Neurocomputing 341 (2019) 1–9

Contents lists available at ScienceDirect

Neurocomputing

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

Social recommendation based on users' attention and preference

Jiawei Chen, Can Wang*, Qihao Shi, Yan Feng, Chun Chen

ZJU-LianlianPay Joint Research Center College of Computer Science, Zhejiang University, Hangzhou 310027, China

ARTICLE INFO

Article history:
Received 30 May 2018
Revised 11 February 2019
Accepted 21 February 2019
Available online 4 March 2019

Communicated by Dr. Guan Ziyu

Keywords:
Social recommendation
Preference
Attention
Graphical model

ABSTRACT

Attention is the behavioral and cognitive process of selectively concentrating on small fraction of information while ignoring other perceivable information. Thus, user's attention will influence his decision on the consumption and consequently the performance of recommender systems. When presented a long list of recommendations, users will focus more on the items that draw their attention and skip others even if they suit their tastes. However, existing works in social recommendation mainly generate recommendations based on users' preference and overlook the importance of users' attention. To overcome this limitation, in this paper we propose a probabilistic model HTPF that considers both users' attention and preference in social recommendation. Besides, by analyzing four real-world social recommendation datasets, we observe that social relations have more explicit impact on users' attention than on their item rating values. That is, user's attention on items are easily affected by his social relations while his rating values seem somehow indifferent to social influence. It will be more effective to use social network information to infer users' attention rather than users' preference. Thus, different from existing methods that model the social influence on user's rating values, HTPF focuses on exploiting social influence on user's attention. Further, a scalable variational inference algorithm has been developed for our model to predict how the specific items will draw users' attention and how the items will suit users' tastes. Then HTPF generates the recommendation results based on the weighted combination of these two aspects. Extensive experiments conducted on large real-world datasets validate the superiority of our method over state-of-the-art methods.

1. Introduction

Social recommendation has attracted considerable research attention in recent years as it effectively addresses data sparsity and cold start issues in traditional recommendation algorithms. One major assumption in social recommendation is that users' behaviors in a social network are heavily affected by their social relations. Thus it is possible to infer users' preference via their social relations for more precise recommendation. However, existing methods in social recommendation mainly generate recommendations based on users' preference and overlook the importance of users' attention. Besides, we observe that the influence of social relations dwells more on users' attention than on their rating values in four real-world datasets.

Attention is the behavioral and cognitive process of selectively concentrating on small fraction of information, while ignoring other perceivable information [1–3]. Therefore our attention is selective and limited by its nature [4–6]. Users tend to capture information that supports their biased pre-existing views from their prior experience and social network [7,8]. Thus it is not surprising when presented a list of recommendations, we will focus more on the items that draw our attention and skip some other items even if they suit our tastes. This psycho-social effect motivates us to model users' attention in social recommendation for better recommendation accuracy.

To better understand the characteristics of users' attention in social recommendation, we conduct empirical analyses based on four large real-world datasets in this paper. Specifically, in recommendation, we explore users' attention based on their *attention behavior*, i.e. which items the user have spent mental effort on consuming and rating. An important conclusion can be drawn: Users' *attention behavior* is heavily affected by their trust relations, while their rating values stay relatively indifferent to them. Thus, on the one hand, social information can be used more effectively to infer users' attention than users' preference. On the other hand, each of us belongs to some content-shared communities and our attention will inevitably be influenced by our social relations [2,9],

* Corresponding author.
E-mail addresses: sleepyhunt@zju.edu.cn (J. Chen), wcan@zju.edu.cn (C. Wang), shiqihao321@zju.edu.cn (Q. Shi), fengyan@zju.edu.cn (Y. Feng), chenc@zju.edu.cn (C. Chen).

https://doi.org/10.1016/j.neucom.2019.02.045
0925-2312/

which makes that our attention is not coincident with our preference. Thus, just considering one aspect for recommendation may not produce effective recommendation results. Specifically, when presented a list of recommendations, we usually focus more on the items that draw our attention, and are unwilling to spend effort to learn the information about the rest. This means that the recommended items just suiting our tastes may be ignored because of our limited and selective attention [4,5]. Meanwhile, the items which just draw our attention may not be accepted, since when we make the decisions on the consumption, we usually observe ourselves and judge whether we will like them. Therefore, to improve the recommendation accuracy, in modern recommender systems, how items draw users' attention and how items suit users' tastes should be both taken into consideration.

Based on the above analyses, we propose a new social recommendation model HTPF that explicitly considers both users' attention and preference for better recommendation accuracy. Besides, different from existing methods that model social influence on users' rating values, HTPF differs from these methods in that it uses social information to better infer users' attention, which is more suspectable to social network. Considering Poisson factorization [10] is particularly efficient on sparse data and usually achieves better performance than traditional methods (such as PMF [11], similar-based methods [12], random walk [13]), HTPF employs an integrated latent factor model based on Poisson factorization [10]: HTPF fuses the Poisson factorization of *attention behavior* information and trust information by sharing common latent vectors of users' attention, since users who have similar social relations tend to pay attention to similar items. At the same time, HTPF fuses the Poisson factorization of *attention behavior* information and rating information by sharing common latent vectors of items, because of constant item attributes. A scalable variational inference algorithm for our HTPF model has been developed to predict how the specific items draw users' attention and how the items suit users' tastes. Targeting at high recommendation accuracy, HTPF generates recommendations based on the weighted combination of these two aspects. We also conduct extensive experiments and demonstrate that by combining users' attention and preference our method outperforms state-of-the-art social recommendation methods.

It is worthwhile to highlight the following contributions of our work:

- We introduce user's attention in the social recommendation and uncover that the influence of trust relations dwells more on users' attention than on their preference.
- We propose a novel probabilistic model HTPF that explicitly models both user's attention and preference for social recommendation. Besides, We develop an efficient inference method for HTPF.
- Our experiments on real-world datasets clearly demonstrate the superiority of our method over existing social recommendation methods.

Before proceeding further, we now formalize the problem definition. Suppose we have a recommender system with users set U (including n users) and items set I (including m items). Rating information is represented as $n \times m$ matrix R whose entries (R_{ui}) denote the rating value of item i given by user u. *Attention behavior* information indicates observed *attention behaviors* that users are willing to spend mental effort on consuming and rating some specific items [1,14], represented as $n \times m$ matrix Ω. For each entry in Ω, $\Omega_{ui} = 1$ denotes that user u have consumed and rated item i, and $\Omega_{ui} = 0$ means not. Trust information indicates the trust relations between users, represented as $n \times n$ matrix G. For each entry in G, $G_{uv} = 1$ denotes user u trusts user v, $G_{uv} = 0$ denotes not (Note that we can treat undirected friend relations as bi-directed trust relations in this paper). The task of social recommendation is to give each user an item list that will be accepted with pleasure based on these information.

