Multi-Modal Mention Topic Model for mentionee recommendation

Kai Wang\textsuperscript{a,}, Weiyi Meng\textsuperscript{b,}, Shijun Li\textsuperscript{a,\ast,}, Sha Yang\textsuperscript{c}

\textsuperscript{a}Computer School, Wuhan University, Wuhan 430072, China
\textsuperscript{b}Department of Computer Science, Binghamton University, Binghamton 13902, USA
\textsuperscript{c}School of Computer Science and Technology, Hnkou University, Wuhan 430072, China

\begin{abstract}
As one of the most commonly used communication functions in Twitter-like social media systems, mention is playing an important role in users’ online interactions. With the dramatic increase in the number of social media users, the problem of mentionee recommendation, i.e., recommending mentionees (mentioned users) when mentioners (mentioning users) attempt to mention others, has received considerable attention in recent years. While an increasing line of work has studied this problem, the existing efforts focus only on the contribution of the non-visual data like post text. In fact, many social media posts contain not only textual but also visual contents like images, and these two heterogeneous data sources both describe users’ mentioning tendencies. In this work, we proposed a novel generative model, named Multi-modal Mention Topic Model (MMTM), to tackle the mentionee recommendation problem by learning users’ semantic patterns and the correlations between contents in different modalities of users’ multi-modal mentioning documents in a unified way. Extensive experiments were conducted on a real-world dataset to evaluate the performance of our method. The experiment results demonstrated the superiority of our method in terms of making more effective recommendations compared with other state-of-the-art methods.
\end{abstract}

\section{Introduction}

With the rapid development of the social networks, more and more people are beginning to use social media as their daily communication tool. In the meantime, it impels social media services to develop new interaction functions to support users’ online communications. As one of the most commonly used ones, mention is playing a significant role in users’ online interactions due to its unique advantages in alleviating information overload issues \cite{5,36}. In social media like Twitter, Facebook and Sina Weibo, a mention is occurred when individual mentioner\textsuperscript{1} add the “@” symbol before a mentionee’s username in a post. Nowadays, social media is holding a huge and growing number of users. In late 2017, there were 330 million monthly active users on Twitter, who have published over 500 million tweets per day, while the average number of friends per user was 707.\textsuperscript{2} Hence, when people intend to mention others in a post, a small list of suggested mentionees would definitely be helpful.

While an increasing line of effort have been undertaken to tackled this user recommendation problem, all of them considered only the contribution of the non-visual data like text \cite{14,28,34,36}. In fact, along with the enrichment of information uploading channels, one significant trend of social media lies on the rapid increasing amount of multi-modality information. According to a recent investigation, there were more than 40% of posts in Twitter contain not only text but also images.\textsuperscript{3} Studies on Sina Weibo showed a similar trend that 34% of Weibo posts contain at least one image, and a large number of posts with images do not contain relevant textual descriptions \cite{13}. After analyzing several social media services, we observed that the visual resources can also provide valuable information to reveal users’ mentioning tendencies. For example, Fig. 1 illustrates two multi-modal mentioning posts from Twitter, in which the mentionee “GGBrige” is the official account of the Golden Gate Bridge operator in San Francisco, US. We can see that with only the text, the correct mentionee is hardly to be identified.

In this paper, we aim to construct a mentionee recommender for common social media users based on a joint model which explores the impact of both the textual and the visual resources on

\textsuperscript{1} Corresponding author.
\textsuperscript{2} For simplicity and clarity, we use the terms “mentioner” and “mentionee” to denote the mentioning and mentioned user, respectively.
\textsuperscript{3} https://thenextweb.com/socialmedia/2015/11/03/what-analyzing-1-million-tweets-taught-us/

\url{https://doi.org/10.1016/j.neucom.2018.10.024}
0925-2312/© 2018 Elsevier B.V. All rights reserved.
users’ mentioning tendencies. To achieve that, there are several challenges we have to face. First, social media users are allowed to select mentionees beyond their neighborhood, which means anyone can be the candidate and leads to an extremely large candidate space. This further disables the traditional classification methods like Naive Bayes on this task. Second, the recommendation scenario of this task should be highly personalized since users have different preferences. It is very challenging to capture individual user’s characters precisely from the massive and knowledge-sparse social media data. The problem is more severe when considering the sparsity of user-word matrix in social media services [30,31]. Third, instead of processing user-generated data of the two modalities separately and synthesizing the results, we try to model users' mentioning activities by exploiting the data with different modalities in a unified way. According to the best of our knowledge, little attention has been paid on this task.

To address these problems, we propose a latent-class generative model, named Multi-modal Mention Topic Model (MMTM), to simulate the generating process of users’ mentioning activities by synthetically exploiting the textual and visual contents. Specifically, MMTM adopts the bag-of-words method for both texts and images to alleviate the data sparsity while describing multi-modal social media posts. Two key latent features are proposed in MMTM – textual topic and visual topic – which are responsible for generating the multi-modal attributes (i.e., textual words and visual features) of users' mentioning activities. Based on these two latent features, MMTM jointly learns users’ semantic patterns and the correlations between contents in different modalities. After training the MMTM to obtain a knowledge model containing the necessary insights about users’ mentioning tendencies, we retrieve the top-k mentionees for a multi-modal mentioning post.

Compared with the existing studies, our contributions can be summarized as follows:

- We studied the task of mentionee recommendation based on the multi-modal property of social media posts. To the best of our knowledge, this is the first work focusing on this task.
- We proposed a novel generative approach to model users’ mentioning activities by exploiting the textual and visual contents in a unified way.
- We constructed extensive experiments on a large real-world dataset. The experiment results demonstrated the effectiveness of our approach.

2. Related work

This work is related to two research threads: social media user recommendation and multi-modal Latent Dirichlet Allocation. In this section, we first summarize the related work and then give a brief description of the differences between the existing work and our approach.

2.1. User recommendation on social media

User recommendations have become a rich research area within the broad recommender systems community and social recommendations in particular. Extensive work has been undertaken to tackle a variety of user recommendation problems on social media such as friend and follower recommendation [16,17], expert recommendation [10,12], retweet recommendation [29] and mention-related recommendations [14,22,27,28,34,36]. In particular, the mentionee recommendation problem has been studied from different aspects in the past few years. For example, Wang et al. [28] tried to find the right persons to mention to enhance a tweet’s diffusion. They formulated the task as a learning-to-rank problem and trained a ranking function with multiple features. The main goal of their work was to make a tweet spread more quickly and widely. Zhou et al. [36] studied the mentionee ranking problem from the perspective of information overload issues. The CAR framework developed by Tang et al. in [27] is a context-aware approach aiming to find target audiences who can help a promotion-oriented post get a higher response rate. They employed the Support Vector Machine model to resolve the ranking based mentionee recommendation problem after extracting related features. More recently, Gong et al. proposed A-UUTTM [14], a topical translation model incorporating the current text and the histories of mentionees to perform the task, which is also the state-of-the-art approach used for mentionee recommendation problem.

