The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Dynamic User Profiling for Streams of Short Texts

Shangsong Liang
Department of Computer Science,
University College London, United Kingdom
shangsong.liang@ucl.ac.uk

Abstract

In this paper, we aim at tackling the problem of dynamic user profiling in the context of streams of short texts. Profiling users' expertise in such context is more challenging than in the case of long documents in static collection as it is difficult to track users' dynamic expertise in streaming sparse data. To obtain better profiling performance, we propose a streaming profiling algorithm (SPA). SPA first utilizes the proposed user expertise tracking topic model (UET) to track the changes of users' dynamic expertise and then utilizes the proposed streaming keyword diversification algorithm (SKDA) to produce top-k diversified keywords for profiling users' dynamic expertise at a specific point in time. Experimental results validate the effectiveness of the proposed algorithms.

Introduction

Microblogging platforms such as Twitter provide a lightweight, easy form of communication that enables users to broadcast and share information about their recent activities, opinions and status via short texts (Kwak et al. 2010). To effectively capture how users' expertise underlying their posts evolves over time is critical to the success of further design of applications such as identifying a list of users who are knowledgeable about a given topic (Balog et al. 2012). In this paper, we study the problem of *identifying the skills and knowledge of a user and tracking how they change over time in the context of streams of short texts*. Our goal is to infer users' topic distributions over time and dynamically profile their expertise with a set of keywords in the context of streams of short texts.

After the launch of expert finding task at TREC 2005 enterprise track (Craswell, de Vries, and Soboroff 2005), the study of *user profiling*, also called *expert profiling*, has generated a lot of interests in expertise retrieval. Most previous work on user profiling uses collection of static, long documents, and hence makes the assumption that users' expertise does not change over time. The task of temporal expertise profiling was first introduced recently in (Rybak, Balog, and Nørvåg 2014), where users' expertise is assumed to be changed over time, and was further studied in (Fang and Godavarthy 2014). However, both of these recent work still work with a set of long documents. To identify and track users' expertise for user profiling in the context of short text streams is more challenging than in the context of long document streams, as documents are short and thus sparse information for inferring users' expertise distributions.

To tackle the problem of dynamic user profiling for streams of short texts, we propose a **S**treaming **P**rofiling **A**lgorithm, abbreviated as **SPA**. SPA algorithm first utilizes our proposed **U**ser **E**xpertise **T**racking topic model (**UET**) to track the changes of users' dynamic expertise. It then utilizes our proposed **S**treaming **K**eyword **D**iversification **A**lgorithm (**SKDA**) to produce top-k diversified keywords as the profiles of users' expertise at a specific point in time.

Our proposed UET is able to capture the evolution of latent topics for users in streams of short texts. Most previous topic models make the assumption that the content of documents is rich enough to infer per-document multinomial distribution of topics. This assumption is not held in the context of streams of short documents where the length of each document is no more than a predefined number, e.g., 140 characters in Twitter platform. In our UET model, to effectively tackle sparsity challenge and infer each user's latent topic distribution as their expertise at a specific point in time, we propose a collapsed Gibbs sampling algorithm where we assign a single topic to all the words of a short document.

In addition, most previous work on user profiling just simply retrieves a list of top-k keywords as their profiles that may be semantically similar to each other and thus redundant. However, users' expertise may be broad and the top-k keywords retrieved by the models should be diversified and cover as many aspects of their expertise as possible. To achieve this goal, our proposed SKDA algorithm, a streaming version of the PM-2 diversification algorithm (Proportionality-based diversification Method–2nd version (Dang and Croft 2012)), works with the output of UET to diversify and return top-k keywords as users' profiles at time t.

The contributions of the paper are fourfold: (i) We propose a user expertise tracking topic model, UET, that can track the changes of users' expertise distributions over time in the context of streams of short texts. (ii) We propose a collapsed Gibbs sampling algorithm to effectively infer users' dynamic expertise distributions in the context of streams of short texts. (iii) We propose a streaming keyword diversification algorithm, SKDA, to diversify the top-k keywords for users' profiling. (iv) We systematically analyze the proposed streaming profiling algorithm, SPA, that consists of UET and SKDA, and find that we achieve better performance compared to the state-of-the-art non-streaming and streaming user profiling models.

Related Work

Two lines of work are related to ours, *user profiling* and *topic models*.

User Profiling. User profiling, also called expert profiling, has been gaining attention after the launch of expert finding task at TREC 2005 enterprise track (Craswell, de Vries, and Soboroff 2005). Balog and de Rijke (2007) worked with a static long document corpus and modeled the profile of an expert as a vector, where each element of the vector corresponds the person's skills on the given knowledge area. Later, Balog et al. (Balog et al. 2007) proposed a generative language modeling algorithm for the task and the experiments were conducted on, again, static long document corpora. (Berendsen et al. 2013) provided a critical assessment and analysis for the evaluation of user profiling systems with static long documents. Recent work was aware of the importance of temporal user profiling. Temporal expertise profiling for long documents was first introduced in (Rybak, Balog, and Nørvåg 2014), where topical areas were organized in a predefined taxonomy and expertise was represented as a weighted unchanged tree built directly by the ACM computing classification system. A probabilistic model was proposed in (Fang and Godavarthy 2014), where experts' academic publications were used to investigate and predict how personal expertise evolves over time. To the best of our knowledge, none of existing user profiling algorithms works with streams of short documents and diversifies the keywords for profiling.

Topic Models. Topic models provide a suite of algorithms to discover hidden thematic structure in a collection of documents. A topic model takes a set of documents as input, and discovers a set of "latent topics"—recurring themes that are discussed in the collection—and the degree to which each document exhibits those topics (Blei, Ng, and Jordan 2003). Since the well-known topic models, PLSI (Probabilistic Latent Semantic Indexing) (Hofmann 1999) and LDA (Latent Dirichlet Allocation) (Blei, Ng, and Jordan 2003), were proposed, topic models with dynamics have been widely studied. These include the Dynamic Topic Model (DTM) (Blei and Lafferty 2006), Dynamic Mixture Model (DMM) (Wei, Sun, and Wang 2007), Topic over Time (ToT) (Wang and McCallum 2006), Topic Tracking Model (TTM) (Iwata et al. 2009). Most of these previous dynamic topic models worked in the context of long text streams. Recent work realized the importance of dynamically modeling topics of short documents in streams, which includes Dynamic Clustering Topic model (DCT) (Liang, Yilmaz, and Kanoulas 2016), dynamic User Clustering Topic model (UCT) (Liang et al. 2017b; Zhao et al. 2016), Dynamic Dirichlet Multinomial Mixture topic Model (D2M3) (Liang et al. 2017c), and User Collaborative Interest Tracking topic model (UCIT) (Liang et al. 2017a). However, these models aimed at different applications rather than user profiling, and how to employ them into user profiling is still unknown. To our knowledge, none of existing dynamic topic models has considered the problem of user profiling in the context of streams of short texts.

