

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection School of Information Systems

School of Information Systems

4-2017

Factored similarity models with social trust for top-N item recommendation

Guibing GUO

Jie ZHANG

Feida ZHU

Singapore Management University, fdzhu@smu.edu.sg

Xingwei WANG

DOI: <https://doi.org/10.1016/j.knosys.2017.01.027>

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research_all

Part of the [Databases and Information Systems Commons](#), and the [E-Commerce Commons](#)

Citation

GUO, Guibing; ZHANG, Jie; ZHU, Feida; and WANG, Xingwei. Factored similarity models with social trust for top-N item recommendation. (2017). *Knowledge-based Systems*. 122, 17-25. Research Collection School of Information Systems.

Available at: https://ink.library.smu.edu.sg/sis_research_all/12

This Journal Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Factored similarity models with social trust for top-N item recommendation

Guibing Guo^{a,*}, Jie Zhang^b, Feida Zhu^c, Xingwei Wang^a

^aSoftware College, Northeastern University, China

^bNanyang Technological University, Singapore

^cSingapore Management University, Singapore

ARTICLE INFO

Article history:

Received 6 September 2016

Revised 16 January 2017

Accepted 20 January 2017

Available online 29 January 2017

Keywords:

Recommender systems

Matrix factorization

Social trust

Trust influence

ABSTRACT

Trust-aware recommender systems have attracted much attention recently due to the prevalence of social networks. However, most existing trust-based approaches are designed for the recommendation task of *rating prediction*. Only few trust-aware methods have attempted to recommend users an ordered list of interesting items, i.e., *item recommendation*. In this article, we propose three factored similarity models with the incorporation of social trust for item recommendation based on implicit user feedback. Specifically, we introduce a matrix factorization technique to recover user preferences between rated items and unrated ones in the light of both user-user and item-item similarities. In addition, we claim that social trust relationships also have an important impact on a user's preference for a specific item. Experimental results on three real-world data sets demonstrate that our approach achieves superior ranking performance to other counterparts.

1. Introduction

At the age of Web 2.0, information has been exponentially growing and challenging the capability of (e-commerce systems) discovering useful knowledge towards user preference. It is well-known as the *information overload* problem. By providing users with quality personalized recommendations, recommender systems have become an essential component of many e-commerce applications. They learn user preference from their behaviors towards and interactions with different items. In particular, trust-aware recommender systems have attracted much attention in the literature due to the advent of online social networks. The underlying assumption is that a user's preferences can be influenced by the recommendations (or ratings) of her social friends, both explicitly and implicitly [1,5,12,22].

Two types of recommendation tasks have been well recognized in the field of recommender systems, namely *rating prediction* and *top-N item recommendation* (a.k.a, item ranking). The former task is to predict the rating value that a user is likely to give towards a certain unrated item, whereas the latter task is to suggest a list of ranked items (e.g., top-N items) that a user has not consumed but

most likely tends to like. Rating prediction requires the existence of users' explicit ratings over a number of items.¹ In contrast, item recommendation especially based on implicit feedback (e.g., purchase, browse, click) is more pervasive in real-world situations. This task is also known as *one-class collaborative filtering* since only one class (positive feedback) is available.² Till now, many trust-based approaches have been proposed [1,5,24,25] and applied in different domains [15,21], and demonstrated the effectiveness of social trust in enhancing the accuracy of recommendations. However, most of them are designed for rating prediction and only very few [25] are advised for item recommendation, a more natural problem in many scenarios. In addition, we have empirically noted that the improvements achieved by the only few trust-based approaches (for item recommendation) are quite limited, and their trust-based assumptions may be invalid in some social networks. Therefore, the utility of social trust for item recommendation requires further understanding and investigation, indicating the motivation of our work.

In this article, we propose a novel trust-based approach for the task of top-N item recommendation. Specifically, we introduce three factored similarity models based on a matrix factorization

* Corresponding author.

E-mail addresses: guogb@swc.neu.edu.cn, guoguibing@gmail.com (G. Guo), zhangj@ntu.edu.sg (J. Zhang), fdzhu@smu.edu.sg (F. Zhu), wangxw@mail.neu.edu.cn (X. Wang).

¹ There are also some works in the literature that convert implicit feedback into explicit ratings, if no ratings are supported by e-commerce systems.

² We do not consider the case of multiple types of positive feedback in this article, but leave it as a part of our future work.

technique. We claim that the ranking score of an item (for a user) is not only influenced by the similarities with the other items she rated, but also influenced by the similarities with the other users who rated the item. Furthermore, we contend that a user's social trust relationships will also have an important impact on the ranking score. Two variants of trust influence will be investigated in this article. The proposed approach, i.e., Factored user and item Similarity model with social Trust (FST), considers all the three factors along with item biases. Experimental results on three real-world data sets (i.e., FilmTrust, Epinions, Ciao) show that our approach FST achieves better ranking performance than other counterparts.

In summary, the main contributions of this article are:

1. We propose a ranking-based recommendation model that formulates a user's preference over an item based on both the influence of other users who rated the same item, and the influence of other items rated by the active user. That is, factored similarities are gained from both the perspectives of users and items rather than either one of them, which has not been investigated before.
2. We extend the factored similarity models with the impact of social trust, and integrate their influence in a unified recommendation model. Alternative approaches to model social trust are studied and compared in our work. Our work aims to resolve the problem of top-N item recommendation based on implicit feedback, which is different from many other trust-based methods for rating prediction based on explicit ratings. Top-N item recommendation in question is a more prevalent task than rating prediction.
3. We conducted a series of experiments to verify the effectiveness of our approach along with the parameter sensitivity and comparison with other models. The results on three real-world data sets demonstrate that social trust has an important impact on items' ranking scores, and thus improve the performance of top-N item recommendation.

The rest of this article is organized as follows. [Section 2](#) overviews related research in the literature. [Section 3](#) then elaborates the proposed three factored similarity models with social trust. Experimental evaluation is conducted in [Section 4](#) based on three real-world datasets. Finally, [Section 5](#) concludes the present work and outlines future research.

2. Related work

Trust-aware recommendation methods can be broadly classified into two types: memory- and model-based approaches. Memory-based approaches aim to form the best nearest neighborhood from the whole user space based on user similarity or social trust. Then, the behaviors of nearest neighbors will be integrated to form recommendations. For example, Golbeck [2] proposes the *TidalTrust* to predict movie's rating by aggregating the preferences of trust users, where user similarity is substituted by social trust. However, as a matter of fact, social trust is even more sparser than user ratings, and thus such a replacement may cause even severer *data sparsity* problem. Instead, Guo et al. [3] adopt social trust to enhance user similarity, and thus make use of both ratings and trust. Specifically, a user's original preference vector will be complemented by those of her trusted neighbors. In this way, her preference can be better modeled and thus improve the recommendation accuracy. However, researchers gradually recognize that memory-based approaches cannot function well given the requirements of large-scale data and real-time prediction.

