

Recommending Flickr groups with social topic model

Jingdong Wang · Zhe Zhao · Jiazhen Zhou ·
Hao Wang · Bin Cui · Guojun Qi

Received: date / Accepted: date

Abstract The explosion of multimedia content in social media networks raises a great demand of developing tools to facilitate producing, sharing and viewing media content. Particularly, Flickr groups, self-organized communities with declared, common interests, are able to help users to conveniently participate in social media network. In this paper, we address the problem of automatically recommending groups to users. We propose to simultaneously exploit media contents and link structures between users and groups. To this end, we present a probabilistic latent topic model

Jingdong Wang
Microsoft Research Asia
No. 5 Danling Street, Haidian District, Beijing 100080, P.R. China
E-mail: jingdw@microsoft.com

Zhe Zhao
Electrical Engineering & Computer Science, University of Michigan
260 Hayward Ave. Ann Arbor, MI 48109 US
E-mail: zhezhao@umich.edu

Jiazhen Zhou
Department of Computer Science Peking University
Beijing 100871 P.R. China
E-mail: jiazhenzhou@yahoo.com

Hao Wang
Beijing University of Posts and Telecommunications
P.O. Box 250, No. 10 Xi Tu Cheng Road Beijing 100876, P.R. China
E-mail: jiehuazi@bupt.edu.cn

Bin Cui
State Key Laboratory of Virtual Reality Technology and Systems, Beihang University
Beijing 100191, P.R. China
School of EECS, Peking University
Beijing 100871, P.R. China
E-mail: bin.cui@pku.edu.cn

Guojun Qi
Beckman Institute, University of Illinois at Urbana-Champaign
405 North Mathews Avenue Urbana, IL 61801-2300, US
E-mail: qi4@illinois.edu

to model them in an integrated framework, expecting to jointly discover the latent interests for users and groups and simultaneously learn the recommendation function. We demonstrate the proposed approach on the dataset crawled from Flickr.com.

Keywords Flickr group · Recommendation · Social topic model

1 Introduction

Recent years have witnessed an explosion of multimedia contents, involving images, videos and text, in social media networks. Users are able to produce, view, share and reproduce content in a number of social scenarios, and even interact with media to create additional metadata such as tags and comments. To facilitate users to conveniently digest media content, a lot of tools, such as tag-based image and video search in Flickr, and Flickr groups, have been developed, and a lot of research efforts have been conducted. Particularly, in this paper, by investigating the varied facilities in social media networks, we study Flickr groups and address the problem of automatically recommending Flickr groups to users.

Flickr groups, a social connection feature on Flickr, are self-managed communities with common interests, where users can share and comment on photos, most on specific topics, such as landmarks, animals and sports. An example of Flickr group is shown in Figure 1 and example photos from its members are shown in Figure 2. They are not only containers of media contents but also bridges to connect users for social media. Groups are created spontaneously but not randomly: people participate in groups for specific intentions (e.g., interested in the visual content or introduced by other users with similar interest), and the photos in the group are with a common theme. The current system is able to index and retrieve the groups to help users conveniently search and discover groups of interest. However, in order to make users access to desired groups more easily, it is necessary to deploy an automatic group recommending system upon the social media infrastructure.

Our principle for recommending groups to users is that the users and the recommended groups have large probability of sharing the similar latent interests, which can be discovered and mined from the rich information available in social media networks including metadata, uploaded and shared images in the groups. Moreover, rather than separately discovering the latent interests and subsequently learning the recommendation function from the available links between users and groups, we propose a *social topic model* to simultaneously discover latent interests for users and groups and estimate the recommendation function. Specifically, we first present a probabilistic latent topic model to capture the interests of users and groups. To match interests of users and groups, we impose the restriction on these two topic models to share the common latent topic bases. Next, we explore the latent discovery process by capturing the social link structure to connect the common interest topics between users and groups. Simultaneously, a recommendation function is embedded over the social connections underlying the discovered latent topics. An effective inference approach based on Gibbs sampling is applied to efficiently build the latent topics and the recommendation functions.

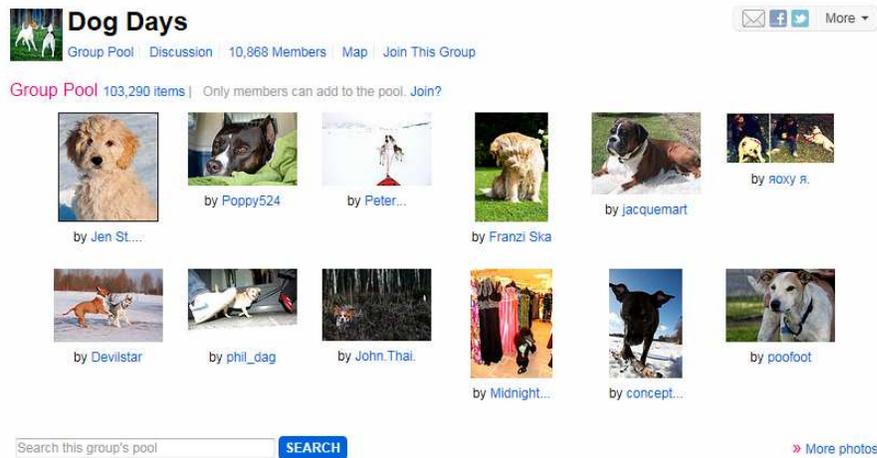


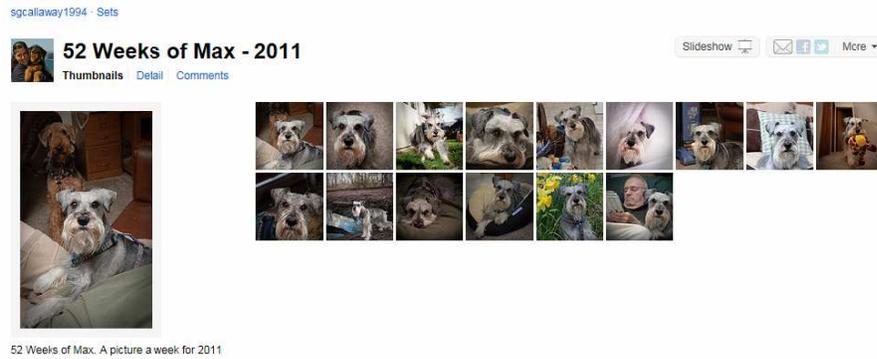
Fig. 1 An example of Flickr group: Dog Days.

