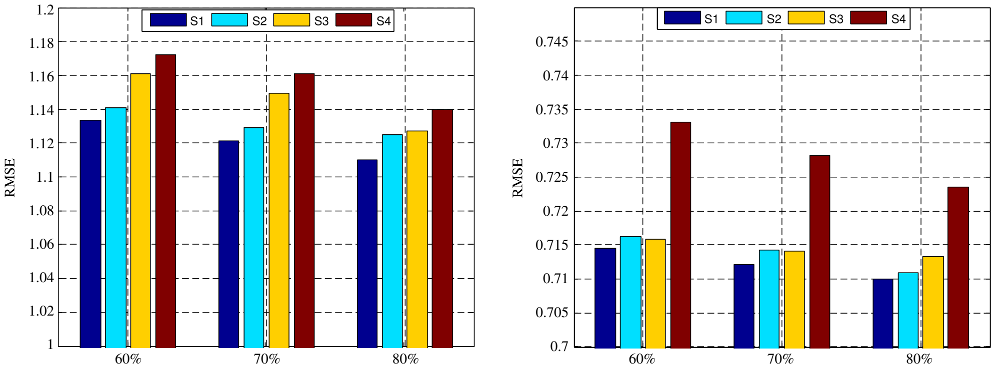
We choose two regular parameters, namely, user reputation C_i and similarity of users based on ratings $sim(i, f)$, as the research objects. We utilize these parameters to train the different object functions in training sets. By comparing RMSE and MAE in the testing sets, we investigate the effect of these regular parameters in recommendation accuracy.

We denote the $S_1 \sim S_4$ four different models of selecting these regular parameters as follows:

(1) $S_1(C_i, sim(i, f) = sim(i, f)_{RB})$: The user reputation and local rating similarity are calculated according to Eq. (4) in Section 2.2 and Eq. (8) in Section 2.3, i.e., the proposed SoRS algorithm. The algorithm not only considered the effect of user reputation, but also considered the effect of local rating similarity in social network.
(2) $S_2(C_i = 1, sim(i, f) = sim(i, f)_{RB})$:In this model, we only consider the effect of local rating similarity and disregard the effect of user reputation because the values are set to 1. This approach is the SoReg method proposed in Ref. [14].

Table 3
Performance comparison of different recommender systems in Douban datasets.

Training	Metric	PMF	SocialMF	SoRec	SoReg	SoRS
60%	RMSE	0.7336	0.7177	0.7181	0.7183	**0.7145**
	Improve	**2.60%**	0.44%	**0.5**%	**0.53**%	
	MAE	0.5788	0.5705	0.5708	0.5709	0.5675
	Improve	1.95%	0.53%	0.58%	0.6%	
70%	RMSE	0.7282	0.7136	0.7148	0.7142	**0.7121**
	Improve	2.21%	0.21%	0.38%	0.29%	
	MAE	0.5749	0.5675	0.5681	0.5671	0.5637
	Improve	1.95%	0.67%	0.77%	0.60%	
80%	RMSE	0.7235	0.7105	0.7113	0.7109	**0.7100**
	Improve	1.87%	0.07%	0.18%	0.13%	
	MAE	0.5693	0.5649	0.5657	0.5650	0.5647
	Improve	0.81%	0.03%	0.18%	0.05%	



(a) Partitions in Epinions datasets.
(b) Partitions in Douban datasets.

Fig. 1. The performance of different models of SoRS framework in two datasets.

(3) $S_3(C_i, sim(i,f) = 0)$: In this case, we calculate the effect of user reputation without considering the effect of similarity.
(4) $S_4(C_i = 1, sim(i,f) = 0)$: We train the object function disregarding the user reputation and local rating similarity. This approach is the probabilistic matrix factorization (PMF) method proposed in Ref. [30].

We use the notation $RMSE(S_i)$ to express the RMSE value of the i_{th} different models, use the notations $RMSE(sim(i,f) = sim(i,f)_{RB})$ and $RMSE(sim(i,f) = 0)$ to express the RMSE values with or without local rating similarity. As same as $MAE(S_i)$, $MAE(sim(i,f) = sim(i,f)_{RB})$ and $MAE(sim(i,f) = 0)$. For fairness, all parameters, namely λ, λ_1, and learning rate Δ are set to the optimal values in different training models. RMSEs under four models and different datasets partitions are shown in Fig. 1(a) and (b).