On the contrary, model-based approaches have attracted much attention due to the recommendation accuracy and capability of

adapting to large-scale data. These approaches attempt to learn users' behavior patterns from their historic data including not only their own past behaviors but also those of other (potentially associated) users. Many trust-aware recommendation models have been proposed to date. For example, Pham et al. [18] cluster users by social trust, and then make prediction by accumulating the ratings of similar users within the same cluster. Guo et al. [4] propose a multiview clustering approach for recommendations, aiming to resolve the low accuracy and coverage of traditional clustering-based recommendation methods. Specifically, similarity and trust are used iteratively to cluster users, and thus users in a cluster contain the information of nearest similar and trust neighbors. Recently, matrix factorization based models have become more and more popular due to their efficiency and effectiveness. In particular, Jamali and Ester [8] propose the SocialMF approach where an active user's feature vector is affected by those of her trusted users and thus a prediction is further influenced. Similarly, Ma et al. [14] also contend that a user's feature vector should be close to the average of her trusted users, but differ in that trust is used to regularize the generation of user and item feature vectors rather than rating prediction. Yang et al. [22] introduce the influence of trusters and trustees of an active user to the prediction of rating values, while Yao et al. [24] adopt similar idea to regularize the learning of a matrix factorization model. Fang et al. [1] decompose social trust into four fine-grained trust facets which are then integrated into a matrix factorization model for accurate rating prediction. Guo et al. [5] propose a TrustSVD model with the incorporation of both the explicit and implicit influence of social trust other than user ratings. These trust-based models are specifically designed for the recommendation task of rating prediction, whereas our focus in this article is to generate an ordered list of interesting items for active users, i.e., item recommendation. Although straightforward methods would be to order items by predicted ratings, Yang et al. [23] have empirically shown that the well-performing trust-based models for rating prediction perform poorly for top-N item recommendation.

Although many studies have been conducted on the top-N recommender systems on the basis of implicit user feedback, very few have incorporated social trust relationships for item recommendation. Jamali and Ester [7] propose the *TrustWalker* method, which is possibly the first trust-based ranking method by adapting a nearest neighborhood approach (aiming for rating prediction) to item recommendation. However, this method has been demonstrated poor performance in the case of implicit feedback in [23]. A state-of-the-art approach for item recommendation is known as Bayesian personalized ranking (BPR) [20]. The underlying assumption is that a rated item for an active user is preferred to an unrated item. Other than the assumption on items, Krohn-Grimberghe et al. [11] impose a similar assumption on user connections and propose the multi-relational BPR (MR-BPR) for item recommendation. However, such an assumption may not be always true in real cases since users could be unaware of an item rather than dislike it. Pan and Chen [17] relax the assumption by constraining the rated items with the support of a group of other users, and thus propose the Group Bayesian Personalized Ranking (GBPR) approach. Zhao et al. [25] further distinguish the preference relationships that an item consumed by an active user is preferred to that consumed by her friends which is superior to the item consumed by other users. This approach is termed as SBPR, which to our best knowledge is the only available trust-based matrix factorization model for item recommendation. Nonetheless, the assumption of SBPR can be error-prone (or even invalid) in some situations, especially for users with only a few trusted users. Our work opts to better define an item's ranking score for an active user by incorporating her trust connections, by which the preference relationships between a rated item and an

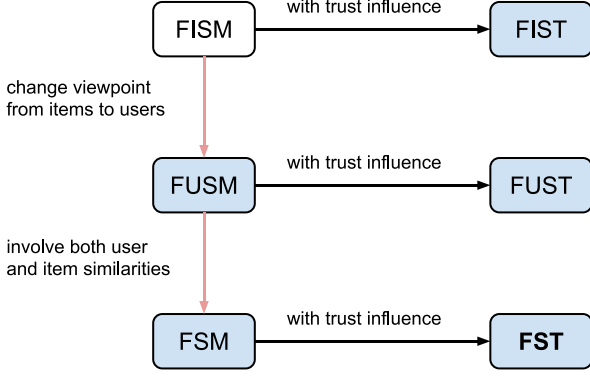


Fig. 1. Factored similarity models where FISM is a previous model and the others are the new models we propose.

unrated one can be better revealed. In this way, we shade light on a new manner to design trust-based ranking models.

3. Factored similarity models with trust

To facilitate discussion, we introduce a number of notations. Let $\mathcal{R} = [r_{u,i}]_{m \times n}$ represent a binary matrix of user behaviors over a number of items, where $r_{u,i} = 1$ indicates that user u has consumed or rated item i (otherwise $r_{u,i} = 0$), and m, n refer to the number of users and items, respectively. For clarity, we preserve symbols u, v, w for users and i, j, k for items. Let $\mathcal{T} = [t_{u,v}]_{m \times m}$ represent the social trust network of users, where $t_{u,v}$ is the trust value that user u has towards user v . In general, only binary trust values are available, i.e., $t_{u,v} = 1$ if user u trusts user v and $t_{u,v} = 0$ otherwise. Hence, in social rating networks the task of top- N item recommendation can be formalized as follows: giving a set of historical user behaviors (i.e., matrix \mathcal{R}) and a set of social trust relationships (i.e., matrix \mathcal{T}), for each user u , recommend her a small list (N) of ordered items from all the candidate items that she has not yet consumed or rated.³

3.1. Factored similarity models

In this section, we will first review the factored item similarity model (FISM) proposed by Kabbur et al. [9], inspired by which we then propose two new factored similarity models and their extensions with trust influence (see Fig. 1).