The rest of this paper is organized as follows. we briefly review related works in Section 2. The social empirical analyses are conducted in Section 3. In Section 4, we introduce the details of our proposed model HTPF. The experiment results and discussions are presented in Section 5. Finally, we conclude the paper and present some directions for future work in section 6.

2. Related works

With the exponential growth of information generated on consumer review websites and e-commerce websites, recommender systems are drawing more attention from both academia and industry [15–21]. For the system with explicit feedback (numerical ratings), substantial works have been done about collaborative filtering (CF) model for its accuracy and scalability during the past two decades [22–26]. To address the limitation of the traditional CF methods such as cold-start problem, substantial works has been done about social recommendation model. Here we briefly review the most related works, more details can refer to some excellent surveys and monographs [15,27].

Recent works in social recommendation integrated trust information into rating information by modeling social influence on user's rating values in different ways. Some methods assume that connected users will share similar preference [28–30]. For example, Sorec [31], TrustMF [32], PSLF [33], which jointly factorize rating matrix and trust(social) matrix by sharing a common latent user space; Soreg [28], SocialMF [34] and circle-based methods [29] extends classical PMF (probabilistic matrix factorization[35]) model by constraining user's latent preference close to the average of his trusted friends.

In some other methods [36–39], users' ratings are considered as synthetic results of their preference and social influence. For example, RSTE [36] was proposed by modeling users' rating values based on their own taste and friends' preference, where a specific parameter was used to control the effect of two parts; Xin et.al [40] notice the different roles of strong ties and weak ties in social recommendation, thus they extend RSTE to PTPMF by incorporating the distinction of strong and weak ties for improving recommendation performance; SPF [38] also extends RSTE by integrating social influence into the poisson factorization [10] of users' rating matrix; Bao et.al [39] notice the gap between users' trust and their preference-similarity. Thus, they consider social trust as a multi-faceted phenomenon and attempt to decompose trust into four aspects for better rating prediction; Guo [41] extends SVD++ [23] by further incorporating both the explicit and implicit influence of friends on users' ratings in their TrustSVD model. On the one hand, the preference of trustees(friends) has been considered to generate user's rating values. On the other hand, they constrained that the user-specific vectors decomposed from the rating matrix are the same as those decomposed from the trust(social) matrix.

However, these methods ignore the role of user's attention in social recommendation. Many psycho-social literatures [4,6] suggested the importance of user's attention and our analyses in Section 3 show that the influence of the social relations on user's ratings is smaller than their influence on the user's attention. Thus, different from existing methods, the HTPF proposed in this paper captures user's attention in the social network, which can effectively bridge the trust-preference gap and achieve better performance than state-of-the-art social recommendation methods.

Table 1
Statistics of the four datasets.

Features	Epinions	Ciao	Flixster	Douban
Number of users	49,289	7,375	147,612	129,490
Number of items	139,738	106,797	48,794	58,541
Number of ratings	664,824	280,391	8,196,077	16,830,839
Density of ratings	0.0097%	0.0356%	0.1138%	0.2220%
Number of trusts	487,183	111,781	2,430,868	1,692,952
Density of trusts	0.0201%	0.2055%	0.0112%	0.0101%
Connection type	Directed	Directed	Undirected	Undirected

3. Analyses of social influence on datasets

In this section, we conduct empirical analyses on four large real-world datasets: Epinions[1], Ciao[2], Flixster[3]and Douban[4]. The four datasets contain several kinds of information. The rating information denotes the rating values that the user gave to some items. The rating values in Epinions, Ciao and Douban are integers from 1 to 5, while those in Flixster are real values from 0.5 to 5.0 with step 0.5; Also, there exist two kinds of social relations between users in these datasets: The trust relations are directed in Epinions and Ciao, while the friend relations are undirected in Flixster and Douban. We treat undirected friend relations as bi-directed trust relations in this paper since friends trust each other; Besides, the four datasets contain the information about user's attention. Considering the phenomenon that users are willing to spend mental effort on consuming and rating some specific items, we could learn user's attention from the information about which items the user have rated [1,14]. In summary, Epinions, Ciao, Flixster and Douban are ideal datasets that have been used widely in recent works for analyses and experiments on social recommendation. The datasets statistics are illustrated in Table 1.

In order to show the influence of social relations on users' attention and their preference, two empirical analyses were conducted as follows: (1) We calculate the average rating similarity and attention (behavior) similarity for each truster-trustee pair, where rating similarity denotes the Pearson correlation coefficient [42] between their rating values of common items and attention (behavior) similarity denotes the average Jaccard coefficient between the sets of items that have been rated respectively by them (similarity between users' attention behavior). In comparison, we calculate the average rating similarity and attention (behavior) similarity between arbitrary two users. The results are shown in Fig. 1(a)(b). (2) We divide pairs of users into 8 groups according to the number of their common trustees. We then calculate the average rating similarity and the average attention (behavior) similarity between every user pairs in each group. The results are shown in Fig. 1(c)(d). An important observation is concluded from these results.

Observation. Comparing with user's rating values, user's attention is more susceptible to social network.

As showed in Fig. 1(a)(b), socially connected users tend to pay attention to more common items than average user pairs, but their rating values do not exhibit too much more similarity than average user pairs. Meanwhile, as showed in Fig. 1(c)(d), as the number of common trustees increases, users tend to have more commonality in their attention behavior while their rating values remain mostly unaffected by those more similar social relations. In content-sharing social networks, items purchased by trustees are more likely brought to our attention and stored in episodic memory [43]. Thus, user's attention behavior is heavily affected by social relations while his rating values stay relatively indifferent to the contents shared by his trustees. This observation motivates us to learn user's attention instead of his preference from social network.

Besides, though we may wish to see ourselves as unique individuals, each of us belongs to some content-shared communities and our attention will be influenced by our friends/trustees [2,9]. Considering the different sensitivity of user's attention and his preference to social relations as shown in Fig. 1, we can conclude that our attention is not coincident with our preference. On the one hand, when presented a list of recommendations, we usually focus on the items that draw our attention, and are unwilling to spend effort to learn the information about the rest. This means that the recommended items just suiting our taste may be skipped up because of our limited and selective attention [4,5]. On the other hand, the recommended item, which draws our attention, may not be accepted because we dislike it. To improve recommendation accuracy, we need to combine both users' attention and their preference in social recommendation.

