

# Who Will You “@”?

Yeyun Gong, Qi Zhang, Xuyang Sun, Xuanjing Huang  
Shanghai Key Laboratory of Data Science  
School of Computer Science, Fudan University  
825 Zhangheng Road, Shanghai, P.R.China  
{12110240006, qz,13210240106, xjhuang}@fudan.edu.cn

## ABSTRACT

In Twitter-like social networking services, people can use the “@” symbol to mention other users in tweets and send them a message or link to their profiles. In recent years, social media services are rapidly growing with thousands of millions of users participating in them every day. When the “@” symbol is entered, there should be an automatic suggestion function which recommends a small list of candidates in order to help users to easily identify and input usernames. In this paper, we present our work on building a recommendation system for the mention function in microblogging services. The recommendation strategy we used takes into consideration not only content of the microblog but also histories of candidate users. To better handle these textual information, we propose a novel method that extends the translation-based model. Experimental results on the dataset we collected from a real world microblogging service demonstrate that the proposed method outperforms state-of-the-art approaches.

## Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval - Information Search and Retrieval;  
H.3.7 [Digital Libraries]: Collection, Systems Issues

## Keywords

Microblog; Recommendation; Topical Model

## 1. INTRODUCTION

In Twitter-like services, the “@” symbol, which means what it appears to mean: “at”, representing a connection, is one of the most important and common components. Once we put the “@” symbol in front of a username, an alert will be sent to that user signifying that a microblog is commenting directly to him. Since microblogging services have become one of the most important communication

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.  
CIKM'15, October 19–23, 2015, Melbourne, Australia.  
© 2015 ACM. ISBN 978-1-4503-3794-6/15/10 ...\$15.00.  
DOI: <http://dx.doi.org/10.1145/2806416.2806458>.

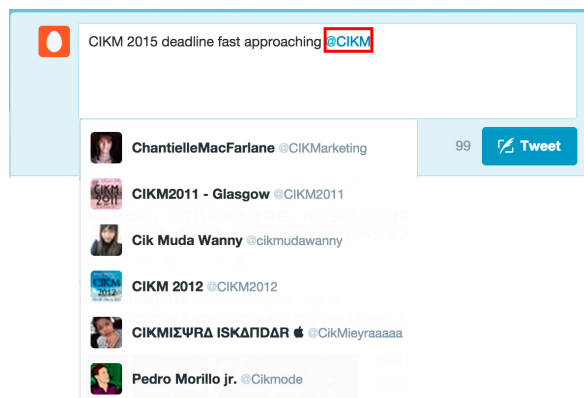


Figure 1: An example of username candidates given by Twitter.

methods for people, there are a huge number of active users in real online services. According to a statistic on Twitter, there are 284 million monthly active users and the average number of followers for each account is 208<sup>1</sup>. Thus, if a recommendation system could recommend a smaller list of candidates, it would help users to efficiently identify and input usernames.

Along with the development of social media, a variety of recommendation tasks have been proposed for different problems, such as content recommendation [4, 7, 11, 16, 35], community recommendation [22, 25, 27, 47], tag recommendation [8, 33], music recommendation [3, 15, 37], news recommendation [19, 38], and so on. Information used for recommendations can be roughly divided into the following types: textual information [8], structural information [16, 25], spatial information [36, 45], and temporal information [18, 46]. However, only a few of attention has been paid on the task of recommending usernames when users input the “@” in microblog posts [42]. The recommendations given by online microblogging services are usually based on either partial inputs or recent histories of users. Figure 1 illustrates an example of username candidates when user input “CIKM” after the “@” symbol. We can see that although the text input has clearly indicated the microblog to connect to CIKM 2015, the username “CIKM2015” is not in the top candidates.

<sup>1</sup><https://about.twitter.com/company>

In this paper, we focus on the task of recommending usernames for users, when they try to use the “@” symbol to mention other users in their posts. To be able to make this task a reality we have to face several challenges. First, as we mentioned above, there are a huge number of active users in microblogging services. This means that the recommendation target space is extremely large. Second, personalization is an inherent requirement of this task. Since users have their specific preferences, different usernames should be recommended even when the microblog posts have the exact same content. Third, due to the limitation of character length in microblog posts, there is brevity in information and inadequate context and structure [6]. Wang et al. [42] formulated the task as a learning to rank problem and studied several features describing the interest of users, content-dependent user relationship, and user influence. However, the main aim of that work is to recommend the users who may retweet the microblog. Different from it, the aim of mention actions we focused on are not limited to the spread of a microblog. We analysed and evaluated on the microblogs posted by real online users.

Motivated by the advantage of translation based methods in capturing the content information in different recommendation tasks [14, 46], in this work, we also follow this framework and assume that the usernames and textual content in a microblog are parallel descriptions of the same thing in different languages. Hence, the username recommendation task can be regarded as a translation process from the content of microblogs to the usernames. Since the existing translation model is sometimes vague without the aid of background knowledge, we integrated latent topical information into the translation model to facilitate the translation process. As we mentioned above, users have their own preferences, histories of users are important factor in this task, so we also incorporated the posts of both the user and candidates into the model. To evaluate the proposed method, we constructed a dataset which contains more than 10 million microblogs from an online microblog service and selected a subset from it as golden standards. We compared the proposed method with several state-of-the-art methods on the constructed dataset. Experimental results demonstrate that the proposed approach can achieve significantly better performance than the other methods.

The contributions of this work can be summarized as follows:

- We proposed a novel topical translation-based method to predict the users who the authors try to mention. Both content of the microblog and histories of candidate users are taken into consideration.
- We constructed a large data set which contains more than ten million words. This dataset can also be used for other related tasks.
- Through various experiments in the real online dataset, we demonstrated the effectiveness of the proposed method.

## 2. RELATED WORK

The proposed method relates to the following two research areas: recommendation on social medias and topic model. In this section, we briefly describe the related works on these areas.

### 2.1 Recommendation on Social Medias

In twitter-like services, users usually have to scan microblogs presented in chronological order to find what they are interested in. Hence, the task of recommending microblogs has received considerable attentions in recent years [4, 7, 11, 16, 30, 35, 44]. Chen et al.[4] proposed to use collaborative ranking to capture personal interests. They incorporate the topics of tweets, social relations factors and some explicit features (such as authority of the publisher and quality of the tweet) to perform the task. Guy et al. [11] proposed to use both related people and related tags to recommend social media items. Pan et al.[30] introduced a joint model to combine collaborative filtering and diffusion processes for recommending microblogs. They also studied several diffusion features to capture the characteristic of diffusion processes. Graph based method was also proposed to perform the microblog recommendation task [44]. Microblogs and authors were ranked following a co-ranking algorithm.

Social interaction among people is the essential character of social media. To help users efficiently find their interested communities or friends, many studies have been proposed to perform the task [5, 22, 25, 27, 47]. Lo and Lin [22] proposed a weighted minimum-message ratio (WMR) method, which uses the real message interaction number among members, to generate personalized friend lists. Hao et al. [25] employed the opinions of trusted friends to make recommendations for the users under a probabilistic matrix factorization framework. In [47], the task of event-based group recommendation was introduced. They proposed a matrix factorization method, which considers location features, social features, and implicit patterns, to perform the task. Chen et al. [5] proposed a collaborative filtering based method to perform personalized community recommendations by considering multiple types of co-occurrences in social data at the same time.