3.1.1. FISM: Factored Item Similarity Model

Kabbur et al. [9] propose the FISM model, a state-of-the-art method that performs better than other well-known models such as SLIM [16] and BPR [20]. It computes a ranking score $\hat{r}_{u,i}$ for a user u towards item i by aggregating the item similarities with the other item j she rated in the past. An item similarity is learned as an inner product of two low-rank matrices X and Y , where $X \in \mathbb{R}^{n \times d}$ and $Y \in \mathbb{R}^{n \times d}$, $d \ll n$ indicates the number of latent features associated with an item. Hence, the similarity of items i and j is derived by $x_j^\top y_i$, where x_j, y_i represent the item-specific latent feature vectors for item j and item i , respectively. The ranking score for user u on an unrated item i is predicted by:

$$\hat{r}_{u,i} = b_i + |I_{u-i}|^{-\alpha} \sum_{j \in I_{u-i}} x_j^\top y_i,$$

where b_i is the bias of item i , $I_{u-i} = I_u^+ \setminus \{i\}$ is the set of items rated by user u (denoted by I_u^+) except the current estimate item i if being rated, and $\alpha \in [0, 1]$ is a user specified parameter. Note

that we omit user bias b_u from the FISM model as it will be cancelled out during learning. The matrices X, Y can be learned by recovering item ranking relationships—a rated item i is preferred to an unrated item j for user u . The objective function to minimize is given by:

$$\mathcal{J} = \frac{1}{2} \sum_u \sum_{i \in I_u^+, j \in I_u^-} \|(r_{u,i} - r_{u,j}) - (\hat{r}_{u,i} - \hat{r}_{u,j})\|_F^2 + \frac{\lambda}{2} (\|X\|_F^2 + \|Y\|_F^2 + \|b\|_F^2),$$

where I_u^- is the set of items user u has not rated, $\|\cdot\|_F$ is the Frobenius norm, and λ is a weight parameter for regularization terms to avoid over-fitting. Note that for simplicity, we adopt the same regularization parameter λ for all variables, which is also used throughout this article. Better performance may be achieved by assigning and tuning different parameters for all the variables.

The merit of FISM is to generalize the item-based nearest neighborhood approach for item ranking into a matrix factorization model, which is generally more effective for model learning and item recommendation. The demerit of FISM is to ignore the perspective of users and not to take into account user similarity for item ranking. This directs the way how we can further improve the performance of item recommendation by additionally considering user perspective and the correlations among users, which we will elaborate in next sections.

3.1.2. FUSM: Factored User Similarity Model

For a given rating matrix \mathcal{R} , FISM computes a ranking score from the viewpoint of items, i.e., how strongly a target item is correlated with the items that a user has rated. Inspired by FISM, we can also yield a ranking score from the viewpoint of users, i.e., how strongly an active user is correlated with the users who have rated the target item. A higher score indicates that the user is more likely to prefer the target item as the like-minded users do. Based on this intuition, we propose a factored user similarity model (FUSM), where two users' similarity is computed by an inner product of two low-rank matrices $P \in \mathbb{R}^{m \times d}$, $Q \in \mathbb{R}^{m \times d}$, and $d \ll m$ is the number of latent factors that a user is associated with. Therefore, we define a user u 's ranking score on a target item i as follows:

$$\hat{r}_{u,i} = b_i + |U_{i-u}|^{-\beta} \sum_{v \in U_{i-u}} p_v^\top q_u, \quad (1)$$

where $U_{i-u} = U_i \setminus \{u\}$ is the set of users who have rated item i (denoted by U_i) except the current estimate user u if she has rated. The similarity of users u and v is computed by the inner product of p_v^\top and q_u , where p_v, q_u denote the user-specific feature vectors for users v and u , respectively. Similarly, we define $\beta \geq 0$ as a parameter to consider the number of users involved, but we do not restrict β to the value range $[0, 1]$ as α in FISM. When $\beta = 0$, Eq. (1) turns to the voting of user similarity; when $\beta = 1$, it becomes the average of user similarities while another value often used in the literature is $\beta = 0.5$; and when $\beta > 1$, the influence of the number of users may outweigh that of the sum of user similarities.

3.1.3. FSM: ranking with both user and item similarities

FISM and FUSM resolve the item ranking problem from the different perspectives of users and items, respectively. To generalize the application of factored similarity models into different scenarios, it is necessary to consider both perspectives of users and items. Hence, we propose a generic factored similarity model (FSM) where a ranking score for user u on item i is composed of three parts: (1) item bias b_i ; (2) the similarity between user u and any other user v who also rated item i : $p_v^\top q_u$; and (3) the

³ An item already rated is of less interest to recommend in our case.

similarity between item i and any other item j rated by the same user u : $x_j^\top y_i$, given as follows:

$$\hat{r}_{u,i} = b_i + s|U_{i-u}|^{-\beta} \sum_{v \in U_{i-u}} p_v^\top q_u + (1-s)|I_{u-i}|^{-\alpha} \sum_{j \in I_{u-i}} x_j^\top y_i,$$

where $\beta, \alpha \geq 0$ are the user-specified parameters, giving the flexibility to be adapted to various scenarios; and $s \in [0, 1]$ denotes the importance of user similarity for ranking scores.

3.2. Ranking with social trust

In a typical social rating network, active users not only consume or rate items but also connect with a number of users as social friends or trusted neighbors. Well-known examples include Ciao (www.ciao.co.uk) and Epinions (www.epinions.com) where a user can specify other users as trustworthy and add them to a trust list if their reviews on products are deemed valuable consistently. These systems are designed to originally support the concept of trust. We next proceed to show how social trust can be incorporated to enhance the factored similarity models (i.e., FISM, FUSM, FSM) we introduced in the previous section. Formally, assume that user u has specified a set of trusted users $T_u = \{w | t_{u,w} = 1\}$, and our objective is to predict a ranking score on item i for user u . The ranking scores are then adopted to generate a list of top-N recommendation items for each active user.

3.2.1. FIST: FISM with social trust

The basic idea of FIST is that social trust will influence an item's ranking score for active users. A real-life example is that a user may prefer to watch a movie suggested by her social friends rather than a movie without the suggestions of friends. By adding trust influence to FISM, the new ranking score for FIST is given by:

$$\hat{r}_{u,i} = b_i + |I_{u-i}|^{-\alpha} \sum_{j \in I_{u-i}} x_j^\top y_i + |T_u|^{-z} \sum_{w \in T_u} p_w^\top y_i,$$

where $\alpha, z \geq 0$ are parameters to consider the number of items and trusted users, respectively. For each trusted user $w \in T_u$, the influence to the ranking score is modelled as the inner product of user w 's feature vector and item i 's feature vector, i.e., $p_w^\top y_i$. Hence, the overall trust influence is the summation of all trusted users' influence weighted by the number of trusted users.