Algorithm 1: Overview of the proposed SPA algorithm.

```
Input  : A set of users $\mathbf{u}_t$ along with their tweets $\mathbf{D}_t$.
Output: Profiling results of users at time $t$, $\mathbf{W}_t$.
1 Use UET model to track each user's expertise as $\boldsymbol{\theta}_{t,u}$,
  which is inferred by the proposed Gibbs algorithm.
2 Use SKDA algorithm to retrieve top-$k$ diversified
  keywords for each user's dynamic profile based on his
  topic distribution at time $t$ generated from UET.
```

Problem Formulation

The problem we address is to track users' dynamic expertise and identify top-k keywords for their profiles over time in the context of streams of short texts. The dynamic user profiling algorithm is essentially a function h that satisfies:

$$\mathbf{D}_t, \mathbf{u}_t \xrightarrow{h} \mathbf{W}_t,$$

where $\mathbf{D}_t = \{\ldots, \mathbf{d}_{t-2}, \mathbf{d}_{t-1}, \mathbf{d}_t\}$ represents the *stream* of short documents generated by the users $\mathbf{u}_t$ up to time t with $\mathbf{d}_t$ being the most recent set of short documents arriving at time period t, $\mathbf{u}_t = \{u_1, u_2, \ldots, u_{|\mathbf{u}_t|}\}$ represents a set of users appearing in the stream up to time t, with u_i being the i-th user in $\mathbf{u}_t$ and $|\mathbf{u}_t|$ being the total number of users in the user set, and $\mathbf{W}_t = \{\mathbf{w}_{t,u_1}, \mathbf{w}_{t,u_2}, \ldots, \mathbf{w}_{u_{t,|\mathbf{u}_t|}}\}$ represents all users' profiling results at time t with $\mathbf{w}_{t,u_i} = \{w_{t,u_i,1}, w_{t,u_i,2}, \ldots, w_{t,u_i,k}\}$ being the expertise profiling result, i.e., the top-k diversified keywords, for user u_i at time t. We assume that the length of a document d in $\mathbf{D}_t$ is no more than a predefined small length (for instance, 140 characters in the case of Twitter).

Method

In this section, we detail our proposed **S**treaming **P**rofiling **A**lgorithm (**SPA**) for the dynamic user profiling task.

Overview

We provide an overview of our proposed SPA in Algorithm 1 that consists of our proposed **U**ser **E**xpertise **T**racking topic model (**UET**) and the proposed **S**treaming **K**eyword **D**iversification **A**lgorithm (**SKDA**), where we use Twitter as our default setting of streams of short texts. We represent each user's expertise by topics. Thus, the expertise of each user $u \in \mathbf{u}_t$ at time period t is represented as a multinomial distribution $\boldsymbol{\theta}_{t,u} = \{\theta_{t,u,z}\}_{z=1}^{Z}$, where Z is the total number of topics. Our proposed UET tracking topic model captures each user's dynamic expertise $\boldsymbol{\theta}_{t,u}$, which is inferred by our proposed collapsed Gibbs sampling algorithm (step 1 of Algorithm 1). Then, our proposed SKDA diversification algorithm identifies top-k keywords for profiling users' expertise at time t (step 2 of Algorithm 1). In the following, we detail the UET, the Gibbs sampling, and the SKDA algorithms, respectively.

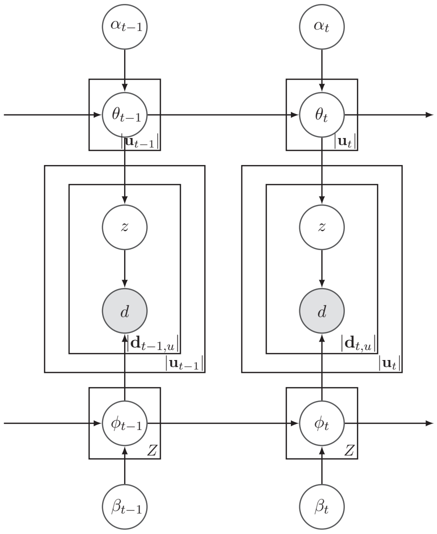

Figure 1: Graphical representation of our UET model. Shaded nodes represent observed variables, whereas other nodes represent random variables.

User Expertise Tracking Topic Model

Modeling Expertise over Time. Our proposed UET topic model aims at inferring the dynamic topic distribution, i.e., expertise, of each user, $\boldsymbol{\theta}_{t,u} = \{\theta_{t,u,z}\}_{z=1}^{Z}$, in stream of short documents at a given time period t. The graphical representation of our UET model is shown in Fig. 1.

To track the dynamics of a user u's expertise, we make the assumption that the mean of his current expertise at t is the same as that at the previous time period $t-1$, unless otherwise newly arrived documents at the current time period are observed. Specifically, we follow the work of previous dynamic topic models (Gao et al. 2017; Wei, Sun, and Wang 2007; Iwata et al. 2009; 2010), and use the following Dirichlet prior with a set of precision $\boldsymbol{\alpha}_t = \{\alpha_{t,z}\}_{z=1}^{Z}$, where we let the mean of the current distribution $\boldsymbol{\theta}_{t,u}$ depend on the mean of the previous distribution $\boldsymbol{\theta}_{t-1,u}$:

$$P(\boldsymbol{\theta}_{t,u}|\boldsymbol{\theta}_{t-1,u}, \boldsymbol{\alpha}_t) \propto \prod_{z=1}^{Z} \theta_{t,u,z}^{\alpha_{t,u,z}\theta_{t-1,u,z}-1}, \qquad (1)$$

where the precision value $\alpha_{t,z} = \{\alpha_{t,u,z}\}_{u=1}^{|\mathbf{u}_t|}$ represents users' topic persistency, which is how saliency topic z is at time t compared to that at $t-1$ for the users. As the distribution is a conjugate prior of the multinomial distribution, the inference can be performed by Gibbs sampling (Liu 1994).