4. Model

In this section, we present our proposed model HTPF and its inference algorithm.

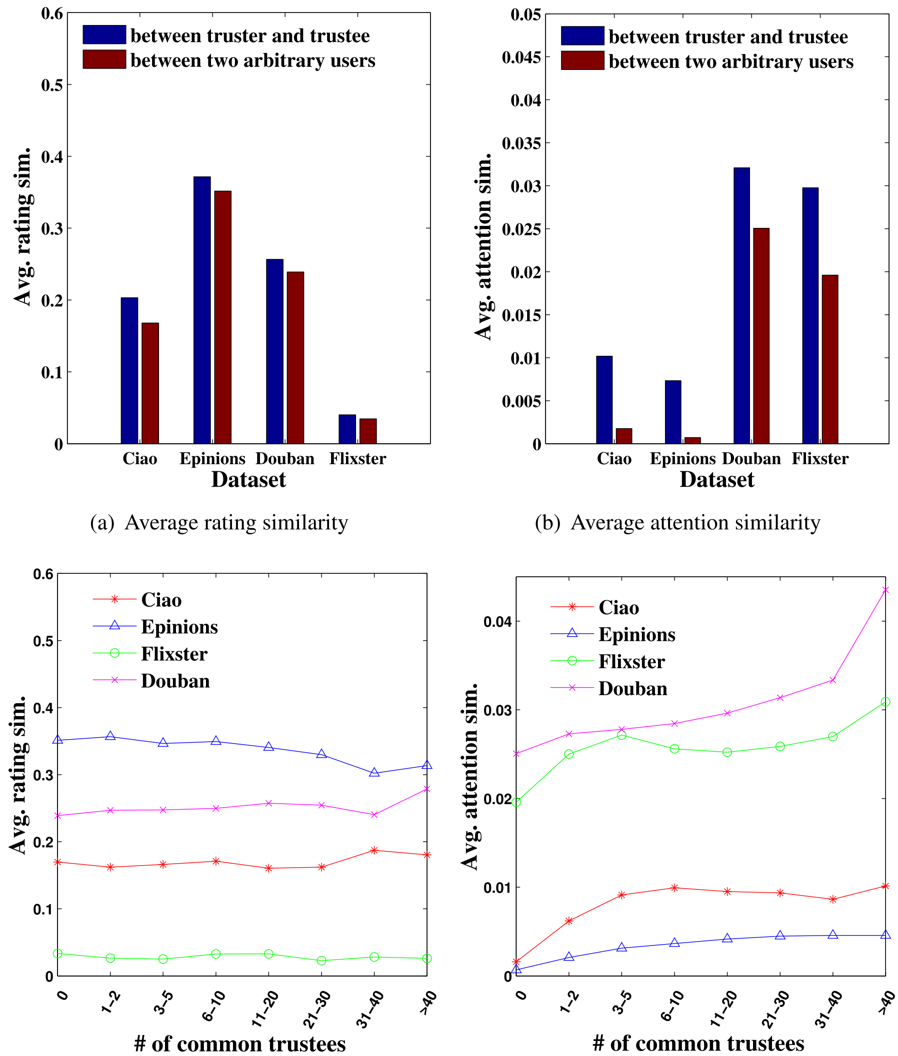
4.1. Hierarchical Trust-based Poisson Factorization (HTPF)

The key to a good social recommendation model is how it exploits trust information. Different from existing methods, we consider trust information as a helpful complemental information to capture users' attention instead of users' preference, since user's attention is important in recommendation as suggested by many psycho-social literatures and is more susceptible to social network as shown in Fig. 1. That is, the trust relations in a large degree indicate what information users pay attention to in the social network. Thus the observed trust relations were modeled as follows:

Considering the advantages of poisson factorization on sparse data [10], we model trust relations as poisson distribution, parameterized by the inner product of the K-dimensional vector of latent attention of user u (g_u) and the K-dimensional vector of latent features of trustee v (h_v), $G_{uv} = poisson(g_u^\top h_v)$. Where g_u represents "what topics or features user u pays attention to" and h_v represents "the topics of the information shared by user v". In the social websites, we tend to trust or follow the users whose shared information draws our attention. Thus, we model trust relations G_{uv} based on the $g_u h_v$, which captures "how users pay attention to the shared information". Similarly, users' attention behavior information and rating information can be modeled by poisson factorization too based on "how users pay attention to the items" and "how they like the items", $\Omega_{ui} = poisson(g_u^\top \beta_i)$, $R_{ui} = poisson(\theta_u^\top \beta_i)$. Where β_i denotes the K-dimensional vector of latent attribute of item i, θ_u denotes the K-dimensional vector of latent preference of user u.

Besides, we place hierarchical gamma priors on these latent variables (g_u, h_v, θ_u, β_i) to control the average size of representation, which can capture the diversity of users and items: (1) users' activity γ_u, where some users who are active tend to pay attention to more items and trustees than others; (2) users' rating habit α_u, where some users are generous and tend to give relatively high ratings than those users who are critical about their ratings; (3) influence power η_u, where some users tend to have more fans than others; (4) items' quality σ_i, where the items with higher quality tend to be more popular and be rated with higher scores than others. HTPF can capture such heterogeneity across users and items,

[1] http://www.trustlet.org/epinions
[2] http://www.public.asu.edu/~jtang20/datasetcode/truststudy
[3] http://www.cs.ubc.ca/~jamalim/datasets/
[4] https://www.cse.cuhk.edu.hk/irwin.king.new/pub/data/douban



(a) Average rating similarity (b) Average attention similarity

(c) Average rating similarity (d) Average attention similarity

Fig. 1. The left two plots (a)(b) show the average rating/attention similarity between socially connected users and arbitrary two users; The right two plots (c)(d) show the users' rating/attention similarity with their varying common trustees.

which was suggested an important factor for the recommendation model [24].

In all, our method has following generative process and the graphical model shows in Fig. 2:

1. For each user $u \in U$:
   - Sample activity: $\gamma_u \sim gamma(a_\gamma^s, a_\gamma^r)$.
   - Sample rating habit: $\alpha_u \sim gamma(a_\alpha^s, a_\alpha^r)$.
   - Sample influence power: $\eta_u \sim gamma(a_\eta^s, a_\eta^r)$.
   - For each component $k (k = 1, 2, \ldots, K)$:
   - Sample latent attention: $g_{uk} \sim gamma(a_g, \gamma_u)$.
   - Sample latent preference: $\theta_{uk} \sim gamma(a_\theta, \alpha_u)$.
   - Sample features of trustee: $h_{uk} \sim gamma(a_h, \eta_u)$.
2. For each item $i \in I$:
   - Sample quality: $\sigma_i \sim gamma(a_\sigma^s, a_\sigma^r)$.
   - For each component $k (k = 1, 2, \ldots, K)$:
   - Sample attribute: $\beta_{ik} \sim gamma(a_\beta, \sigma_i)$.
3. For each item $i \in I$ and user $u \in U$:
   - Sample attention behavior: $\Omega_{ui} \sim poisson(g_u^\top \beta_i)$.
   - Sample rating: $R_{ui} \sim poisson(\theta_u^\top \beta_i)$.
4. For each user $u \in U$ and user $v \in U$:
   - Sample trust: $G_{uv} \sim poisson(g_u^\top h_v)$.