3.2.2. FISTa: an alternative model

An alternative way to measure trust influence (from the perspective of items) is to follow the similar way as FISM. That is, we measure the similarity between an item rated by trusted users and the target item. Specifically, the new ranking score for the alternative model is given by:

$$\hat{r}_{u,i} = b_i + |I_{u-i}|^{-\alpha} \sum_{j \in I_{u-i}} x_j^\top y_i + |I_{T_u-i}^N|^{-\gamma} \sum_{k \in I_{T_u-i}^N} x_k^\top y_i,$$

where $\alpha, \gamma \geq 0$ are parameters to consider the number of items rated by user u and u 's trust users, respectively; $I_{T_u-i}^N$ denotes the set of top-N most popular items rated by the user u 's trust users T_u except the current estimate item i . Although on average a user may have only specified a small number of trust users, the number of items rated by these trusted users can be up to thousands. It may cause the following three issues if all items are used: (1) time complexity will increase exponentially and even be prohibitively expensive; (2) higher chance to include niche items (receiving few ratings only) that may deteriorate recommendation performance; (3) the effect of these items may outweigh the effect of items rated by user u herself. Therefore, we select the top-N most popular items upon which user u are more likely to act.

Note that our empirical results, presented in the latter section, show that FIST outperforms FISTa in terms of the ranking accuracy.

One possible explanation is that the manner in FIST measures trust influence (user's effect on target items) more directly than that in FISTa (implicitly via the items rated by trusted users). Hence, we opt to select the manner of FIST to model trust influence for the subsequent models.⁴

3.2.3. FUST: FUSM with social trust

Similarly, we obtain a new ranking prediction approach by incorporating the trust influence (from the perspective of users) into FUSM, given by:

$$\hat{r}_{u,i} = b_i + |U_{i-u}|^{-\beta} \sum_{v \in U_{i-u}} p_v^\top q_u + |T_u|^{-z} \sum_{w \in T_u} y_i^\top q_w,$$

where $\beta, z \geq 0$ are parameters to consider the number of similar users and trusted users, respectively. For each trusted user $w \in T_u$, the inner product $y_i^\top q_w$ is treated as the amount of influence made by user w on the target item i .

3.2.4. FST: FSM with social trust

Lastly, we add similar trust influence (from both the perspectives of users and items) to FSM and yield a new ranking prediction approach for FST, given by:

$$\hat{r}_{u,i} = b_i + s|U_{i-u}|^{-\beta} \sum_{v \in U_{i-u}} p_v^\top q_u + (1-s)|I_{u-i}|^{-\alpha} \sum_{j \in I_{u-i}} x_j^\top y_i + |T_u|^{-z} \sum_{w \in T_u} p_w^\top y_i, \quad (2)$$

where $\alpha, \beta, z \geq 0$ are parameters for the number of rated items, similar users and trusted users, respectively; and $s \in [0, 1]$ denotes the relative importance of user similarity. For each trusted user $w \in T_u$, the inner product $p_w^\top y_i$ is regarded as the amount of influence made by user w on the target item i .

Till now, we have added similar types of trust influence to FISM, FUSM and FSM, and obtained new formulas for the ranking prediction. We are aware that other kinds of trust influence can be designed in the future, and possibly distinct ones for different factored similarity models. In present work, we adopt the similar trust influence (1) for the ease and fairness of model comparison; and (2) for the verification of trust value for top-N item recommendation other than rating prediction.

The variables of b, P, Q, X, Y can be learned by minimizing the following objective function:

$$\mathcal{J} = \frac{1}{2} \sum_{u \in C} \sum_{i \in I_u^+, j \in I_u^-} \|(r_{u,i} - r_{u,j}) - (\hat{r}_{u,i} - \hat{r}_{u,j})\|_F^2 + \frac{\lambda}{2} (\|P\|_F^2 + \|Q\|_F^2 + \|X\|_F^2 + \|Y\|_F^2 + \|b\|_F^2), \quad (3)$$

where C is a set of all users, the same λ is used to reduce the model complexity for presentation, and we can easily specify different regularization parameters for each variable in the implementation. Note that Eq. (3) computes loss over all possible item pairs (i, j) of entries in $i \in I_u^+$ and $j \in I_u^-$ for item recommendation, whereas rating prediction attempts to minimize the errors between predictive and real ratings merely on I_u^+ , i.e., $\sum_{u,i} \|r_{u,i} - \hat{r}_{u,i}\|_F^2$. To reduce the computational cost, we randomly sample a number ρ of unrated items from I_u^- for each user u and item i ($\rho = 10$ in our case).

⁴ Other alternative approaches to model trust influence are also studied, e.g., trust as a weighting factor for rated items. Since they generally perform worse than the FIST and FISTa models, we omit them for the sake of clarity.

3.3. Model comparison

Fig. 1 illustrates the relationships among different models discussed and introduced in this paper. Specifically, FIST is obtained by integrating trust influence to the FISM model, while FUST is similarly generated with the incorporation of trust influence based on the FUSM model. Finally, FST is the recommendation model that combines the merits of both the FIST and FUST models. To be specific, if we set $s = 0$ in Eq. (2), FST will be degraded to FIST, while a value of $s = 1$ results in an equivalent model as FUST. When $s \in (0, 1)$, FST can make the best use of both user and item correlations as well as social influence. As we stressed in Section 3.1.1, it is important to provide more aspects of user preferences and to enhance model learning. In this regard, FST can produce better performance than FIST and FUST. The experimental results in Section 4 also confirm the superiority of the FST model.

3.4. Model learning

A popular technique to achieve an optimal solution to Eq. (3) is stochastic gradient descent (SGD). Algorithm 1 provides

Algorithm 1: The learning algorithm of FST.

Input : $\alpha, \beta, z, \rho, \lambda, \eta$ (learning rate)