2.2. Latent Dirichlet Allocation for multi-modal documents

Topic models provide us an unsupervised tool to analyze not only content but many types of discrete data. As one of the most important topic models, Latent Dirichlet Allocation (LDA) [4] and its many variants have been demonstrated to be very useful in discovering topical structures from large datasets and generating high-quality knowledge models for recommendation systems.
However, the traditional LDA was designed to represent document themes on textual corpora, which may not be appropriate for analyzing multi-modal data. To cope with that, some extended models were proposed to describe the probabilistic distributions of both images and texts in multi-media documents, such as multi-modal LDA (mm-LDA) [1], correspondence LDA (Corr-LDA) [3], topic-regression multi-modal LDA (tr-mmLDA) [23], multi-modal event topic model (mmETM) [24], hashtag multi-modal LDA [13], sparse relational multi-modal topical coding (SR-MMTC) [25] and so on. Different from the multi-modality data fusion methods [33], all of these multi-modal topic models have nice property of mining multi-modal knowledge in a more semantically interpretable way. Specifically, the mm-LDA [1] and Corr-LDA [3] were proposed to capture the topic-level relevance and learn the joint distribution between image and annotated words. They assumed a one-to-one correspondence between the topic of each modality. In contrast, the tr-mmLDA [23] used two separate sets of hidden topics to represent the different data modalities and introduced a linear regression module to correlate them. The SR-MMTC [25] is a non-probabilistic formulation of the relational topic model, which models both long multi-modal documents and the corresponding link information among them. Based on the traditional mm-LDA, the mmETM [24] was proposed to capture the multi-modal topics of social events with long text and related images by considering non-visual-representative topics. More recently, to recommend hashtags for the multi-modal microblog posts, the hashtag multi-modal LDA [13] is designed to model the generation process of hashtags in multi-modal posts.

2.3. Differences of our method from existing approaches

The distinction between our method and the existing approaches can be summarized as follows. First, the existing mentionee recommenders has mainly focused on the information propagation aspect of mention mechanism. In other words, they were aiming to find target users who have the maximal capability and possibility to spread a post faster and further. Nevertheless, mention functions as not only an information propagation channel but also a communication tool in social media [27]. Intuitively, the communication aspect of the mention is more valuable to common users than to the media workers and advertisers. Moreover, existing studies have suggested that information propagation relies more on retweet than the mention in social media [20,21]. Hence in this work, the aim of mention we focus on is not limited to the spreading of a post but its generalized attribute of common users’ online interactions. Second, we observed that both of the textual and visual content of posts can provide important clues on users’ mentioning tendencies. Instead of simply employing non-visual features, we address the problem by exploiting the visual and textual modalities synchronously. Experimental results demonstrated that the visual resources of users’ mentioning activities are very helpful for the performance improvement of mentionee recommender. Third, due to the randomness of social users communications through mention utility, the well used collaborative filtering method [7,26,29] may not be appropriate for tackling our problem. In this work, we learn users’ mentioning tendencies based on a novel multi-modal topic model. Most of the existing multi-modal topic models were constructed on a key provision that there should be a strong correlation between the visual and textual resources. However, a lot of multi-modal documents from social media systems does not satisfy the constraint due to the length limitation and the irregularity of the user-generated content as the example shown in Fig. 1. Further, most of the former studies assumed a one-to-one correspondence between latent variables and multi-modal features, which restricted the model to handling either documents with long text like news or document with discrete words like image tags. In this work, the proposed MMTM adopts different topic sets to represent the two modalities and assumes a single textual topic and a single visual topic for each document as a single post is more likely to talk about one topic [11,35]. Through these designs, MMTM can better describe the user-generated multi-modal documents in social media, especially when mentioning instances are incorporated.

3. Our proposed method

In this section, we first introduce the key notations and concepts used in our work and then formulate the problem definition, followed by the model structure and the parameter inference of MMTM.

3.1. Problem formulation

Table 1 shows the key notations of our proposed MMTM. The concepts are defined as follows.

Definition 1 (Mentioner and mentionee). We denote a user \( u \) as a mentioner if she/he has mentioned others (i.e., mentionees) in at least one mentioning post. Clearly, "mentioner" and "mentionee" are relative concepts, i.e., a user can be both a mentioner in one post and a mentionee in another. We use \( U \) and \( M \) to denote the sets of all mentioners and mentionees, respectively.

Definition 2 (User mentioning activity). In this work, a user mentioning activity \( d \) is represented by a 4-tuple \( (u, W_d, F_d, M_d) \) denoting that a mentioner \( u \) mentions mentionees \( M_d \) with words \( W_d \) of related text and visual features \( F_d \) of related images. Note that \( M_d \) denotes a set of mentionees since a mentioner may mention multiple mentionees in a single post. \( W \) and \( F \) are used to denote the sets of all textual words and visual features.

Definition 3 (User mentioning document). For each mentioner \( u \), the mentioning document \( D_u \) refers to a collection of all user mentioning activities of \( u \). We use \( D \) to denote all the mentioning documents, i.e., \( D = \{ D_u | u \in U \} \).

Based on the above definitions, the problem can be formally defined as follows. Given a dataset \( D \) as a collection of all users’ mentioning documents, for a querying post published by \( u_q \) with words \( W_q \) and visual features \( F_q \), i.e., the query is \( q = (w_q, W_q, F_q) \), our goal is to recommend a set of top-k mentionees \( M_q \) that \( u_q \) most likely to mention.

3.2. Multi-modal Mention Topic Model

In this work, we propose a latent-class generative model, named Multi-modal Mention Topic Model (MMTM), to simulate the generating process of users’ mentioning activities. Fig. 2 shows the graphical representation of MMTM, where the observations, e.g., words, are shown as shaded circles, and the hidden variables e.g., topics are shown as unshaded ones. In this section, we give a detail description of the model structure of MMTM.