2 Related work

Flickr group has become one of the most representative social media networks connecting hundreds of thousands of interest groups. Its commercial success has been attracting more research attention to study many interesting social phenomena occurring in this platform. One of the pioneer works is analyzing the Flickr group ecosystem [23,25] and designing new group-search strategies by a topic-based representation for groups computed from group tags. Topickr [24] explores discovered topic-based representations for users and groups and furthermore ranks users and groups for each topic for another Flickr exploration experience. Hypergroup [22] presents an approach to cluster Flickr groups in order to help search Flickr groups. The semantic hierarchies of Flickr groups are also exploited in [18]. Flickr groups are also used to help finding landmark photos [1].

All these work encourages to develop and deploy automatic group recommending system for Flickr. The approaches in [30,31] are proposed to discover the latent events or topics by mining the visual contents and tags from images and then recommend photos to groups by matching their latent events or topics. The Sheep-Dog system [6] recommends Flickr groups to photos by detecting photo concepts. A semi-automatic approach was presented to loop human into the process of suggesting Flickr groups to users [5] that combines the group classifiers. Recommending Flickr groups to users is more challenging because user's profile contains a set of photos and extra existing user-group relations. This motivates us to investigate a new way to perform group recommendation.

Suggesting groups to users was ever studied in [32,33] by simply casting it into the collaborative filtering framework and directly applying tensor-based approaches for recommendation. Those approaches neglect the rich visual information associated with the photos. In contrast, our approach explores both visual contents and



(a)



(b)

Fig. 2 Example photos from the two members of the *Dog Days* Flickr group.

social relations to discover and match the latent interests of users and groups for recommendation.

Besides social media networks, social communities have also been investigated and studied in other social networks, including Facebook, Orkut and so on. An approach to model social groups is presented in [21] by aggregating individual user models, in order to cope with the potentially conflicting preferences of multiple users when selecting items for recommendation. A group recommendation approach is presented in [4] by detecting intrinsic communities of users whose preferences are similar. The group personality composition is investigated in [26] for group recommendation. A combinational collaborative filtering approach [7] is proposed to fuse both semantic and user information for community suggestion. Essentially most current approaches can be cast into the collaborative filter framework. In contrast, additional rich visual content in social media network implies the interests of users and groups and can help group recommendation.

General recommendation systems have been well investigated and a comprehensive survey can be found in [2, 10]. Roughly speaking, recommendation techniques are divided into three categories. The first one is content-based recommendation, in which the user (or other modules) will be recommended items similar to the ones the

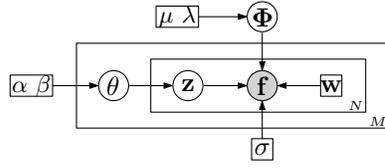


Fig. 3 The graphical representation of latent Beta composition.

user preferred in the past. The widely-used methods are based on topic models, such as latent semantic analysis [8], probabilistic latent semantic analysis [11, 12, 24, 22], and latent Dirichlet allocation [3]. The second one is collaborative filtering, in which the user will be recommended items that people with similar tastes and preferences liked in the past. Most methods are based on matrix factorization [10, 15, 13, 14, 28], and some probabilistic matrix factorization methods, such as [20, 19], are proposed to deal with large scale data and noises. The last one is hybrid recommendation that combines collaborative and content-based methods and essentially benefits from both the two methods. The proposed approach in this paper can be viewed as a hybrid approach and exploits both visual contents and the existing links between users and Flickr groups.

3 Latent Beta composition

Before moving further, we first propose a generative probabilistic model - Latent Beta composition (LBC), based upon which the group recommendation model is built. The basic idea is that users are represented as mixtures over latent topics, where each topic is characterized by a Gaussian distribution. Its graphical representation is shown in Figure 3. LBC assumes the following generative process for each user u .

1. Sample a random vector $\theta = [\theta_1 \cdots \theta_D]^T$ so that $\theta_i \sim \text{Beta}(\alpha, \beta)$.
2. Sample each element ϕ_{ij} of Φ so that $\phi_{ij} \sim \text{Gaussian}(\mu, \lambda^2)$.
3. For each of the N images I_i ,
 - (a) Sample a topic vector $\mathbf{z}_i = [z_{i1} \cdots z_{iD}]^T$ so that $z_{ij} \sim \text{Bernoulli}(\theta_j)$.
 - (b) Sample an image I_i so that $\mathbf{f}_i \sim \text{Gaussian}(\Phi \text{Diag}(\mathbf{z}_i) \mathbf{w}_i, \sigma^2 \mathbf{I})$.

The variables in this model consist of three categories: model parameters including $\alpha, \beta, \sigma, \mu, \lambda$ and \mathbf{w}_i , hidden variables including θ and \mathbf{z}_i , and observation variables \mathbf{f}_i . Compared with topic modeling for a corpus of document, a user (or group) in our problem corresponds to a document, and an image corresponds to a word.

The assumption in the topic model is that the interest of a user (or a group) can be characterized by the latent variables θ and an image can be composed of a few of topics selected from a dictionary $\Phi = [\phi_1 \cdots \phi_D]$. In the mathematical form, $\mathbf{f} = \Phi \text{Diag}(\mathbf{z}) \mathbf{w}$, \mathbf{z} is a topic vector, and $\text{Diag}(\mathbf{z})$ is a diagonal matrix with the diagonal entries valued from \mathbf{z} , and serves the role of selecting the atoms from the dictionary. \mathbf{w} is the projection coefficient vector in which the entries corresponding to the selected atoms are valid. Each entry in \mathbf{z} is valued as 1 if the i -th topic is used to compose the image and valued as 0 otherwise.

Each entry of the D -dimensional random vector $\boldsymbol{\theta}$ satisfies a Beta distribution, and then the vector essentially satisfies a joint distribution,

$$\boldsymbol{\theta} \sim \prod_{d=1}^D \text{Beta}(\alpha, \beta) = \prod_{d=1}^D \frac{1}{\text{B}(\alpha, \beta)} \theta_d^{\alpha-1} (1 - \theta_d)^{\beta-1}, \quad (1)$$

where the beta function, $\text{B}(\alpha, \beta)$, appears as a normalization constant to ensure that the total probability integrates to unity. Here we use a Beta distribution instead of the Dirichlet distribution in Latent Dirichlet Allocation [3]. The basic unit in document analysis is a single word and often has a single meaning in the context of the document. Thus, the Dirichlet distribution is enough. In contrast, the basic unit in our problem, a feature vector \mathbf{f} , which is composed of an image and its textual feature represented by the vector space model, is itself also a mixture of topics. This is natural because an image may not always contain a single concept and moreover it cannot be represented by a single topic, even if it only contains a single concept due to the various appearance of the same concept.