Similarly, to model the dynamic changes of the multinomial distribution of words specific to topic z, we assume a Dirichlet prior, in which the mean of the current distribution $\boldsymbol{\phi}_{t,z} = \{\phi_{t,z,v}\}_{v=1}^{V}$ evolves from the mean of the previous distribution $\boldsymbol{\phi}_{t-1,z}$:

$$P(\boldsymbol{\phi}_{t,z}|\boldsymbol{\phi}_{t-1,z}, \boldsymbol{\beta}_t) \propto \prod_{v=1}^{V} \phi_{t,z,v}^{\beta_{t,z,v}\phi_{t-1,z,v}-1}, \qquad (2)$$

where V is the total number of words in a vocabulary $\mathbf{v} = \{v_i\}_{i=1}^{V}$ and $\boldsymbol{\beta}_t = \{\beta_{t,v}\}_{v=1}^{V}$, with $\beta_{t,v} = \{\beta_{t,z,v}\}_{z=1}^{Z}$ being the persistency of the word v in all topics at t, a measure of how consistently the word belongs to the topics at time period t compared to that at the previous time period $t-1$. Later in this section, we propose a collapsed Gibbs sampling algorithm to infer all users' dynamic expertise distributions $\boldsymbol{\Theta}_t = \{\boldsymbol{\theta}_{t,u}\}_{u=1}^{|\mathbf{u}_t|}$ and the words' dynamic topic distributions $\boldsymbol{\Phi}_t = \{\boldsymbol{\phi}_{t,z}\}_{z=1}^{Z}$, and describe the update rules of the persistency values $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$.

Assuming that we know all users' topic distribution at period $t-1$, $\boldsymbol{\Theta}_{t-1}$, and the words' topic distribution, $\boldsymbol{\Phi}_{t-1}$, the proposed user expertise tracking model is a generative topic model that depends on $\boldsymbol{\Theta}_{t-1}$ and $\boldsymbol{\Phi}_{t-1}$. For initialization, i.e., $t=0$, we let $\theta_{0,u,z} = 1/Z$ and $\phi_{0,z,v} = 1/V$. Let $\mathbf{d}_{t,u}$ ($\mathbf{d}_{t,u} \in \mathbf{d}_t$) be the set of documents posted by user u at time period t. The generative process (used by the Gibbs sampler for parameter estimation) of our model at t, is the following:

i. Draw Z multinomials $\boldsymbol{\phi}_{t,z}$, one for each topic z, from Dirichlet distributions $\boldsymbol{\beta}_{t,z}\boldsymbol{\phi}_{t-1,z}$;

ii. Draw $|\mathbf{u}_t|$ multinomials $\boldsymbol{\theta}_{t,u}$, one for each user $u \in \mathbf{u}_t$, from Dirichlet distributions $\boldsymbol{\alpha}_{t,u}\boldsymbol{\theta}_{t-1,u}$;

iii. For each document $d \in \mathbf{d}_{t,u}$, draw a single topic z_d form the multinomial distribution $\boldsymbol{\theta}_{t,u}$ and for each word v_d in the short document d:

   (a) Draw a word $v_d \in d$ from multinomial $\boldsymbol{\phi}_{t,z_d}$;

The graphical representation of our dynamic UET model is shown in Fig. 1. Given the documents in streams are short, and because most of the short documents are likely to talk about one single topic only (Yin and Wang 2014), we let all the words in the same document d be drawn from the multinomial distribution associated with the same topic z_d; see Fig. 1 and the above generative process of the model.

Inferring Expertise Distributions. We propose a collapsed Gibbs sampling algorithm developed from the basic collapsed Gibbs sampler (Griffiths and Steyvers 2004) to approximately infer distribution parameters of our UET. As shown in Fig. 1 and the generative process, we adopt a conjugate prior (Dirichlet) for the multinomial distributions, and thus we can easily integrate out the uncertainty associated with multinomials $\boldsymbol{\theta}_{t,u}$ and $\boldsymbol{\phi}_{t,z}$. In this way, we enable sampling since we do not need to sample these multinomials.

The overview of our proposed collapsed Gibbs sampling algorithm is shown in Algorithm 2, where $m_{t,u,z}$ and $n_{t,z,v}$ are the number of documents assigned to topic z and the number of times word v assigned to topic z for user u at t, respectively. In the Gibbs sampling procedure we need to calculate the conditional distribution $P(z_{t,u,d}|\mathbf{z}_{t,-(u,d)}, \mathbf{d}_t, \boldsymbol{\Theta}_{t-1}, \boldsymbol{\Phi}_{t-1}, \mathbf{u}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t)$ at time t, where $\mathbf{z}_{t,-(u,d)}$ represents the topic assignments for all the documents in $\mathbf{d}_t$ except the document $d \in \mathbf{d}_{t,u}$ associated with user u at t. For obtaining this conditional distribution, we begin with the joint probability of the current document set, $P(\mathbf{d}_t, \mathbf{z}_t|\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Phi}_{t-1}, \mathbf{u}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t)$ at time t:

$$P(\mathbf{d}_t, \mathbf{z}_t|\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Phi}_{t-1}, \mathbf{u}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t) =$$

Algorithm 2: Inference for our UET model at time t.

```
Input  : Distributions $\mathbf{\Theta}_{t-1}$ and $\mathbf{\Phi}_{t-1}$ at $t-1$; Initialized $\boldsymbol{\alpha}_t$
         and $\boldsymbol{\beta}_t$; Number of iterations $N_{iter}$.
Output: Current distributions $\mathbf{\Theta}_t$ and $\mathbf{\Phi}_t$.
1 Initialize topic assignments randomly for all documents in $\mathbf{d}_t$.
2 for $iteration$ = 1 to $N_{iter}$ do
3     for $user$ = 1 to $|\mathbf{u}_t|$ do
4         for $d$ = 1 to $\mathbf{d}_{t,u}$ do
5             Draw $z_{t,u,d}$ from
              $P(z_{t,u,d}|\mathbf{z}_{t,-(u,d)}, \mathbf{d}_t, \mathbf{\Theta}_{t-1}, \mathbf{\Phi}_{t-1}, \mathbf{u}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t)$
6             Update $m_{t,u,z_{t,u,d}}$ and $n_{t,z_{t,u,d},v}$
7     Update $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$.
8 Compute the posterior estimates $\mathbf{\Theta}_t$ and $\mathbf{\Phi}_t$.
```