In Fig. 1, different partitions have two groups experiment data, user reputations are taken into account and $sim(i,f) = sim(i,f)_{RB}$ in first data, and user reputation and similarity are set to 1 and 0 respectively in second data. From the experimental results of the two groups, the following conclusions are drawn:

(1) The effects are better with user reputation than those without user reputation.

$$RMSE(S_1) < RMSE(S_2)$$
$$RMSE(S_3) < RMSE(S_4).$$

These results show that introduce user reputation into the social recommendation really can improve the recommendation accuracy.
(2) The effects are also remarkably better with local rating similarity than those without local rating similarity.

$$RMSE(sim(i,f)_{RB}) < RMSE(sim(i,f) = 0).$$

The results suggest that the local rating similarity $sim(i,f)_{RB}$ can strengthen the effect of social relations.

In order to investigate the effect of user reputation in general recommendation system, we perform the experiments on the two Movielens benchmark datasets, that is Movielens 100 K and Movielens 1 M. For the MovieLens datasets without social relation, the parameter of $sim(i,f)$ set to zero in the object function Eq. (9), that is S_3 according to the above denotations. The experiment results are shown as following Table 4.

Table 4
Performance comparison in Movielens.

Training	Metric	User-Mean	Item-Mean	PMF	SoRS under S_3
5-fold cross validation on 100 K	RMSE	1.045	1.021	0.9426	0.9135
	Improve	12.58%	10.53%	3.09%	
	MAE	0.831	0.817	0.7421	0.7156
	Improve	13.89%	12.41%	3.57%	
5-fold cross validation on 1 M	RMSE	1.036	0.983	0.9016	0.8538
	Improve	17.56%	13.14%	5.3%	
	MAE	0.827	0.783	0.7115	0.668
	Improve	19.23%	14.69%	6.11%	

Table 5
Different similarity performance comparison on Epinions.

Training	Metric	SoRS with cosine similarity	SoRS with Jaccard similarity	SoRS with $sim(i,f)_{RB}$ similarity
60%	RMSE	1.136	1.1333	1.1333
	MAE	0.8806	0.8818	0.8806
70%	RMSE	1.1245	1.1240	1.1213
	MAE	0.872	0.872	0.871
80%	RMSE	1.1173	1.1099	1.1099
	MAE	0.8753	0.8715	0.8709

Table 6
Different similarity performance comparison on Douban.

Training	Metric	SoRS with cosine similarity	SoRS with Jaccard similarity	SoRS with $sim(i,f)_{RB}$ similarity
60%	RMSE	0.7148	0.716	0.7145
	MAE	0.5681	0.5682	0.5675
70%	RMSE	0.71485	0.7155	0.7121
	MAE	0.5657	0.5659	0.5637
80%	RMSE	0.7112	0.7124	0.7100
	MAE	0.5657	0.5655	0.5647

From Table 4, we can draw the conclusion that the user reputation really can considerably improve the recommendation accuracy on Movielens benchmark datasets.

The similarity function $sim(i,f)$ measures how similar user i and user f are in SoRS framework. In this is paper, we employ $sim(i,f)_{RB}$ similarity function defined as Eq. (8) in SoRS. In order to examine how much the similarity function $sim(i,f)$ contributions to the SoRS framework, we also add some experiments on the other two similarities: Cosine and Jaccard. And the results are listed as Tables 5 and 6.

From Tables 5 and 6, we can see that SoRS with $sim(i,f)_{RB}$ similarity are generally better than the other two similarities. This observation demonstrates that $sim(i,f)_{RB}$ is a better measure of similarity.

3.2.3. The characters analysis of reputation in SoRS

Using the proposed SoRS algorithm, we find that the effect on Epinions datasets is significantly better than that on Douban datasets in the improvement extent of recommendation accuracy from Fig. 1. Compared to the worst result S_4, the improvement rate of the best result S_1 is 2.6% in 80% training set/20% testing set of Epinions datasets, while 2.1% in Douban. Apart from the intrinsically different characteristics of the two datasets, a few reasons behind this phenomenon are attributed to the distribution of user reputation in the datasets. The distributions of user reputation obtained from 80% training sets in Epinions and Douban datasets are described in Figs. 2 and 3.

The threshold of the reputation is 0.5 according to the Eqs. (3) and (4). If the user reputation is lower than the threshold, it indicates that the ratings of the user are not correlated or have negative correlations with intrinsic qualities of items. This kind of user with low reputation is viewed as noise user which is very loose on giving ratings. While, the user with high reputation can be regarded as an expert.