Generally, MMTM is a generative probabilistic model jointly over the text and images. To alleviate the user-word sparsity, we adopt the well-used bag-of-word method for both texts and images like existing topic model based approaches [8,13,24,25]. Specifically, the textual vocabulary \( W = \{ w_1, w_2, \ldots, w_{|W|} \} \) denoted by a collection of \( |W| \) distinct textual words, and the visual vocabulary \( F = \{ f_1, f_2, \ldots, f_{|F|} \} \) denoted by a collection of \( |F| \) distinct visual features. Most of existing multi-modal topic models assumed that each document contains a mixture of topics and assigned a topic to each feature of one document \([1,3,13,23,24]\). While this assumption makes sense when modeling documents with long texts like news or documents with discrete words like image tags, it does not
Table 1
Key notations in MMTM.

<table>
<thead>
<tr>
<th>Notations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u, m, z, v, w, f$</td>
<td>Index of mentioner, mentionee, textual topic, visual topic, textual word and visual word, respectively.</td>
</tr>
<tr>
<td>$K_u, K_v$</td>
<td>The numbers of textual and visual topics, respectively.</td>
</tr>
<tr>
<td>$\theta_u, \theta_v$</td>
<td>The multinomial distributions over textual and visual topics for user $u$, respectively.</td>
</tr>
<tr>
<td>$\beta_u, \beta_v$</td>
<td>The multinomial distribution over textual words for topic $z$, and the background multinomial distribution over textual words.</td>
</tr>
<tr>
<td>$y$</td>
<td>A switch that determines which distribution the word is generated from, $y = 0$ or $1$.</td>
</tr>
<tr>
<td>$\phi_v$</td>
<td>The multinomial distribution over visual features for topic $v$.</td>
</tr>
<tr>
<td>$\pi_{z, w}$</td>
<td>The multinomial distribution over mentionees specific to textual topic $z$ and word $w$.</td>
</tr>
<tr>
<td>$\pi'_{v, f}$</td>
<td>The multinomial distribution over mentionees specific to visual topic $v$ and visual feature $f$.</td>
</tr>
<tr>
<td>$x$</td>
<td>A switch that determines which distribution the mentionee is generated from, $x = 0$ or $1$.</td>
</tr>
<tr>
<td>$\tau_u$</td>
<td>The mixing weight specific to user $u$, representing the influence probability of the textual modality.</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Beta priors to generate $\tau_u$, denoted by $\lambda = (\lambda_1, \lambda_2)$.</td>
</tr>
</tbody>
</table>

As shown in Fig. 2, MMTM is a mixed model which learns users’ topical interest distributions over both the textual and visual topics. Specifically, we use a smoothing parameter $\tau_u$ to denote the influence probability of the textual content on $u$’s mentionee selection. In this way, give a query $q = (u, W, F)$, the likelihood that user $u$ chooses mentionee $m$ with a serial of textual words $W$ and visual features $F$ is calculated as follows:

$$P(m|u, W, F, \Phi) = \tau_u P(\theta_u, y, \beta_v, \pi_{z, w}) + (1 - \tau_u) P(m|\theta'_u, \phi_v, \pi'_{v, f})$$

(1)

where $\Phi$ denotes the parameter set of MMTM (i.e., $\Phi = \{\theta_u, \theta'_v, \beta_v, \phi_v, \pi_{z, w}, \rho, \tau\}$). $P(m|\theta_u, y, \beta_v, \pi_{z, w})$ is the probability that mentionee $m$ is generated according to the textual content derived semantic pattern of $u$, and $P(m|\theta'_u, \phi_v, \pi'_{v, f})$ is the probability that mentionee $m$ is generated according to $u$’s visual content derived semantic pattern. In other words, the generation of mentionee $m$ for the multi-modal post $d$ is influenced by $u$’s semantic pattern on the textual content of $d$ with probability $\tau_u$ and on the visual content of $d$ with probability $1 - \tau_u$. Note that, the weight $\tau_u$ is specific to the mentionee $u$. Considering the difference between users in personalities, MMTM holds personalized mixture weights for individual mentioners. Based on that, we can learn the correlations between the content information in different modalities of users’ mentioning documents.

Accordingly, the generative process of MMTM can be summarized as Algorithm 1.

3.3. Model inference

In this section, we show how to learn the parameters of our model, i.e., $\Phi$. In general, we need to learn parameters that maximize the marginal log-likelihood of the observed random variables $w, m$ and $f$. The marginalization is performed with respect to the latent random variables $z$ and $v$. However, the marginal log-likelihood cannot be computed tractably due to the coupling among the latent variables. Therefore, we devise an approximate learning method based on collapsed Gibbs sampling [15] to maximize the complete data likelihood as shown in Eq. (2). We assume that the priors follow symmetric Dirichlet, which are conjugate priors for multinomial (or beta for Bernoulli). For simplicity, we use fixed values for the hyperparameters, i.e., we set $\alpha = 50/K_z, \alpha' = 50/K_v, \beta = \gamma = \xi = \xi' = 0.01, \lambda_1 = \lambda_2 = 0.5$ as [15](5) suggested.

$$P(z, v, W_d, F_d, M_d, x, y|\alpha, \alpha', \beta, \gamma, \epsilon, \lambda, \xi, \xi') = P(z|\alpha)P(y|z)P(W_d|z, y, \beta)P(v|\alpha')P(F_d|v, \gamma) \times P(x|\lambda)P(M_d|z, v, W_d, F_d, x, \xi, \xi').$$

(2)

For Gibbs sampler, we need to derive the posterior probabilities for sampling latent textual topic $z$, latent visual topic $v$, latent
Algorithm 1: The generative Process of MMTM

