

## Research Article

# A Dynamic Users' Interest Discovery Model with Distributed Inference Algorithm

Shuo Xu,<sup>1</sup> Qingwei Shi,<sup>1,2</sup> Xiaodong Qiao,<sup>3</sup> Lijun Zhu,<sup>1</sup> Han Zhang,<sup>1</sup> Hanmin Jung,<sup>4</sup> Seungwoo Lee,<sup>4</sup> and Sung-Pil Choi<sup>4</sup>

<sup>1</sup> Information Technology Support Center, Institute of Scientific and Technical Information of China, No. 15 Fuxing Road, Haidian District, Beijing 100038, China

<sup>2</sup> School of Software, Liaoning Technical University, No. 188 Longwan Street South, Huludao, Liaoning 125105, China

<sup>3</sup> College of Software, Northeast Normal University, 5268 Renmin Street, Changchun, Jilin 130024, China

<sup>4</sup> Department of Computer Intelligence Research, Korea Institute of Science and Technology Information, 245 Daehak-ro, Yuseong-gu, Daejeon 305-806, Republic of Korea

Correspondence should be addressed to Xiaodong Qiao; qiaox@istic.ac.cn

Received 6 December 2013; Accepted 27 February 2014; Published 22 April 2014

Academic Editor: Gorette Marreiros

Copyright © 2014 Shuo Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One of the key issues for providing users user-customized or context-aware services is to automatically detect latent topics, users' interests, and their changing patterns from large-scale social network information. Most of the current methods are devoted either to discovering static latent topics and users' interests or to analyzing topic evolution only from intrafeatures of documents, namely, text content, without considering directly extrafeatures of documents such as authors. Moreover, they are applicable only to the case of single processor. To resolve these problems, we propose a dynamic users' interest discovery model with distributed inference algorithm, named as Distributed Author-Topic over Time (D-AToT) model. The collapsed Gibbs sampling method following the main idea of MapReduce is also utilized for inferring model parameters. The proposed model can discover latent topics and users' interests, and mine their changing patterns over time. Extensive experimental results on NIPS (Neural Information Processing Systems) dataset show that our D-AToT model is feasible and efficient.

## 1. Introduction

With a dynamic users' interest discovery model, one can answer a range of important questions about the content of information uploaded or shared to social network service (SNS), such as which topics each user prefers, which users are similar to each other in terms of their interests, which users are likely to have written documents similar to an observed document, and who are influential users at different stages of topic evolution, and it also helps characterize users as pioneers, mainstream, or laggards in different subject areas.

Users' interests have shown their increasing importance for the development of personalized web services and user-centric applications [1, 2]. Hence, users' interest modeling has been attracting extensive attentions during the past few years, such as (a) Author-Topic (AT) model [3–5], (b)

Author-Recipient-Topic (ART) model [6–8], Role-Author-Recipient-Topic (RART) model [6–8], and Author-Persona-Topic (APT) model [9], (c) Author-Interest-Topic (AIT) model [10] and Latent-Interest-Topic (LIT) model [11], and (d) Author-Conference-Topic (ACT) model [12].

In fact, when people enjoy SNS with their smart devices including phones and tablets, each user's interest is usually not static. However, the above models are devoted to discovering static latent topics and user's interests. Moreover, they are applicable only to the case of single processor. Of course, one can perform some post hoc or pre hoc analysis [4, 13] to discover changing patterns over time, but this misses the opportunity for time to improve topic discovery [14], and it is very difficult to align corresponding topics [15]. Currently, attention for dynamic models is mainly focused on analyzing topic evolution only from text content, such as Dynamic

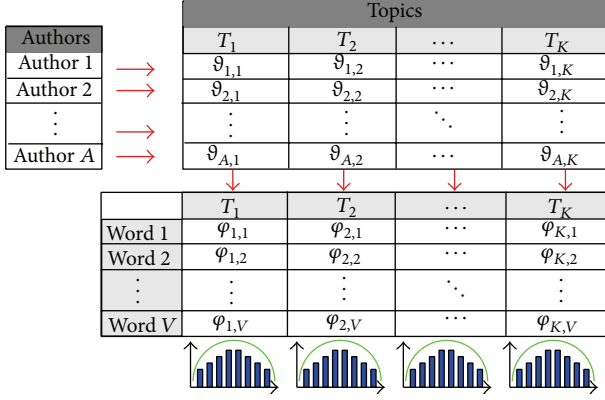


FIGURE 1: The illustration for discovering dynamic users' interests.