Similarly, the topic vector \mathbf{z}_i satisfies the following joint distribution,

$$\mathbf{z}_i \sim \prod_{d=1}^D \text{Bernoulli}(\theta_j) = \prod_{d=1}^D \theta_j^{z_{id}} (1 - \theta_j)^{1-z_{id}}. \quad (2)$$

In the Gaussian distribution, $\text{Gaussian}(\boldsymbol{\Phi} \text{Diag}(\mathbf{z}_i) \mathbf{w}_i, \sigma^2 \mathbf{I})$, \mathbf{w}_i is computed as $\mathbf{w}_i = \boldsymbol{\Phi}^T \mathbf{f}_i$. Then \mathbf{f}_i satisfies the following distribution,

$$\mathbf{f}_i \sim \text{Gaussian}(\boldsymbol{\Phi} \text{Diag}(\mathbf{z}_i) \mathbf{w}_i, \sigma^2 \mathbf{I}) \quad (3)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{D}{2}}} e^{-\frac{1}{2\sigma^2} \|\mathbf{f}_i - \boldsymbol{\Phi} \text{Diag}(\mathbf{z}_i) \mathbf{w}_i\|_2^2}. \quad (4)$$

4 Formulation

Let \mathcal{U} represent the set of m users. Each user $u \in \mathcal{U}$ is associated with a set of photos $\mathcal{I}^u = \{I_1^u, \dots, I_{N^u}^u\}$. Let g represent a Flickr group, and \mathcal{G} represent the set of n groups. Each group g is associated with a set of photos $\mathcal{I}^g = \{I_1^g, \dots, I_{N^g}^g\}$. Let \mathbf{f} (\mathbf{f}_i^u or \mathbf{f}_j^g) represent the content feature of an image I (I_i^u or I_j^g). Denote the user-group relation by an $m \times n$ matrix \mathbf{R} , in which $r_{ug} = 1$ means that the user u is a member of the group g . Initially, only a part of the entries are valued as 1 and the rest are unknown. The task is to find the pairs (u, g) that have large probabilities to be 1, which indicates that the user u will tend to join the group g .

4.1 Joint topic model

We present a social topic model to jointly discover both interests for users and groups and find the prediction function that checks the matching degree of users to groups. The joint model links users and groups together by building a prediction edge over

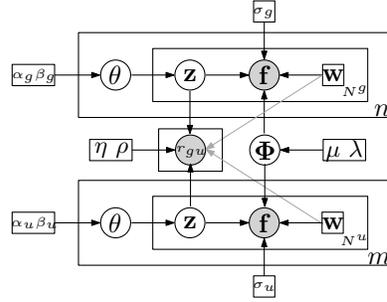


Fig. 4 The graphical representation of joint latent Beta composition over groups and users. The meanings of the variables and the relations in this model can be found from Section 3.

related users and groups to bridge their interests. There are several advantages in the joint model. On the one hand, the bridges between users and groups will benefit the interest discovery for each other. The join-in relation of a user to a group implies that they share common interests, and the joint topic model with this connect will result in that the interests discovered are more consistent. On the other hand, the prediction function and interest discovery will help each other since these two are simultaneously computed. The graphical representation is shown in Figure 4. The joint LBC assumes the following generative process.

1. Sample each element ϕ_{ij} of Φ so that $\phi_{ij} \sim \text{Gaussian}(\mu, \lambda^2)$.
2. The generative process for each user u
 - (a) Sample a random vector $\theta^u = [\theta_1^u \cdots \theta_D^u]^T$ so that $\theta_i^u \sim \text{Beta}(\alpha^u, \beta^u)$.
 - (b) For each of the N^u images I_i^u ,
 - i. Sample a topic vector $\mathbf{z}_i^u = [z_{i1}^u \cdots z_{iD}^u]^T$ so that $z_{ij}^u \sim \text{Bernoulli}(\theta_j^u)$.
 - ii. Sample an image I_i^u so that $\mathbf{f}_i^u \sim \text{Gaussian}(\Phi \text{Diag}(\mathbf{z}_i^u) \mathbf{w}_i^u, \sigma_u^2 \mathbf{I})$.
3. The generative process for each group g
 - (a) Sample a random vector $\theta^g = [\theta_1^g \cdots \theta_D^g]^T$ so that $\theta_i^g \sim \text{Beta}(\alpha^g, \beta^g)$.
 - (b) For each of the N^g images I_i^g ,
 - i. Sample a topic vector $\mathbf{z}_i^g = [z_{i1}^g \cdots z_{iD}^g]^T$ so that $z_{ij}^g \sim \text{Bernoulli}(\theta_j^g)$.
 - ii. Sample an image I_i^g so that $\mathbf{f}_i^g \sim \text{Gaussian}(\Phi \text{Diag}(\mathbf{z}_i^g) \mathbf{w}_i^g, \sigma_g^2 \mathbf{I})$.
4. Sample the relations between users and groups $\mathbf{R} = [r_{ug}]$ so that $r_{ug} \sim \text{Bernoulli}(h(\mathbf{y}^u, \mathbf{y}^g; \eta, \rho))$.

In the above process, the superscripts u and g denote the user-related or group-related variables. Several additional descriptions of this joint model are as follows. We model the topics for users and groups by assuming that they share the same latent topic dictionary Φ , which results in that the comparison of interests can be directly obtained by comparing the latent topic variables. Using different dictionaries however will lead to the complexity of the model and the difficulty of the optimization. We propose to use different conjugate prior parameters $\alpha \beta$ for users and groups based on the observation that users and groups essentially have different interest distribution. The variance parameters σ_u and σ_g in the Gaussian distribution are also different because the photos from users and groups have different diversity.

The prediction function $h(\mathbf{y}^u, \mathbf{y}^g; \eta, \rho)$ is defined over the latent topics of users and groups. The user interest \mathbf{y}^u is the averaged aggregation, $\mathbf{y}^u = \frac{1}{N^u} \sum_{i=1}^{N^u} \text{Diag}(\mathbf{z}_i^u) \mathbf{w}_i^u$, and the interest for a group is similarly defined. Besides the relations between users and groups, we also aim to discover the latent interests from the image content and in addition explore the similarity of the latent interests between users and groups. We observed that the photos from a user may contain multiple topics while the photos from a group often cover a single or a few topics. Figures 1, 2 and 5 show such an observation. Based on this point, we design the prediction function using the max-min criterion,

$$h(\mathbf{y}^u, \mathbf{y}^g; \eta, \rho) = \frac{1}{1 + e^{-\eta \max_{d \in \{1, \dots, D\}} \min\{y_d^u, y_d^g\} - \rho}}. \quad (5)$$

Here the inner min operator aims to find the consensus over the d -th topic between the user u and the group g , and the outer max operator aims to find the dominant common topic between the user and the group.

5 Inference

In this section, we present a Gibbs-sampling algorithm to infer the model parameters. In our inference algorithm, we sample the latent variables \mathbf{z}_i . Let's first see how the other latent variables are integrated out. For the convenience of the description, we may drop the superscripts u or g and subscripts $_i$ if not affecting the understanding.