1. Draw $\rho \sim \text{Beta}(\bar{\epsilon}, \bar{\delta})$, $\theta_d \sim \text{Dirichlet}(\beta)$;
2. for each textual topic $z \in T_d$ do
   3. Sample a distribution over textual words $\theta_z \sim \text{Dirichlet}(\cdot | \beta)$;
   4. for each textual word $w \in W$ do
      5. Sample a distribution over textual topics and words
         $\pi_{z,w} \sim \text{Dirichlet}(\cdot | \xi_z)$;
   6. for each visual topic $v \in T_v$ do
      7. Sample a distribution over visual features $\phi_v \sim \text{Dirichlet}(\cdot | \psi)$;
      8. for each visual features $f \in F$ do
         9. Sample a distribution over visual topics and features
            $\pi'_{v,f} \sim \text{Dirichlet}(\cdot | \xi'_v)$;
   10. for each $u \in U$ do
       11. Sample her distribution over textual topics $\theta_{hu} \sim \text{Dirichlet}(\cdot | \alpha)$;
       12. Sample her distribution over visual topics $\theta'_{hu} \sim \text{Dirichlet}(\cdot | \alpha')$;
   13. for each $D_a \in D$ do
      14. for each mentioning activity $d \in D_a$ do
         15. Sample a textual topic $z \sim \text{Multi}(\cdot | \theta_d)$;
         16. for each token $w \in W_d$ do
            17. Toss a coin $\nu_{d,w}$ according to Bernoulli($\rho$);
            18. if $\nu_{d,w} = 0$ then
               19. Sample word $w \sim \text{Multi}(\cdot | \theta_z)$;
            20. else
               21. Sample word $w \sim \text{Multi}(\cdot | \theta_{hu})$;
            22. Sample a visual topic $v \sim \text{Multi}(\cdot | \theta_v)$;
         23. for each feature $f \in F$ do
            24. Sample feature $f \sim \text{Multi}(\cdot | \phi_v)$;
         25. for each mentionee mentioned in $d$, $m \in M_d$ do
            26. Toss a coin $x_{d,m}$ according to Bernoulli($\tau_m$) $\sim \text{Beta}(\lambda)$;
            27. if $x_{d,m} = 0$ then
               28. Sample mentionee $m \sim P(\cdot | z, W_d, \pi_{z,w})$;
            29. else
               30. Sample mentionee $m \sim P(\cdot | v, F_d, \pi'_{v,f})$;

switch $x$ and latent switch $y$ for each user mentioning activity. Due to space limitations, we omit the derivation details here. Given the joint probability distribution of all variables as shown in Eq. (2), the sampling probability of a latent textual topic $z$ for user mentioning activity $d$ is calculated as in Eq. (3) according to the Bayes chain rule.

\[
P(Z_{(u,d)}) = k|z_{(u,d)}, v, W_d, F_d, M_d, x, y, u)
\]

\[
\propto n_{D_a,k}^{D_a,k} \cdot \alpha \cdot \prod_{w \in W_d} n_{k,w}^{k,w} \cdot \sum_{m \in M_d} \frac{n_{m,w}^{m,w} \cdot \xi_m}{\sum_{w' \in W_d} n_{k,w'}^{k,w'}}
\]

\[
\times \prod_{m \in M_d} \frac{n_{D_a,m}^{D_a,m} \cdot \xi'}{\sum_{w' \in W_d} n_{m,w'}^{m,w'}}
\]

(3)

where $n_{D_a,k}^{D_a,k}$ is the number of activities assigned to topic $k$ of $u$, $n_{k,w}^{k,w}$ is the number of times that word $w$ is generated by topic $k$, $n_{m,k}^{m,k}$ is the number of times that word $w$ and mentionee $m$ co-occur in the same activity under the topic $k$, $-(u, d)$ denotes that all the counts are calculated without taking account of the current activity $d$ of $u$.

Similarly, we can sample visual topic $v$ according to the following posterior probability:

\[
P(v_{(u,d)} = k|v_{-(u,d)}, z, W_d, F_d, M_d, x, y, u)
\]

\[
\propto n_{D_a,k}^{D_a,k} + \alpha' \cdot \prod_{f \in F_d} \sum_{f \in F_d} n_{f}^{f} \cdot \sum_{m \in M_d} \frac{n_{m,k}^{m,k} \cdot f}{\sum_{f' \in F_d} n_{m,k}^{m,k} \cdot f' + \xi'}
\]

\[
\times \prod_{m \in M_d} \frac{n_{D_a,m}^{D_a,m} \cdot \xi'}{\sum_{w' \in W_d} n_{m,w'}^{m,w'}}
\]

(4)

where $n_{D_a,k}^{D_a,k}$ is the number of activities assigned to topic $k$ of $u$, $n_{f}^{f}$ is the number of times that visual feature $f$ is generated by topic $k$, $n_{m,k}^{m,k}$ is the number of times that feature $f$ and mentionee $m$ co-occur in the same activity under the topic $k$, the number $-(u, d)$ denotes a quantity excluding the current instance.

Then, the textual word generation coin $y$ can be sampled according to the following equations:

\[
P(y_{(u,d,w)} = 0|y_{-(u,d,w)}, z, v, W_d, F_d, M_d, x, y, u)
\]

\[
\propto n_{D_a,0}^{D_a,0} \cdot \sum_{y' \in \{0, 1\}} \frac{n_{y',y}^{y',y} \cdot y + \xi \cdot \sum_{w' \in W_d} n_{y',y}^{y',y} \cdot w + \beta \cdot \xi}{\sum_{y' \in \{0, 1\}} \frac{n_{y',y}^{y',y} \cdot y + \xi \cdot \sum_{w' \in W_d} n_{y',y}^{y',y} \cdot w + \beta \cdot \xi}}
\]

(5)

\[
P(y_{(u,d,w)} = 1|y_{-(u,d,w)}, z, v, W_d, F_d, M_d, x, y, u)
\]

\[
\propto n_{D_a,1}^{D_a,1} \cdot \sum_{y' \in \{0, 1\}} \frac{n_{y',y}^{y',y} \cdot y + \xi \cdot \sum_{w' \in W_d} n_{y',y}^{y',y} \cdot w + \beta \cdot \xi}{\sum_{y' \in \{0, 1\}} \frac{n_{y',y}^{y',y} \cdot y + \xi \cdot \sum_{w' \in W_d} n_{y',y}^{y',y} \cdot w + \beta \cdot \xi}}
\]

(6)

where the number $n_{y=0}$ is a count of topic words (i.e., words are generated from the topic-word distribution) and $n_{y=1}$ is a count of background words (i.e., words are generated from the background word distribution), $n_{w,y=0}$ is the number of times that word $w$ appears as a topic word and $n_{w,y=1}$ is the number of times that word $w$ occurs as a background word, $-(u, d)$ denotes that all the counts are calculated without taking account of the current word $w$ in mentioning activity $d$ of $u$.