Topic Model (DTM) [16], continuous time DTM (cDTM) [17], and Topic over Time (ToT) [14].

This paper mainly focuses on the dynamic users' interest discovery model, especially collapsed Gibbs sampling following the main idea of MapReduce [18]. Figure 1 gives a detailed illustration for discovering dynamic users' interests. Our previous work [19, 20] is limited to inference algorithm on single-processor.

The organization of the rest of this work is as follows. In Section 2, we firstly discuss two related generative models, Author-Topic (AT) model and Topic over Time (ToT) model, and then introduce in detail our proposed Author-Topic over Time (AToT) model. Sections 3 and 4 describe the collapse Gibbs sampling methods used for inferring the model parameters and distributed inference algorithm version, respectively. In Section 5, extensive experimental evaluations are conducted, and Section 6 concludes this work.

## 2. Generative Models for Documents

Before presenting our Author-Topic over Time (AToT) model, we first describe two related generative models: AT model and ToT model. The notation is summarized in Table 1.

**2.1. Author-Topic (AT) Model.** Rosen-Zvi et al. [3–5] propose an Author-Topic (AT) model for extracting information about authors and topics from large text collections. Rosen-Zvi et al. model documents as if they were generated by a two-stage stochastic process. An author is represented by a probability distribution over topics, and each topic is represented as a probability distribution over words. The probability distribution over topics in a multiauthor paper is a mixture of the distributions associated with the authors.

The graphical model representations for AT model are shown in Figure 2. The AT model can be viewed as a generative process, which can be described as follows.

- (1) For each topic  $k \in [1, K]$ ,
  - (i) draw a multinomial  $\phi_k$  from  $\text{Dirichlet}(\beta)$ ;
- (2) for each author  $a \in [1, A]$ ,

TABLE 1: Notation used in the generative models.

Symbol	Description
$K$	Number of topics
$M$	Number of documents
$V$	Number of unique words
$A$	Number of unique authors
$N_m$	Number of word tokens in document $m$
$A_m$	Number of authors in document $m$
$a$	Single author index, $a \in [1, A]$
$k$	Single topic index, $k \in [1, K]$
$m$	Single document index, $m \in [1, M]$
$n$	Single word token index, $n \in [1, N_m]$
$v$	Single word index, $v \in [1, V]$
$\mathbf{a}_m$	Authors in document $m$ , $\mathbf{a}_m \subseteq [1, A]$
$\theta_a$	Multinomial distribution of topics specific to the author $a$ .
$\theta_m$	Multinomial distribution of topics specific to the document $m$ .
$\phi_k$	Multinomial distribution of words specific to the topic $k$ .
$\psi_k$	Beta distribution of timestamp specific to the topic $k$ .
$z_{m,n}$	Topic associated with the $n$ th token in the document $m$
$w_{m,n}$	$n$ th token in document $m$
$x_{m,n}$	Chosen author associated with the word token $w_{m,n}$
$t_{m,n}$	Timestamp associated with the $n$ th token in the document $m$
$\alpha$	Dirichlet priors (hyperparameter) to the multinomial distribution $\theta$
$\beta$	Dirichlet priors (hyperparameter) to the multinomial distribution $\phi$

- (i) draw a multinomial  $\theta_a$  from  $\text{Dirichlet}(\alpha)$ ;
- (3) for each word  $n \in [1, N_m]$  in document  $m \in [1, M]$ ,
  - (i) draw an author assignment  $x_{m,n}$  uniformly from the group of authors  $\mathbf{a}_m$ ;
  - (ii) draw a topic assignment  $z_{m,n}$  from  $\text{Multinomial}(\theta_{x_{m,n}})$ ;
  - (iii) draw a word  $w_{m,n}$  from  $\text{Multinomial}(\phi_{z_{m,n}})$ .