Integrating out θ

Given a user or a group, we have the likelihood $p(z_{ij}|\theta_j) = \text{Bernoulli}(\theta_j)$ and the conjugate prior distribution, $p(\theta_j; \alpha, \beta) = \text{Beta}(\alpha, \beta)$. Our goal is to integrate out θ ,

$$\begin{aligned} & p(\{\mathbf{z}_i\}; \alpha, \beta) \\ &= \int p(\theta; \alpha, \beta) p(\{\mathbf{z}_i\}|\theta) d\theta \\ &= \int \left(\prod_{j=1}^D \frac{1}{\text{B}(\alpha, \beta)} \theta_j^{\alpha-1} (1-\theta_j)^{\beta-1} \right) \prod_{i=1}^N \prod_{j=1}^D \theta_j^{z_{ij}} (1-\theta_j)^{1-z_{ij}} d\theta \\ &= \int \prod_{j=1}^D \frac{1}{\text{B}(\alpha, \beta)} \theta_j^{\alpha + \sum_{i=1}^N z_{ij} - 1} (1-\theta_j)^{\beta + N - \sum_{i=1}^N z_{ij} - 1} d\theta \\ &= \prod_{j=1}^D \frac{\text{B}(\alpha + \sum_{i=1}^N z_{ij}, \beta + N - \sum_{i=1}^N z_{ij})}{\text{B}(\alpha, \beta)}. \end{aligned} \quad (6)$$

Integrating out Φ

We have $p(\mathbf{f}_i | \mathbf{z}_i, \Phi; \mathbf{w}_i, \sigma) = \text{Gaussian}(\Phi \text{Diag}(\mathbf{z}_i) \mathbf{w}_i, \sigma^2 \mathbf{I})$ and $p(\phi_{ij}; \mu, \lambda) = \text{Gaussian}(\mu, \lambda^2)$. Our goal is to integrate out Φ . The conditional distribution over



(a)



(b)

Fig. 5 Other example photos from the same two members in Figure 2.

$\{\mathbf{f}_i^u\}$ is

$$\begin{aligned}
 & p(\{\mathbf{f}_i^u\} | \{\mathbf{z}_i^u\}, \{\mathbf{w}_i^u\}, \Phi; \sigma_u) \\
 &= \prod_{i=1}^{P_u} p(\mathbf{f}_i | \mathbf{z}_i, \mathbf{w}_i, \Phi; \sigma_u) \\
 &= \prod_{i=1}^{P_u} \frac{1}{(2\pi\sigma_u^2)^{\frac{D}{2}}} e^{-\frac{1}{2\sigma_u^2} \|\mathbf{f}_i - \Phi \text{Diag}(\mathbf{z}_i) \mathbf{w}_i\|_2^2} \\
 &= \frac{1}{(2\pi\sigma_u^2)^{\frac{DP_u}{2}}} e^{-\frac{1}{2\sigma_u^2} \|\mathbf{F}_u - \Phi \mathbf{X}_u\|_2^2} \\
 &= \frac{1}{(2\pi\sigma_u^2)^{\frac{DP_u}{2}}} e^{-\frac{1}{2\sigma_u^2} \text{Tr}((\mathbf{F}_u - \Phi \mathbf{X}_u)(\mathbf{F}_u - \Phi \mathbf{X}_u)^T)}. \tag{7}
 \end{aligned}$$

where P_u is the number of all images that belong to users, $\mathbf{F}_u = [\mathbf{f}_1 \cdots \mathbf{f}_{P_u}]$, $\mathbf{X}_u = [\text{Diag}(\mathbf{z}_1)\mathbf{w}_1 \cdots \text{Diag}(\mathbf{z}_{P_u})\mathbf{w}_{P_u}]$ are matrices of which each column matches one image of a user, $\text{Tr}(\cdot)$ is the trace operator.

Similarly, the conditional distribution over $\{\mathbf{f}_i^g\}$ is

$$\begin{aligned}
& p(\{\mathbf{f}_i^g\}|\{\mathbf{z}_i^g\}, \{\mathbf{w}_i^g\}, \Phi; \sigma_g) \\
&= \prod_{i=1}^{P_g} p(\mathbf{f}_i|\mathbf{z}_i, \mathbf{w}_i, \Phi; \sigma_g) \\
&= \prod_{i=1}^{P_g} \frac{1}{(2\pi\sigma_g^2)^{\frac{D}{2}}} e^{-\frac{1}{2\sigma_g^2} \|\mathbf{f}_i - \Phi \text{Diag}(\mathbf{z}_i)\mathbf{w}_i\|_2^2} \\
&= \frac{1}{(2\pi\sigma_g^2)^{\frac{DP_g}{2}}} e^{-\frac{1}{2\sigma_g^2} \|\mathbf{F}_g - \Phi \mathbf{X}_g\|_2^2} \\
&= \frac{1}{(2\pi\sigma_g^2)^{\frac{DP_g}{2}}} e^{-\frac{1}{2\sigma_g^2} \text{Tr}((\mathbf{F}_g - \Phi \mathbf{X}_g)(\mathbf{F}_g - \Phi \mathbf{X}_g)^T)}, \tag{8}
\end{aligned}$$

where P_g is the number of all images that belong to groups, $\mathbf{F}_g = [\mathbf{f}_1 \cdots \mathbf{f}_{P_g}]$, $\mathbf{X}_g = [\text{Diag}(\mathbf{z}_1)\mathbf{w}_1 \cdots \text{Diag}(\mathbf{z}_{P_g})\mathbf{w}_{P_g}]$ are matrices of which each column matches one image of a group.

The distribution over Φ is

$$\begin{aligned}
& p(\Phi; \mu, \lambda) \\
&= \prod p(\phi_{ij}; \mu, \lambda) \\
&= \prod \frac{1}{\sqrt{2\pi\lambda^2}} e^{-\frac{1}{2\lambda^2}(\phi_{ij} - \mu)^2} \\
&= \frac{1}{(2\pi\lambda^2)^{\frac{D^2}{2}}} e^{-\frac{1}{2\lambda^2} \|\Phi - \mu \mathbf{E}\|_2^2} \\
&= \frac{1}{(2\pi\lambda^2)^{\frac{D^2}{2}}} e^{-\frac{1}{2\lambda^2} \text{Tr}((\Phi - \mu \mathbf{E})(\Phi - \mu \mathbf{E})^T)}. \tag{9}
\end{aligned}$$

where \mathbf{E} is a $D \times D$ matrix whose elements are all equal to 1.