- 1 Initialize b, X, Y, P, Q with random values in $(0, 0.01)$;
- 2 **while** \mathcal{J} not converged **do**
- 3 **foreach** $u \in C$ **do**
- 4 **foreach** $i \in I_u^+$ **do**
- 5 $Z \leftarrow \text{sample}(\rho, I_u^-)$
- 6 $m_{ki} \leftarrow \sum_{k \in I_{u-i}} x_k, w_{ki} \leftarrow |I_{u-i}|^{-\alpha}$
- 7 $m_{vi} \leftarrow \sum_{v \in U_{i-u}} p_v, w_{vi} \leftarrow |U_{i-u}|^{-\beta}$
- 8 $m_t \leftarrow \sum_{w \in T_u} p_w, w_t \leftarrow |T_u|^{-z}$
- 9 $g \leftarrow 0, h \leftarrow 0, l \leftarrow 0$
- 10 **foreach** $j \in Z$ **do**
- 11 $m_{kj} \leftarrow \sum_{k \in I_{u-j}} x_k, w_{kj} \leftarrow |I_{u-j}|^{-\alpha}$
- 12 $m_{vj} \leftarrow \sum_{v \in U_{j-u}} p_v, w_{vj} \leftarrow |U_{j-u}|^{-\beta}$
- 13 compute $\hat{r}_{u,i}, \hat{r}_{u,j}$ by Equation 2;
- 14 $r_{u,j} \leftarrow 0$
- 15 $e \leftarrow (r_{u,i} - r_{u,j}) - (\hat{r}_{u,i} - \hat{r}_{u,j})$
- 16 $b_i \leftarrow b_i + \eta(e - \lambda b_i)$
- 17 $b_j \leftarrow b_j - \eta(e - \lambda b_j)$
- 18 $q_u \leftarrow q_u - \eta(e(w_{vj}m_{vj} - w_{vi}m_{vi}) + \lambda q_u)$
- 19 $y_i \leftarrow y_i + \eta(e(w_{ki}m_{ki} + w_t m_t) - \lambda y_i)$
- 20 $y_j \leftarrow y_j - \eta(e(w_{kj}m_{kj} + w_t m_t) - \lambda y_j)$
- 21 $g \leftarrow g - ew_{ki}q_u$
- 22 $h \leftarrow h + e(w_{kj}y_j - w_{ki}y_i)$
- 23 $l \leftarrow l + ew_t(y_j - y_i)$
- 24 **foreach** $v \in U_{j-u}$ **do**
- 25 $p_v \leftarrow p_v - \eta(ew_{vj}q_u - \lambda p_v)$
- 26 **foreach** $v \in U_{i-u}$ **do**
- 27 $p_v \leftarrow p_v - \eta(g/\rho + \lambda p_v)$
- 28 **foreach** $k \in I_{u-i}$ **do**
- 29 $x_k \leftarrow x_k - \eta(h/\rho + \lambda x_k)$
- 30 **foreach** $w \in T_u$ **do**
- 31 $p_w \leftarrow p_w - \eta(l/\rho + \lambda p_w)$
- 32 **return** b, P, Q, X, Y ;

the detailed procedure and the gradient descent rules to update the variables of our approach FST. Specifically, all the variables

Table 1

Specification of the used data sets, where the density is computed by $\text{Density} = \#Ratings / (\#Users \times \#Items)$.

Data Set	#Users	#Items	#Ratings	#Trust	Density
Epinions	40,163	139,738	664,824	487,183	0.01%
Ciao	7,375	99,746	278,483	111,781	0.04%
FilmTrust	1,508	2,071	35,497	1,853	1.14%

are initialized with random small values in $(0, 0.01)$ (line 1). For each iteration (lines 3-31), we randomly sample a set Z of negative examples with the sampling factor ρ (line 5) to train the model. The variables are updated according to the SGD rules (lines 16-31). This process is repeated until the loss value has converged or the maximum number of iterations has been reached. Lastly, all the learned variables will be returned as output (line 32). Since the other algorithms to train FUSM, FSM, FISTa and FUST basically follow similar procedures, we omit the details for simplicity. Note that the complexity and scalability of our algorithms will be deferred later and discussed in Section 4.5.

4. Evaluation and results

In this section, we will conduct a series of experiments on three real-world data sets to investigate: (1) the usefulness and sensitivity of parameters α, β, z, s in Eq. (2); and (2) the effectiveness of our approaches in comparison with other counterparts in terms of precision.

4.1. Experimental setup

4.1.1. Data sets

Three real-world data sets are used in our experiments, namely Epinions,⁵ Ciao⁶ and FilmTrust.⁷ Both Epinions and Ciao are review sharing websites where users can write textual reviews and issue numerical ratings on a variety of products; and FilmTrust is a movie sharing website where users share movie ratings with their friends. All these data sets include both user-item ratings and user-user social trust connections. With the built-in support of the concept of social trust, these datasets are often used as benchmarks for many trust-aware recommender systems. The specification of the three data sets is illustrated in Table 1. All the positive rating values are preprocessed and binarized to 1 (and 0 if otherwise), indicating that a user has consumed or rated a specific item (i.e., implicit feedback). The table shows that all the data sets are very sparse and distinct in nature.

4.1.2. Comparison methods

The following methods (11 in total) are used for comparison which are designed for top-N item recommendation based on implicit feedback. The approaches include:

- **MostPop** is the baseline approach that computes the ranking score of an item by its popularity, i.e., how many times the item is rated or consumed by other users;
- **BPR** is proposed by Rendle et al. [20] in which the pairwise assumption is adopted for item ranking. BPR is a state-of-the-art method for top-N item recommendation.
- **GBPR** is proposed by Pan and Chen [17] in which the BPR assumption is relaxed by the group preference;
- **SBPR** is proposed by Zhao et al. [25] in which social connections are used to strengthen the BPR assumption. Since

⁵ http://www.trustlet.org/extended_epinions.html

⁶ <http://www.cse.msu.edu/~tangjili/trust.html>

⁷ <http://www.librec.net/datasets.html>

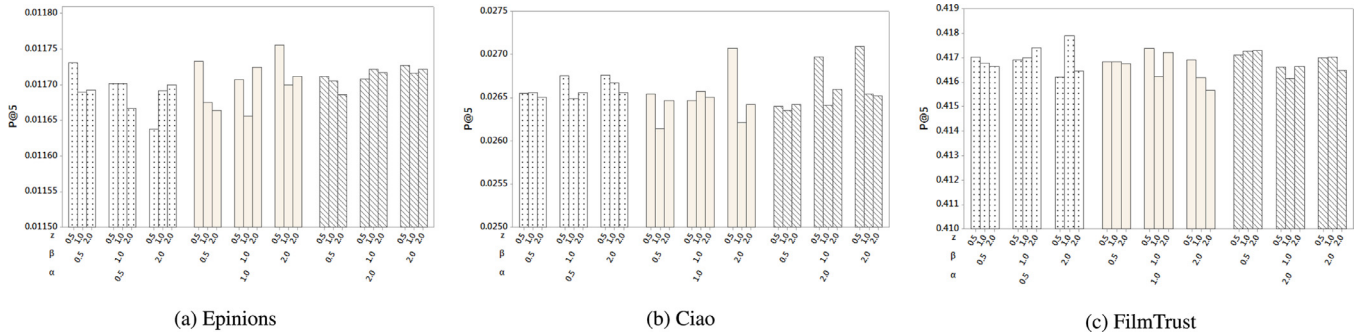


Fig. 2. The effect of parameters α , β , z on our approach FST in terms of precision at 5 (i.e., P@5)

SBPR performs better than MR-BPR [11], we will not compare with the latter method. Given that GBPR and SBPR incorporate additional information to improve recommendation as our approach does, it is necessary to compare with them.