**2.2. Topic over Time (ToT) Model.** Unlike other dynamic topic models that rely on Markov assumptions or discretization of time, each topic in Topic over Time (ToT) model [14] is associated with a continuous distribution over timestamps, and, for each generated document, the mixture distribution over topics is influenced by both word cooccurrences and the document's timestamp. Thus, the meaning of a particular topic can be relied upon as constant, but the topics' occurrence and correlations change significantly over time.

The graphical model representations for ToT model are shown in Figure 3. The ToT is a generative model of timestamps and the words in the timestamped documents. The generative process can be described as follows.

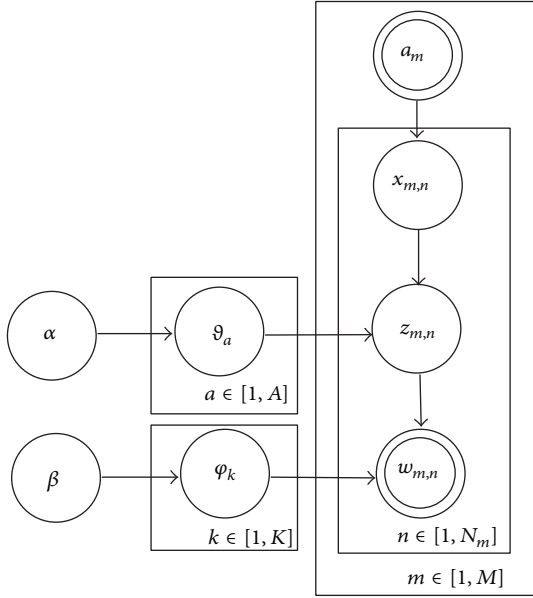


FIGURE 2: The graphical model representation of the AT model.

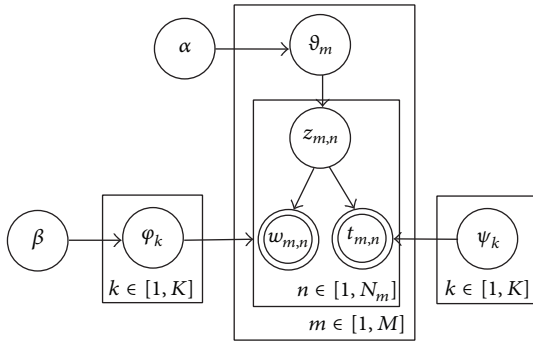


FIGURE 3: The graphical model representation of the ToT model.

- (1) For each topic  $k \in [1, K]$ ,
  - (i) draw a multinomial from  $\text{Dirichlet}(\beta)$ ;
- (2) for each document  $m \in [1, M]$ ,
  - (i) draw a multinomial  $\theta_m$  from  $\text{Dirichlet}(\alpha)$ ;
  - (ii) for each word  $n \in [1, N_m]$  in document  $m$ ,
    - (a) draw a topic assignment  $z_{m,n}$  from  $\text{Multinomial}(\theta_m)$ ;
    - (b) draw a word  $w_{m,n}$  from  $\text{Multinomial}(\phi_{z_{m,n}})$ ;
    - (c) draw a timestamp  $t_{m,n}$  from  $\text{Beta}(\psi_{z_{m,n}})$ .

**2.3. Author-Topic over Time (AToT) Model.** The graphical model representations for AToT model are shown in Figure 4. The AToT model can be viewed as a generative process, which can be described as follows.

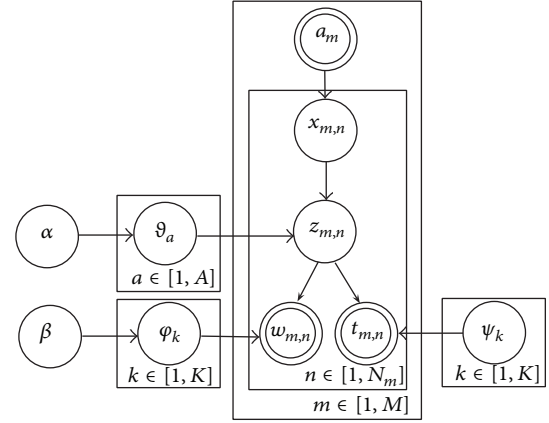


FIGURE 4: The graphical model representation of the AToT model.