As we can see from our model, comparing with existing methods that direct fusing trust information into rating information to recommend items, our method uses attention behavior information as medium for combining trust information with rating information. Specifically, our method combines poisson factorization of attention behavior information and rating information by sharing common latent vectors of items, because of constant item attributes. At the same time, we combine poisson factorization of attention behavior information and trust information by sharing common latent vectors of users' attention, because the users who have similar social relations tend to pay attention to similar items.

Also, different from recent works which recommend items just based on predicted ratings and overlook the importance of users' attention, we combine users' preference and their attention for better recommendation accuracy. For each user u, we rank items based on the score as follow:

$$score_{ui} = (\tilde{R}_{ui}^{w})(\tilde{\Omega}_{ui}^{1-w}) \tag{1}$$

Where $\tilde{R}_{ui}$ means the predicted rating of item i given by user u, depicting how user u likes item i. $\tilde{\Omega}_{ui}$ means how user u pays attention to item i. $w \in [0, 1]$ controls the importance of two aspects. We just employ users' attention to recommend items when

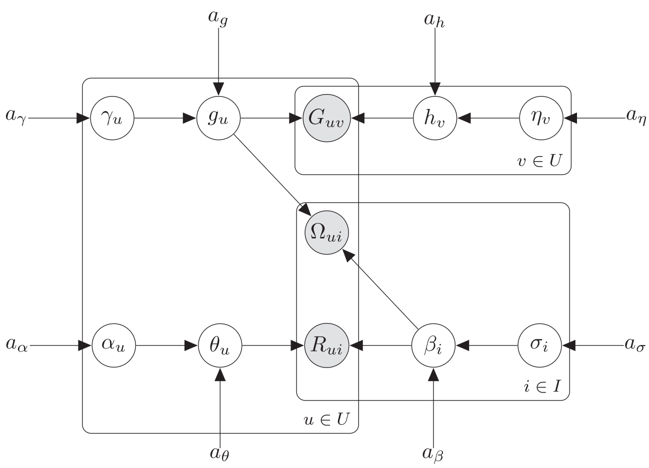


Fig. 2. A conditional directed graphical model of HTPF to show considered dependencies.

$w = 0$, while we just consider users' preference when $w = 1$. Once the posterior is fit, $\tilde{R}_{ui}$, $\tilde{\Omega}_{ui}$ could be calculated by their posterior expected poisson parameters: $\tilde{R}_{ui} = \mathrm{E}[\theta_u^\top \beta_i]$, $\tilde{\Omega}_{ui} = \mathrm{E}[g_u^\top \beta_i]$.

4.2. Approximate inference

Considering that the posterior probability is intractable, we develop an efficient approximate method to compute the posterior based on the variational inference [44]. The mean field theory drives us to partition latent variables into disjoint groups and these variables are governed by their own variational parameters [45]. Also, we add auxiliary latent variables $Z^R_{uik} \sim poisson(\theta_{uk}\beta_{ik})$, $Z^\Omega_{uik} \sim poisson(g_{uk}\beta_{ik})$, $Z^G_{uvk} \sim poisson(g_{uk}h_{vk})$ to facilitate derivation and update [38]. Note that the sum of poisson random variables is itself a poisson with rate equal to the sum of the rates [44]. Thus, these variables can be thought of as the contribution from component k to the total observations R_{ui}, Ω_{ui}, G_{uv} according to the property of poisson distribution.

$$\sum_k Z^R_{uik} = R_{ui}, \sum_k Z^\Omega_{uik} = \Omega_{ui}, \sum_k Z^G_{uik} = G_{ui} \tag{2}$$

Besides, according to the theory of variational inference, we specify the form of factored variational distribution of each variable as same as its corresponding conditional distribution [45]: The variational distributions of latent variables β, θ, g, h, γ, α, η, σ are gamma distributions with their own parameters and Z^R, Z^G, Z^Ω are multinomial distributions. So, we define the variational distribution as follows:

$$\begin{aligned} q(\beta, \theta, g, h, Z^R, Z^G, Z^\Omega, \gamma, \alpha, \eta, \sigma) &= \textstyle\prod_{i,k} q(\beta_{ik}|\lambda^s_{ik}, \lambda^r_{ik}) \\ &\times \textstyle\prod_{u,k} q(h_{uk}|\tau^s_{uk}, \tau^r_{uk}) q(\theta_{uk}|\kappa^s_{uk}, \kappa^r_{uk}) q(g_{uk}|\varphi^s_{uk}, \varphi^r_{uk}) \\ &\times \textstyle\prod_i q(\sigma_i|c^s_i, c^r_i) \prod_u q(\alpha_u|b^s_u, b^r_u) q(\gamma_u|d^s_u, d^r_u) q(\eta_u|e^s_u, e^r_u) \\ &\times \textstyle\prod_{u,i} q(Z^R_{ui}|\phi^R_{ui}) q(Z^\Omega_{ui}|\phi^\Omega_{ui}) \prod_{u,v} q(Z^G_{uv}|\phi^G_{uv}) \end{aligned} \tag{3}$$

where we use superscript 's' and 'r' represent the shape parameter and the rate parameter respectively. Note that minimizing the KL divergence between the variational distribution and the true posterior is equivalent to optimizing an *evidence lower bound* (ELBO) $L(q)$, a bound on the log likelihood of the observations [45]. The ELBO is