Due to the usefulness of tag recommendation, a variety of methods have been proposed from different perspectives [12, 17, 34, 21, 8]. Heymann et al. [12] investigated the tag recommendation problem using the data collected from social bookmarking system. They introduced an entropy-based metric to capture the generality of a particular tag. In [39], a Poisson Mixture Model based method is introduced to perform the tag recommendation task. Krestel et al. [17] introduced a Latent Dirichlet Allocation to elicit a shared topical structure from the collaborative tagging effort of multiple users for recommending tags. Based on the observation that similar webpages tend to have the same tags, Lu et al. proposed a method taking both tag information and page content into account to perform the task [24]. Ding et al. proposed to use translation process to model this task [8]. They extended the translation based method and introduced a topic-specific translation model to process the various meanings of words in different topics. In [40], discriminative-term-weights were used to establish topic-term relationships, of which users’ perception were learned to suggest suitable hashtags for users. To handle the vocabulary problem in keyphrase extraction task, Liu et al. proposed a topical word trigger model, which treated the keyphrase extraction problem as a translation process with latent topics [21].

The most similar work to ours is that of Wang et al. [42] which proposed the work of *whom-to-mention*. They

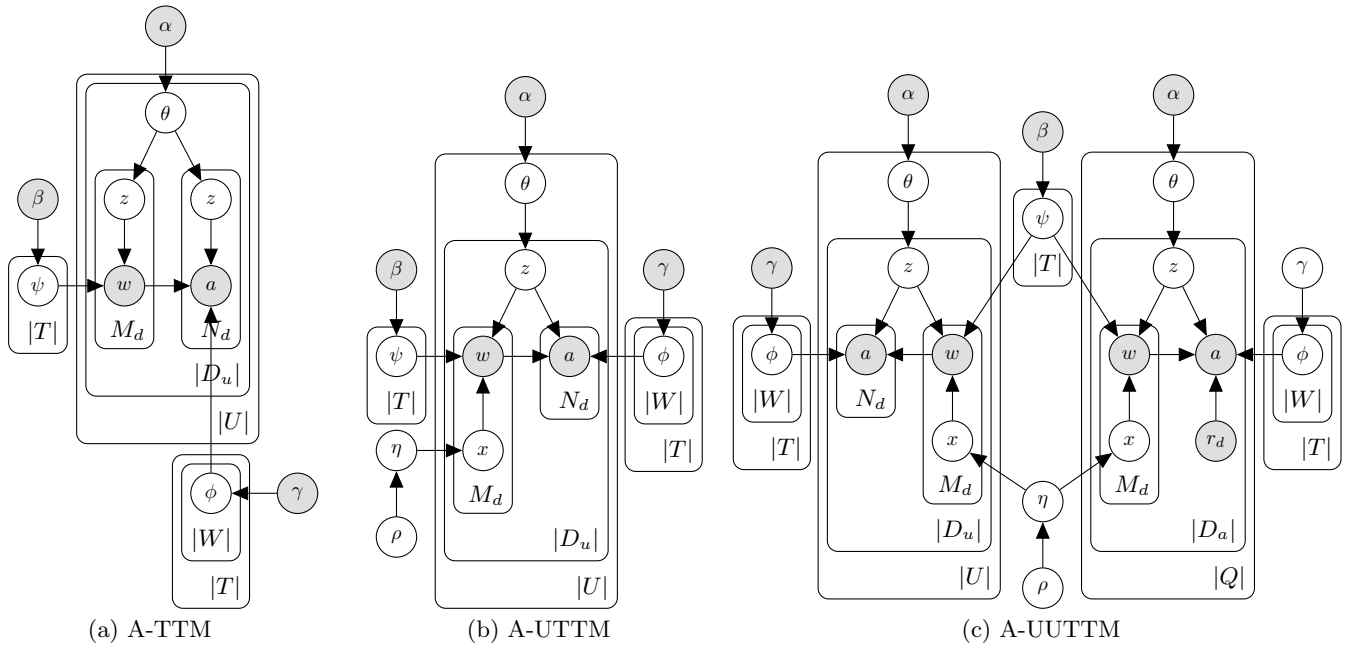


Figure 2: The graphical representation of the proposed model. Shaded circles are observations or constants. Unshaded ones are hidden variables.

try to find the users who have big influences and high probabilities to retweet the microblog. The main goal of their recommendation is to make the microblog to spread more quickly. Hence, they formulated the task as a learning to rank problem, and studied several features describing the interest of users, content-dependent user relationship, and user influence. Different from them, in this work, we use the microblogs which contain mention users and are posted by real users. Hence, we need to process different kinds of motivations when users input “@” in a microblog.

## 2.2 Topic Models

Topic models, such as probabilistic latent semantic indexing (pLSI) [13], latent Dirichlet allocation (LDA) [1], hierarchical dirichlet processes (HDP) [41], and so on, have proven to be useful in various tasks. In [1], latent Dirichlet allocation (LDA), a three-level hierarchical Bayesian model, was described for modeling collections of discrete data. Griffiths and Steyvers proposed a Markov chain Monte Carlo algorithm for inference LDA and used it to analyze abstracts of scientific papers [9]. Teh et al. [41] described Hierarchical dirichlet processes (HDP), a nonparametric approach, to model of groups of data. Each group was characterized by a mixture model and was desirable to allow mixture components to be shared between groups. McAuliffe and Blei [26] introduced supervised latent Dirichlet allocation (sLDA) to model labelled documents.

Due to the capability of representing document themes, topic models have also been successfully used in natural language processing, image annotation, and various other applications [20, 32, 21, 8]. In [20], LDA-based image representation is proposed to classify the patches of the large image into the semantic concepts. To handle the vocabulary problem in keyphrase extraction task, Liu et al. [21] proposed a topical word trigger model, which treated

the keyphrase extraction problem as a translation process with latent topics. Pennacchiotti and Gurumurthy [31] introduced a LDA based method to recommend friends with similar interests for users. For analysing text streams such as a sequence of posts from the same author, Wang et al. proposed Temporal-LDA to model the topic transitions that naturally arise in these data [43].

In this work, we also incorporate LDA with translation models to integrate topical information into them. The most similar work to ours is that of Lu et al. [23], which use translation methods to recommend citations. However, they did not take the topical information into consideration.

## 3. APPROACH

### 3.1 Preliminaries

Given a microblog  $d$  and its author  $u$ , the “@” recommendation task is to discover a list of candidate authors. In this work, we first use generative models to learn the joint distribution of the topics  $\mathbf{z}$ , the microblogs  $\mathbf{w}$  and the mentioned users  $\mathbf{a}$ ,  $p(\mathbf{z}, \mathbf{w}, \mathbf{a})$ . Then, we can use the probability  $p(a|\mathbf{z}, w_d)$ , which can be influenced based on the learned joint distribution, to generate candidate lists. The notations to be used throughout this paper are as follows. We use  $W$  and  $U$  to denote the word vocabulary set and the user set respectively.  $D_u$  denotes the set of microblogs posted by the user  $u$ .  $Q$  denotes the set of users who have been mentioned in one or more microblogs in the corpus.  $D_a$  is the set of microblogs which are posted by the mentioned user  $a$ ,  $r_d$  is the rank of microblog  $d$  in the extracted microblogs which is based on the time of the microblog. Let  $T$  denote the topic set, and  $\psi_w^z$  be the probabilistic distribution over words for each topic. We use  $\phi_a^{z,w}$  to denote the probabilistic distribution over users for each topic and each word which indexes the probability of a user being mentioned given

a word and a topic. At the microblog level, we assume that each microblog has one topic and each user contains a mixture of topics. Let  $\theta_u$  and  $\theta_d$  denote the topic distribution of user  $u$  and topic distribution of microblog  $d$  respectively,  $w_d = \{w_m\}_{m=1}^{M_d}$  are the words in microblog  $d$ ,  $M_d$  is the number of words in microblog  $d$ .  $a_d = \{a_n\}_{n=1}^{N_d}$  are the users mentioned in microblog  $d$ ,  $N_d$  is the number of users mentioned in microblog  $d$ .  $x_d = \{x_m\}_{m=1}^{M_d}$  are the topic word or background word indicate variables. Fig. 2 illustrates the graphical model representation of the proposed models.