At last, given the variable states other than the latent switch $x$, the sampling probability is calculated as follows:

\[
P(x_{(u,d,m)} = 0|x_{-(u,d,m)}, z, v, W_d, F_d, M_d, x, y, u)
\]

\[
\propto \frac{n_{x=0}^{x=0} \cdot \sum_{w \in W_d} n_{x=0}^{x=0} \cdot w + \beta + \lambda_1 \cdot \lambda_2 \cdot \sum_{m \in M_d} \frac{n_{W,m}^{w,0} \cdot x + \xi}{\sum_{w'=w} n_{W,m}^{w,0} \cdot w' + \xi}}
\]

\[
\times \prod_{m \in M_d} \frac{n_{x=0}^{x=0} \cdot \sum_{w \in W_d} n_{x=0}^{x=0} \cdot w + \beta + \lambda_1 \cdot \lambda_2 \cdot \sum_{m \in M_d} \frac{n_{W,m}^{w,0} \cdot x + \xi}{\sum_{w'=w} n_{W,m}^{w,0} \cdot w' + \xi}}
\]

(7)

\[
P(x_{(u,d,m)} = 1|x_{-(u,d,m)}, z, v, W_d, F_d, M_d, x, y, u)
\]

\[
\propto \frac{n_{x=1}^{x=1} \cdot \sum_{w \in W_d} n_{x=1}^{x=1} \cdot w + \beta + \lambda_1 \cdot \lambda_2 \cdot \sum_{m \in M_d} \frac{n_{W,m}^{w,1} \cdot x + \xi}{\sum_{w'=w} n_{W,m}^{w,1} \cdot w' + \xi}}
\]

\[
\times \prod_{m \in M_d} \frac{n_{x=1}^{x=1} \cdot \sum_{w \in W_d} n_{x=1}^{x=1} \cdot w + \beta + \lambda_1 \cdot \lambda_2 \cdot \sum_{m \in M_d} \frac{n_{W,m}^{w,1} \cdot x + \xi}{\sum_{w'=w} n_{W,m}^{w,1} \cdot w' + \xi}}
\]

(8)

where the number $n_{x=0}$ is the number of times that mentionees are generated under the influence of $u$'s semantic pattern on the textual content of $d$, $n_{x=1}$ is the number of times that mentionees are generated under the influence of $u$'s semantic pattern on the visual content of $d$, $n_{W,m}^{w,0}$ is the number of times that mentionee $m$ is generated from textual topic-word pairs (i.e., $(z, w)$, $w \in W_d$), $n_{W,m}^{w,1}$ is the number of times that mentionee $m$ is generated from visual topic-feature pairs (i.e., $(v, f) \in F_d$), $-(u, d, m)$ denotes that all the counts are calculated without taking account of the current mentionee $m$ in user activity $d$ of $u$.

After a sufficient number of sampling iterations, we can make the following estimation of the model parameters with
the collapsed Gibbs sampler by using the calculated approximate posteriors.

\[
\hat{\beta}_{u.z} = \frac{\eta^{u.z} + \alpha}{\sum_{z \in T} (\eta^{u.z} + \alpha)},
\]

(9)

\[
\hat{\gamma}_{u} = \frac{\eta^{u} + \alpha'}{\sum_{z \in T} (\eta^{u.z} + \alpha')},
\]

(10)

\[
\hat{\beta}_{w} = \frac{\eta^{w} + \beta}{\sum_{w \in D} (\eta^{w} + \beta)},
\]

(11)

\[
\hat{\nu}_{f} = \frac{\eta^{f} + \gamma}{\sum_{f \in T} (\eta^{f} + \gamma)},
\]

(12)

\[
\hat{\tau}_{z,m} = \frac{\eta^{z,m} + \xi}{\sum_{m \in M} (\eta^{z,m} + \xi)},
\]

(13)

\[
\hat{\tau}_{v,f,m} = \frac{\eta^{v,f,m} + \xi'}{\sum_{m \in M} (\eta^{v,f,m} + \xi')},
\]

(14)

\[
\hat{\lambda}_{m} = \frac{\eta^{m} + \lambda_{1} + \lambda_{2}}{\sum_{m \in M} (\eta^{m} + \lambda_{1} + \lambda_{2})}.
\]

(15)

Moreover, to alleviate data sparsity problem when estimating the topic-word-mentionee matrices \( \pi \) and \( \pi' \), we adopt a linear interpolation with the topic-free word alignment probability following the work [11,18]. Specifically, we smooth the translation probability using a balance parameter \( \mu \) as follows:

\[
\begin{align*}
\pi^{\text{f}} & = \mu \pi + (1 - \mu) P(m|w), \\
\pi^{\text{t}} & = \mu \pi' + (1 - \mu) P(m|f),
\end{align*}
\]

(16)

where \( P(m|w) \) is the topic-free word alignment probability between the word \( w \) and mentionee \( m \), \( P(m|f) \) is the topic-free word alignment probabilities between the visual feature \( f \) and mentionee \( m \). In this work, we adopt the well-used word alignment model IBM-1 [6] to calculate these two topic-free word alignment probabilities. As for the smoothing parameter, we fixed it at \( \mu = 0.8 \) following the empirical studies in [13,14,18].

3.4. Mentionee recommendation

After learning the MMTM to obtain a knowledge model containing the necessary insights about users’ mention tendencies (i.e., the parameter set \( \Phi = \{ \beta, \hat{\beta}, \hat{\nu}, \hat{\tau}, \hat{\pi}, \hat{\lambda} \} \), we retrieve the top-k mentionees for a multi-modal mentionee post.

Given a query \( q = [u_{q}, W_{q}, F_{q}] \), i.e., mentionee \( u_{q} \) publishes a mention post with words \( W_{q} \) and visual features \( F_{q} \), the ranking score of a mentionee \( m \) for query \( q \) can be calculated as follows:

\[
P(m|u_{q}, W_{q}, F_{q}, \Phi) = \sum_{z \in T} \sum_{v \in T} P(z|u_{q}) P(v|u_{q}) \left[ \hat{\lambda}_{z} P(m|z, W_{q}, \Phi) \right] + (1 - \hat{\lambda}_{z}) P(m|v, F_{q}, \Phi) \]

(1)

\[
\times \sum_{z \in T} \sum_{v \in T} \theta_{z,v} \phi_{w} \left[ \hat{\lambda}_{z} \prod_{w \in W_{q}} P(w|W_{q}) \pi_{t,m} \right] + (1 - \hat{\lambda}_{z}) \prod_{f \in F_{q}} P(f|F_{q}) \pi_{v,f,m} \]

(17)

where the component \( P(w|W_{q}) \) represents the weight of the textual word \( w \) in query words \( W_{q} \), as well as \( P(f|F_{q}) \) denotes the weight of visual feature \( f \) in query features \( F_{q} \). In this work, we use the Inverse Document Frequency (IDF) score of the word or the visual feature to estimate them. The IDF is a simple but effective metric to evaluate the importance of a keyword in a document and can be applied seamlessly to estimate the weight of an individual word in query words in our problem. Then, for query \( q \), we can retrieve the top-k mentionees with the highest ranking scores as the recommendation result. It is worth to mention that although the ranking score of individual mentionee for a query is computed online, the main components of \( P(m|u_{q}, W_{q}, F_{q}, \Phi) \) (i.e., \( \hat{\lambda}_{z}, \phi_{w}, \theta_{z,v}, \pi_{t,m}, \pi_{v,f,m} \)) can be pre-computed offline. At the query time, after calculating the weights of the words and visual features of the query post (this process does not take an expensive time cost), the online computation is just a simple synthesis process.