With the above equations we have

$$\begin{aligned}
& p(\{\mathbf{f}_i^u\}, \{\mathbf{f}_i^g\}|\{\mathbf{z}_i^u\}, \{\mathbf{z}_i^g\}; \{\mathbf{w}_i^u\}, \{\mathbf{w}_i^g\}, \sigma_u, \sigma_g, \mu, \lambda) \\
&= \int p(\{\mathbf{f}_i^u\}|\{\mathbf{z}_i^u\}, \{\mathbf{w}_i^u\}, \Phi; \sigma_u) p(\{\mathbf{f}_i^g\}|\{\mathbf{z}_i^g\}, \{\mathbf{w}_i^g\}, \Phi; \sigma_g) p(\Phi; \mu, \lambda) d\Phi \\
&= \frac{1}{(2\pi\sigma_u^2)^{\frac{DP_u}{2}} (2\pi\sigma_g^2)^{\frac{DP_g}{2}} (2\pi\lambda^2)^{\frac{D^2}{2}}} \int e^{-\frac{1}{2} \text{Tr}(\Phi \mathbf{A} \Phi^T - 2\Phi \mathbf{B} + \mathbf{C})} d\Phi \\
&= \frac{(2\pi)^{\frac{D^2}{2}}}{(2\pi\sigma_u^2)^{\frac{DP_u}{2}} (2\pi\sigma_g^2)^{\frac{DP_g}{2}} (2\pi\lambda^2)^{\frac{D^2}{2}}} |\mathbf{A}|^{-\frac{D}{2}} e^{-\frac{1}{2} \text{Tr}(\mathbf{C} - \mathbf{B} \mathbf{A}^{-1} \mathbf{B}^T)}, \tag{10}
\end{aligned}$$

where $\mathbf{A} = \frac{\mathbf{X}_u \mathbf{X}_u^T}{\sigma_u^2} + \frac{\mathbf{X}_g \mathbf{X}_g^T}{\sigma_g^2} + \frac{\mathbf{I}}{\lambda^2}$, $\mathbf{B} = \frac{\mathbf{X}_u \mathbf{F}_u^T}{\sigma_u^2} + \frac{\mathbf{X}_g \mathbf{F}_g^T}{\sigma_g^2} + \frac{\mu \mathbf{E}}{\lambda^2}$, $\mathbf{C} = \frac{\mathbf{F}_u \mathbf{F}_u^T}{\sigma_u^2} + \frac{\mathbf{F}_g \mathbf{F}_g^T}{\sigma_g^2} + \frac{\mu^2 \mathbf{E}^2}{\lambda^2}$.

Joint distribution

The joint distribution over $\{\mathbf{z}_i^u\}$ is computed as follows,

$$p(\{\mathbf{z}_i^u\}; \alpha_u, \beta_u) = \prod_{u=1}^U p(\{\mathbf{z}_i\}^u; \alpha_u, \beta_u). \quad (11)$$

Similarly, the joint distribution over $\{\mathbf{z}_i^g\}$ is

$$p(\{\mathbf{z}_i^g\}; \alpha_g, \beta_g) = \prod_{g=1}^G p(\{\mathbf{z}_i\}^g; \alpha_g, \beta_g). \quad (12)$$

The joint distribution over $\{\mathbf{z}_i^u\}, \{\mathbf{z}_i^g\}$ is computed as follows,

$$p(\{\mathbf{z}_i^u\}, \{\mathbf{z}_i^g\}; \alpha_u, \alpha_g, \beta_u, \beta_g) = p(\{\mathbf{z}_i^u\}; \alpha_u, \beta_u) p(\{\mathbf{z}_i^g\}; \alpha_g, \beta_g). \quad (13)$$

The joint distribution over $\{\mathbf{z}_i^u, \mathbf{f}_i^u\}, \{\mathbf{z}_i^g, \mathbf{f}_i^g\}$ is computed as follows,

$$\begin{aligned} & p(\{\mathbf{z}_i^u, \mathbf{f}_i^u\}, \{\mathbf{z}_i^g, \mathbf{f}_i^g\}; \Theta) \\ &= p(\{\mathbf{z}_i^u\}, \{\mathbf{z}_i^g\}; \Theta) p(\{\mathbf{f}_i^u\}, \{\mathbf{f}_i^g\} | \{\mathbf{z}_i^u\}, \{\mathbf{z}_i^g\}; \Theta). \end{aligned} \quad (14)$$

where $\Theta = \{\alpha_u, \beta_u, \{\mathbf{w}_i^u\}, \sigma_u, \mu, \lambda, \alpha_g, \beta_g, \{\mathbf{w}_i^g\}, \sigma_g, \eta, \rho\}$ represents the variables to be estimated.

The distribution over the links is as follows,

$$p(\{r_{ug}\} | \{\mathbf{z}_i^u\}, \{\mathbf{z}_j^g\}; \{\mathbf{w}_i^u\}, \{\mathbf{w}_j^g\}) = \prod_{(u,g)} p(r_{ug} | \mathbf{y}^u, \mathbf{y}^g; \eta, \rho). \quad (15)$$

Therefore, the joint distribution over $\{\mathbf{z}_i^u, \mathbf{f}_i^u\}, \{\mathbf{z}_i^g, \mathbf{f}_i^g\}, \{r_{ug}\}$ is written as the following,

$$\begin{aligned} & p(\{\mathbf{z}_i^u, \mathbf{f}_i^u\}, \{\mathbf{z}_i^g, \mathbf{f}_i^g\}, \{r_{ug}\}; \Theta) \\ &= p(\{\mathbf{z}_i^u, \mathbf{f}_i^u\}, \{\mathbf{z}_i^g, \mathbf{f}_i^g\}; \Theta) p(\{r_{ug}\} | \{\mathbf{z}_i^u, \mathbf{f}_i^u\}, \{\mathbf{z}_i^g, \mathbf{f}_i^g\}; \Theta). \end{aligned} \quad (16)$$

5.1 Algorithm

The algorithm consists of two steps: conditional sampling and parameter estimation. The first step samples latent topics according to the conditional distribution, and the second step mainly estimates the parameters.

Conditional sampling

Given the current topic parameters $\{\mathbf{z}_i^u\}, \{\mathbf{z}_i^g\}$, the conditional distribution with respect to the d -th topic z_{id}^u of the i -th image of user d can be written as follows,

$$p(z_{id}^u | -) \propto p(\{\mathbf{z}_i^u, \mathbf{f}_i^u\}, \{\mathbf{z}_i^g, \mathbf{f}_i^g\}; \Theta) p(\{r_{ug}\} | \{\bar{\mathbf{z}}_i^u, \mathbf{f}_i^u\}, \{\mathbf{z}_i^g, \mathbf{f}_i^g\}; \Theta), \quad (17)$$

where $\bar{z}_{ik}^u = z_{ik}^u$ for $k \neq d$ and $\bar{z}_{id}^u = \tilde{z}_{id}^u$. Similarly, we can get the conditional distribution $p(z_{jd}^g | -)$.