- **FISM** is proposed by Kabbur et al. [9] in which a ranking score is composed of item similarities. Since our approaches are inspired by FISM, it is necessary to study if our approaches can (and to what extent) reach improvements by integrating social trust.
- **FUSM, FSM, FIST, FISTa, FUST, FST** are a set of our approaches, where FST is our highly suggested approach.

Since we target the recommendation task of top-N item recommendation, we opt not to compare with other trust-based models for rating prediction (e.g., RSTE [13], SoReg [14], FM [19]). For other ranking-based models, we skip the ones that work better with explicit ratings, including Trust-CF [7], WRMF [6] and SLIM [16]. Our work is suitable for the case of implicit feedback.

4.1.3. Evaluation metrics

The 5-fold cross validation approach is adopted. That is, a data set is randomly split into five folds, and for each iteration, four of which are used as the training set and the rest as the test set. The average results of five executions are reported as the final performance. We adopt two popular ranking metrics to evaluate recommendation performance, namely precision and F1-measure at N (i.e., P@N, F1@N), where the cutoff N is chosen in {5, 10}, i.e., the number of recommended items.

$$P@N = \frac{1}{|U'|} \sum_{u \in U'} \frac{|R_N(u) \cap I'_u|}{N}$$

$$R@N = \frac{1}{|U'|} \sum_{u \in U'} \frac{|R_N(u) \cap I'_u|}{|I'_u|}$$

$$F1@N = \frac{2 \cdot P@N \cdot R@N}{P@N + R@N}$$

where I'_u is the set of items rated by user u , U' is the set of users in the test set, and $R@N$ denotes the measurement of recall at N. F1-measure represents a trade-off between ranking precision and recall. Higher values of P@N and F1@N indicate better top-N recommendation performance.⁸ Note that evaluation metrics like mean absolute error (MAE) and root mean square error (RMSE) are not applicable in this work as they are used for the task of rating prediction, which is a distinct recommendation task from top-N item recommendation.

⁸ We have also considered other ranking metrics, such as NDCG, MRR and MAP. However, they generally vary less significantly than precision and F1-measure, and thus are not present in this article.

4.1.4. Parameter settings

Parameter settings in our experiments are either obtained by empirical results or suggested by literature works. Specifically, the number of latent factors is set to $d = 10$, the same setting as Yang et al. [23] and Zhao et al. [25]. For GBPR, we fix the size of group users to 5 and adjust the parameter $\rho \in [0, 1]$ which is the tradeoff between group and individual preferences. For SBPR, there are no additional parameters to tune with. For all the BPR-based approaches, we employ a uniform sampling strategy to select unrated items for model learning, as suggested by Zhao et al. [25]. For all the factored similarity models, the values of parameters α , β , z are searched in a small set of typical values, i.e., {0.5, 1, 2}, and the sampling factor ρ (see Algorithm 1) is fixed by 10 since Kabbur et al. [9] suggest that a small value in the range [3, 15] will suffice for FISM. Regarding FISTa, we select the top 20 most popular items rated by trust users, i.e., $N = 20$. For all the matrix factorization-based approaches, we employ grid search in {0.000001, 0.00001, 0.0001, 0.01, 0.1} to find out the optimal settings for regularization parameters λ .

4.2. Effect of parameters α , β , z

The three parameters control the impact of item similarities, user similarities and social influence on the ranking prediction (see Eq. (2)). Some of these parameters are also used in other approaches FIST, FISTa, FUST and FSM. To save space, we focus on the effect of these parameters on our approach FST across all the data sets. Specifically, in the experiments we tune the values of each parameter in a small set: {0.5, 1, 2} while fixing the values of parameter s to be 0.5. The results are illustrated in Figs. 2 and 3 in terms of P@5 and P@10 respectively, where the x-axis accommodates three rows of settings corresponding to the values of parameters α , β and z , respectively. A number of observations can be noted from the results. The most straightforward one is that different groups of settings (α , β , z) produce distinct results, and thus it is necessary to tune a proper combination for them. The best results are often observed when $\beta = 2$ and $\alpha, z < 2$. It indicates that the impact of item similarities and trust influence should be more weighted while that of user similarities should be less counted for top-N item recommendation. Note that even better performance can be achieved if not restricting α , β , z in the small set {0.5, 1, 2}.

4.3. Effect of parameters

Parameter s in Eq. (2) controls the impact of similar users on the ranking prediction. We adopt the best values of parameters α , β , z reported by the previous subsection when the number of latent factors is fixed at 10, and then vary the values of parameter s in the range of [0, 1] with step 0.1. The results are illustrated in Fig. 4, where the best values for Epinions, Ciao and FilmTrust are

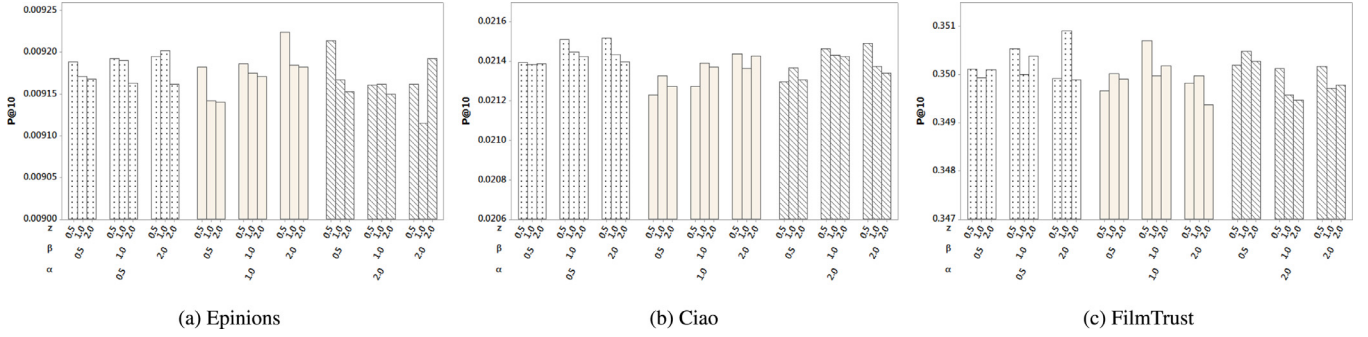


Fig. 3. The effect of parameters α , β , z on our approach FST in terms of precision at 10 (i.e., P@10)