Algorithm 3: SKDA algorithm to generate top-k keywords for dynamically profiling each user's expertise.

```
Input  : Current distributions $\mathbf{\Theta}_t$ and $\mathbf{\Phi}_t$
Output: All users' profiling results at time $t$, $\mathbf{W}_t$
1 for $u = 1, \ldots, |\mathbf{u}_t|$ do
2     $\mathbf{w}_{t,u} \leftarrow \varnothing$                    /* $\mathbf{w}_{t,u} \in \mathbf{W}_t$ */
3     $\widetilde{\mathbf{v}} \leftarrow \mathbf{v}$
4     for $z = 1, \ldots, Z$ do
5         $e_{z|t,u} \leftarrow P(z|t,u)$
6         $s_{z|t,u} \leftarrow 0$
7     for all positions in the rank list $\mathbf{w}_{t,u}$ do
8         for $z = 1, \ldots, Z$ do
9             $qt[z|t,u] = \frac{e_{z|t,u}}{2s_{z|t,u}+1}$
10        $z^* \leftarrow \arg\max_z qt[z|t,u]$
11        $v^* \leftarrow \arg\max_{v \in \widetilde{\mathbf{v}}} \lambda_1 \times qt[z^*|t,u] \times$
          $P(v|t,z^*) + \lambda_2 \sum_{z \neq z^*} qt[z|t,u] \times P(v|t,z) +$
          $(1-\lambda_1-\lambda_2) \times \text{tfidf}(v|t,u)$
12        $\mathbf{w}_{t,u} \leftarrow \mathbf{w}_{t,u} \cup \{v^*\}$     /* append $v^*$ to
          $\mathbf{w}_{t,u}$ */
13        $\widetilde{\mathbf{v}} \leftarrow \widetilde{\mathbf{v}} \backslash \{v^*\}$  /* remove $v^*$ from $\widetilde{\mathbf{v}}$ */
14        for $z = 1, \ldots, Z$ do
15            $s_{z|t,u} \leftarrow s_{z|t,u} + \frac{P(v^*|t,u)}{\sum_{z'=1}^{Z} P(v^*|t,z')}$
```

$$\prod_{z=1}^{Z} \frac{\Gamma(\sum_{v=1}^{V} \beta_{t,z,v}\overline{\phi})}{\prod_{v=1}^{V} \Gamma(\beta_{t,z,v}\overline{\phi})} \prod_{z=1}^{Z} \frac{\prod_{v=1}^{V} \Gamma(n_{t,z,v} + \beta_{t,z,v}\overline{\phi})}{\Gamma(\sum_{v=1}^{V} n_{t,z,v} + \beta_{t,z,v}\overline{\phi})} \times$$

$$\prod_{u=1}^{|\mathbf{u}_t|} \frac{\Gamma(\sum_{z=1}^{Z} \alpha_{t,u,z}\overline{\theta})}{\prod_{z=1}^{Z} \Gamma(\alpha_{t,u,z}\overline{\theta})} \prod_{u=1}^{|\mathbf{u}_t|} \frac{\prod_{z=1}^{Z} \Gamma(m_{t,u,z} + \alpha_{t,u,z}\overline{\theta})}{\Gamma(\sum_{z=1}^{Z} m_{t,u,z} + \alpha_{t,u,z}\overline{\theta})}, \quad (3)$$

where we let $\overline{\theta}$ and $\overline{\phi}$ abbreviate for $\theta_{t-1,u,z}$ and $\phi_{t-1,z,v}$, respectively, $\Gamma(\cdot)$ is a gamma function, and $\mathbf{z}_t$ is the topic assignments to all documents in $\mathbf{d}_t$. Based on the above joint probability and using the chain rule, we can obtain the following conditional probability conveniently for the proposed Gibbs sampling (step 5 of Algorithm 2) as:

$$P(z_{t,u,d} = z|\mathbf{z}_{t,-(u,d)}, \mathbf{d}_t, \mathbf{\Theta}_{t-1}, \mathbf{\Phi}_{t-1}, \mathbf{u}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t) =$$

$$\frac{m_{t,u,z} + \alpha_{t,u,z}\overline{\theta} - 1}{\sum_{z=1}^{Z}(m_{t,u,z} + \alpha_{t,u,z}\overline{\theta}) - 1} \times$$

$$\frac{\prod_{v \in d} \prod_{j=1}^{N_{d,v}} (n_{t,z,v,-(u,d)} + \beta_{t,z,v}\overline{\phi} + j - 1)}{\prod_{i=1}^{N_d} (n_{t,z,-(u,d)} + i - 1 + \sum_{v=1}^{V} \beta_{t,z,v}\overline{\phi})}, \quad (4)$$

where N_d, $N_{d,v}$, $\mathbf{z}_{t,-(u,d)}$, $n_{t,z,v,-(u,d)}$ and $n_{t,z,-(u,d)}$ are the length of document d, the number of word v appearing in d, topic assignments for all documents except the document d from user u at t, the number of word v assigned to topic z in all documents except the one from user u at t, and the number of documents assigned to z in all documents except the one from user u at t, respectively. At each iteration during the sampling, the precision parameters $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ can be estimated by maximizing the joint distribution, i.e., (3). We apply fixed-point iterations to obtain the optimal $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$. By applying the two bounds in (Minka 2000), we can derive the following update rules of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\beta}_t$ for maximizing the joint distribution in our fixed-point iterations:

$$\alpha_{t,u,z} \leftarrow \frac{\alpha_{t,u,z}\left(\varPsi(m_{t,u,z} + \alpha_{t,u,z}\overline{\theta}) - \varPsi(\alpha_{t,u,z}\overline{\theta})\right)}{\varPsi(\sum_{z=1}^{Z} m_{t,u,z} + \alpha_{t,u,z}\overline{\theta}) - \varPsi(\sum_{z=1}^{Z} \alpha_{t,u,z}\overline{\theta})},$$

$$\beta_{t,z,v} \leftarrow \frac{\beta_{t,z,v}\left(\varPsi(n_{t,z,v} + \beta_{t,z,v}\overline{\phi}) - \varPsi(\beta_{t,z,v}\overline{\phi})\right)}{\varPsi(\sum_{v=1}^{V} n_{t,z,v} + \beta_{t,z,v}\overline{\phi}) - \varPsi(\sum_{v=1}^{V} \beta_{t,z,v}\overline{\phi})}, \quad (5)$$

where $\varPsi(x) = \frac{\partial \log \varGamma(x)}{x}$ is a Digamma function.

Once the Gibbs sampling has been done, with the fact that Dirichlet distribution is conjugate to multinomial distribution, we can conveniently infer each user's expertise distribution $\boldsymbol{\theta}_{t,u}$ at t and the words' topic distributions $\boldsymbol{\phi}_{t,z}$ at t, respectively as:

$$\theta_{t,u,z} = \frac{m_{t,u,z} + \alpha_{t,u,z}}{\sum_{z'=1}^{Z} m_{t,u,z'} + \alpha_{t,u,z'}},$$

$$\phi_{t,z,v} = \frac{n_{t,z,v} + \beta_{t,z,v}}{\sum_{v'=1}^{V} n_{t,z,v'} + \beta_{t,z,v'}}. \quad (6)$$

Streaming Keyword Diversification Algorithm

After we obtain each user's expertise distribution and the words' topic distributions, inspired by the PM-2 diversification algorithm (Dang and Croft 2012), we propose a streaming keyword diversification algorithm, i,e., SKDA, for dynamically profiling each user's expertise at time t (step 2 of Algorithm 1); see Algorithm 3.