From Figs. 2 and 3, the noise from user ratings in Epinions datasets is significantly more than that in Douban datasets. And there are also many users with high reputation in Epinions. We can draw the conclusion that the polarization of users is more serious in Epinions. And the users in Douban are generally more rigorous than those in Epinions. SoRS can weaken the effect of users with low reputations and improve the effect of users with high reputation. This ability of the proposed algorithm can explain why recommendation accuracy in different datasets improves in different extents.

Figs. 2 and 3 also show that the distributions of user reputation obtained from different percent training set have relatively stable properties in the certain dataset.

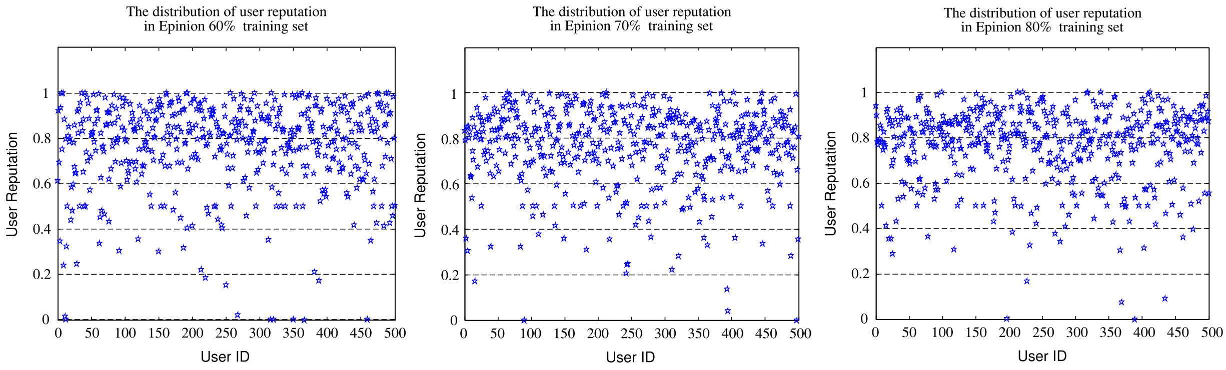


Fig. 2. Distribution of user reputation in Epinions datasets.

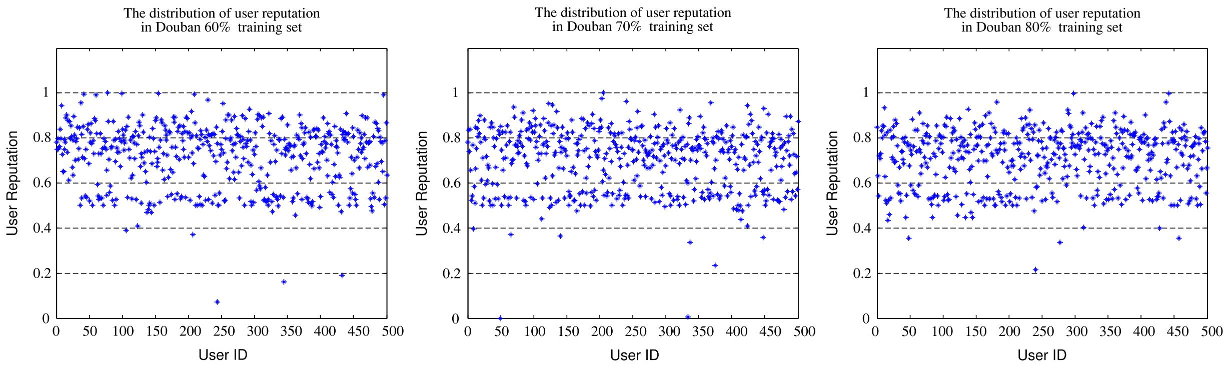


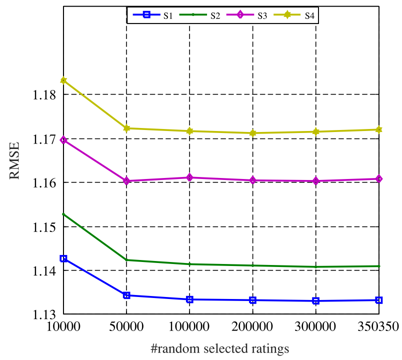
Fig. 3. Distribution of user reputation in Douban datasets.