Parameter estimation

The parameters are estimated in a coordinate-wise manner and for each coordinate the gradient descent algorithm is adopted. The parameters α_U and β_U for users can be estimated by maximizing the likelihood,

$$\begin{aligned} & \log L(\alpha_U, \beta_U | \{\mathbf{z}_i^u\}) \\ &= \sum_{u,d} (\log B(\alpha_U + \sum_{i=1}^{N_u} z_{id}^u, \beta_U + N_u - \sum_{i=1}^{N_u} z_{id}^u) - \log B(\alpha_U, \beta_U)). \end{aligned} \quad (18)$$

The partial derivatives are written as follows,

$$\begin{aligned} & \frac{\partial}{\partial \alpha_U} \log L(\alpha_U, \beta_U | \{\mathbf{z}_i^u\}) \\ &= \sum_{u,d} (\psi(\alpha_U + \sum_{i=1}^{N_u} z_{id}^u) - \psi(\alpha_U + \beta_U + N_u) + \psi(\alpha_U + \beta_U) - \psi(\alpha_U)) \quad (19) \\ & \frac{\partial}{\partial \beta_U} \log L(\alpha_U, \beta_U | \{\mathbf{z}_i^u\}) \\ &= \sum_{u,d} (\psi(\beta_U + N_u - \sum_{i=1}^{N_u} z_{id}^u) - \psi(\alpha_U + \beta_U + N_u) + \psi(\alpha_U + \beta_U) - \psi(\beta_U)), \end{aligned} \quad (20)$$

where $\psi(\cdot)$ is the digamma function. Then a gradient descent algorithm can be adopted to estimate the two parameters.

The parameters, μ , λ , σ_U and σ_G , are then sequentially estimated by maximizing the loglikelihood computed from Eqn. 10 in a gradient descent manner. The parameters \mathbf{w}_i^u and \mathbf{w}_j^g are estimated by jointly considering Eqns. 5 and 10 in a gradient descent manner. The prediction function $h(\cdot; \eta, \rho)$ with the Bernoulli distribution is essentially equivalent to the logistic regression with $\max_{d \in \{1, \dots, D\}} \min\{y_d^u, y_d^g\}$ as the training features. At the first iteration of the algorithm, this prediction function is involved into the process. After getting the first estimation of the latent topics, we compute $\max_{d \in \{1, \dots, D\}} \min\{y_d^u, y_d^g\}$ for each pair (u, g) , and update the relation matrix \mathbf{R} by setting entries zeros or ones and leaving the rest known according to the maxmin values. In the later steps, the pairs with known values are used to estimate the parameters.

6 Experiment

In this section, we conduct experiments to evaluate the performance of our proposed approach for Flickr group recommendation and demonstrate its effectiveness by comparing with other widely-used techniques.

6.1 Setup

6.1.1 Dataset

There is no public benchmark of Flickr group recommendation that is available for performance evaluation. We collect the dataset using Flickr API¹. We crawled 200 popular Flickr groups based on the tag-based group search in Flickr, by selecting 20 popular tags and selecting top 10 groups for each tag. Then we collect the information of 53,858 users in these 200 groups, including the photos and the associated tags they uploaded as well as their profile information. The number of photos in our data set is 6,156,124, including 3,711,319 user-uploaded photos, 2,892,271 user-favored photos, and 142,495 photos in the group pools.

For each photo, we extract two types of features: visual feature and textual feature. We extract the dense sampling-based visual words as the visual feature that has been shown to yield better performance than raw visual features such as color or texture as reported. Specifically, we divide the image into uniformly distributed blocks. Then raw SIFT features [17] are extracted for each block, and quantized to visual words [27] with a vocabulary of 1024 visual words obtained by k-means clustering. A histogram of visual words is formed to represent the visual content information. We also extract the textual features from the tags associated with the images. As pre-processing, we remove stop words with the snowball stop word list, and filter the tags whose frequency is less than 5 in the whole corpus, which are viewed as noise or typo. After modeling the continuous latent topic representation for the textual features by latent Dirichlet allocation, we concatenate the textual and visual features together as a whole vector to represent the image.

6.1.2 Evaluation

The performance evaluation metrics of group recommendation [10] consist of three classes, predictive accuracy metric, classification accuracy metric and rank accuracy metric. In this paper, we choose classification accuracy as the evaluation metric, which measures the frequency of a recommender system making correct or incorrect decisions. Specifically, we adopt one popular metric, i.e., the precision at the first N position Precision@N to measure the percentage of groups that the user would like to join. We randomly sample 10% of the user-group links (i.e., the user has joined the group) for prediction, sample another 10% for validation, and use the rest user-group links as the training set.

6.2 Illustration

We show experimental results on the Flickr group recommendation. We first present visual illustration, then show the effectiveness in convergence and correctness of the inference process.

¹ <http://www.flickr.com/services/api/>

6.2.1 Visual illustration

Figure 6 shows the latent topics learned from the joint topic model. We show the representative images from the four latent topics of different users and the associated tags and owners. We have the following observations. On the one hand, they are classified into different semantic meanings. On the other hand, the photos in the same topic may be similar in low-level visual contents (e.g., color and texture) but with missing or noisy tags, or similar in textual information with dissimilar visual contents, or even dissimilar in both visual and textual information but with the same semantic meanings which is thanks to the user-group relations. In summary, this result justified that both user-group relations and content feature are helpful for latent topic discovery.

6.2.2 Convergence

The following illustrates the convergence of the inference process. Figure 7(a) shows the log-joint probability at each iteration and the maximum log-joint probability among this iteration and the previous iterations. We can see that the joint probability reaches a very high value in the early 10-20 iterations, and then becomes a little stable. The theoretical analysis on the convergence of Gibbs-sampling has been discussed in [16]. Our experiments validate that the Gibbs-sampling based inference is effective for our model.

Figure 7(b) shows the mean absolute error (MAE) in each iteration (MAE@iteration) and the minimum MAE among this iteration and the previous iterations (Minimum MAE). The MAE value is calculated by $E \|r_{gu} - r'_{gu}\|$. It describes the error between the actual result r_{gu} and predicted score r'_{gu} , with the lower value showing a better performance. One can see that consistent to Figure 7(a) the MAE value drops much in the early 10-20 iterations and is stably decreasing in the later learning progress.

6.2.3 Latent topics

The best dimensionality of the latent space, i.e., the number of latent topics D , is achieved from the validation set. We check different D from 6, 8, 10 to 50 and test the results over the validation set by calculating the P@N for each K . The results are shown in Figure 8. One can see that the best performance is reached when K is set to 20. K with a smaller value may be not able to discriminate the topics with different meanings, and K with a value higher may generate too many latent topics that separate photos with similar meaning in different topics. Although the validation can help find the optimal D , this process is a little time-consuming for large scale data sets as various D have to be checked. In the future, we will investigate more efficient determination schemes to determine the number of topics.