### 3.2 A-TTM

The ‘‘At’’ Topic Translation Model (A-TTM) assumes that each microblog contains a mixture of topics. For the generation of the words in the microblog, each topic corresponds to a multinomial distribution over words. For the generation of the users mentioned in the microblog, each topic and word corresponds to a user distribution matrix  $\phi^{z,w}$ . The A-TTM is based on the following assumptions. When a user posts a microblog, she first chooses the words of the microblog and then selects the users she wants to mention. For each word, she first chooses a topic from the topic distribution and then chooses a word from the topic-word distribution. Finally, she chooses users to mention according to the topics and words in the microblog. The generation process of A-TTM is shown in Alg. 1.

---

**Algorithm 1** The generation process of A-TTM model

---

```

for each topic  $z \in T$  do
  Draw  $\psi^z \sim Dir(\beta)$ 
  for each word  $w \in W$  do
    Draw  $\phi^{z,w} \sim Dir(\gamma)$ 
  end for
end for
for each user  $u \in U$ : do
  for each microblog  $d \in D_u$ : do
    Draw  $\theta_d \sim Dir(\cdot|\alpha)$ 
    for each word in microblog  $d, w_m \in w_d$ : do
      Draw a topic  $z_m \sim Mult(\cdot|\theta_d)$ 
      Draw a word from topic-word distribution  $w_m \sim Mult(\cdot|\psi^{z_m})$ 
    end for
    for each user mentioned in microblog  $d, a_n \in a_d$ : do
      Draw a topic  $z_n \sim Mult(\cdot|\theta_d)$ 
      Draw a user  $a_n \sim p(\cdot|z_n, w_d, \phi^{z_n, w})$ 
    end for
  end for
end for

```

---

### 3.3 A-UTTM

The A-TTM assumes that a microblog contains a mixture of topics, and each topic corresponds to a multinomial distribution over words. While this assumption is reasonable for long documents, for short microblog posts, a single post is most likely to be about a single topic. Therefore, in the ‘‘At’’ User Topic Translation Model (A-UTTM), we assume that a user contains a mixture of topics, represented by a topic distribution, and each microblog has a single topic label. The A-UTTM is based on the following assumptions. When a user posts a microblog, she first chooses a topic from her topic distribution and then chooses a sequence of words from the topic-word distribution or the background-

word distribution one by one. Finally, she chooses a user to mention according to the topic and topic words in the microblog. The generation process of the A-UTTM is described in Alg. 2.

---

**Algorithm 2** The generation process of A-UTTM

---

```

Draw  $\psi^B \sim Dir(\beta), \eta \sim Beta(\rho)$ 
for each topic  $z \in T$  do
  Draw  $\psi^z \sim Dir(\beta)$ 
  for each word  $w \in W$  do
    Draw  $\phi^{z,w} \sim Dir(\gamma)$ 
  end for
end for
for each user  $u \in U$ : do
  Draw  $\theta_u \sim Dir(\cdot|\alpha)$ 
  for each microblog  $d \in D_u$ : do
    Draw a topic  $z_d \sim Mult(\cdot|\theta_u)$ 
    for each word in microblog  $d, w_m \in w_d$ : do
      Draw  $x_m \sim Bern(\eta)$ 
      if  $x_m = 1$  then
        Draw a word from topic-word distribution  $w_m \sim Mult(\cdot|\psi^{z_d})$ 
      else
        Draw a word from background-word distribution  $w_m \sim Mult(\cdot|\psi^B)$ 
      end if
    end for
    for each user mentioned in microblog  $d, a_n \in a_d$ : do
      Draw a user  $a_n \sim p(\cdot|z_d, w_d, \phi^{z_d, w})$ 
    end for
  end for
end for

```

---

### 3.4 A-UUTTM

In the A-UTTM, we use a user level topic model to model the original microblogs. Although the A-UTTM takes into account the content of the original microblogs, usually the microblogs of the mentioned user are also important factors in influencing the mention behavior of the user. Hence, we incorporate the microblogs of the mentioned users into the A-UTTM, and we propose the ‘‘At’’ User-User Topic Translation Model (A-UUTTM). In the A-UUTTM, we assume that when a user posts a microblog, she first generates the words in the microblog. The generation process is similar as A-UTTM. She then chooses a user to mention based on not only the topic and topic words in the microblog, but also the microblogs of the mentioned user. The generation process of the A-UUTTM is described in Alg. 3.

### 3.5 Inference using Gibbs Sampling

To learn the parameters of all the models (A-TTM, A-UTTM and A-UUTTM), we use collapsed Gibbs sampling [10] to obtain samples of hidden variable assignment.

#### 3.5.1 Inference Estimation for A-TTM

According to the generation process of Alg. 1, we can factorize the joint probability distribution of the topic  $\mathbf{z}$ , the microblog words  $\mathbf{w}$  and the mentioned users  $\mathbf{a}$  as follows:

$$p(\mathbf{w}, \mathbf{a}, \mathbf{z}|\alpha, \beta, \gamma) = p(\mathbf{z}|\alpha)p(\mathbf{w}|\mathbf{z}, \beta)p(\mathbf{a}|\mathbf{z}, \mathbf{w}, \gamma). \quad (1)$$

---

**Algorithm 3** The generation process of A-UUTTM

---

Draw  $\psi^B \sim \text{Dir}(\beta), \eta \sim \text{Beta}(\rho)$   
**for** each topic  $z \in T$  **do**  
  Draw  $\psi^z \sim \text{Dir}(\beta)$   
  **for** each word  $w \in W$  **do**  
    Draw  $\phi^{z,w} \sim \text{Dir}(\gamma)$   
  **end for**  
**end for**  
**for** each user mentioned in any microblog,  $a \in Q$ : **do**  
  Draw  $\theta_a \sim \text{Dir}(\cdot|\alpha)$   
  **for** each microblog  $d \in D_a$ : **do**  
    Draw a topic  $z_d \sim \text{Mult}(\cdot|\theta_a)$   
    **for** each word in microblog  $d$ ,  $w_i \in w_d$  **do**  
      Draw  $x_i \sim \text{Bern}(\eta)$   
      **if**  $x_i == 1$  **then**  
        Draw a topic word  $w_i \sim \text{Mult}(\cdot|\psi^z)$   
      **else**  
        Draw a background word  $w_i \sim \text{Mult}(\cdot|\psi^B)$   
      **end if**  
    **end for**  
    Draw  $a \sim p(\cdot|z_d, w_d, r_d, \phi^{z,w})$   
  **end for**  
**end for**  
**for** each user  $u \in U$ : **do**  
  Draw  $\theta_u \sim \text{Dir}(\cdot|\alpha)$   
  **for** each microblog  $d \in D_u$ : **do**  
    Draw a topic  $z_d \sim \text{Mult}(\cdot|\theta_u)$   
    **for** each word in microblog  $d$ ,  $w_m \in w_d$ : **do**  
      Draw  $x_m \sim \text{Bern}(\eta)$   
      **if**  $x_m == 1$  **then**  
        Draw a topic word  $w_m \sim \text{Mult}(\cdot|\psi^z)$   
      **else**  
        Draw a background word  $w_m \sim \text{Mult}(\cdot|\psi^B)$   
      **end if**  
    **end for**  
  **for** each user mentioned in microblog  $d$ ,  $a_n \in a_d$ : **do**  
    Draw a user  $a_n \sim p(\cdot|z_d, w_d, \phi^{z,w})$   
  **end for**  
**end for**  
**end for**