- (1) For each topic  $k \in [1, K]$ ,
  - (i) draw a multinomial  $\phi_k$  from  $\text{Dirichlet}(\beta)$ ;
- (2) for each author  $a \in [1, A]$ ,
  - (i) draw a multinomial  $\theta_a$  from  $\text{Dirichlet}(\alpha)$ ;
- (3) for each word  $n \in [1, N_m]$  in document  $m \in [1, M]$ ,
  - (i) draw an author assignment  $x_{m,n}$  uniformly from the group of authors  $\mathbf{a}_m$ ;
  - (ii) draw a topic assignment  $z_{m,n}$  from  $\text{Multinomial}(\theta_{x_{m,n}})$ ;
  - (iii) draw a word  $w_{m,n}$  from  $\text{Multinomial}(\phi_{z_{m,n}})$ ;
  - (vi) draw a timestamp  $t_{m,n}$  from  $\text{Beta}(\psi_{z_{m,n}})$ .

From the above generative process, one can see that AToT model is parameterized as follows:

$$\begin{aligned}
 \theta_a &| \alpha \sim \text{Dirichlet}(\alpha) \\
 \phi_k &| \beta \sim \text{Dirichlet}(\beta) \\
 z_{m,n} &| \theta_{x_{m,n}} \sim \text{Multinomial}(\theta_{x_{m,n}}) \\
 w_{m,n} &| \phi_{z_{m,n}} \sim \text{Multinomial}(\phi_{z_{m,n}}) \\
 x_{m,n} &| A_m \sim \text{Multinomial}(1/A_m) \\
 t_{m,n} &| \psi_{z_{m,n}} \sim \text{Beta}(\psi_{z_{m,n}}).
 \end{aligned}$$

As a matter of fact, a paper is usually written by the first author and reprint author. If one wants to differentiate the contributions of the first author and reprint author from those of other coauthors, it is very easy for AToT model to set different weights for different authors. But since there are no criteria to guide the corresponding weights, we just set the equal weights for all coauthors in this work; that is to say,  $x_{m,n} | A_m$  follows the uniform distribution.

### 3. Inference Algorithm

For inference, the task is to estimate the sets of the following unknown parameters in the ATotT model: (1)  $\Phi = \{\varphi_k\}_{k=1}^K$ ,  $\Theta = \{\vartheta_a\}_{a=1}^A$ , and  $\Psi = \{\psi_k\}_{k=1}^K$  and (2) the corresponding topic and author assignments  $z_{m,n}, x_{m,n}$  for each word token  $w_{m,n}$ . In fact, inference cannot be done exactly in this model. A variety of algorithms have been used to estimate the parameters of topics models, such as variational EM (expectation maximization) [21, 22], expectation propagation [23, 24], belief propagation [25], and Gibbs sampling [19, 20, 26, 27]. In this work, collapsed Gibbs sampling algorithm [26] is used, since it provides a simple method for obtaining parameter estimates under Dirichlet priors and allows combination of estimates from several local maxima of the posterior distribution.

In the Gibbs sampling procedure, we need to calculate the conditional distribution  $P(z_{m,n}, x_{m,n} \mid \mathbf{w}, \mathbf{z}_{-(m,n)}, \mathbf{x}_{-(m,n)}, \mathbf{t}, \mathbf{a}, \alpha, \beta, \Psi)$ , where  $\mathbf{z}_{-(m,n)}, \mathbf{x}_{-(m,n)}$  represents the topic and author assignments for all tokens except  $w_{m,n}$ , respectively. We begin with the joint distribution  $P(\mathbf{w}, \mathbf{z}, \mathbf{x}, \mathbf{t} \mid \mathbf{a}, \alpha, \beta, \Psi)$  of a dataset, and, using the chain rule, we can get the conditional probability conveniently as

$$\begin{aligned} &P(z_{m,n}, x_{m,n} \mid \mathbf{w}, \mathbf{z}_{-(m,n)}, \mathbf{x}_{-(m,n)}, \mathbf{t}, \mathbf{a}, \alpha, \beta, \Psi) \\ &\propto \frac{n_{z_{m,n}}^{(w_{m,n})} + \beta_{w_{m,n}} - 1}{\sum_{v=1}^V (n_k^{(v)} + \beta_v) - 1} \times \frac{n_a^{(z_{m,n})} + \alpha_{z_{m,n}} - 1}{\sum_{k=1}^K (n_a^{(k)} + \alpha_k) - 1} \quad (1) \\ &\quad \times \text{Beta}(\psi_{z_{m,n}}), \end{aligned}$$

where  $n_k^{(v)}$  is the number of times tokens of word  $v$  are assigned to topic  $k$  and  $n_a^{(k)}$  represents the number of times author  $a$  is assigned to topic  $k$ . Detailed derivation of Gibbs sampling for ATotT is provided in the appendix.