4. Experiments

In this section, we detail an experimental study of the mentionee recommendation problem and evaluate the performance of our methods compared with other state-of-the-art methods. The goal of our experiments is to understand: (1) whether our proposed method is effective; (2) whether the visual resources can help improve the performance of mentionee recommendation; (3) what impact the data of different modalities have on mentionee recommendation; (4) how the parameters of MMTM affect the quality of recommendation.

4.1. Dataset collection and configuration

Our experiments are conducted on a real dataset crawled from Twitter. We started by collecting user profiles based on snowball sampling strategy following the work [14]. Along with the following-follow networks of 10 selected initial users (each of them has about 500 followers), we crawled the profiles of each user and their followers. Then, the posts published by these users between October 15, 2016 and March 15, 2017 were collected. Through these steps, we had an initial dataset consisting of 1.7 million user profiles and 127.1 million posts. Since this work focuses on analyzing users’ multi-modal mentioning posts, we constructed the mentioning network by selecting posts with at least one mentioning instance (i.e., “username”) and one image. In this step, we filtered out mentionees who have published fewer than 5 multi-modal mentioning posts, and mentionees who have been mentioned fewer than 5 times. At last, the experimental dataset we constructed contained 957,441 multi-modal mentioning posts published by 106,842 mentionees, and 157,285 distinct mentionees.

Table 2 lists the descriptive statistics of our dataset.

Table 2
Statistics of our constructed dataset.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td># of multi-modal mentioning posts</td>
<td>957,441</td>
</tr>
<tr>
<td># of mentioners</td>
<td>106,842</td>
</tr>
<tr>
<td># of mentionees</td>
<td>157,285</td>
</tr>
<tr>
<td># of mention-mentioned relations</td>
<td>1,284,074</td>
</tr>
<tr>
<td># of Avg. mentioning cases per mentioner</td>
<td>12.02</td>
</tr>
<tr>
<td># of Avg. mentioning cases per post</td>
<td>1.34</td>
</tr>
</tbody>
</table>

For textual data, we simply used word frequency statistics to extract textual features. We first counted the number of words and filtered out those that less than 10. After removing stop words, we build a textual vocabulary by using the remaining words. For visual data, we preprocessed the images following the work [13,19]. Specifically, we first resized the images to the size of 224 x 224 and segment each image into some 20 x 20 patches by sliding a window with a 20-pixel interval. Then, we quantized a sampled set of the SIFT descriptors in [19] into 10,000 clusters using k-means.
Each center of clusters was treated as a codeword in the visual vocabulary.

Note that, for a mentioning activity \( d = (u, W_d, F_d, M_d) \), since the norm of \( M_d \) may be greater than 1, we denoted each \( (u, W_d, F_d, m) \), \( m \in M_d \) as a training/test case. In other words, the number of mentioning cases in the dataset was decided by the number of mention-mentioned relations rather than the number of posts. To construct the training and testing sets, we divide our dataset into 5 folds, with the goal of performing a 5-fold cross-validation.

4.2. Evaluation metrics and comparative methods

For each test case \((u, W_d, F_d, m)\) in the test set, we formed a recommendation list \( V_b \) by selecting the k mentionees with the top-ranked scores. For the ground-truth mentionee \( m \), if \( m \in V_b \), we considered it as a hit; otherwise, we considered it as a miss. To evaluate the performance of all methods, we used the well-known Precision, Recall, and F1-Score measures for the highest ranked results. Moreover, we also used the Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (NDCG) metrics to measure the rank of the recommended results. At last, we adopted the measurement Accuracy@k to evaluate whether mentionees were correctly recommended from the top k results as follows:

\[
\text{Accuracy@k} = \frac{\text{Hits@k}}{N_{\text{test case}}}
\]

where we use Hits@k to denote the number of hit times for a given k in the test set, and \( N_{\text{test case}} \) represents the number of all test cases.

For baselines, we evaluated the following approaches as the competitive methods.

- History based mentionee recommendation (HIS): HIS is a simple heuristic method that recommends mentionees who were frequently mentioned. Mentionees in the results list are sorted in descending order of the frequency they were mentioned. If two or more mentionees share the same frequency, the latest mentioned one is ranked first.

- Context-aware At Recommendation (CAR): CAR [27] is a context-aware approach which recommends mentionees for promotion-oriented posts based on learning-to-rank techniques. In this work, we implemented CAR using the same features in [27]. Hence, in addition to the original data, we also crawled the historical posts of each mentionee and the interaction network between users (i.e., the replying and retweeting histories between users).

- "At" User-User Topic Translation Model (A-UUUTM): A-UUUTM [14] is a translation model that takes into account the textual content of both current post and the histories of mentionees, which is also the state-of-the-art approach for mentionee recommendation. Since the histories of mentionees are important to A-UUUTM, we extracted the latest 4 posts published by each mentionee before she/he was mentioned.

- MMTM with no visual information (MMTM-NV): MMTM-NV is a variant of our proposed MMTM, which considers only the textual words of the posts. MMTM-NV produces recommendations based on the basic topic-translation model from the textual content, i.e., a mentioner chooses a mentionee \( m \) according to the probability \( P(z \mid W_d, \pi_{u,f}) \) given the textual topic \( z \) and words \( W_d \).

- MMTM with no textual information (MMTM-NT): MMTM-NT is another variant of the proposed MMTM that takes only the images of the posts into account. It produces recommendations based on the basic topic-translation model from the visual content, i.e., a mentioner chooses a mentionee \( m \) according to the probability \( P(v \mid W_d, \pi_{v,f}) \) given the visual topic \( v \) and features \( F_d \).