---

Let  $N_z^{w,a}$  denote the number of co-occurrence times of user  $a$  and word  $w$  under topic  $z$ . According to the multinomial assumption on occurrences users, we obtain:

$$\begin{aligned} p(\mathbf{a}|\mathbf{z}, \mathbf{w}, \phi) &= \prod_{u \in U} \prod_{d \in D_u} \prod_{a_n \in a_d} p(a_n|z_n, \mathbf{w}_d) \\ &= \prod_{z \in T} \prod_{a \in Q} \prod_{w \in W} (\phi_a^{z,w})^{N_z^{w,a}}, \end{aligned} \quad (2)$$

where  $\phi_a^{z,w}$  is proportional to the probability that user  $a$  to be mentioned with the topic  $z$ . The target posterior distribution for user generation, i.e.  $p(\mathbf{a}|\mathbf{z}, \mathbf{w}, \gamma)$ , can be obtained by integrating over all possible values of  $\phi$ :

$$\begin{aligned} p(\mathbf{a}|\mathbf{z}, \mathbf{w}, \gamma) &= \int \prod_{z \in T} \prod_{w \in W} \frac{1}{\Delta(\gamma)} \prod_{a \in Q} (\phi_a^{z,w})^{N_z^{w,a} + \gamma_a - 1} d_{\phi^{z,w}} \\ &= \prod_{z \in T} \prod_{w \in W} \frac{\Delta(\mathbf{N}_{\phi^{z,w}} + \gamma)}{\Delta(\gamma)}, \end{aligned} \quad (3)$$

where  $\Delta(\gamma) = \frac{\prod_{a=1}^{|Q|} \Gamma(\gamma_a)}{\Gamma(\sum_{a=1}^{|Q|} \gamma_a)}$  and  $\mathbf{N}_{\phi^{z,w}} = \{N_z^{w,a}\}_{a \in Q}$ .

A similar derivation holds for  $p(\mathbf{w}|\mathbf{z}, \beta)$  and  $p(\mathbf{z}|\alpha)$  leading to the expression for joint distribution:

$$\begin{aligned} p(\mathbf{w}, \mathbf{a}, \mathbf{z}|\alpha, \beta, \gamma) &= \prod_{z \in T} \prod_{w \in W} \frac{\Delta(\mathbf{N}_{\phi^{z,w}} + \gamma)}{\Delta(\gamma)} \\ &\prod_{z \in T} \frac{\Delta(\mathbf{N}_{\psi^z} + \beta)}{\Delta(\beta)} \prod_{u \in U} \prod_{d \in D_u} \frac{\Delta(\mathbf{N}_d + \alpha)}{\Delta(\alpha)}. \end{aligned} \quad (4)$$

For Gibbs sampler, we need to derive  $p(z_i = k|\mathbf{z}_{-i}, \mathbf{w}, \mathbf{a})$ , where  $\mathbf{z}_{-i}$  denotes the entire state space of  $z$  except the  $i$ th token and  $i$  iterates over each word in the corpus.

The sampling probability of a latent topic for word  $w_m$  in the microblog  $d$  of the user  $u$  is sampled from:

$$\begin{aligned} p(z_m = k|\mathbf{z}_{-m}, \mathbf{w}, \mathbf{a}) &\propto \frac{N_{k,-m}^{w_m} + \beta}{N_{k,-m}^{(\cdot)} + \beta|W|} \\ &\frac{N_{k,-m} + \alpha}{N_{(\cdot),-m} + \alpha|T|} \prod_{a \in a_d} \frac{N_{k,-m}^{w_m,a} + \gamma}{N_{k,-m}^{w_m,(\cdot)} + \gamma|U|}, \end{aligned} \quad (5)$$

where  $N_{k,-m}^{w_m}$  is the number of word  $w_m$  assigned to topic  $k$ .  $N_{k,-m}^{w_m,a}$  is the number of word  $w_m$  and mentioned user  $a$  co-occurrence in the same microblog under the topic  $k$ .  $N_{k,-m}$  is the total number of topic  $k$ . All the count with  $-m$  is calculated without considering the current word  $w_m$  in the microblog  $d$ . And  $(\cdot)$  represents the count is calculated of all conditions, for example,  $N_{(\cdot),-m} = \sum_{k \in T} N_{k,-m}$ .

The sampling probability of a latent topic for the mentioned user  $a_n$  in the microblog  $d$  is sampled from:

$$\begin{aligned} p(z_n = k|\mathbf{z}_{-n}, \mathbf{w}, \mathbf{a}) &\propto \\ &\frac{N_{k,-m} + \alpha}{N_{(\cdot),-m} + \alpha|T|} \cdot \left( \sum_{w_m \in w_d} \frac{N_{k,-m}^{w_m,a_n} + \gamma}{N_{k,-m}^{(\cdot),a_n} + \gamma|U|} \right). \end{aligned} \quad (6)$$

### 3.5.2 Inference Estimation for A-UTTM

According to the generation process of Alg. 2, we can factorize the joint probability distribution of the topic  $\mathbf{z}$ , the microblog words  $\mathbf{w}$ , the topic or background word indicate variable  $\mathbf{x}$  and the mentioned users  $\mathbf{a}$  as follows:

$$\begin{aligned} p(\mathbf{w}, \mathbf{a}, \mathbf{z}, \mathbf{x}|\alpha, \beta, \gamma, \rho) &= \\ p(\mathbf{z}|\alpha)p(\mathbf{x}|\rho)p(\mathbf{w}|\mathbf{z}, \mathbf{x}, \beta)p(\mathbf{a}|\mathbf{z}, \mathbf{w}, \mathbf{x}, \gamma). \end{aligned} \quad (7)$$

The derivation in previous section applies here which leads to following algebraic expression:

$$\begin{aligned} p(\mathbf{w}, \mathbf{a}, \mathbf{z}, \mathbf{x}|\alpha, \beta, \gamma, \rho) &= \frac{\Delta(\mathbf{N}_\eta + \rho)}{\Delta(\rho)} \prod_{z \in T} \frac{\Delta(\mathbf{N}_{\psi^z} + \beta)}{\Delta(\beta)} \\ &\frac{\Delta(\mathbf{N}_{\psi^B} + \beta)}{\Delta(\beta)} \prod_{z \in T} \prod_{w \in W} \frac{\Delta(\mathbf{N}_{\phi^{z,w}} + \gamma)}{\Delta(\gamma)} \prod_{u \in U} \frac{\Delta(\mathbf{N}_d + \alpha)}{\Delta(\alpha)}. \end{aligned} \quad (8)$$

Given the other variable state except the variable  $x_m$ , the sampling probability is calculated as follows:

$$\begin{aligned} p(x_m = p|\mathbf{w}, \mathbf{a}, \mathbf{z}, \mathbf{x}_{-m}) &= \\ &\propto \frac{N_{-m,p} + \rho}{N_{-m,(\cdot)} + 2\rho} \cdot \frac{N_{-m,l}^{w_m} + \beta}{N_{-m,l}^{(\cdot)} + \beta|W|}, \end{aligned} \quad (9)$$

where  $l = B$  if  $p = 0$  and  $l = z$  if  $p = 1$ .  $N_{-m,1}$  is a count of topic words and  $N_{-m,0}$  is a count of background words.  $N_{-m,B}^{w_m}$  is the times of word  $w_m$  assigned to background word.  $N_{-m,z}^{w_m}$  is the times of word  $w_m$  occurs as a topic word.  $-m$  represents taking no account of the current word  $w_m$ .