$$\begin{aligned} L(q) &= E_q[\ln p(R, G, \Omega, \Theta)] - E_q[\ln q(\Theta)] \\ &= \sum_{u \in U, i \in I} \big(E_q[p(R_{ui}|\theta_u, \beta_i)] + E_q[p(\Omega_{ui}|g_u, \beta_i)]\big) \\ &+ \sum_{u,v \in U} E_q[p(G_{uv}|g_u, h_v)] + \sum_{i \in I, k=1,\dots K} E_q[\beta_{ik}|a_\beta, \sigma_i] \\ &+ \sum_{u \in U, k=1,\dots K} \big(E_q[p(\theta_{uk}|a_\theta, \alpha_u)] + E_q[p(g_{uk}|a_g, \gamma_u)]\big) \\ &+ \sum_{u \in U, k=1,\dots K} E_q[p(h_{uk}|a_h, \eta_u)] + \sum_{i \in I} E_q[p(\sigma_i|a_\sigma)] \\ &+ \sum_{u \in U} \big(E_q[p(\alpha_u|a_\alpha) + E_q[p(\gamma_u|a_\gamma)] + E_q[p(\eta_u|a_\eta)]\big) \\ &- E_q[\ln q(\Theta)] \end{aligned} \tag{4}$$

where $\Theta \equiv \{\beta, \theta, g, h, Z^R, Z^G, Z^\Omega, \gamma, \alpha, \eta, \sigma\}$ denotes latent variables. We can use the coordinate ascent method to optimize variational parameters in turns by optimizing the lower bound $L(q)$. The variational inference of HTPF is illustrated in Algorithm 1.

Algorithm 1 The variational inference of HTPF.

```
initialize λ, κ, φ, τ, b^r, c^r, d^r, e^r randomly;
while not converge do
    for each observations Ω_ui > 0, R_ui > 0, G_uv > 0 do
        update variational parameters of Z^R_ui, Z^Ω_ui, Z^G_uv. {Equation (5)-(7)}
    end for
    for each user do
        update variational parameters of θ_u, g_u, γ_u, α_u, h_v, η_v. {Equation (8)-(15)}
    end for
    for each item do
        update variational parameters of β_i, σ_i. {Equation (16)-(19)}
    end for
end while
```

As we can see from Algorithm 1, in the beginning, we initialize variables to their priors with small offsets, where offsets can be sampled from uniform distribution over the interval [-0.1,0.1]. Then, we calculate the gradient of $L(q)$ on each parameter and iteratively optimize each parameter while holding the others fixed. For the parameters of observations Ω_{ui}, R_{ui}, G_{uv}, we just consider those non-zero observations. Since when observations are zero, these variational variables are not random. The posterior distributions of Z^Ω_{ui}, Z^R_{ui}, Z^G_{uv} will place all their mass on the zero vector. We calculate the optimal variational parameters of those non-zero observations as follows:

$$\phi^R_{ui} \propto \exp\{\Psi(\kappa^s_{uk}) - \log \kappa^r_{uk} + \Psi(\lambda^s_{ik}) - \log \lambda^r_{ik}\} \tag{5}$$

$$\phi^\Omega_{ui} \propto \exp\{\Psi(\varphi^s_{uk}) - \log \varphi^r_{uk} + \Psi(\lambda^s_{ik}) - \log \lambda^r_{ik}\} \tag{6}$$

$$\phi^G_{uv} \propto \exp\{\Psi(\varphi^s_{uk}) - \log \varphi^r_{uk} + \Psi(\tau^s_{vk}) - \log \tau^r_{vk}\} \tag{7}$$

where $\Psi()$ is the digamma function (the first derivative of the log gamma function).

Then, for each user we calculate the optimal variational parameters of θ_u, g_u, γ_u, α_u, h_v, η_v as follows:

$$\kappa^s_{uk} = a_\theta + \sum_{i \in I} R_{ui}\phi^R_{uik}; \kappa^r_{uk} = \frac{b^s_u}{b^r_u} + \sum_{i \in I} \frac{\lambda^s_{ik}}{\lambda^r_{ik}} \tag{8}$$

$$\varphi^s_{uk} = a_g + \sum_{i \in I} \Omega_{ui}\phi^\Omega_{uik} + \sum_{v \in U} G_{uv}\phi^G_{uvk} \tag{9}$$

$$\varphi^r_{uk} = \frac{d^s_u}{d^r_u} + \sum_{i \in I} \frac{\lambda^s_{ik}}{\lambda^r_{ik}} + \sum_{v \in U} \frac{\tau^s_{vk}}{\tau^r_{vk}} \tag{10}$$

$$b_u^s = Ka_\theta + a_\alpha^s; d_u^s = Ka_p + a_\gamma^s \tag{11}$$

$$b_u^r = \sum_{k=1,\dots K} \frac{\kappa_{uk}^s}{\kappa_{uk}^r} + a_\alpha^r; d_u^r = \sum_{k=1,\dots K} \frac{\varphi_{uk}^s}{\varphi_{uk}^r} + a_\gamma^r \tag{12}$$

$$\tau_{vk}^s = a_h + \sum_{u \in U} G_{uv}\phi_{uvk}^G; \tau_{vk}^r = \frac{e_v^s}{e_v^r} + \sum_{u \in U} \frac{\varphi_{uk}^s}{\varphi_{uk}^r} \tag{13}$$

$$e_v^s = Ka_h + a_\eta^s \tag{14}$$

$$e_v^r = \sum_{k=1,\dots K} \frac{\tau_{vk}^s}{\tau_{vk}^r} + a_\eta^r \tag{15}$$

Similarly, for each item we calculate the optimal variational parameters of β_i, σ_i as follows:

$$\lambda_{ik}^s = a_\beta + \sum_{u \in U} R_{ui} Z_{uik}^R + \sum_{u \in U} \Omega_{ui} Z_{uik}^\Omega \tag{16}$$

$$\lambda_{ik}^r = \frac{c_i^s}{c_i^r} + \sum_{u \in U} \frac{\kappa_{uk}^s}{\kappa_{uk}^r} + \sum_{u \in U} \frac{\varphi_{uk}^s}{\varphi_{uk}^r} \tag{17}$$

$$c_i^s = Ka_\sigma + a_\sigma^s \tag{18}$$

$$c_i^r = \sum_{k=1,\dots,K} \frac{\lambda_{ik}^s}{\lambda_{ik}^r} + a_\sigma^r \tag{19}$$

Note that the optimal shape parameters $b_u^s, d_u^s, c_v^s, e_i^s$ are independent on other variational parameters. Thus, we can get those optimal values at once without iteration (Eq. (11), (14), (18)). Finally, we terminate iteration by observing the change in the average predicted log likelihood of the validation set.