To generate top-k diversified keywords for each user u at t, SKDA starts with an empty keyword set $\mathbf{w}_{t,u}$ with k empty seats (step 2 of Algorithm 3), and a set of candidate keywords (step 3), $\widetilde{\mathbf{v}}$, which is the whole words $\mathbf{v}$ in the vocabulary, i.e., initially let $\widetilde{\mathbf{v}} = \mathbf{v}$. For each of the seats, it computes the quotient $qt[z|t,u]$ for each topic z given a user u at time t by the Sainte-Laguë formula (step 9):

$$qt[z|t,u] = \frac{e_{z|t,u}}{2s_{z|t,u} + 1}, \quad (7)$$

where $e_{z|t,u}$ is the probability of the user u has expertise on topic z at t and is set to be $P(z|t,u)$ (step 5), and $s_{z|t,u}$ is the "number" of seats occupied by topic z (in initialization, $s_{z|t,u}$ is set to 0 for all topics (step 6)). We obtain $P(z|t,u)$ by our proposed UET algorithm such that we have $P(z|t,u) = \theta_{t,u,z}$. According to the Sainte-Laguë method, seats should be awarded to the topic with the largest quotient in order to best maintain the proportionality of the result list. Therefore, our SKDA assigns the current seat to the topic z^* with the largest quotient (step 10). The keyword to fill this seat is the one that is not only relevant to topic z^* but to other topics and should be specific to the user, and thus we propose to obtain the keyword v^* for profiling as (step 11):

$$
\begin{aligned}
v^* \leftarrow \arg\max_{v \in \widetilde{\mathbf{v}}} & \lambda_1 \times qt[z^*|t,u] \times P(v|t,z^*) + \\
& \lambda_2 \sum_{z \neq z^*} qt[z|t,u] \times P(v|t,z) + \\
& (1 - \lambda_1 - \lambda_2) \times \text{tfidf}(v|t,u), \qquad (8)
\end{aligned}
$$

where $0 \leq \lambda_1, \lambda_2 \leq 1$ are two free parameters that satisfy $0 \leq \lambda_1 + \lambda_2 \leq 1$, $P(v|t,z)$ is the probability that v is associated with topic z at time t and is set to be $P(v|t,z) = \phi_{t,z,v}$, and $\text{tfidf}(v|t,u)$ is a time-sensitive term frequency-inverse document frequency function for user u at t. We define it as:

$$
\text{tfidf}(v|t,u) = \text{tf}(v|\mathbf{d}_{t,u}) \times \text{idf}(v|u,\mathbf{d}_t), \qquad (9)
$$

where $\text{tf}(v|\mathbf{d}_{t,u}) = \frac{|\{d \in \mathbf{d}_{t,u} : v \in d\}|}{|\mathbf{d}_{t,u}|}$ is the term frequency function that computes how many percents of the documents that contain the word v in the whole document set $\mathbf{d}_{t,u}$, and $\text{idf}(v|u,\mathbf{d}_t) = \log \frac{|\mathbf{d}_t|}{|\{d \in \mathbf{d}_t : v \in d\}| + \epsilon}$ is the inverse document frequency function with ϵ being set to 1 to avoid the division-by-zero error. According to (9), if the word v frequently appears in the document set $\mathbf{d}_{t,u}$ generated by user u but not frequently appears in the document set $\mathbf{d}_t$ generated by all the users in $\mathbf{u}_t$, $\text{tfidf}(v|t,u)$ will return a high score. After the word v^* is selected, SKDA adds v^* as a result keyword to $\mathbf{w}_{t,u}$ for profiling the user u at time t, i.e., $\mathbf{w}_{t,u} \leftarrow \mathbf{w}_{t,u} \cup \{v^*\}$ (step 12), removes it from the candidate word set $\widetilde{\mathbf{v}}$, i.e., $\widetilde{\mathbf{v}} \leftarrow \widetilde{\mathbf{v}} \backslash \{v^*\}$ (step 13), and increases the "number" of seats occupied by each of the topics z by its normalized relevance to v^* (step 15):

$$
s_{z|t,u} \leftarrow s_{z|t,u} + \frac{P(v^*|t,u)}{\sum_{z'=1}^{Z} P(v^*|t,z')}. \qquad (10)
$$

The process (steps 7 to 15) repeats until we get k keywords for $\mathbf{w}_{t,u}$. The order in which a keyword is appended to $\mathbf{w}_{t,u}$ determines its ranking for the profiling. After the process is done, we obtain a set of diversified keywords $\mathbf{w}_{t,u}$ that profile the expertise of a user, $\boldsymbol{\theta}_{t,u}$, at time t.

Our SKDA differs from PM-2 diversification algorithm (Dang and Croft 2012) in at least four aspects: (i) SKDA aims at retrieving top-k diversified keywords; whereas PM-2 aims at retrieving top-k diversified documents. (ii) SKDA is a time-sensitive algorithm–the results change over time, and works for streams of short texts; whereas PM-2 works with a static set of long documents. (iii) SKDA proposed the tfidf scores for generating the keywords; whereas no tfidf scores are applied for diversifying documents in PM-2. (iv) The input of SKDA is the users' expertise distributions and the words' distributions over topics; PM-2 considers the aspects of the input query being uniform. Details about PM-2 can be found in (Dang and Croft 2012). Obviously, we can not directly apply PM-2 for our propose of dynamic user profiling in streams.

Experiments and Results

In this section, we describe our experimental setup, report and analyze the results.

Experimental Setup

Research Questions. The research questions guiding the remainder of the paper are : (**RQ1**) How does SPA perform for user profiling compared to state-of-the-art methods? (**RQ2**) How does the contribution of the proposed expertise tracking topic model UET to the overall performance of SPA compared to the contribution of other topic models? (**RQ3**) What is the impact of the length of the time intervals, $t_i - t_{i-1}$, in SPA? (**RQ4**) How is the generalization performance of UET compared to other topic models?

Dataset. In order to answer our research questions, we work with a dataset collected from Twitter.[1] The dataset contains 1,375 active users and their tweets that were posted from the beginning of their registration up to May 31, 2015. In total, we have 7.52 million tweets with each tweet having its own timestamp. The average length of the tweets is 12 words.