If one further manipulates the above (1), one can turn it into separated update equations for the topic and author of each token, suitable for random or systematic scan updates:

$$P(x_{m,n} \mid \mathbf{x}_{-(m,n)}, \mathbf{z}, \mathbf{a}, \alpha) \propto \frac{n_{x_{m,n}}^{(z_{m,n})} + \alpha_{z_{m,n}} - 1}{\sum_{k=1}^K (n_{x_{m,n}}^{(k)} + \alpha_k) - 1}, \quad (2)$$

$$\begin{aligned} &P(z_{m,n} \mid \mathbf{w}, \mathbf{z}_{-(m,n)}, \mathbf{x}, \mathbf{t}, \alpha, \beta, \Psi) \\ &\propto \frac{n_{z_{m,n}}^{(w_{m,n})} + \beta_{w_{m,n}} - 1}{\sum_{v=1}^V (n_k^{(v)} + \beta_v) - 1} \times \frac{n_a^{(z_{m,n})} + \alpha_{z_{m,n}} - 1}{\sum_{k=1}^K (n_{x_{m,n}}^{(k)} + \alpha_k) - 1} \quad (3) \\ &\quad \times \text{Beta}(\psi_{z_{m,n}}). \end{aligned}$$

During parameter estimation, the algorithm keeps track of two large data structures: an  $A \times K$  count matrix  $n_a^{(k)}$  and a

$K \times V$  count matrix  $n_k^{(v)}$ . From these data structures, one can easily estimate the  $\Phi$  and  $\Theta$  as follows:

$$\varphi_{k,v} = \frac{n_k^{(v)} + \beta_v}{\sum_{v=1}^V (n_k^{(v)} + \beta_v)}, \quad (4)$$

$$\vartheta_{a,k} = \frac{n_a^{(k)} + \alpha_k}{\sum_{k=1}^K (n_a^{(k)} + \alpha_k)}. \quad (5)$$

As for  $\Psi$ , similar to [14], for simplicity and speed, we update it after each Gibbs sample by the method of moments [28]:

$$\begin{aligned} \psi_{k,1} &= \bar{t}_k \left( \frac{\bar{t}_k (1 - \bar{t}_k)}{s_k^2} - 1 \right), \\ \psi_{k,2} &= (1 - \bar{t}_k) \left( \frac{\bar{t}_k (1 - \bar{t}_k)}{s_k^2} - 1 \right), \end{aligned} \quad (6)$$

where  $\bar{t}_k$  and  $s_k^2$  indicate the sample mean and biased sample variance of the timestamps belonging to topic  $k$ , respectively. The readers are invited to consult [28] for details. In fact, similar to [14], since the Beta distribution with the support  $[0, 1]$  can behave many more shapes including the bell curve than Gaussian distribution, it is utilized to model the timestamps. But Wang and McCallum [14] did not provide much detail on how to handle documents with 0 and 1 timestamps so that they have some probability, so the time range of the data is normalized to  $[0.01, 0.99]$  in the paper.

With (2)–(6), Gibbs sampling algorithm for ATotT model is summarized in Algorithm 1. The procedure itself uses only seven larger data structures, the count variables  $n_a^{(k)}$  and  $n_k^{(v)}$ , which have dimension  $A \times K$  and  $K \times V$ , respectively, their row sums  $n_a$  and  $n_k$  with dimensions  $A$  and  $K$ , Beta parameters  $\Psi$  with dimension  $K \times 2$ , and the state variable  $z_{m,n}, x_{m,n}$  with dimension  $W = \sum_{m=1}^M N_m$ .