4.2.1. Complexity analysis

We can find that our algorithm is very efficient on sparse data. When updating variational parameters ϕ_{ui}^R, ϕ_{ui}^Ω, ϕ_{uv}^G, we just consider those non-zero observations, since when observations are zero these variational variables are not random. The time to update these variational parameters is $O(K(|R| + |G|))$, where $|R|$, $|G|$ means the number of non-zero ratings and social relations, K means the length of latent vector. When updating κ^s, λ^s, φ^s, τ^s, we just sum over non-zero observations with complexity $O(K(|R| + |G|))$. When updating κ^r, λ^r, φ^r, τ^r, we can preprocess the sum of $\frac{\lambda_{ik}^s}{\lambda_{ik}^r}$, $\frac{\kappa_{uk}^s}{\kappa_{uk}^r}$, $\frac{\varphi_{uk}^s}{\varphi_{uk}^r}$, $\frac{\tau_{vk}^s}{\tau_{vk}^r}$ over all items or users to speed up our algorithm. Thus, the complexity of our algorithm is $O(K(|R| + |G|))$. Generally, the overall computational time is linear with respect to the number of observed non-zero entries and our model has potential to scale up to large datasets.

5. Experiments and results

In this section, we evaluate our algorithm on real-world datasets. We start with the description of four datasets.

5.1. Datasets and evaluation metrics

The four datasets, Epinions, Ciao, Flixster and Douban, presented in Section 3 are used in our experiment. Specially, we double the rating values in Flixster to get integral values. A 5-fold cross-validation for learning and testing is used in our experiment where the datasets are divided into 5 folds: four folds for training and the remaining one for testing in each iteration. Also, we sample 1% data of training set as validation set to measure convergence and tune parameters. The final result is averaged over 5 iterations where all folds are tested. To quantitatively evaluate the experimental results, we used two metrics:

Normalized precision@M (Pre@M): We recommend each user M items and calculate the fraction of relevant items in the test set as follow. As recent works [10,38], highly rated items in the test set will be considered as relevant items for each user (larger than 3 in a 5-star system).

$$\frac{1}{|U|} \sum_{u \in U} \frac{hit_u}{\min(M, t_u)} \tag{20}$$

where hit_u denotes the number of relevant items in the user's top-M recommendations, t_u denotes the number of relevant items of the user u. We adjust denominator to $\min(M, t_u)$, because traditional measurement of precision will artificially deflate this measurement for the users who have fewer than M relevant items in the test set [10].

Normalized Discounted Cumulative Gain (NDCG) [27]: This is widely used in information retrieval and it measures the quality of ranking through discounted importance based on positions. In recommender systems, NDCG is computed as following:

$$NDCG = \frac{1}{|U|} \sum_{u \in U} \frac{DCG_u}{IDCG_u} \tag{21}$$

where DCG_u is defined as follow and $IDCG_u$ is the ideal value of DCG_u coming from the best ranking.

$$DCG_u = \sum_{i \in \mathrm{Re}(u)} \frac{1}{\log_2(ran_{ui} + 1)} \tag{22}$$

where ran_{ui} represents the rank of the item i in the recommended list of the user u and $\mathrm{Re}(u)$ denotes the set containing all relevant items of user u.

5.2. Compared algorithms

To demonstrate the effectiveness of our HTPF algorithm, we compared with some methods including:

- The baseline POP: In POP, we rank items by their universal popularity.
- HPF [10]: The state-of-the-art method based on rating information only. HPF conducts poisson factorization on rating matrix to get user's latent preference for recommendation. Note that HPF is a special case of our model HTPF, where user's attention is left out in the model.
- SPF [38], TrustSVD [41]: The state-of-the-art social recommendation methods mentioned in Section 2, which model social influence on the rating values.
- We designed another method HTPF-a as comparison. HTPF-a, a simple version of HTPF, generates recommendations just based on users' attention by integrating poisson factorization of trust information with attention attentive information, while rating information is left out in HTPF-a. Note that HTPF-a is one of our contributions.

The optimal experimental settings for all methods are determined either by our experiments or suggested by previous works. Specifically, the number of latent components K are set to 30 across all datasets for our methods (HTPF-a, HTPF), while the best values of w on different datasets are showed in Table 3. For hyper-parameters about prior distributions, we set a_β to 3 and others to 0.3.

Table 2
The recommended performance on four datasets. The bold terms denote the best performance among all methods.

Datasets	Ciao		Epinions		Flixster		Douban	
Evaluation	Pre@20 (%)	NDCG	Pre@20 (%)	NDCG	Pre@20 (%)	NDCG	Pre@20 (%)	NDCG
POP	3.83	0.1551	3.06	0.1335	13.43	0.2956	9.96	0.3193
TrustSVD	2.37	0.1336	2.23	0.1194	14.68	0.3023	10.45	0.3223
HPF	4.68	0.1600	4.18	0.1446	32.97	0.3991	19.69	0.4086
SPF	4.69	0.1629	4.27	0.1449	33.93	0.3994	20.19	0.4104
HTPF-a	4.86	0.1650	4.31	0.1480	31.93	0.3921	18.10	0.3942
HTPF	**5.29**	**0.1692**	**4.86**	**0.1520**	**35.92**	**0.4068**	**20.53**	**0.4144**