In order to investigate whether the reputation and local rating similarity does work in less training set, we randomly select a certain quantity ratings from different training set. It is noteworthy that the user reputation is obtained by Eq. (4) through different training sets, that is 60%, 70%, 80% of total ratings record dataset. We do this random selections process 5 times, and obtain the mean values. From Figs. 4 and 5, when the number of random selections is less, the RMSE values under different similarity values show that values with reputation information are better than the values without reputation information, that is $RMSE(S_1) < RMSE(S_2)$; $RMSE(S_3) < RMSE(S_4)$. This is due to the implement of reputation information can help to improve the recommendation precision with the limited training sets. As the number of random selection is increased, the difference between values with repudiation information and values without reputation information gets decreased. But on the whole, the former ones are still better. We do multiple random selections and set to the mean values. Up to the certain extent, the recommendation is tending towards stability. Thus, using the reputation as complement can compensate the lack of training sets and improve the recommendation precision.

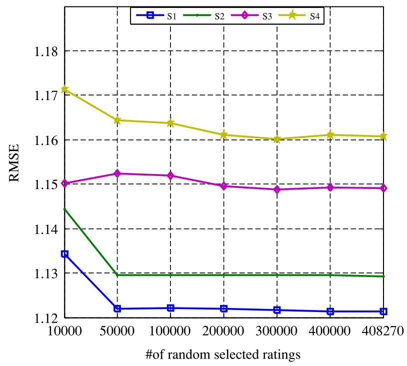
3.2.4. Impact of parameters λ_1

In our proposed framework, the parameter λ_1 controls how much the information of local ratings similarity should be incorporated in the recommendation system. If we use an extremely small value of λ_1, then we mine only the user-item rating matrix for matrix factorization. Moreover, if we employ a large value of λ_1, then information of local rating similarity would control the learning processes. In this section, we analyze how the change of λ_1 can affect the final recommendation accuracy.

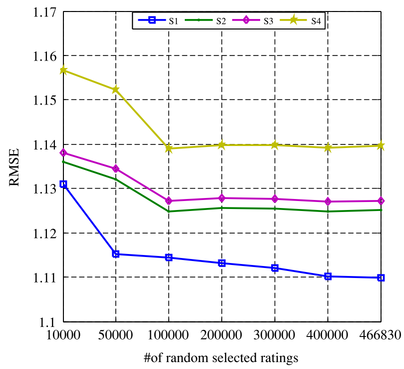
Fig. 6 shows the effects of λ_1 on RMSE in our proposed framework on the two real datasets. The value of λ_1 significantly influences the recommendation results. This phenomenon demonstrates that incorporating the social relation information significantly improves the recommendation accuracy. According to Fig. 6, no matter which datasets are used under different training data setting, as λ_1 decrease, the RMSE values initially decrease, but when λ_1 drops below a certain threshold such as 0.0001 on Epinions dataset and 0.001 on Douban dataset, the RMSE values then increase with further increase of the value of λ_1. The existence of the yielding point indicates that combining user reputation and local rating similarity (calculating the user reputation and using matrix factorization for recommendation) can generate better performance than using only one of the aforementioned information sources for recommendation.



(a) 60% Epinions RMSE.



(b) 70% Epinions RMSE.



(c) 80% Epinions RMSE.

Fig. 4. Different proportion of random ratings in Epinions datasets.

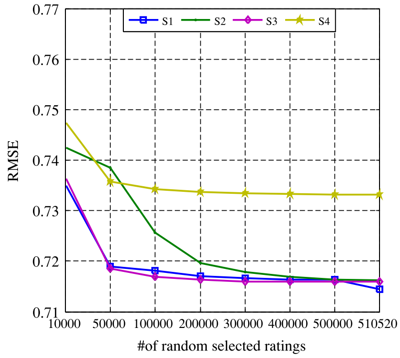
4. Conclusion

Considering that the social phenomenon of the recommendation from high reputation users and social friends is often being at work, we proposed a recommendation algorithm that combines user reputation and social relation, namely, SoRS. The SoRS algorithm obtains the user reputation based on historical ratings through the correlation calculation of user ratings and intrinsic qualities of item. This method of calculating reputation is completely independent of network topology, and the predication performance has the characteristic of good understandability. If the user has high reputation with high correlation of ratings and qualities, he would be more serious and professional on giving his rating. Conversely, the user would provide ratings on items randomly and loosely, if he has low reputation with low or negative correlation of ratings and qualities. We introduce the reputation as the regular parameter into the training model to improve the model prediction performance. This task was conducted by strengthening the impact of users with high reputation and reducing the impact of users with low reputation. Furthermore, we strengthen the interaction between high similar friends by using the similarity based on ratings. Besides the experiments on Movielens benchmark datasets, we operate the prediction performance experiments with our proposed approach on undirected and directed social networks, namely, Douban and Epinions, respectively. Empirical results demonstrate that SoRS framework has a considerable improvement. SoRS also does well work under relatively small training set with reputation as additional information.