### 4. Distributed Inference Algorithm

Our distributed inference algorithm, named as D-ATotT, is inspired by AD-LDA algorithm [29, 30], following the main idea of the well-known distributed programming model, MapReduce [18]. The overall distributed architecture for ATotT model is shown in Figure 5.

As stated in Figure 5, the master firstly distributes  $M$  training documents over  $P$  mappers, with nearly equal number  $M/P$  of documents on each mapper. Specifically, D-ATotT partitions document  $\{\mathbf{w}\}$ ,  $\{\mathbf{a}\}$ , and  $\{\mathbf{t}\}$  into  $\{\{\mathbf{w}_{|p}\}\}_{p=1}^P$ ,  $\{\{\mathbf{a}_{|p}\}\}_{p=1}^P$ , and  $\{\{\mathbf{t}_{|p}\}\}_{p=1}^P$  and corresponding topic and author assignments  $\{\mathbf{z}\}$  and  $\{\mathbf{x}\}$  into  $\{\{\mathbf{z}_{|p}\}\}_{p=1}^P$  and  $\{\{\mathbf{x}_{|p}\}\}_{p=1}^P$ , where  $\{\mathbf{w}_{|p}\}$ ,  $\{\mathbf{a}_{|p}\}$ ,  $\{\mathbf{t}_{|p}\}$ ,  $\{\mathbf{z}_{|p}\}$ , and  $\{\mathbf{x}_{|p}\}$  exist only on mapper  $p$ . The Author-Topic count  $\{n_a^{(k)}\}$  and topic-word count  $\{n_k^{(v)}\}$  are likewise distributed, denoted as  $\{n_{a|p}^{(k)}\}$  and  $\{n_{k|p}^{(v)}\}$  on mapper  $p$ , which are used to temporarily store local Author-Topic and topic-word counts.

**Algorithm** ATOTGibbs( $\{\mathbf{w}\}, \{\mathbf{a}\}, \{\mathbf{t}\}, \alpha, \beta, \psi, K$ )

**Input:** word vectors  $\{\mathbf{w}\}$ , author vector  $\{\mathbf{a}\}$ , time vector  $\{\mathbf{t}\}$ , hyperparameters  $\alpha, \beta$ , Beta parameters  $\psi$ , topic number  $K$

**Global data:** count statistics  $\{n_a^{(k)}\}, \{n_k^{(v)}\}$  and their sums  $\{n_a\}, \{n_k\}$

**Output:** topic associations  $\{\mathbf{z}\}$ , author associations  $\{\mathbf{x}\}$ , multinomial parameters  $\Phi$  and  $\Theta$ , Beta parameter estimates  $\psi$ , hyperparameter estimates  $\alpha, \beta$

// initialization  
 zero all count variables,  $n_a^{(k)}, n_a, n_k^{(v)}, n_k$   
**for** all documents  $m \in [1, M]$  **do**  
   **for** all words  $n \in [1, N_m]$  in document  $m$  **do**  
 sample topic index  $z_{m,n} \sim \text{Multinomial}(1/K)$   
  
 sample author index  $x_{m,n} \sim \text{Multinomial}(\mathbf{p})$  with  $p_a = \begin{cases} \frac{1}{A_m}, & a \in \mathbf{a}_m \\ 0, & \text{otherwise} \end{cases}$   
  
 // increment counts and sums  
 $n_{x_{m,n}}^{(z_{m,n})} + 1; n_{x_{m,n}} + 1; n_{z_{m,n}}^{(w_{m,n})} + 1; n_{z_{m,n}} + 1$   
 // Gibbs sampling over burn-in period and sampling period  
**while** not finished **do**  
   **for** all documents  $m \in [1, M]$  **do**  
**for** all words  $n \in [1, N_m]$  in documents  $m$  **do**  
   // decrement counts and sums  
 $n_{x_{m,n}}^{(z_{m,n})} - 1; n_{x_{m,n}} - 1; n_{z_{m,n}}^{(w_{m,n})} - 1; n_{z_{m,n}} - 1$   
   sample author index  $\tilde{a}$  according to (2)  
   sample topic index  $\tilde{z}$  according to (3)  
   // increment counts and sums  
 $n_{\tilde{a}}^{(\tilde{z})} + 1; n_{\tilde{a}} + 1; n_{\tilde{z}}^{(w_{m,n})} + 1; n_{\tilde{z}} + 1$   
   update  $\psi$  according to (6)  
**if** converged and  $L$  sampling iterations since last read out **then**  
   // different parameters read outs are averaged  
   read out parameter set  $\Phi$  according to (4)  
   read out parameter set  $\Theta$  according to (5)

ALGORITHM 1: Gibbs sampling algorithm for ATOT model.