4.3. Results and discussion

In this section, we first represent and analyze the performance of all the approaches mentioned in Section 4.2 as well as our proposed MMTM with well-tuned model parameters, and then study the impact of different model parameters.

Table 3 reports the performance of all methods on our datasets in terms of Precision, Recall, F1-Score, MRR, NDCG@5, Accuracy@3, and Accuracy@5. Figure 3 reports the Precision, Recall and F1-Score of all methods with different numbers of recommended mentionees. We list metric values with \( k \) varying from 1 to 7 since the values do not change significantly when \( k > 7 \). As shown in the results, the methods had significant performance disparity, and our proposed MMTM outperformed all competitors significantly and consistently in all measures. For example, MMTM achieved 0.602 in Precision, 0.577 in Recall, and 0.589 in F1-score. Compared with the state-of-the-art approach A-UUUTM, MMTM performed better by 13.8% in Precision, 14.03% in Recall and 13.9% in F1-Score. The MRR and NDCG@5 results of MMTM were also the best among the rest, which demonstrated that MMTM is capable of producing not only more precise recommendations but also a better ranking of the results. Moreover, we report the top-k recommendation accuracies of our proposed MMTM and the comparative methods in Figure 4. We only show the performance where \( k \) is less than 10 since larger values of \( k \) are not practically useful for this recommendation task. From the results, we can clearly see that the performance of our proposed MMTM is still the highest among all methods in terms of recommendation accuracy. For example, the results of MMTM on metrics Accuracy@3 and Accuracy@5 showed that the 63.5% of ground truth mentionees were correctly found in the top 3 and 66.1% of the mentionees were correctly recommended in the top 5. All of these results demonstrated the effectiveness of our proposed MMTM.

From the results, several additional observations can be made:

1. HIS, which is used by many real services, performs poorly on all metrics.
2. Both MMTM and A-UUUTM performed much better than CAR, which validated the advantages of a well-designed probabilistic generative model in the representation and generalization for mentionee recommendation over the general feature-based learning-to-rank method.
3. The proposed MMTM performed significantly better than the single-factor aware models MMTM-NT and MMTM-NV, showing the advantages of the method that jointly exploits the textual and visual contents of users mentioning activities.

For example, the recommendation accuracy of MMTM is 0.697 when \( k = 10 \). Compared to MMTM-NV and MMTM-NV, MMTM made a relative improvement of 16.4% and 65.2%, respectively.

(4) MMTM achieved significantly better results than A-UUUTM and CAR, although both A-UUUTM and CAR were constructed based on multiple data sources. Specifically, A-UUUTM models user mention behavior by analyzing the content information of both mentioners and related mentionees, and CAR measures the relevance among users and posts by extracting features based on the textual content, user interactive network and spatiotemporal context information. In contrast, MMTM modeled users’ mentioning activities based purely on the content information of mentioners’ multimodal documents. Still, compared to A-UUUTM and CAR, MMTM made a relative improvement of 11.7% and 20.6% in terms of top-5 recommendation accuracy, respectively. Since both A-UUUTM and CAR ignored the impact of the visual resources, the results clearly demonstrated the necessity of capturing users’ mentioning tendencies by exploiting the visual information of users’ mentioning documents.
Table 3

<table>
<thead>
<tr>
<th>Methods</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>MRR</th>
<th>No5</th>
<th>A@3</th>
<th>A@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIS</td>
<td>0.347</td>
<td>0.336</td>
<td>0.342</td>
<td>0.575</td>
<td>0.169</td>
<td>0.382</td>
<td>0.410</td>
</tr>
<tr>
<td>CAR</td>
<td>0.487</td>
<td>0.472</td>
<td>0.479</td>
<td>0.533</td>
<td>0.177</td>
<td>0.515</td>
<td>0.548</td>
</tr>
<tr>
<td>A-UITTM</td>
<td>0.529</td>
<td>0.506</td>
<td>0.517</td>
<td>0.562</td>
<td>0.162</td>
<td>0.570</td>
<td>0.592</td>
</tr>
<tr>
<td>MMTM-NV</td>
<td>0.504</td>
<td>0.481</td>
<td>0.492</td>
<td>0.544</td>
<td>0.156</td>
<td>0.540</td>
<td>0.563</td>
</tr>
<tr>
<td>MMTM-NT</td>
<td>0.315</td>
<td>0.270</td>
<td>0.291</td>
<td>0.391</td>
<td>0.127</td>
<td>0.354</td>
<td>0.375</td>
</tr>
<tr>
<td>MMTM</td>
<td>0.602</td>
<td>0.577</td>
<td>0.589</td>
<td>0.611</td>
<td>0.191</td>
<td>0.635</td>
<td>0.661</td>
</tr>
</tbody>
</table>

![Fig. 3. Precision, Recall and F1-Score with different values of k.](image)

Specifically, we show how the Accuracy@5 performance of the proposed MMTM changes as $K_r$ varies from 10 to 35 and $K_v$ varies from 20 to 45. As for the hyperparameters, we fixed them at $\alpha = 50/K_r$, $\alpha' = 50/K_v$, $\beta = \gamma = \xi = \xi' = 0.01$, $\lambda_1 = \lambda_2 = 0.5$ as [15,30] suggested. Table 4 reports the performance of MMTM on varying numbers of visual and textual topics. We can observe that the recommendation accuracy of MMTM increased quickly first with the increasing values of $K_r$ and $K_v$ and the increase slowed down when $K_r$ and $K_v$ exceed certain thresholds (i.e., $K_r = 25$ and $K_v = 35$). The reason is that the numbers of the textual and visual topics reflect the model complexity. When $K_r$ and $K_v$ were too small, the model has limited ability to describe the data. On the other hand, when the values of $K_r$ and $K_v$ reached certain thresholds, the model was complex enough to handle the data and further increasing the numbers did not help much.