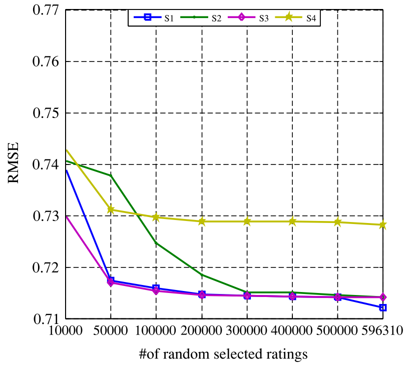
With the development of social media, the derived social network and different network topology structures have become research hotspots. We will further mine the inherent information in social relation with the ratings and the context information such as user and item profiles. The methods which are utilized the different social network topology structures and the network structure evolutionary processes will be emphasized in our future research. Along with the above research idea, we will design the more reasonable user reputation and user social relation influence models to improve the recommendation accuracy.

Acknowledgments

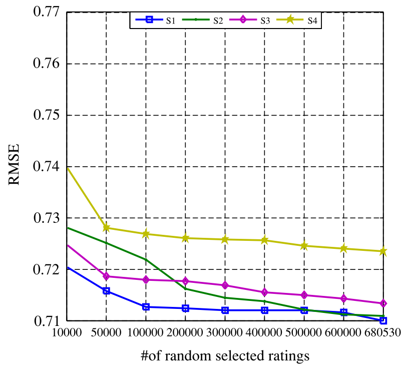
This work is partially supported by the National Natural Science Foundation of China (Grant Nos. 61175046 and 61402006), the Humanities and Social Science Research on Youth Fund Project, Ministry of Education (No. 14YJC860020), and the Provincial Natural Science Research Program of Higher Education Institutions of Anhui province (No. KJ2013A016).



(a) 60% Douban RMSE.



(b) 70% Douban RMSE.



(c) 80% Douban-RMSE.

Fig. 5. Different proportion of random ratings in Douban datasets.

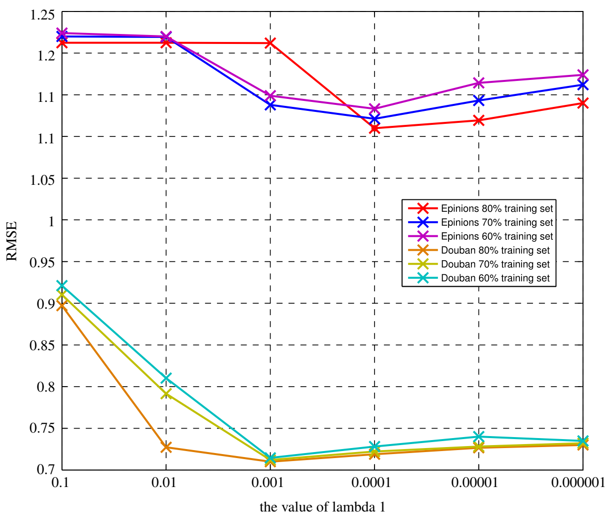


Fig. 6. The value of λ_1 in Epinions and Douban datasets.

References

[1] D.-C. Nie, Z.-K. Zhang, J.-L. Zhou, Y. Fu, K. Zhang, Information filtering on coupled social networks, PloS One 9 (2014) e101675.
[2] T. Qiu, T.-T. Wang, Z.-K. Zhang, L.-X. Zhong, G. Chen, Alleviating bias leads to accurate and personalized recommendation, Europhys. Lett. 104 (2013) 48007.
[3] C.-X. Zhang, Z.-K. Zhang, L. Yu, C. Liu, H. Liu, X.-Y. Yan, Information filtering via collaborative user clustering modeling, Physica A 396 (2014) 195–203.
[4] T. Qiu, Z.-K. Zhang, G. Chen, Information filtering via a scaling-based function, PloS One 8 (2013).
[5] Z.-K. Zhang, C. Liu, Y.-C. Zhang, T. Zhou, Solving the cold-start problem in recommender systems with social tags, Europhys. Lett. 92 (2010) 28002.
[6] Y. Guan, S. Cai, M. Shang, Recommendation algorithm based on item quality and user rating preferences, Front. Comput. Sci. 8 (2014) 289–297.