We use this dataset as our stream of short texts. We obtain two categories of ground truth: the **G**round **T**ruth from **M**anual judgements, abbreviated as MGT, and the **A**utomatically generated one, abbreviated as AGT. To create the MGT ground truth, each annotator (totally 20 annotators) was asked to generate a rank list of 10 keywords for one Twitter user (randomly chosen), respectively, after examining the content of the tweets at specific time periods, resulting in 20 users' profiles being labeled. We also propose a process to automatically obtain the ground truth for all users, i.e., the AGT ground truth: for each user at a specific time period, we rank the hashtags by the number of times they appear in the user's posts and assume the top-k hashtags as the keywords for his profile in the ground truth, resulting in all the 1,375 Twitter users having their own keywords as profiles in AGT at that time period. We transfer the format of each hashtag in this way: simply remove the first character '#', convert any capital letters to lowercase ones and keep the content as final hashtag, e.g., transferring the original hashtag "#SocialMedia" to "socialmedia". We found that the maximum cosine similarity between the word embeddings of the keywords in MGT and those in AGT is as high as 0.82. In total, we obtain the MGT and the AGT ground truths for 5 different partitions of time periods, i.e., a week, a month, a quarter, half a year and a year, respectively.

Baselines. We make comparisons among our SPA, the baselines and the following state-of-the-art algorithms:

tfidf. It simply utilizes (9), i.e., the content of users' documents to retrieve top-k keywords as profiles for the users.

Predictive Language Model (PLM). It models the dynamics of personal expertise via a probabilistic language model (Fang and Godavarthy 2014).

Latent Dirichlet Allocation (LDA). This model (Blei, Ng, and Jordan 2003) infers topic distributions specific to each document via the LDA model.

[1] Crawled from https://dev.twitter.com/.

Author Topic Model (AuthorT). This model (Rosen-Zvi et al. 2004) infers topic distributions specific to each user in a static dataset.

Dynamic Topic Model (DTM). This model (Blei and Lafferty 2006) utilizes a Gaussian distribution for inferring topic distribution of long documents in streams.

Topic over Time model (ToT). This model (Wang and McCallum 2006) normalizes timestamps of long documents in a collection and then infers topics distribution for each document.

Topic Tracking Model (TTM). This model (Iwata et al. 2009) captures the dynamic topic distributions of long documents arriving at time t in streams based on the content of the documents and the previous estimated topic distributions.

GSDMM. This is a Gibbs Sampling-based Dirichlet Multinomial Mixture model that assigns one topic for each short document in a static collection (Yin and Wang 2014).

The only difference between our SPA and the other expertise profiling baseline topic models, LDA, AuthorT, DTM, ToT, TTM and GSDMM, is that SPA utilizes our proposed UET topic model while the baseline models utilize their own topic models for obtaining users' expertise distributions. The baselines, tfidf, PLM and AuthorT, are static profiling algorithms, while the others are dynamic. Other baselines such as modified revisions of previous dynamic topic models, AuthorT, DTM, ToT and TTM etc., that apply the same inference strategy as that in our SPA, i.e., drawing one topic per document during the sampling, would be possible. However, their inference and the parameters' estimation in the revisions would be totally different from those of the corresponding original models. To keep focused, we keep the research on applying the drawing strategy of one topic per text into previous models as future work. For fair comparisons, SPA and all the other topic models use our SKDA algorithm to obtain the top-k keywords. We set the number of topics $Z = 50$ in all the topic models. For tuning parameters λ_1 and λ_2 in (8), we use a 70%/20%/10% split for our training, validation and test sets, respectively. In the training we vary the parameters λ_1 and λ_2 from 0.0 to 1.0. The best parameters are then chosen on the validation set, and evaluated on the test set. The train/validation/test splits are permuted until all users were chosen once for the test set. We repeat the experiments 10 times and report the average results.

Evaluation Metrics. Standard evaluation metrics, Pre@k (Precision at k), NDCG@k (Normalized Discounted Cumulative Gain at k), MRR@k (Mean Reciprocal Rank at k), and MAP@k (Mean Average precision at k), (Croft, Metzler, and Strohman 2015), are used for evaluation. We also propose semantic versions of the standard metrics, denoted as Pre-S@k, NDCG-S@k, MRR-S@k, and MAP-S@k, respectively. Here the only difference between the standard metrics and the corresponding semantic ones is the way to obtain the relevance score of a retrieval keyword v^* and the keyword in the ground truth v_{gt}. In standard metrics, we let the relevance score be 1 if and only if $v^* = v_{gt}$, otherwise be 0; whereas in the semantic versions, we let the relevance score be the cosine similarity between the word embedding vectors of v^* and v_{gt}, computed as $\cos(\mathbf{c}(v^*), \mathbf{c}(v_{gt}))$. Here $\mathbf{c}(v)$ is the word embedding of v pre-trained on a Twitter dataset (Mikolov et al. 2013), and the size of which is set to 300.[2] Since we usually choose not too many keywords to describe a user's profile, we compute the scores at depth 10, i.e., let $k = 10$. For all the metrics we abbreviate $M@k$ as M, where M is one of the metrics. Additionally, we adopt Perplexity to evaluate the generalization performance of the models. This metric, used by convention in many topic models (Blei and Lafferty 2006), is monotonically decreasing the likelihood of the test data, and is algebraically equivalent to the inverse of the geometric mean per-word likelihood.

Results and Analysis

Overall Performance. To begin, we answer the research question **RQ1**. Table 1 compares the performance of our SPA to that of the baseline models against the two ground truths, MGT and AGT, using time periods of a month on the standard metrics, Pre, NDCG, MRR, MAP, and the semantic versions, Pre-S, NDCG-S, MRR-S, MAP-S, respectively.

The following findings can be observed from Table 1: (i) All the topic model-based profiling algorithms, i.e., SPA, GSDMM, ToT, TTM, DTM, AuthorT and LDA, outperform traditional algorithms, i.e., PLM and tfidf, which demonstrates that topic modeling does help to profile users' expertise. (ii) SPA outperforms all the baseline models in terms of the two ground truths and all the metrics, which confirms its effectiveness for the expertise profiling task. (iii) The ordering of the models, SPA $>$ GSDMM $>$ ToT $\sim$ TTM $\sim$ DTM $\sim$ AuthorT $\sim$ LDA $>$ PLM $>$ tifdf, is mostly consistent across the two ground truths and the evaluation metrics. Here A $>$ B denotes statistically significantly higher performance and A $\sim$ B denotes that we did not observe a significant difference between A and B. This, once again, confirms the fact that the proposed method SPA outperforms all the baseline models. (iv) Both our SPA and GSDMM topic models draw one single topic for each document during the inference, whereas other baseline topic models, LDA, AuthorT, DTM, TTM, ToT, draw multiple topics for each document during the inference. According to Table 1, both SPA and GSDMM outperform other baseline topic models, which demonstrates the merit of our one topic per document sampling strategy that aims at tackling sparsity problem in streams of short texts for inference in our SPA model.