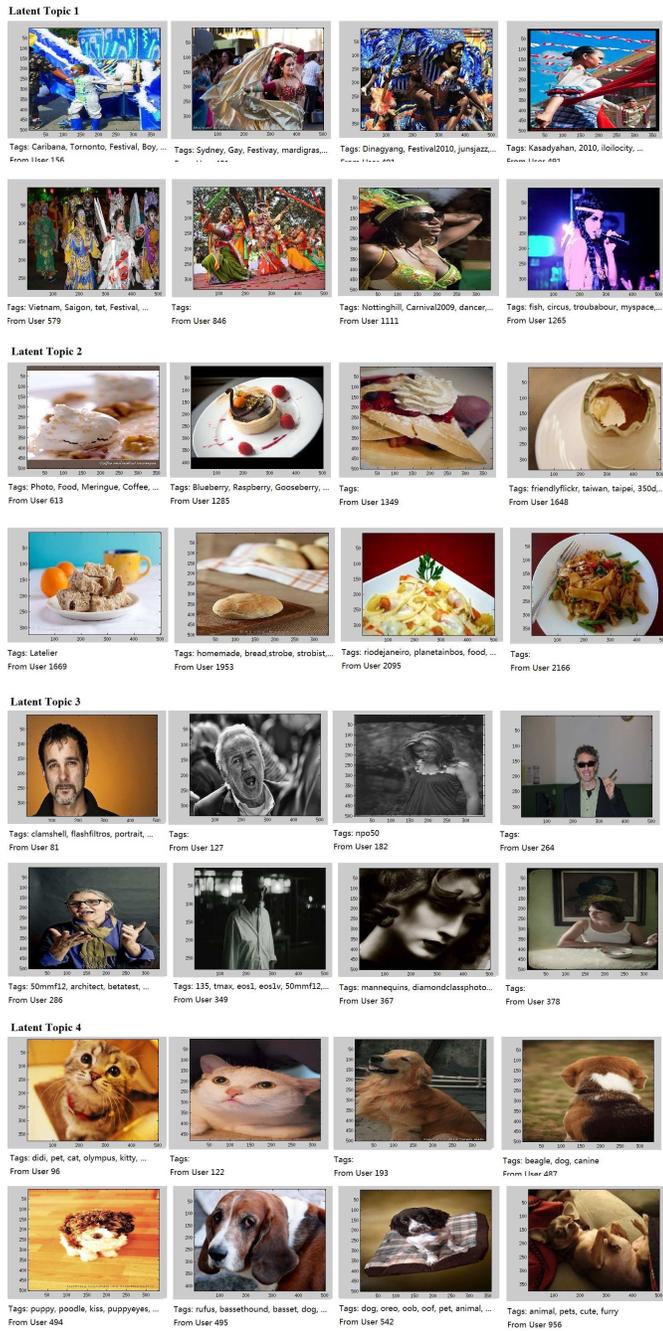


Fig. 6 Visual examples for four random-selected latent topics.

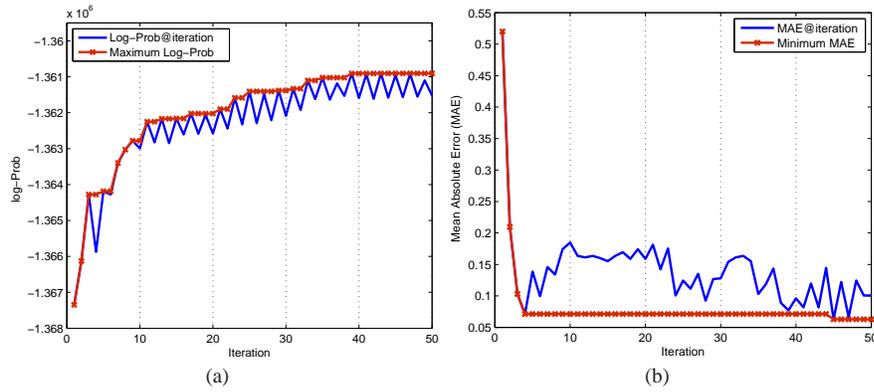


Fig. 7 (a) The change of the log-joint probability with the number of the iterations increases. (b) The change of the mean absolute error with the number of the iterations increases.

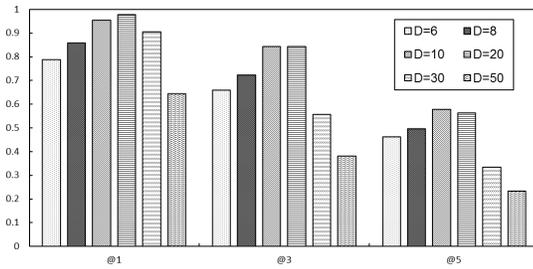


Fig. 8 The performance with different numbers of latent topics.

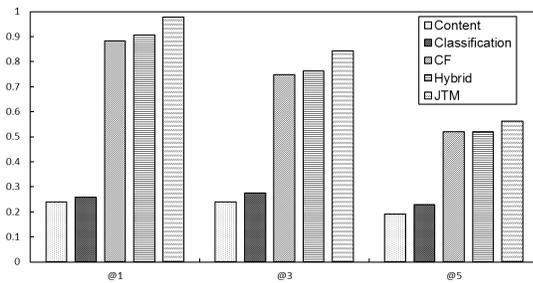


Fig. 9 Performance comparison of our approach (JTM) with other state-of-the-art recommendation methods

6.3 Comparison

We compare the joint topic model with several state-of-the-art techniques adopted in other recommendation applications for various social media applications, which can be directly adopted to solve this problem. According to the taxonomy of recommendation in related work, the competitors fall in the following categories.

- *Content*. We compare our approach with the content analysis approaches, which try to convert all sources of features to an unified space using some methods, such as latent semantic analysis [8], probabilistic latent semantic analysis [11, 12, 24, 22], and latent Dirichlet allocation [3]. We tested all the three types of methods and found their performances are similar. In the comparison, we select the best result from the three methods and name it as content. Besides, we also directly train a classifier for each group over photos using support vector machines. To recommend a user to a group, we class the photos into each group and then use the classification results to vote the group for the user.
- *Collaborative Filtering*. We also compare the result from collaborative filtering methods [10, 15, 28]. and report the result from the recent state-of-the art collaborative filtering approach, called singular value decomposition++ (SVD++) [13, 14]. SVD++ can improve the low rank matrix factorization based methods and handle the cases with missing data.
- *Hybrid*. We also compare our approach with one representative method, Rank-Boost [9]. This method has been applied to Flickr tag recommendation [29] by combining the recommendations from content and collaborative filtering and has been shown to achieve the best result. In the comparison, the method is denoted by hybrid.