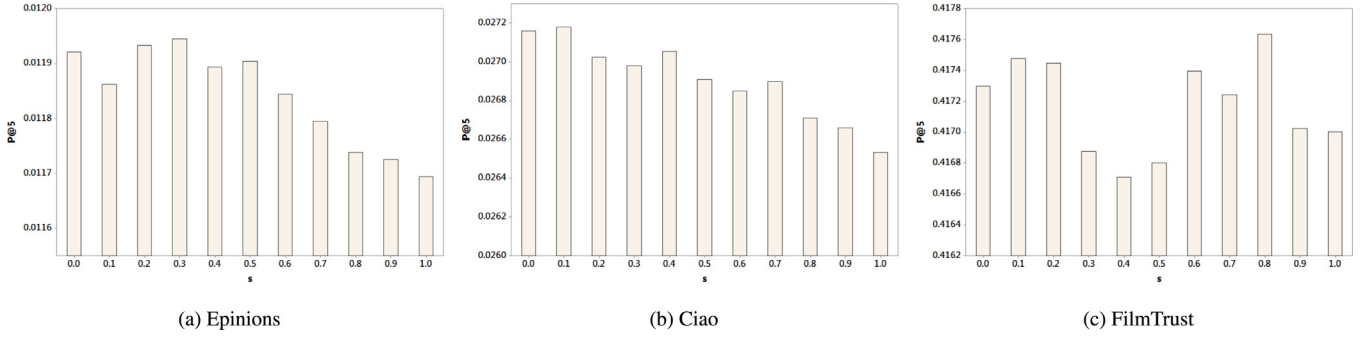


Fig. 4. The effect of parameter s on our approach FST in terms of precision at 5 (i.e., P@5)

Table 2

The experimental results on three data sets in terms of **precision**, where the best performance of the first five approaches is highlighted in bold as well as that of our approaches for a clearer comparison. Note that the values in parentheses indicate the scale of the corresponding values.

Data Set	d	MostPop	BPR	GBPR	SBPR	FISM	FUSM	FSM	FIST	FISTa	FUST	FST
Epinions	5	0.1169	0.09263	0.09353	0.05892	0.1147	0.1134	0.1173	0.1175	0.1170	0.1171	0.1179
	($\times 10^{-1}$) 10	0.09171	0.07690	0.07560	0.05529	0.09102	0.08979	0.09173	0.09183	0.09182	0.08979	0.09187
Ciao	5	0.2677	0.1855	0.2228	0.1206	0.2704	0.2656	0.2670	0.2706	0.2680	0.2703	0.2741
	($\times 10^{-1}$) 10	0.2142	0.1671	0.1827	0.1007	0.2141	0.2137	0.2139	0.2142	0.2144	0.2169	0.2174
FilmTrust	5	0.4170	0.4147	0.4124	0.4058	0.4171	0.4179	0.4180	0.4174	0.4181	0.4182	0.4191
	10	0.3503	0.3494	0.3470	0.3433	0.3503	0.3510	0.3503	0.3503	0.3510	0.3510	0.3514

Table 3

The experimental results on three data sets in terms of **F1-measure**, i.e., the balance between precision and recall, where the values in parentheses indicate the scale of the corresponding values.

Data Set	d	MostPop	BPR	GBPR	SBPR	FISM	FUSM	FSM	FIST	FISTa	FUST	FST
Epinions	5	0.1298	0.1100	0.1103	0.0651	0.1307	0.1290	0.1298	0.1304	0.1299	0.1299	0.1330
	($\times 10^{-1}$) 10	0.1305	0.1134	0.1111	0.07830	0.1315	0.1287	0.1303	0.1304	0.1305	0.1304	0.1328
Ciao	5	0.2436	0.1729	0.2063	0.08536	0.2495	0.2418	0.2437	0.2487	0.2458	0.2488	0.2523
	($\times 10^{-1}$) 10	0.2662	0.2116	0.2292	0.1070	0.2687	0.2664	0.2663	0.2676	0.2680	0.2719	0.2720
FilmTrust	5	0.4095	0.4095	0.4051	0.3972	0.4087	0.4087	0.4093	0.4099	0.4094	0.4092	0.4099
	10	0.4518	0.4505	0.4458	0.4416	0.4516	0.4516	0.4507	0.4517	0.4523	0.4521	0.4521

0.3, 0.1 and 0.8, respectively. A straightforward conclusion is that similar users may have distinct impacts on different applications. We may also conclude that trust are more important in trust networks than in trust-alike networks [5]. In summary, the best performance can be achieved when a proper portion of similar users and trust users is adopted in our approach.

4.4. Comparison with other methods

Tables 2 and 3 present the recommendation performance of all the comparison methods across the three data sets in terms

of P@N and F1@N, respectively. The best performance of the first five methods is bolded as well as that of our approaches for comparison purposes. Generally, our approach FST obtains the best performance in comparison with all the other methods.

First, the most basic, non-personalized approach MostPop is able to achieve comparable, or in some cases even the best (bolded) results among all the baselines. It may imply that users tend to consume popular items to some extent.

Second, BPR-based approaches do not present good performance even compared with the MostPop approach. Zhao et al. [25] also reported that MostPop outperformed BPR in three (out of

four) social rating networks, whereas Pan and Chen [17] obtained the opposite results in four different data sets without social connections. One indication drawn from these reported results is the necessity to revisit the assumption of the BPR method. That is, to what extent a rated item will be preferred to an unrated one, and how it differs in different kinds of data sets. This may help explain why SBPR fails to work in our tested data sets whereas Zhao et al. [25] claimed a success in their experiments. We attribute it to the ill-fitness of their assumption in our data sets.

Third, purely factored similarity models (namely FISM, FUSM and FSM) generally perform better than BPR-based approaches, indicating their usefulness for item recommendation. Specifically, FISM by Kabbur et al. [9] achieves better performance than other baselines. Although factored user similarity model (i.e., FUSM) works worse than FISM, combining both user similarities and item similarities (i.e., FSM) can help gain better ranking precision.

Lastly, social trust is noted to impose important influence on the ranking performance by comparing the performance of method pairs (FISM, FIST), (FUSM, FUST) and (FSM, FST). Among the two different manners to model trust influence, FIST works more reliably and effectively than FISTa in that the former approach consistently outperforms FISM whereas the latter does not. This may be due to the fact that FIST models trust influence explicitly by the inner products of user and item vectors whereas FISTa does implicitly via the items rated by trusted users. Most importantly and consistently, our approach FST achieves the best performance in comparison with all the other approaches. Although the relative improvements are small, it may be explained by the fact that the parameters α , β , z are only tuned in a small set. Koren [10] has also justified that even relatively small improvements may lead to big difference in practice.