We further make the performance comparison with keywords from the AGT ground truth and those generated by SPA and the best baseline GSDMM for a randomly selected user, respectively. Table 2 shows the top 6 keywords of an example user's dynamic profile with time being five quarters from April 2014 to May 2015. As can be seen in the table, the keywords generated by SPA are semantically closer to those from the AGT ground truth compared to those generated by the baseline, GSDMM, which again demonstrates the effectiveness of the proposed SPA algorithm.

[2] Embeddings of both regular words and all hashtags are publicly available from https://nlp.stanford.edu/projects/glove/.

Table 1: Performance of SPA and the baselines using a time period of a month. Statistically significant differences between SPA and the best baseline, GSDMM, are marked in the upper right hand corner of SPA's scores, respectively. The statistical significance is tested using a two-tailed paired t-test and is denoted using $^{\blacktriangle}$ for $\alpha = .01$, and $^{\triangle}$ for $\alpha = .05$.

		Pre	NDCG	MRR	MAP	Prec-S	NDCG-S	MRR-S	MAP-S
MGT	tfidf	.269	.235	.674	.149	.447	.431	.867	.221
	PLM	.290	.249	.674	.155	.451	.433	.875	.230
	LDA	.298	.264	.674	.160	.453	.443	.878	.237
	AuthorT	.304	.268	.674	.160	.457	.440	.898	.238
	DTM	.310	.282	.694	.171	.463	.452	.877	.245
	TTM	.316	.290	.735	.175	.465	.458	.877	.247
	ToT	.329	.299	.755	.177	.469	.459	.878	.248
	GSDMM	.339	.318	.755	.181	.474	.467	.878	.254
	SPA	.365$^{\blacktriangle}$	.350$^{\blacktriangle}$	.813$^{\blacktriangle}$	.199$^{\blacktriangle}$	.488$^{\blacktriangle}$	.481$^{\blacktriangle}$	.918$^{\blacktriangle}$	.262$^{\triangle}$
AGT	tfidf	.216	.212	.633	.100	.314	.292	.816	.157
	PLM	.228	.215	.653	.102	.331	.309	.857	.165
	LDA	.239	.222	.674	.102	.349	.316	.877	.169
	AuthorT	.247	.238	.674	.107	.357	.326	.895	.175
	DTM	.257	.240	.694	.108	.367	.334	.898	.178
	TTM	.262	.239	.710	.107	.374	.343	.897	.180
	ToT	.267	.241	.714	.108	.384	.350	.898	.183
	GSDMM	.275	.257	.724	.116	.394	.359	.925	.188
	SPA	.304$^{\blacktriangle}$	.282$^{\blacktriangle}$	.735$^{\blacktriangle}$	.125$^{\blacktriangle}$	.425$^{\blacktriangle}$	.385$^{\blacktriangle}$	.939$^{\blacktriangle}$	.206$^{\blacktriangle}$

Contribution of UET. Next, we turn to answer research question **RQ2**. Recall that the only difference between our SPA and the baselines is that SPA utilizes our proposed UET topic model to track users' dynamic expertise and then the SKDA for diversifying the keywords, whereas other topic models utilize different topic models and then the SKDA for keyword diversification. As can be seen in Table 1, SPA outperforms all the other topic model-based baselines, i.e., GSDMM, ToT, TTM, DTM, AuthorT and LDA, which illustrates that the proposed topic model, UET, does be effective and has significant contribution to the performance of the streaming profiling algorithm.

Impact of Time Period Length. We now address research question **RQ3**. To understand the influence on SPA and the baselines of the length of the time period used for evaluation, we compare the performance for different time periods using AGT ground truth on standard metrics: a week, a month, a quarter, half a year and a year in Fig. 2, respectively. For the baselines, we take GSDMM and ToT as representatives only and use the standard metrics only, as the performance of other baselines is worse than that of GSDMM and ToT and has similar pattern on the semantic metrics.

As is shown in Fig. 2, SPA beats the baselines for time periods of all lengths, which illustrates the fact that SPA works better than the state-of-the-art algorithms for dynamic user profiling regardless of period length in the context of streams of short texts. The performance of SPA and the baselines improves significantly on all the metrics when the period length increases from a week to a month, whereas it reaches a plateau as the time periods further increase. In all the cases SPA significantly outperforms the baselines. These findings illustrate that the performance of the proposed SPA is robust and is able to maintain significant improvements over the state-of-the-art algorithms.

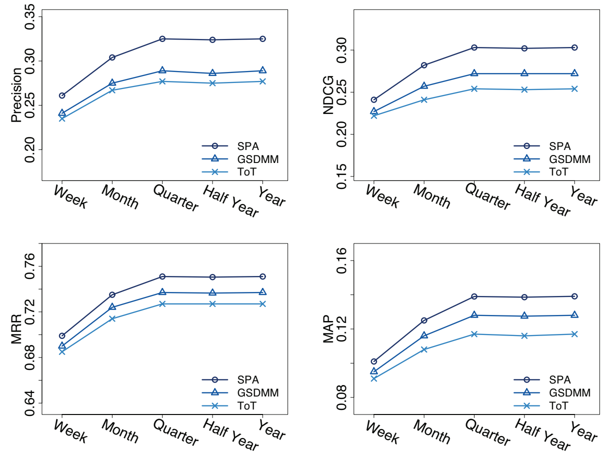

Figure 2: Performance of SPA and the best baselines on time periods of a week, a month, a quarter, half a year, and a year evaluated by the standard metrics, respectively.

Perplexity Comparisons. Finally, we turn to answer **RQ4** for understanding the generalization performance of UET and the baseline topic models. We use perplexity as evaluation metric for the comparisons. Fig. 3 shows the performance comparisons on perplexity. A lower perplexity score indicates better generalization performance. As is shown in the figure, UET outperforms all the baseline topic models.

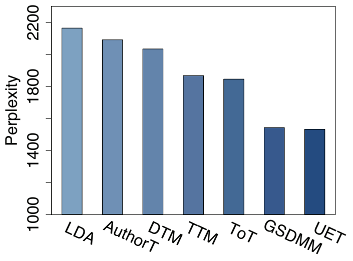

Figure 3: Generalization performance comparisons between UET and the baseline topic models.