Given the other variable state except the hidden topic  $z_d$  of microblog  $d$  in user  $u$ , the sampling probability is calculated as follows:

$$p(z_d = k | \mathbf{w}, \mathbf{a}, \mathbf{z}_{-d}, \mathbf{x}) \propto \frac{N_{k,-d}^u + \alpha}{N_{(\cdot),-d}^u + \alpha |T|} \prod_{w_m \in w_d} \frac{N_{k,-m}^{w_m} + \beta}{N_{k,-m}^{(\cdot)} + \beta |W|} \prod_{a_n \in a_d} \sum_{w_m \in w_d} \frac{N_{k,-d}^{w_m, a_n} + \gamma}{N_{k,-d}^{w_m, (\cdot)} + \gamma |Q|}, \quad (10)$$

where  $N_{k,-d}^u$  is the number of microblogs assigned to topic  $k$  of user  $u$ .  $-d$  denotes the count is calculated without taking account of the microblog  $d$ .

### 3.5.3 Inference Estimation for A-UUTTM

According to the generation process shown in Alg. 3, we can factorize the joint probability distribution of the topic  $\mathbf{z}$ , the microblog words  $\mathbf{w}$ , the topic or background word indicate variable  $\mathbf{x}$  and the mentioned users  $\mathbf{a}$  as follows:

$$p(\mathbf{w}, \mathbf{a}, \mathbf{z}, \mathbf{x} | \alpha, \beta, \gamma, \rho) = p(\mathbf{z} | \alpha) p(\mathbf{x} | \rho) p(\mathbf{w} | \mathbf{z}, \mathbf{x}, \beta) p_u(\mathbf{a} | \mathbf{z}, \mathbf{w}, \mathbf{x}, \gamma) p_a(\mathbf{a} | \mathbf{z}, \mathbf{w}, \mathbf{x}, \gamma). \quad (11)$$

The derivation in previous section applies here which leads to following algebraic expression:

$$p(\mathbf{w}, \mathbf{a}, \mathbf{z}, \mathbf{x} | \alpha, \beta, \gamma, \rho) = \frac{\Delta(\mathbf{N}_\eta + \rho)}{\Delta(\rho)} \frac{\Delta(\mathbf{N}_{\psi \mathbf{B}} + \beta)}{\Delta(\beta)} \prod_{z \in T} \prod_{w \in W} \frac{\Delta(\mathbf{N}_{\phi \mathbf{z}, \mathbf{w}}^U + \mathbf{N}_{\phi \mathbf{z}, \mathbf{w}}^Q + \gamma)}{\Delta(\gamma)} \prod_{z \in T} \frac{\Delta(\mathbf{N}_{\psi \mathbf{z}} + \beta)}{\Delta(\beta)} \prod_{u \in U} \frac{\Delta(\mathbf{N}_d^U + \alpha)}{\Delta(\alpha)} \prod_{a \in Q} \frac{\Delta(\mathbf{N}_d^Q + \alpha)}{\Delta(\alpha)}, \quad (12)$$

where  $\mathbf{N}_{\phi \mathbf{z}, \mathbf{w}}^U = \{N_z^{w,a}\}_{a \in Q}$  are the counts calculated in the original microblogs.  $\mathbf{N}_{\phi \mathbf{z}, \mathbf{w}}^Q$  are the counts calculated in the microblogs of mentioned users with decay factors, which can be calculated as follow:

$$\mathbf{N}_{\phi \mathbf{z}, \mathbf{w}}^Q = \left\{ \sum_{d \in D_a} N_w^d e^{-\frac{r_d}{\lambda}} \delta(z_d = z) \right\}_{a \in Q}, \quad (13)$$

where  $N_{w_i}^d$  is the number of word  $w_i$  in microblog  $d$ .  $\delta$  is the indicator function.

The inference process of A-UUTTM for the latent variables in the original microblogs is similar to A-UTTM model. Due to the space limit, we only show the inference process of the latent variables in the microblogs extracted from the mentioned users.

Given the other variable state except the variable  $x_{-i}$ , the sampling probability is calculated as follows:

$$p(x_i = p | \mathbf{w}, \mathbf{a}, \mathbf{z}, \mathbf{x}_{-i}) \propto \frac{N_{-i,p} + \rho}{N_{-i,(\cdot)} + 2\rho} \cdot \frac{N_{-i,l}^{w_i} + \beta}{N_{-i,l}^{(\cdot)} + \beta |W|}, \quad (14)$$

where  $-i$  represents taking no account of the current word  $w_i$ .

Given the other variable state except the topic  $z_j$  of microblog  $j$  in user  $a$ , the sampling probability is calculated

as follows:

$$p(z_j = k | \mathbf{w}, \mathbf{a}, \mathbf{z}_{-j}, \mathbf{x}) \propto \frac{N_{k,-j}^a + \alpha}{N_{(\cdot),-j}^a + \alpha |T|} \prod_{w_i \in w_j} \frac{N_{k,-i}^{w_i} + \beta}{N_{k,-i}^{(\cdot)} + \beta |W|} \sum_{w_i \in w_j} \frac{N_{U,k,-j}^{w_i,a} + N_{Q,k,-j}^{w_i,a} + \gamma}{N_{U,k,-j}^{w_i,(\cdot)} + N_{Q,k,-j}^{w_i,(\cdot)} + \gamma |Q|}, \quad (15)$$

where  $N_{U,k}^{w_i,a}$  is the times of word  $w_i$  and user  $a$  co-occurrence under the topic  $k$  in the recommending users.  $N_{Q,k}^{w_i,a}$  can be calculated by the equation  $N_{Q,k}^{w_i,a} = \sum_{d \in D_a} N_{w_i}^d e^{-\frac{r_d}{\lambda}} \delta(z_d = k)$ .

## 3.6 Prediction

We first run the inference algorithm of all the methods as described in previous section using the training set. Then we extend the Gibbs sampler state with the samples from the test set with following updates:

$$p(x_m = p | \mathbf{w}, \mathbf{a}, \mathbf{z}, \mathbf{x}_{-m}) \propto \frac{N_{-m,p}^t + N_{-m,p} + \rho}{N_{-m,(\cdot)}^t + N_{-m,(\cdot)} + 2\rho} \cdot \frac{N_{-m,l}^{w_m^t} + N_{-m,l}^{w_m} + \beta}{N_{-m,l}^{(\cdot)t} + N_{-m,l}^{(\cdot)} + \beta |W|}, \quad (16)$$

where  $(\cdot)^t$  denotes the count calculated in the test set.

$$p(z_d = k | \mathbf{w}, \mathbf{a}, \mathbf{z}_{-d}, \mathbf{x}) \propto \quad (17)$$

$$\frac{N_{k,-d}^t + N_{k,-d} + \alpha}{N_{(\cdot),-d}^t + N_{(\cdot),-d} + \alpha |T|} \prod_{w_m \in w_d} \frac{N_{k,-m}^{w_m^t} + N_{k,-m}^{w_m} + \beta}{N_{k,-m}^{(\cdot)t} + N_{k,-m}^{(\cdot)} + \beta |W|}.$$

The parameters  $\theta$  and  $\psi$  can be estimated as follows:

$$\theta_k^u = \frac{N_k^u + \alpha}{N_{(\cdot)}^u + \alpha |T|}; \psi_w^k = \frac{N_k^{w^t} + N_k^w + \beta}{N_k^{(\cdot)t} + N_k^{(\cdot)} + \beta |W|}, \quad (18)$$

and the parameter  $\phi$  can be obtained as:

$$\phi_a^{k,w} = \frac{N_k^{w,a^t} + N_{U,k}^{w,a} + N_{Q,k}^{w,a} + \gamma}{N_k^{w,(\cdot)t} + N_{U,k}^{w,(\cdot)} + N_{Q,k}^{w,(\cdot)} + \gamma |Q|}. \quad (19)$$

The possibility table  $\phi$  has a potential size of  $|W| \cdot |Q| \cdot |T|$ . The data sparsity problems may cause difficulties in estimating  $\phi_a^{k,w}$ . Hence, we employ a linear interpolation with the topic-free word alignment probability  $p(a|w)$  to reduce the problem:

$$\Phi = \mu \phi + (1 - \mu) p(a|w). \quad (20)$$

To estimate the topic-free alignment probability  $p(a|w)$ , we explore IBM Model-1 [2] here.  $\mu$  is trade-off of two probabilities ranging from 0.0 to 1.0. When  $\mu = 0.0$ ,  $\Phi$  will be reduce to topic-free word alignment probability; and when  $\mu = 1.0$ , there will be no smoothing in  $\Phi$ .