Figure 9 shows the comparison results, in which we can clearly see the competitive results of our approach compared with other methods. More specifically, the content-based approach gets the worst performance, due to the low effectiveness of low-level visual features and noises in the tags. However, the collaborative filtering approach performs much better. The hybrid approach, combining both content and collaborative filtering, gains the best results as expected, which is consistent to the results in the other recommendation system.

Our approach simultaneously makes use of the content, and in addition the relations between users and groups, while the content-based approach and the collaborative filtering-based approach only explores one cue. Hence, our approach is more capable to model the problem. In addition, the joint discovery of latent topics for users and groups makes the discovered topics more accurate and more capable of capturing the underlying user-group interaction. Compared with previous hybrid-based methods, our approach models the content using the latent topics, which are learnt from the data. The manner can extract more discriminative and informative features and remove some useless information for our specific problem, the relation mining, while the whole content may affect the performance if useless information is kept. On the other hand, the prediction function and the latent topic learning can benefit from each other. This means that the prediction function is more adaptive to the latent topics, and it can also help on learning latent topics that will furthermore make the prediction stronger. All these make our approach more robust and perform the best among the state-of-the-art approaches.

7 Conclusion

In this paper, we systematically study the problem of recommending groups to users in social media network. A joint topic model is proposed, which makes use of content feature, collaborative information and social relations in an integrated framework. Specifically, the proposed approach discovers the topics for users and groups jointly and simultaneously learns the prediction function of the matching latent topics. All these characteristics of the proposed algorithm make the discovered latent structure more accurately reflect the underlying user-group interaction, which results in a more effective group recommendation. An efficient inference algorithm is proposed based on Gibbs sampling. Experimental results on Flickr group recommendation show that our approach is more effective and efficient compared with the state-of-the-art group recommendation system for social media networks.

Acknowledgements

Bin Cui is supported by the grant of Natural Science Foundation of China (No. 61073019 and 60933004).

References

1. Abbasi, R., Chernov, S., Nejdl, W., Paiu, R., Staab, S.: Exploiting flickr tags and groups for finding landmark photos. In: ECIR, pp. 654–661 (2009)
2. Adomavicius, G., Tuzhilin, A.: 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), 734–749 (2005)
3. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *Journal of Machine Learning Research* **3**, 993–1022 (2003)
4. Boratto, L., Carta, S., Chessa, A., Agelli, M., Clemente, M.L.: Group recommendation with automatic identification of users communities. In: Web Intelligence/IAT Workshops, pp. 547–550 (2009)
5. Cai, J., Zha, Z.J., Tian, Q., Wang, Z.: Semi-automatic flickr group suggestion. In: MMM (2), pp. 77–87 (2011)
6. Chen, H.M., Chang, M.H., Chang, P.C., Tien, M.C., Hsu, W.H., Wu, J.L.: Sheepdog: group and tag recommendation for flickr photos by automatic search-based learning. In: ACM Multimedia, pp. 737–740 (2008)
7. Chen, W., Zhang, D., Chang, E.Y.: Combinational collaborative filtering for personalized community recommendation. In: KDD, pp. 115–123 (2008)
8. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by latent semantic analysis. *JASIS* **41**(6), 391–407 (1990)
9. Freund, Y., Iyer, R.D., Schapire, R.E., Singer, Y.: An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research* **4**, 933–969 (2003)
10. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.: Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* **22**(1), 5–53 (2004)
11. Hofmann, T.: Probabilistic latent semantic indexing. In: SIGIR, pp. 50–57 (1999)
12. Hofmann, T.: Latent semantic models for collaborative filtering. *ACM Trans. Inf. Syst.* **22**(1), 89–115 (2004)
13. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: KDD, pp. 426–434 (2008)
14. Koren, Y.: Factor in the neighbors: Scalable and accurate collaborative filtering. *TKDD* **4**(1) (2010)
15. Koren, Y., Bell, R.M.: Advances in collaborative filtering. In: *Recommender Systems Handbook*, pp. 145–186 (2011)
16. Liu, J.S.: *Monte Carlo Strategies in Scientific Computing*. Springer (2002)

17. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60**(2), 91–110 (2004)
18. Lu, D., Li, Q.: Exploiting semantic hierarchies for flickr group. In: *AMT*, pp. 74–85 (2010)
19. Ma, H., King, I., Lyu, M.R.: Learning to recommend with social trust ensemble. In: *SIGIR*, pp. 203–210 (2009)
20. Ma, H., Yang, H., Lyu, M.R., King, I.: Sorec: social recommendation using probabilistic matrix factorization. In: *CIKM*, pp. 931–940 (2008)
21. McCarthy, K., McGinty, L., Smyth, B.: Case-based group recommendation: Compromising for success. In: *ICCBR*, pp. 299–313 (2007)
22. Negoescu, R.A., Adams, B., Phung, D.Q., Venkatesh, S., Gatica-Perez, D.: Flickr hypergroups. In: *ACM Multimedia*, pp. 813–816 (2009)
23. Negoescu, R.A., Gatica-Perez, D.: Analyzing flickr groups. In: *CIVR*, pp. 417–426 (2008)
24. Negoescu, R.A., Gatica-Perez, D.: Topickr: flickr groups and users reloaded. In: *ACM Multimedia*, pp. 857–860 (2008)
25. Negoescu, R.A., Gatica-Perez, D.: Modeling flickr communities through probabilistic topic-based analysis. *IEEE Transactions on Multimedia* **12**(5), 399–416 (2010)
26. Recio-García, J.A., Jiménez-Díaz, G., Sánchez-Ruiz-Granados, A.A., Díaz-Agudo, B.: Personality aware recommendations to groups. In: *RecSys*, pp. 325–328 (2009)
27. Sivic, J., Zisserman, A.: Efficient visual search of videos cast as text retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(4), 591–606 (2009)
28. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. *Adv. Artificial Intelligence* **2009** (2009)
29. Wu, L., Yang, L., Yu, N., Hua, X.S.: Learning to tag. In: *WWW*, pp. 361–370 (2009)
30. Yu, J., Jin, X., Han, J., Luo, J.: Mining personal image collection for social group suggestion. In: *ICDM Workshops*, pp. 202–207 (2009)
31. Yu, J., Joshi, D., Luo, J.: Connecting people in photo-sharing sites by photo content and user annotations. In: *ICME*, pp. 1464–1467 (2009)
32. Zheng, N., Li, Q., Liao, S., Zhang, L.: Flickr group recommendation based on tensor decomposition. In: *SIGIR*, pp. 737–738 (2010)
33. Zheng, N., Li, Q., Liao, S., Zhang, L.: Which photo groups should i choose? a comparative study of recommendation algorithms in flickr. *J. Information Science* **36**(6), 733–750 (2010)