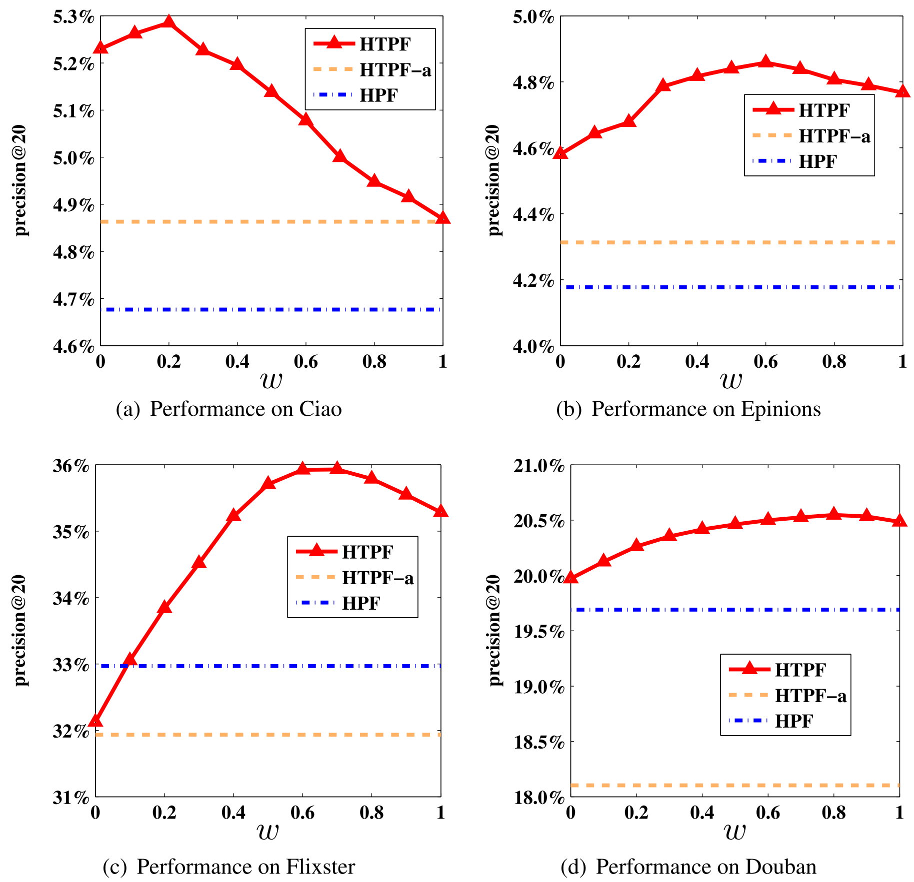


Fig. 3. The performance in terms of normalized precision@20 with varying values of parameter w, where w balances the contribution of users' preference and attention in recommendation.

5.3. Experimental results and analyses

The experimental results of all methods on four datasets are presented in Table 2. HTPF obviously outperforms all the comparison methods on all the four data sets. The results confirm that by modeling users' attention HTPF effectively improves recommendation performance. One reason for this improvement is that users' attention plays an important role in users' selection on recommendation. This can be seen from the good performance of HTPF-a, which only considers users' attention in social recommendation. By combining users' attention and preference in social recommendation, HTPF is able to achieve even better performance than HPF and HTPF-a. Another important reason for this improvement is that comparing with users' preference their attention is more susceptible to social network. Trust information can act as useful complementary information to deduce user attention but it contains limited information about users' preference. By modeling users' attention in social recommendation, HTPF can effectively bridge the trust-preference gap [39] and achieve better performance than state-of-the-art social recommendation methods.

Table 3
Optimal w values on different datasets.

Datasets	Ciao	Epinions	Flixster	Douban
Optimal w	0.2	0.6	0.7	0.8
Density of the network	0.2055%	0.0201%	0.0112%	0.0101%

5.3.1. Impact of parameter w

Another experiment is conducted to investigate how parameter w affects the performance of our model, where w balances the contribution of users' preference and attention in recommendation. The results in terms of precision@20 with varying w are presented in Fig. 3. Also, we present the performance of HPF and HTPF-a in comparison. As we can see, as w becomes larger, the performance becomes better first. This is because when users' preference plays a more important role in rank, more favorite items will be found. However, when w surpasses a threshold, the performance becomes worse with further increase of w. As users' attention becomes unimportant in rank, the recommended items may be refused because these items won't draw users' attention and may be skipped up. We find HTPF has worst result when w is fixed at 0 or 1, which confirms with the idea that combining users' attention and preference performs better than considering only one aspect.

The recommendation generated by HTPF is a balanced result between users' attention and their preference. However, as shown in Fig. 3, even when $w = 0$ (meaning we recommend items based only on users' attention), HTPF still achieves better performance than HTPF-a. The reason is that HTPF has already incorporated trust, attention behavior and rating information in its model training and can capture more accurate features of users and items than HTPF-a. Similar result can be observed when $w = 1$, when HTPF still achieves better performance than HPF.

5.3.2. Optimal w values on different datasets

It will be interesting to explore how the density of network connections will affect the balance parameter w. Table 3 lists the optimal w values in four different networks with different density. As can be seen, more densely connected networks correspond to smaller optimal w values, meaning that user attention plays a more important role. For instance, Ciao has much denser social network than Epinions. The users in Ciao have a higher degree of exposure to social network and are more strongly affected by their trustees/friends. To achieve best recommendation accuracy, HTPF need rely more on users' attention in Ciao, since users' attention is susceptible to social network.

6. Conclusions

In this paper, we propose a new probabilistic model HTPF that explicitly considers both users' attention and preference in social recommendation. Many psycho-social literatures suggest the importance of users' attention in recommendation and our observations in Section 3 show that the influence of trust relations dwells more on users' attention than on their preference. Thus, we propose the model HTPF with a generative process where we use social network as complemental information to deduce user's attention instead of their preference. We also design an efficient stochastic variational inference method for our model that can scale up to large data sets. Our comprehensive experimental results on four real-world datasets clearly demonstrate the effectiveness of our proposed method and its superiority over existing social recommendation methods.

One interesting direction for future work is to further exploit the relations and differences between user's attention and preference based on more supplementary information, such as content data and browsing data. Also, we can consider both users' attention and preference to deal with link prediction problem [46], since attention is more susceptible to the social network.

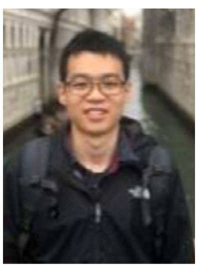
Acknowledgemnt

This work is supported by the National Natural Science Foundation of China (Grant no: U1866602) and by a Discovery Grant from the National Science and Engineering Research Council of Canada. It is also partially supported by ByteDance.