4.5. Complexity and scalability

The computational time of our model FST is mainly taken in two phases: (1) model training where the gradients and update rules are computed for all the variables (see lines 16-31, Algorithm 1). For each iteration, the overall computational cost in Algorithm 1 is around $O(nb)$, where n is the number of ratings in the training set and b is the average number of users rating an item. (2) model test where we compute all the ranking scores for a large volume of candidate items, and identify the user set U_{i-u} , item sets I_{u-i} , I_{u-i} and trust set T_u for each ranking score $\hat{r}_{u,i}$ (see Eq. (2)). To resolve these two issues, Kabbur et al. [9] make the following suggestion: (1) parallelizing the computation of gradients and the update rules of SGD (for training); (2) thresholding the computed ranking score to speedup ranking items (for test). In addition, we find that caching techniques are useful to help efficiently (and repeatedly) retrieve (the same) user or item sets, and thus greatly reduce the test time. By doing so, our approaches can be scaled up to larger data sets. Specifically, our experiments are run on a server with 32 Genuine Intel(R) CPUs (2.6 GHz), 256G memory. For each execution, the average time for FST on FilmTrust, Ciao and Epinions is around 5, 20 and 47 min, respectively.

5. Conclusion and future work

This article proposed three factored similarity models with the incorporation of social trust influence for item recommendation based on implicit user feedback. Both user-user similarities and item-item similarities were factored from the proposed matrix factorization models. In addition, we also incorporated the influence of social trust when estimating a ranking score for an active user on a target item. We conducted experiments on three real-world data sets, and demonstrated that our approach performed the best in comparison with other counterparts. Further, the impact of item

similarities and trust influence should be more weighted than that of user similarities in order to achieve the best performance.

For future work, we intend to incorporate more types of trust influence to the proposed factored similarity models, and investigate the impact of trust on the assumption of Bayesian personalized ranking. In addition, it is also interesting to consider the influence of distrust, which may be distinct from social trust.

Acknowledgments

This work is supported by the National Science Foundation for Distinguished Young Scholars of China under Grant No. 61225012 and No. 71325002; the National Natural Science Foundation of China under Grant No. 61572123 and No. 61472073.

References

- [1] H. Fang, Y. Bao, J. Zhang, Leveraging decomposed trust in probabilistic matrix factorization for effective recommendation, in: Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI), 2014, pp. 30–36.
- [2] J. Golbeck, Generating predictive movie recommendations from trust in social networks, in: Proceedings of the 4th International Conference on Trust Management (iTrust), 2006, pp. 93–104.
- [3] G. Guo, J. Zhang, D. Thalmann, Merging trust in collaborative filtering to alleviate data sparsity and cold start, *Knowl.-Based Syst. (KBS)* 57 (2014) 57–68.
- [4] G. Guo, J. Zhang, N. Yorke-Smith, Leveraging multiviews of trust and similarity to enhance clustering-based recommender systems, *Knowl.-Based Syst. (KBS)* 74 (2015) 14–27.
- [5] G. Guo, J. Zhang, N. Yorke-Smith, TrustSVD: collaborative filtering with both the explicit and implicit influence of user trust and of item ratings, in: Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI), 2015, pp. 123–129.
- [6] Y. Hu, Y. Koren, C. Volinsky, Collaborative filtering for implicit feedback datasets, in: Proceedings of the 8th IEEE International Conference on Data Mining (ICDM), 2008, pp. 263–272.
- [7] M. Jamali, M. Ester, Using a trust network to improve top-n recommendation, in: Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys), 2009, pp. 181–188.
- [8] M. Jamali, M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks, in: Proceedings of the 4th ACM Conference on Recommender Systems (RecSys), 2010, pp. 135–142.
- [9] S. Kabbur, X. Ning, G. Karypis, FISM: factored item similarity models for top-n recommender systems, in: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2013, pp. 659–667.
- [10] Y. Koren, Factorization meets the neighborhood: a multifaceted collaborative filtering model, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2008, pp. 426–434.
- [11] A. Krohn-Grimberghe, L. Drumond, C. Freudenthaler, L. Schmidt-Thieme, Multi-relational matrix factorization using bayesian personalized ranking for social network data, in: Proceedings of the 5th ACM International Conference on Web Search and Data Mining (WSDM), 2012, pp. 173–182.
- [12] H. Ma, On measuring social friend interest similarities in recommender systems, in: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR), 2014, pp. 465–474.
- [13] H. Ma, I. King, M. Lyu, Learning to recommend with social trust ensemble, in: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), 2009, pp. 203–210.
- [14] H. Ma, D. Zhou, C. Liu, M. Lyu, I. King, Recommender systems with social regularization, in: Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM), 2011, pp. 287–296.
- [15] T.T.S. Nguyen, H.Y. Lu, J. Lu, Web-page recommendation based on web usage and domain knowledge, *IEEE Trans. Knowl. Data Eng. (TKDE)* 26 (10) (2014) 2574–2587.
- [16] X. Ning, G. Karypis, SLIM: sparse linear methods for top-n recommender systems, in: Proceedings of the 11th International Conference on Data Mining (ICDM), 2011, pp. 497–506.
- [17] W. Pan, L. Chen, GBPR: group preference based bayesian personalized ranking for one-class collaborative filtering, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013, pp. 2691–2697.
- [18] M.C. Pham, Y. Cao, R. Klamka, M. Jarke, A clustering approach for collaborative filtering recommendation using social network analysis, *J. Universal Comput. Sci.* 17 (4) (2011) 583–604.
- [19] S. Rendle, Factorization machines with libfm, *ACM Trans. Intell. Syst. Technol. (TIIST)* 3 (3) (2012) 57.
- [20] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, BPR: Bayesian personalized ranking from implicit feedback, in: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI), 2009, pp. 452–461.
- [21] Q. Shambour, J. Lu, A hybrid trust-enhanced collaborative filtering recommendation approach for personalized government-to-business e-services, *Int. J. Intell. Syst. 26* (9) (2011) 814–843.

- [22] B. Yang, Y. Lei, D. Liu, J. Liu, Social collaborative filtering by trust, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013, pp. 2747–2753.
- [23] X. Yang, H. Steck, Y. Guo, Y. Liu, On top-k recommendation using social networks, in: Proceedings of the 6th ACM conference on Recommender systems (RecSys), 2012, pp. 67–74.
- [24] W. Yao, J. He, G. Huang, Y. Zhang, Modeling dual role preferences for trust-aware recommendation, in: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR), 2014, pp. 975–978.
- [25] T. Zhao, J. McAuley, I. King, Leveraging social connections to improve personalized ranking for collaborative filtering, in: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM), 2014, pp. 261–270.