In A-UUTTM, we perform the user recommendation as follows. Suppose  $a$  is the user to be mentioned and  $w_d$  are the words in the microblog  $d$ . We can calculate the score for user  $a$  by:

$$p(a|w_d) = \sum_{k \in T} p(a|k, w_d, \Phi) \int p(k|\theta^u) d\theta^u \propto \sum_{k \in T} \sum_{w_m \in w_d} \theta_k^u \Phi_a^{k,w} p(w_m|w_d), \quad (21)$$

where  $p(w_m|w_d)$  is the weight of the word  $w_m$  in the microblog  $d$ , which we use the IDF score of the word. We can recommend the top-ranked users for the microblog according to the ranking scores.

## 4. EXPERIMENTS

### 4.1 Data sets and Settings

To examine the effectiveness of the proposed method, we constructed a dataset from Sina Weibo<sup>2</sup>, which is one of the most popular websites providing a Twitter-like microblogging service in China. It also allows users to follow each other. If user A follows user B, user A is called the *follower* of B and B is called the *followee* of A. We collected both microblogs and *following-followee* relations of users. We constructed the dataset in the following way. First, we randomly selected 200 users as the *central* users. Then we collected the 2-*ego* network for all the central users based on their followees<sup>3</sup>. Through these steps, the crawling process produced a total of 2.07 million users and 299.6 billion following relationships. For each user, we collected the 2,000 most recent microblogs, resulting in 84.8 million microblogs in total.

In this paper, we focus on the “@” behavior. Hence, we randomly selected 10 users from the total sample of 200 central users. All of the microblogs were collected to construct the data set in accordance with the following requirements: 1) they were posted by the selected 10 central users or their followees; 2) the posting date lies between Jan. 1, 2013 and Dec. 31, 2013; 3) at least one “@username” is included in each microblog. Using these criteria, we gathered 178,841 microblogs. For each mentioned user in the set of microblogs, we extracted 4 microblogs matching the information of the mentioned user from the time line. Although the number of followees of the selected 10 central users is 1,286, there are in total 240,191 users involved in the constructed data set. Table 1 lists the statistics of the constructed dataset. We use the microblogs posted from Jan. 1, 2013 to Oct. 31, 2013 as training data. The others are used as test data.

Table 1: Statistics of the constructed data set

Data Set	#Users	#Relations	#Microblogs
SINA	240,191	383,639	178,841

For the evaluation metrics, we use Precision, Recall and F-score to evaluate the performance of user recommendation methods for the highest ranked result. To evaluate whether candidates with the top  $n$  highest scores contain the correct result, in this study, we use Hits@3 and Hits@5, which means percentage of usernames can be correctly identified from the top  $n$  results. Since the rank of the results is usually important, we use the common metrics detailed below to measure the rank of the results:

- **Mean Reciprocal Rank(MRR):** The Reciprocal Rank (RR) calculates the reciprocal of the rank at

<sup>2</sup><http://www.weibo.com>

<sup>3</sup>According to Sina Weibo’s constraints, we can get only the top 200 followees of each user.

which the first correct result occurs. The averaged RR across microblogs is called the MRR, which is calculated as follows:

$$\frac{1}{Z} \sum_{u \in U} \sum_{d \in D_u} \frac{1}{\text{rank}_d},$$

where  $Z$  is the total number of microblogs.

- **Binary Preference Measure(Bpref):** For a microblog  $d$ , suppose an approach recommends a list of users  $A$ .  $J$  is a list of correctly recommended users in  $A$ . Let  $j$  be a correct recommendation and  $i$  be an incorrect recommendation. Bpref is defined as:

$$Bpref = \frac{1}{|J|} \sum_{j \in J} \left(1 - \frac{|\text{i ranked higher than j}|}{|A|}\right)$$

### 4.2 Baselines and Parameters Settings

For comparison with the proposed model, we evaluated the following methods on the constructed corpus:

- **Link-PLSA-LDA:** We implemented the method proposed in [28], where the microblogs content and the mentioned user information are incorporated under the topic model.
- **Frequency Descending (FD):** We recommend user to the microblog depend on the rank of frequency of the candidates in the history. This method can recommend the user who is frequently mentioned by the author of the microblog.
- **Citation Translation Model (CTM):** We implemented the method proposed in [14], which modifies the GIZA++ 1.07 [29] to learn translation probabilities using IBM Model-1. We set the number of training iterations equals to 10.
- **Ranking:** Following the approach proposed in [42], we also formalized the problem as a ranking problem and use support vector regression (SVR)<sup>4</sup> method to model it. We use same features set described in [42].

We run all of the topic-based models with 500 iterations of Gibbs sampling. We set the number of topics to 30 after trying a few different number of topics. We use  $\alpha = 50.0/T$  and  $\beta = 0.1$  as [10] suggested. We set parameter  $\rho$  to 0.01. The smoothing parameter  $\mu$  is set to 0.8. For all compared methods, we use the parameter settings mentioned in Section 4.2, which are designed to give the best performance.

### 4.3 Results and Discussions

Table 2 shows the comparisons of the proposed method with the state-of-the-art methods on the constructed evaluation dataset. From the results, we observe that the A-UUTTM outperforms all the other methods on the dataset across the different evaluation metrics and significantly improves the user’s recommendation robustly. The relative improvement of A-UUTTM over CTM is about 15.7% on MRR. The Precision, Recall and F-Score show us that the proposed method delivers the best result. The Bpref and MRR metrics show that the proposed method generates recommendations which are consistently ranked by users

<sup>4</sup>The toolkit scikit-learn 0.16.1 is used.

Table 2: Precision, Recall, F-Score, MRR, Bpref, Hits@3 and Hits@5 metrics on the dataset

Method	Precision	Recall	F-Score	MRR	Bpref	Hits@3	Hits@5
Link-PLSA-LDA	0.212	0.158	0.181	0.218	0.284	0.236	0.269
FD	0.340	0.253	0.290	0.341	0.385	0.309	0.341
Ranking	0.356	0.265	0.304	0.347	0.425	0.340	0.374
CTM	0.390	0.290	0.332	0.351	0.435	0.359	0.383
A-TTM	0.403	0.299	0.343	0.362	0.450	0.367	0.399
A-UTTM	0.424	0.315	0.362	0.385	0.482	0.394	0.426
A-UUTTM	<b>0.451</b>	<b>0.336</b>	<b>0.385</b>	<b>0.406</b>	<b>0.508</b>	<b>0.414</b>	<b>0.449</b>