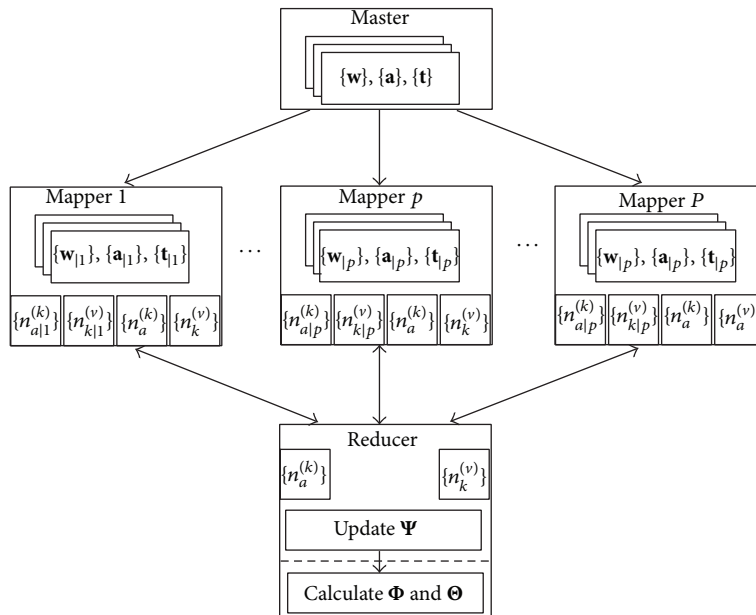


FIGURE 5: The overall distributed architecture for ATOT model.



TABLE 2: Distribution of number of papers over year in NIPS dataset.

Year	Number of papers
1987	90 (5.2%)
1988	95 (5.5%)
1989	101 (5.8%)
1990	143 (8.2%)
1991	144 (8.3%)
1992	127 (7.3%)
1993	144 (8.3%)
1994	140 (8.0%)
1995	152 (8.7%)
1996	152 (8.7%)
1997	151 (8.7%)
1998	151 (8.7%)
1999	150 (8.6%)

In each Gibbs sampling iteration, each mapper  $p$  updates  $\{z_{i|p}\}$  and  $\{x_{i|p}\}$  by sampling  $z_{m,n|p}$  and  $x_{m,n|p}$  from the following posterior distributions:

$$\begin{aligned}
& P(x_{m,n|p} | \mathbf{x}_{-(m,n)|p}, \mathbf{z}_{i|p}, \mathbf{a}_{i|p}, \boldsymbol{\alpha}) \\
& \propto \frac{n_{x_{m,n}}^{(z_{m,n})} + \alpha_{z_{m,n}} - 1}{\sum_{k=1}^K (n_{x_{m,n}}^{(k)} + \alpha_k) - 1}, \\
& P(z_{m,n|p} | \mathbf{w}_{i|p}, \mathbf{z}_{-(m,n)|p}, \mathbf{x}_{i|p}, \mathbf{t}_{i|p}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi) \\
& \propto \frac{n_{z_{m,n}}^{(w_{m,n})} + \beta_{w_{m,n}} - 1}{\sum_{v=1}^V (n_k^{(v)} + \beta_v) - 1} \times \frac{n_a^{(z_{m,n})} + \alpha_{z_{m,n}} - 1}{\sum_{k=1}^K (n_{x_{m,n}}^{(k)} + \alpha_k) - 1} \\
& \times \text{Beta}(\Psi_{z_{m,n}})
\end{aligned} \tag{7}$$

and updates local  $n_{a|p}^{(k)}$  and  $n_{k|p}^{(v)}$  according to the new topic and author assignments. After each iteration, each mapper sends the local counts to the reducer and then the reducer updates  $\Psi$  and broadcasts the global  $n_a^{(k)}$ ,  $n_k^{(v)}$ , and  $\Psi$  to all mappers. After all sampling iterations, the reducer calculates the  $\Phi$  and  $\Theta$  according to (4)-(5).