[7] J.-G. Liu, T. Zhou, B.-H. Wang, Y.-C. Zhang, Q. Guo, Degree correlation of bipartite network on personalized recommendation, Internat. J. Modern Phys. C 21 (2010) 137–147.
[8] P. Victor, M. De Cock, C. Cornelis, Trust and recommendations, in: Recommender Systems Handbook, Springer, 2011, pp. 645–675.
[9] J. Golbeck, Generating Predictive Movie Recommendations From Trust in Social Networks, Springer, 2006.
[10] P. Massa, P. Avesani, Trust-aware recommender systems, in: Proceedings of the 2007 ACM Conference on Recommender Systems, ACM, 2007, pp. 17–24.
[11] M. Jamali, M. Ester, Trustwalker: a random walk model for combining trust-based and item-based recommendation, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2009, pp. 397–406.
[12] H. Ma, H. Yang, M.R. Lyu, I. King, Sorec: social recommendation using probabilistic matrix factorization, in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, ACM, 2008, pp. 931–940.
[13] H. Ma, I. King, M.R. Lyu, Learning to recommend with social trust ensemble, in: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2009, pp. 203–210.
[14] H. Ma, D. Zhou, C. Liu, M.R. Lyu, I. King, Recommender systems with social regularization, in: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, ACM, 2011, pp. 287–296.
[15] Z. Sun, L. Han, W. Huang, X. Wang, X. Zeng, M. Wang, H. Yan, Recommender systems based on social networks, J. Syst. Softw. 99 (2015) 109–119.
[16] L. Guo, J. Ma, Z. Chen, H. Zhong, Learning to recommend with social contextual information from implicit feedback, Soft Comput. (2014) 1–12.
[17] L. Backstrom, J. Leskovec, Supervised random walks: predicting and recommending links in social networks, in: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, ACM, 2011, pp. 635–644.
[18] T. Zhao, C. Li, M. Li, Q. Ding, L. Li, Social recommendation incorporating topic mining and social trust analysis, in: Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management, ACM, 2013, pp. 1643–1648.
[19] G. Guo, J. Zhang, N. Yorke-Smith, TrustSVD: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings, 2015.
[20] M. Medo, J.R. Wakeling, The effect of discrete vs. continuous-valued ratings on reputation and ranking systems, Europhys. Lett. 91 (2010) 48004.
[21] C. De Kerchove, P. Van Dooren, Iterative filtering in reputation systems, SIAM J. Matrix Anal. Appl. 31 (2010) 1812–1834.
[22] R.-H. Li, J.X. Yu, X. Huang, H. Cheng, Robust reputation-based ranking on bipartite rating networks, in: SDM, SIAM, 2012, pp. 612–623.
[23] L. De Alfaro, A. Kulshreshtha, I. Pye, B.T. Adler, Reputation systems for open collaboration, Commun. ACM 54 (2011) 81–87.
[24] F. Hendrikx, K. Bubendorfer, R. Chard, Reputation systems: A survey and taxonomy, J. Parallel Distrib. Comput. 75 (2015) 184–197.
[25] J. Tang, X. Hu, H. Gao, H. Liu, Exploiting local and global social context for recommendation, in: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, AAAI Press, 2013, pp. 2712–2718.
[26] G. Costagliola, V. Fuccella, F.A. Pascuccio, Towards a trust, reputation and recommendation meta model, J. Vis. Lang. Comput. 25 (2014) 850–857.
[27] M. Tang, Y. Xu, J. Liu, Z. Zheng, X. Liu, Combining global and local trust for service recommendation, in: 2014 IEEE International Conference on Web Services, ICWS, IEEE, 2014, pp. 305–312.
[28] P. De Meo, F. Messina, D. Rosaci, G.M. Sarné, Recommending users in social networks by integrating local and global reputation, in: Internet and Distributed Computing Systems, Springer, 2014, pp. 437–446.
[29] Y.-B. Zhou, T. Lei, T. Zhou, A robust ranking algorithm to spamming, Europhys. Lett. 94 (2011) 48002.
[30] A. Mnih, R. Salakhutdinov, Probabilistic matrix factorization, in: Advances in Neural Information Processing Systems, 2007, pp. 1257–1264.