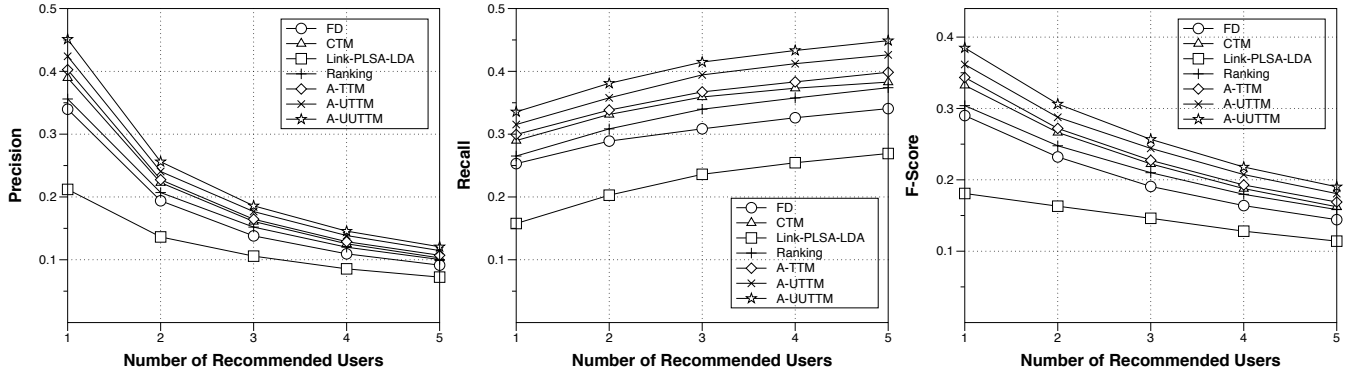


Figure 3: Precision, Recall and F-Score with recommended users varying from 1 to 5

as being superior. The results of Hits@3 and Hits@5 demonstrate that more than 41% of the original users can be found in the top 3 recommendations and around 45% of the original users can be found in the top 5 list.

From the results of CTM and Link-PLSA-LDA, we observe that the topic model and the translation model are effective for this task. Through comparing the results of the LDA-based model with CTM and Link-PLSA-LDA, we see that the A-TTM achieves better performance than either CTM or Link-PLSA-LDA on all the metrics. The results demonstrate that incorporating the topic into the translation model can impact the performance of the recommendation. The performances of ranking based method are worse than CTM. We think that the main reason may be caused by the features which are more suitable for predicting who will retweet it. From the results of the A-TTM and A-UTTM, we observe that the A-TTM, based on the standard LDA, does not perform very well on the short microblogs. On the other hand, the A-UUTTM inherits the advantages of A-TTM and A-UTTM. Viewing the single microblog as one topic and each user corresponding to a topic distribution can improve LDA performance for short microblogs. Comparing the metrics of the A-UTTM with the A-UUTTM, we see that possessing the information of the candidates is helpful for the recommendation task.

Figure 3 shows the performance of different methods on Precision, Recall and F-Score varying with the number of recommended users by the various methods. In these figures, x-axis denotes the number of users recommended and the y-axis indicates the results of Precision, Recall, and F-Score respectively. From the results, we observe that the A-UUTTM achieves consistently better performance than the other methods with different number of recommendations.

Intuitively, since heavy social media users post a variety of microblogs and often mention other users in their microblogs, recommendation for them should be much more easily when then input “@” symbol in a microblog. To investigate this intuition, we split the users into three groups based on the number of occurrence of “@” in their microblogs. Group 1 consists of the users whom microblogs mentioned the other users less than 300 times. The users in whom microblogs the number of mentioned users is more than 300 times and less than 600 times are included in the Group 2. Group 3 includes heavy social media users whom microblogs contains more than 600 “@”. Figure 4 shows the results of A-UUTTM on different groups. We see that the more frequently users mention others, the better performance the model can achieve. The relative improvement of A-UUTTM achieved in Group 3 over the results achieved in Group 1 is around 66.3% in MRR. The main possible reason is that the model can be better estimated with more training data.

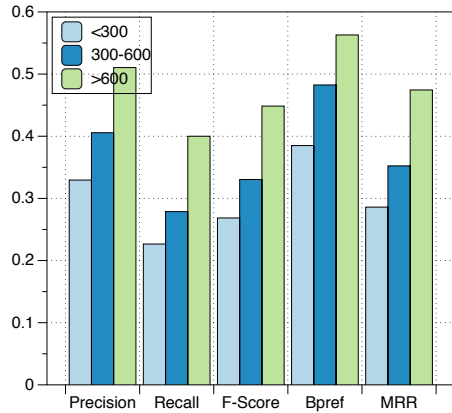
#### 4.4 Parameters Sensitive Analysis

In our model, there are two important parameters, the smoothing parameter  $\mu$  and the number of topics  $|T|$ . In this section, we show how the performance of our model varies as  $\mu$  changes from 0.0  $\sim$  1.0 and  $|T|$  changes from 10  $\sim$  50. When analyzing one parameter, we keep other parameters fixed to the settings mentioned in the previous section. Experimental results validate that the proposed methods can achieve stable and superior performance under a wide range of parameter values.

Figure 5 shows the results of A-UUTTM when the smoothing parameter  $\mu$  varies from 0.0  $\sim$  1.0. From the figure, we observe that our model obtains the best



Figure 4: The results between heavy social media users and normal users.



performance when  $\mu = 0.8$ . Furthermore, when the smoothing parameter  $\mu$  is from 0.2 ~ 0.8 we can achieve better performance than baseline methods. When  $\mu$  equals to 1.0, the performance decrease. This demonstrates that it is necessary to reduce the sparsity problem by exploiting smoothing in our model. Table 3 shows the influences of topic number, we see that when the topic number changes from 10 ~ 30, the performance of our model is much better than that of baseline methods, and we can obtain the best performance when  $|T|$  is around 30. However, as the number of topics increases, the data sparsity problem will be more serious when estimating the topic-specific probability matrix  $\phi$ , so the performance of our model will decrease accordingly. We can choose the topic number  $|T|$  from 10 ~ 30 and  $\gamma$  from 0.2 ~ 0.8 to reduce the difficulty of parameter selection.

Table 3: The influence of topic number  $|T|$  in the A-UUTTM.

$ T $	P	R	F	Bpref	MRR
10	0.445	0.331	0.380	0.505	0.402
20	0.446	0.332	0.381	0.506	0.403
30	<b>0.451</b>	<b>0.336</b>	<b>0.385</b>	<b>0.508</b>	<b>0.406</b>
40	0.434	0.323	0.370	0.488	0.390
50	0.422	0.314	0.360	0.477	0.382

Figure 6 shows the influence of the size of training data. According to the description, we used the posting date lies between Jan. 1, 2013 and Oct. 31, 2013 as training data. We trained A-UUTTM with several subsets selected based on the posting date. Comparing the result achieved by the model trained with the whole data set and part of it, we observe that A-UUTTM can achieve better performance with more training data. However, with only 2 months training data, A-UUTTM can achieve satisfactory results. This means that the more recent microblogs can make the greater contribution to the performance. Microblogs from more than 10 months ago offer little help for the recommendation. We also observe that the proposed method (A-UUTTM) achieves the best performance among all the methods when using only 5 months training microblogs.

Figure 5: The influence of the smoothing parameter  $\mu$  in the A-UUTTM.

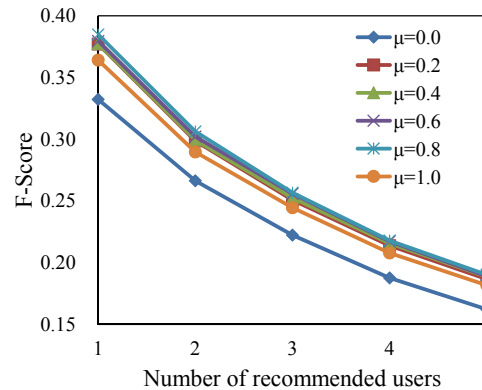


Figure 6: F-Score curve of different training size on this task.