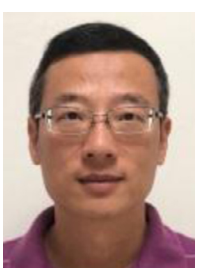
References

[1] J.R. Anderson, Cognitive Psychology and its Implications, WH Freeman/Times Books/Henry Holt & Co, 1990.
[2] D. Kahneman, Attention and Effort, Citeseer, 1973.
[3] N. Lavie, A. Hirst, J.W. De Fockert, E. Viding, Load theory of selective attention and cognitive control., J. Exp. Psychol.: Gen. 133 (3) (2004) 339.
[4] R.A. Rensink, J.K. O'Regan, J.J. Clark, To see or not to see: the need for attention to perceive changes in scenes, Psychol. Sci. 8 (5) (1997) 368–373.
[5] J.-H. Kang, K. Lerman, LA-CTR: a limited attention collaborative topic regression for social media, in: Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence, AAAI Press, 2013, pp. 1128–1134.
[6] S.J. Gershman, J.D. Cohen, Y. Niv, Learning to selectively attend, in: Proceedings of the 32nd Annual Conference of the Cognitive Science Society, 2010.
[7] J.L. Freedman, D.O. Sears, Selective exposure, Adv. Exp. Soc. Psychol. 2 (1965) 57–97.
[8] P. Dayan, S. Kakade, P.R. Montague, Learning and selective attention, Nat. Neurosci. 3 (11s) (2000) 1218.
[9] G. Palla, I. Derényi, I. Farkas, T. Vicsek, in: Uncovering the overlapping community structure of complex networks in nature and society, 2005.
[10] P. Gopalan, J.M. Hofman, D.M. Blei, Scalable recommendation with hierarchical poisson factorization, in: Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence, UAI 2015, July 12-16, 2015, Amsterdam, The Netherlands, 2015, pp. 326–335.
[11] A. Mnih, R. Salakhutdinov, Probabilistic matrix factorization, in: Proceedings of the Advances in Neural Information Processing Systems, 2007, pp. 1257–1264.
[12] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, in: Proceedings of the 10th international Conference on World Wide Web, ACM, 2001, pp. 285–295.
[13] 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.
[14] S. Ohsawa, Y. Obara, T. Osogami, Gated probabilistic matrix factorization: learning users' attention from missing values, Proceedings of the International Joint Conference on Artificial Intelligence, 2016, pp. 1888–1894.
[15] F. Ricci, L. Rokach, B. Shapira, Recommender systems: introduction and challenges, in: Recommender Systems Handbook, Springer, 2015, pp. 1–34.
[16] S. Zhang, L. Yao, A. Sun, Deep learning based recommender system: a survey and new perspectives, CoRR (2017). arXiv: 1707.07435.
[17] Y. Zhu, J. Zhu, J. Hou, Y. Li, B. Wang, Z. Guan, D. Cai, A brand-level ranking system with the customized attention-GRU model, in: Proceedings of the 27th International Joint Conference on Artificial Intelligence, AAAI Press, 2018, pp. 3947–3953.
[18] Y. Zhu, H. Li, Y. Liao, B. Wang, Z. Guan, H. Liu, D. Cai, What to do next: modeling user behaviors by time-LSTM, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, AAAI Press, 2017, pp. 3602–3608.
[19] B. Wang, M. Ester, J. Bu, Y. Zhu, Z. Guan, D. Cai, Which to view: Personalized prioritization for broadcast emails, in: Proceedings of the 25th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2016a, pp. 1181–1190.
[20] B. Wang, M. Ester, Y. Liao, J. Bu, Y. Zhu, Z. Guan, D. Cai, The million domain challenge: Broadcast email prioritization by cross-domain recommendation, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016b, pp. 1895–1904.
[21] Y. Zhu, Z. Guan, S. Tan, H. Liu, D. Cai, X. He, Heterogeneous hypergraph embedding for document recommendation, Neurocomputing 216 (2016) 150–162.
[22] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Trans. Knowl. Data Eng. 17 (6) (2005) 734–749.
[23] 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, ACM, 2008, pp. 426–434.
[24] Y. Koren, R. Bell, C. Volinsky, et al., Matrix factorization techniques for recommender systems, Computer 42 (8) (2009) 30–37.
[25] R. Salakhutdinov, A. Mnih, Probabilistic matrix factorization., in: Proceedings of the NIPS, 1, 2007, pp. 2–11.
[26] F. Ricci, L. Rokach, B. Shapira, Introduction to Recommender Systems Handbook, Springer, 2011.
[27] J. Tang, X. Hu, H. Liu, Social recommendation: a review, Soc. Netw. Anal. Min. 3 (4) (2013) 1113–1133.
[28] 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.
[29] X. Yang, H. Steck, Y. Liu, Circle-based recommendation in online social networks, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2012, pp. 1267–1275.
[30] 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, ACM, 2014, pp. 975–978.
[31] 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.
[32] B. Yang, Y. Lei, D. Liu, J. Liu, Social collaborative filtering by trust, in: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, AAAI Press, 2013, pp. 2747–2753.
[33] Y. Shen, R. Jin, Learning personal+ social latent factor model for social recommendation, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2012, pp. 1303–1311.
[34] M. Jamali, M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks, in: Proceedings of the Fourth ACM Conference on Recommender Systems, ACM, 2010, pp. 135–142.
[35] A. Mnih, R.R. Salakhutdinov, Probabilistic matrix factorization, in: Advances in Neural Information Processing Systems, 2008, pp. 1257–1264.
[36] 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.

[37] J. Tang, H. Gao, H. Liu, MTRUST: discerning multi-faceted trust in a connected world, in: Proceedings of the fifth ACM International Conference on Web Search and Data Mining, ACM, 2012, pp. 93–102.
[38] A.J. Chaney, D.M. Blei, T. Eliassi-Rad, A probabilistic model for using social networks in personalized item recommendation, in: Proceedings of the 9th ACM Conference on Recommender Systems, ACM, 2015, pp. 43–50.
[39] Y. Bao, H. Fang, J. Zhang, Leveraging decomposed trust in probabilistic matrix factorization for effective recommendation, in: Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI), 2014, p. 350.
[40] X. Wang, S.C. Hoi, M. Ester, J. Bu, C. Chen, Learning personalized preference of strong and weak ties for social recommendation, in: Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, 2017, pp. 1601–1610.
[41] 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.
[42] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, Grouplens: an open architecture for collaborative filtering of netnews, in: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, ACM, 1994, pp. 175–186.
[43] E. Tulving, Episodic memory: from mind to brain, Ann. Rev. Psychol. 53 (1) (2002) 1–25.
[44] M.J. Wainwright, M.I. Jordan, Graphical models, exponential families, and variational inference, Found. Trends® Mach. Learn. 1 (1-2) (2008) 1–305.
[45] M.D. Hoffman, D.M. Blei, C. Wang, J. Paisley, Stochastic variational inference, J. Mach. Learn. Res. 14 (1) (2013) 1303–1347.
[46] S. Liu, S. Wang, F. Zhu, J. Zhang, R. Krishnan, Hydra: large-scale social identity linkage via heterogeneous behavior modeling, in: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, ACM, 2014, pp. 51–62.



Jiawei Chen received his B.S. at University of electronic science and technology of China in 2014. he is currently a Ph.D. candidate in the college of computer science at Zhejiang University. His main research topics are recommender systems, graphical model and social networks.



Can Wang received the Ph.D. degree and M.S. degree in computer science and B.S. degree in economics from Zhejiang University, in 2009, 2003 and 1995, respectively. His research interests include data mining, machine learning and information retrieval.

Qihao Shi received his B.S. at Nanjing Normal University of China in 2014. He is currently a Ph.D. candidate in the college of computer science at Zhejiang University. His main research topics are social and information networks, algorithmic game theory and Internet economics.

Yan Feng received the Ph.D. degree in computer application from Zhejiang University in 2004. She is currently an associate professor in the College of Computer Science at Zhejiang University, China. Her research interests include database, data mining etc.

Chun Chen is a professor in the College of Computer Science, Zhejiang University. His research interests include data mining, computer vision, computer graphics and embedded technology.