![Fig. 4. Top-k recommendation accuracy.](image)

Besides, to evaluate the best performance of the model and the performance changes under larger parameter values, we conducted another experiment by varying $K_r$ and $K_v$ on a large scale. Table 5 shows the Accuracy@5 values when $K_r > 50$ and $K_v > 70$. From the results, we can observe that the proposed MMTM achieved its highest recommendation accuracy when $K_r = 60$ and $K_v = 80$. However, the accuracy dropped significantly when $K_r$ and $K_v$ are increased further. A possible reason is that when the values of $K_r$ and $K_v$ were too large, the data sparsity problem became more significant when estimating the visual and textual topic-specific probability matrix (e.g., $\pi$), which led to overfitting and made the learned parameters unreliable. Hence, we adopted $K_v = 25$ and

![Table 4](image)

<table>
<thead>
<tr>
<th>$K_r, K_v$</th>
<th>$K_r = 10$</th>
<th>$K_r = 15$</th>
<th>$K_r = 20$</th>
<th>$K_r = 25$</th>
<th>$K_r = 30$</th>
<th>$K_r = 35$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_v = 20$</td>
<td>0.518</td>
<td>0.603</td>
<td>0.621</td>
<td>0.625</td>
<td>0.627</td>
<td>0.627</td>
</tr>
<tr>
<td>$K_v = 35$</td>
<td>0.560</td>
<td>0.625</td>
<td>0.633</td>
<td>0.638</td>
<td>0.638</td>
<td>0.638</td>
</tr>
<tr>
<td>$K_v = 40$</td>
<td>0.588</td>
<td>0.636</td>
<td>0.646</td>
<td>0.652</td>
<td>0.652</td>
<td>0.652</td>
</tr>
<tr>
<td>$K_v = 45$</td>
<td>0.602</td>
<td>0.641</td>
<td>0.656</td>
<td>0.661</td>
<td>0.661</td>
<td>0.661</td>
</tr>
<tr>
<td>$K_v = 50$</td>
<td>0.604</td>
<td>0.643</td>
<td>0.656</td>
<td>0.661</td>
<td>0.661</td>
<td>0.661</td>
</tr>
</tbody>
</table>

For $K_v > 50$, we used a large scale for $K_v$, ranging from 50 to 80, and observed how the performance changed. We can see that the accuracy increased with increasing $K_v$, which suggests that the model benefits from more visual topics. This is because the visual topics provide additional information that complements the textual topics, thereby improving the recommendation accuracy. However, this increase is not linear and eventually plateaus as $K_v$ becomes too large. Conversely, decreasing $K_v$ leads to a decrease in performance, as the model loses important visual information.

Moreover, we conducted a series of experiments to study the impact on performance of tuning the model parameters, e.g., the number of textual topics $K_r$ and the number of visual topics $K_v$, which were fixed at 50 and 80, respectively.
Table 5 Recommendation Accuracy@5 with large $K_p$ and $K_r$.

<table>
<thead>
<tr>
<th>$K_p$</th>
<th>$K_r = 50$</th>
<th>$K_r = 55$</th>
<th>$K_r = 60$</th>
<th>$K_r = 65$</th>
<th>$K_r = 70$</th>
<th>$K_r = 80$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_r = 70$</td>
<td>0.673</td>
<td>0.673</td>
<td>0.673</td>
<td>0.671</td>
<td>0.668</td>
<td>0.659</td>
</tr>
<tr>
<td>$K_r = 75$</td>
<td>0.673</td>
<td>0.674</td>
<td>0.675</td>
<td>0.671</td>
<td>0.669</td>
<td>0.660</td>
</tr>
<tr>
<td>$K_r = 80$</td>
<td>0.675</td>
<td>0.676</td>
<td>0.677</td>
<td>0.672</td>
<td>0.669</td>
<td>0.661</td>
</tr>
<tr>
<td>$K_r = 85$</td>
<td>0.670</td>
<td>0.672</td>
<td>0.673</td>
<td>0.668</td>
<td>0.660</td>
<td>0.649</td>
</tr>
<tr>
<td>$K_r = 90$</td>
<td>0.605</td>
<td>0.665</td>
<td>0.665</td>
<td>0.659</td>
<td>0.647</td>
<td>0.636</td>
</tr>
<tr>
<td>$K_r = 100$</td>
<td>0.653</td>
<td>0.654</td>
<td>0.654</td>
<td>0.646</td>
<td>0.633</td>
<td>0.616</td>
</tr>
</tbody>
</table>

$K_p = 35$ as the best parameters when we conducted the experiment described in the above, considering the trade-offs between performance and training time-cost.

5. Conclusion

In this work, we constructed a recommendation system that generates mentions when a mentionee attempts to mention others in a multi-modal post. While an increasing line of work has studied this problem from different aspects, they focus only on the use of non-visible information like text. After analyzing several social media services, we observed that the visual resources can also provide valuable information to reveal users' mentioning tendencies. In light of this, we propose a latent-class generative model, named Multi-modal Mention Generation Model (M3MTM), to simulate the generating process of users' mentioning activities by synthetically exploiting the textual and visual contents. Specifically, M3MTM adopts the bag-of-word method for both texts and images to alleviate the data sparsity while describing multi-modal documents. By considering the impact of the two modalities on users' mentioning tendencies in a unified way, M3MTM jointly learns users' semantic patterns and the correlations between contents in different modalities of users' multi-modal mentioning documents. Then, after learning a knowledge model containing the necessary insights about users' mentioning tendencies based on M3MTM, we retrieve the top-k mentionees for a multi-modal mentioning post. We conducted extensive experiments on a large dataset crawled from real-world social media service to evaluate the performance of our proposed solution. The experimental results demonstrated the superiority of our approach over the state-of-the-art methods.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61502350).

References

Kai Wang is a Ph.D. student of the School of Computer, Wuhan University, Wuhan, China. His research interests include social user profiling and spatio-temporal data mining.

Wei Yi Meng is a professor and the chair of the Computer Science Department, State University of New York at Binghamton, Binghamton, USA. He received his Ph.D. in computer science from University of Illinois at Chicago in 1992. His research interests include metasearch engines, Internet-based information retrieval, social network data analysis and database management systems. He has done pioneering work in large-scale metasearch engines. He has published over 150 papers and many of them in top conferences and journals in his research areas, including ACM SIGMOD, VLDB, IEEE ICDE, WWW, KDD, ACM SIGIR, ACM TOIS, ACM TODS, ACM TSK, ACM TWEB, and IEEE TKDE.

Shijun Li is a professor of the School of Computer, Wuhan University, Wuhan, China. He received his Ph.D. in computer science from Wuhan University, China. His research interests include Web data mining, Web information retrieval and social network data analysis. He is member of China Computer Federation.

Sha Yang is an associate professor of the School of Computer Science and Technology, Hankou University, Wuhan, China. She received her Ph.D. in computer science from Wuhan University, China. Her research interests include data quality evaluation and social network data analysis.