## 5. CONCLUSIONS

In this paper, we studied the task of recommending usernames when users input the “@” symbol in their microblogs. We incorporated both the content of the microblog and histories of candidate users into translation based methods to perform the task. A-TTM incorporated the topical information into the translation method. To tackle the short text problem, we proposed the A-UUTTM model, which assumed that one microblog has a single topic and each user corresponds to a topic distribution. Furthermore, we proposed the A-UUTTM to incorporate the microblogs posted by the mentioned users. We evaluated the relative performance of the proposed methods based on a dataset collected from real world microblogging services. The experimental results demonstrated that our proposed algorithms can significantly outperform baseline methods.

## 6. ACKNOWLEDGEMENT

The authors wish to thank the anonymous reviewers for their helpful comments. This work was partially funded by National Natural Science Foundation of China (No. 61473092 and 61472088), the National High Technology Research and Development Program of China (No. 2015AA011802), and Shanghai Science and Technology Development Funds (13dz2260200iiĴ13511504300).

## 7. REFERENCES

- [1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, Mar. 2003.
- [2] P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational linguistics*, 1993.
- [3] J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, L. Zhang, and X. He. Music recommendation by unified hypergraph: combining social media information and music content. In *Proceedings of MM '10*, 2010.
- [4] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu. Collaborative personalized tweet recommendation. In *Proceedings of SIGIR '12*, 2012.
- [5] W.-Y. Chen, D. Zhang, and E. Y. Chang. Combinational collaborative filtering for personalized community recommendation. In *Proceedings of KDD '08*, 2008.
- [6] A. Das and A. Kannan. Discovering topical aspects in microblogs. In *Proceedings of COLING 2014*, 2014.
- [7] S. Debnath, N. Ganguly, and P. Mitra. Feature weighting in content based recommendation system using social network analysis. In *Proceedings of WWW'08*, 2008.
- [8] Z. Ding, X. Qiu, Q. Zhang, and X. Huang. Learning topical translation model for microblog hashtag suggestion. In *Proceedings of IJCAI '13*, 2013.
- [9] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 2004.
- [10] T. L. Griffiths and M. Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 2004.
- [11] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel. Social media recommendation based on people and tags. In *Proceedings of SIGIR' 10*, 2010.
- [12] P. Heymann, D. Ramage, and H. Garcia-Molina. Social tag prediction. In *Proceedings of SIGIR '08*, 2008.
- [13] T. Hofmann. Probabilistic latent semantic indexing. In *Proceedings of SIGIR '09*, 1999.
- [14] W. Huang, S. Kataria, C. Caragea, P. Mitra, C. L. Giles, and L. Rokach. Recommending citations: translating papers into references. In *Proceedings of CIKM '12*, 2012.
- [15] M. Kaminskas and F. Ricci. Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review*, 2012.
- [16] I. Konstas, V. Stathopoulos, and J. M. Jose. On social networks and collaborative recommendation. In *Proceedings of SIGIR '09*, 2009.
- [17] R. Krestel, P. Fankhauser, and W. Nejdl. Latent dirichlet allocation for tag recommendation. In *Proceedings of RecSys '09*, 2009.
- [18] N. Lathia, S. Hailes, L. Capra, and X. Amatriain. Temporal diversity in recommender systems. In *Proceedings of SIGIR '10*, 2010.
- [19] Q. Li, J. Wang, Y. P. Chen, and Z. Lin. User comments for news recommendation in forum-based social media. *Information Sciences*, 2010.
- [20] M. Lienou, H. Maitre, and M. Datcu. Semantic annotation of satellite images using latent dirichlet allocation. *Geoscience and Remote Sensing Letters, IEEE*, 2010.
- [21] Z. Liu, C. Liang, and M. Sun. Topical word trigger model for keyphrase extraction. In *Proceedings of COLING*, 2012.
- [22] S. Lo and C. Lin. Wmr—a graph-based algorithm for friend recommendation. In *Proceedings of WI '06*, 2006.
- [23] Y. Lu, J. He, D. Shan, and H. Yan. Recommending citations with translation model. In *Proceedings of CIKM '11*, 2011.
- [24] Y.-T. Lu, S.-I. Yu, T.-C. Chang, and J. Y.-j. Hsu. A content-based method to enhance tag recommendation. In *Proceedings of IJCAI'09*, 2009.
- [25] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble. In *Proceedings of SIGIR '09*, 2009.
- [26] J. D. McAuliffe and D. M. Blei. Supervised topic models. In *Proceedings of NIPS*, 2008.
- [27] M. Moricz, Y. Dosbayev, and M. Berlyant. Pymk: friend recommendation at myspace. In *Proceedings of SIGMOD '10*, 2010.
- [28] R. Nallapati and W. Cohen. Link-plsa-lda: A new unsupervised model for topics and influence of blogs. In *Proceedings of AAAA '08*, 2008.
- [29] F. J. Och and H. Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 2003.
- [30] Y. Pan, F. Cong, K. Chen, and Y. Yu. Diffusion-aware personalized social update recommendation. In *Proceedings of RecSys '13*, 2013.
- [31] M. Pennacchiotti and S. Gurumurthy. Investigating topic models for social media user recommendation. In *Proceedings of WWW '11*, 2011.
- [32] D. Putthividhy, H. T. Attias, and S. S. Nagarajan. Topic regression multi-modal latent dirichlet allocation for image annotation. In *Proceedings of CVPR '10*, 2010.
- [33] A. Rae, B. Sigurbjörnsson, and R. van Zwol. Improving tag recommendation using social networks. In *Proceedings of Adaptivity, Personalization and Fusion of Heterogeneous Information*, 2010.
- [34] S. Rendle, L. Balby Marinho, A. Nanopoulos, and L. Schmidt-Thieme. Learning optimal ranking with tensor factorization for tag recommendation. In *Proceedings of KDD '09*, 2009.
- [35] I. Ronen, I. Guy, E. Kravi, and M. Barnea. Recommending social media content to community owners. In *Proceedings of SIGIR '14*, 2014.
- [36] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of WWW '10*, 2010.
- [37] M. Schedl and D. Schnitzer. Hybrid retrieval approaches to geospatial music recommendation. In *Proceedings of SIGIR '13*, 2013.
- [38] E. Shmueli, A. Kagian, Y. Koren, and R. Lempel. Care to comment?: recommendations for commenting on news stories. In *Proceedings of WWW '12*, 2012.
- [39] Y. Song, Z. Zhuang, H. Li, Q. Zhao, J. Li, W.-C. Lee, and C. L. Giles. Real-time automatic tag recommendation. In *Proceedings of SIGIR '08*, 2008.
- [40] A. Tariq, A. Karim, F. Gomez, and H. Foroosh. Exploiting topical perceptions over multi-lingual text for hashtag suggestion on twitter. In *FLAIRS Conference*, 2013.
- [41] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical dirichlet processes. *Journal of the american statistical association*, 2006.
- [42] B. Wang, C. Wang, J. Bu, C. Chen, W. V. Zhang, D. Cai, and X. He. Whom to mention: expand the diffusion of tweets by@ recommendation on micro-blogging systems. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1331–1340. International World Wide Web Conferences Steering Committee, 2013.
- [43] Y. Wang, E. Agichtein, and M. Benzi. Tm-lda: Efficient online modeling of latent topic transitions in social media. In *Proceedings of KDD '12*, 2012.
- [44] R. Yan, M. Lapata, and X. Li. Tweet recommendation with graph co-ranking. In *Proceedings of ACL '12*, 2012.
- [45] M. Ye, P. Yin, and W.-C. Lee. Location recommendation for location-based social networks. In *Proceedings of SIGSPATIAL '10*, 2010.
- [46] Q. Zhang, Y. Gong, X. Sun, and X. Huang. Time-aware personalized hashtag recommendation on social media. In *Proceedings of COLING 2014*, 2014.
- [47] W. Zhang, J. Wang, and W. Feng. Combining latent factor model with location features for event-based group recommendation. In *Proceedings of SIGKDD '13*, 2013.