Conclusions

We have studied the problem of dynamic user profiling for streams of short texts. To tackle the problem, we have proposed a streaming profiling algorithm, SPA, that first applies the proposed user expertise tracking topic model, UET, and then the proposed streaming keyword diversification algorithm, SKDA. Our UET dynamically tracks the changes of users' expertise distributions over time in a sequentially organized corpus of short text, and our SKDA diversifies the top-k keywords for profiling users' dynamic expertise. To effectively infer users' dynamic expertise distributions in our UET model, we have proposed a collapsed Gibbs sampling algorithm. We have conduced experiments on a Twitter dataset. We evaluated the performance of our SPA and the baseline algorithms using two categories of ground truth

Table 2: Top 6 keywords of an example user's dynamic profile with the time being five quarters from April 2014 to May 2015. The keywords from the AGT ground truth, generated from GSDMM and SPA are presented for the user, respectively.

	Apr. 2014 to Jun. 2014	Jul. 2014 to Sep. 2014	Oct. 2014 to Dec. 2014	Jan. 2015 to Mar. 2015	Apr. 2015 to May 2015
AGT	Apple Java iPhone Python ApplePay OjectiveC	Apple Git iPad OjectiveC AppleEvent Python	AppleEvent LininProfile openEducation iOS NatsTwitter education	Microblog Students LinkedInProfile ArtsEducation FB AfterSchool	SocialMedia Education NatsTwitter ConnectedLearning FB Courses
GSDMM	Apple Computer iPhone Science Java Technology	Apple Company University Technology iPad Language	Apple Christmas LinkedIn Education iOS Friends	Online Education Students Website Degree Presentation	Courses Online Presentation Digital Learning Education
SPA	Apple Language iPhone Programming Java Computer	Apple Programming iPad Git Event Language	Apple LinkedIn Education iOS Twitter Class	LinkedIn Students Microblog Education FB Art	Education Twitter Learning Media Courses FB

on both the standard metrics and the semantic versions of the metrics. Experimental results show that our SPA is able to profile users' dynamic expertise over time for streams of short texts. As future work, we plan to utilize other techniques, e.g., deep learning, to tackle the task and intent to incorporate other information such as users' social networks into the algorithm to further enhance the performance. Profiling for a group of knowledgeable experts (Liang and de Rijke 2016) would also be of value to investigate.

References

Balog, K., and de Rijke, M. 2007. Determining expert profiles (with and application to expert finding). In *IJCAI*, 2657–2662.

Balog, K.; Bogers, T.; Azzopardi, L.; de Rijke, M.; and van den Bosch, A. 2007. Broad expertise retrieval in sparse data environments. In *SIGIR*, 551–558.

Balog, K.; Fang, Y.; de Rijke, M.; Serdyukov, P.; and Si, L. 2012. Expertise retrieval. *Found. Trends Inf. Retr.* 6:127–256.

Berendsen, R.; Rijke, M.; Balog, K.; Bogers, T.; and Bosch, A. 2013. On the assessment of expertise profiles. *Journal of the Association for Information Science and Technology* 64(10):2024–2044.

Blei, D. M., and Lafferty, J. D. 2006. Dynamic topic models. In *ICML*, 113–120.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3:993–1022.

Craswell, N.; de Vries, A. P.; and Soboroff, I. 2005. Overview of the TREC 2005 enterprise track. In *TREC'05*, 1–7.

Croft, W. B.; Metzler, D.; and Strohman, T. 2015. *Search engines: Information retrieval in practice*. Addison-Wesley Reading.

Dang, V., and Croft, W. B. 2012. Diversity by proportionality: An election-based approach to search result diversification. In *SIGIR*, 65–74.

Fang, Y., and Godavarthy, A. 2014. Modeling the dynamics of personal expertise. In *SIGIR*, 1107–1110.

Gao, L.; Wu, J.; Zhou, C.; and Hu, Y. 2017. Collaborative dynamic sparse topic regression with user profile evolution for item recommendation. In *AAAI*, 1316–1322.

Griffiths, T. L., and Steyvers, M. 2004. Finding scientific topics. *PNAS* 101:5228–5235.

Hofmann, T. 1999. Probabilistic latent semantic indexing. In *SIGIR*, 50–57.

Iwata, T.; Watanabe, S.; Yamada, T.; and Ueda, N. 2009. Topic tracking model for analyzing consumer purchase behavior. In *IJCAI*, volume 9, 1427–1432.

Iwata, T.; Yamada, T.; Sakurai, Y.; and Ueda, N. 2010. Online multiscale dynamic topic models. In *KDD*, 663–672. ACM.

Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is twitter, a social network or a news media? In *WWW*, 499–508.

Liang, S., and de Rijke, M. 2016. Formal language models for finding groups of experts. *Information Processing & Management* 529–549.

Liang, S.; Ren, Z.; Yilmaz, E.; and Kanoulas, E. 2017a. Collaborative user clustering for short text streams. In *AAAI*, 3504–3510.

Liang, S.; Ren, Z.; Zhao, Y.; Ma, J.; Yilmaz, E.; and Rijke, M. D. 2017b. Inferring dynamic user interests in streams of short texts for user clustering. *ACM Trans. Inf. Syst.* 36(1):10:1–10:37.

Liang, S.; Yilmaz, E.; Shen, H.; Rijke, M. D.; and Croft, W. B. 2017c. Search result diversification in short text streams. *ACM Trans. Inf. Syst.* 36(1):8:1–8:35.

Liang, S.; Yilmaz, E.; and Kanoulas, E. 2016. Dynamic clustering of streaming short documents. In *KDD*, 995–1004.

Liu, J. S. 1994. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *J. Am. Stat. Assoc.* 89(427):958–966.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*, 3111–3119.

Minka, T. 2000. Estimating a dirichlet distribution.

Rosen-Zvi, M.; Griffiths, T.; Steyvers, M.; and Smyth, P. 2004. The author-topic model for authors and documents. In *UAI*, 487–494.

Rybak, J.; Balog, K.; and Nørvåg, K. 2014. Temporal expertise profiling. In *ECIR*, 540–546.

Wang, X., and McCallum, A. 2006. Topics over time: a non-markov continuous-time model of topical trends. In *KDD*, 424–433.

Wei, X.; Sun, J.; and Wang, X. 2007. Dynamic mixture models for multiple time-series. In *IJCAI*, 2909–2914.

Yin, J., and Wang, J. 2014. A dirichlet multinomial mixture model-based approach for short text clustering. In *KDD*, 233–242.

Zhao, Y.; Liang, S.; Ren, Z.; Ma, J.; Yilmaz, E.; and de Rijke, M. 2016. Explainable user clustering in short text streams. In *SIGIR*, 155–164.