## 5. Experimental Results and Discussions

NIPS proceeding dataset is utilized to evaluate the performance of our model, which consists of the full text of the 13 years of proceedings from 1987 to 1999 Neural Information Processing Systems (NIPS) Conferences. The dataset contains 1,740 research papers and 2,037 unique authors. The distribution of the number of papers over year is shown in Table 2.

In addition to downcasing and removing stop words and numbers, we also remove the words appearing less than five times in the corpus. After the preprocessing, the dataset contains 13,649 unique words and 2,301,375 word tokens in total. Each document's timestamp is determined by the year of the proceedings. In our experiments,  $K$  is fixed at 100 and

the symmetric Dirichlet priors  $\alpha$  and  $\beta$  are set at 0.5 and 0.1, respectively. Gibbs sampling is run for 2000 iterations.

**5.1. Examples of Topic, Author Distributions, and Topic Evolution.** Table 3 illustrates examples of 8 topics learned by ATOT model. The topics are extracted from a single sample at the 2000th iteration of the Gibbs sampler. Each topic is illustrated with (1) the top 10 words most likely to be generated conditioned on the topic, (b) the top 10 authors which have the highest probability conditioned on the topic, and (c) histograms and fitted beta PDFs which show topics evolution patterns over time.

**5.2. Author Interest Evolution Analysis.** In order to analyze further author interest evolution, it is interesting to calculate

$$P(z, t | a) = P(z | a) p(z | t) = \vartheta_{a,z} \times \text{Beta}(\Psi_z). \tag{8}$$

In this subsection, we take Sejnowski.T as an example, who published 43 papers in total from 1987 to 1999 in the NIPS conferences, as shown in Figure 6(a). The research interest evolution for Sejnowski.T is reported in Figure 6(b), in which the area occupied by a square is proportional to the strength of his research interest.

From Figure 6(b), one can see that Sejnowski.T's research interest focused mainly on Topic 51 (Eye Recognition and Factor Analysis), Topic 37 (Neural Networks), and Topic 58 (Data Model and Learning Algorithm) but with different emphasis from 1987 to 1999. In the early phase (1989–1993), Sejnowski.T's research interest is only limited to Topic 51 and then extended to Topic 37 in 1994 and Topic 58 in 1996 with great research interest strength and finally back to Topic 51 after 1997. Anyway, Sejnowski.T did not change his main research direction, Topic 51, which is verified from his homepage again.

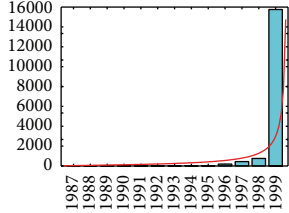
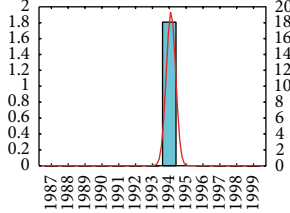
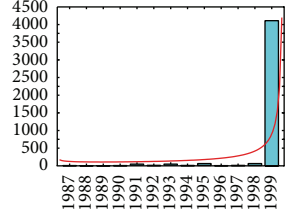
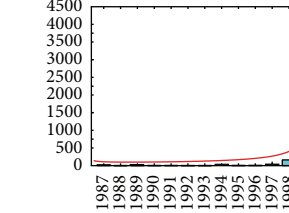
**5.3. Predictive Power Analysis.** Similar to [5], we further divide the NIPS papers into a training set  $\mathcal{D}^{\text{train}}$  of 1,557 papers and a test set  $\mathcal{D}^{\text{test}}$  of 183 papers of which 102 are single-authored papers. Each author in  $\mathcal{D}^{\text{test}}$  must have authored at least one of the training papers. The perplexity, originally used in language modeling [31], is a standard measure for estimating the performance of a probabilistic model. The perplexity of a test document  $\tilde{m} \in \mathcal{D}^{\text{test}}$  is defined as the exponential of the negative normalized predictive likelihood under the model:

$$\begin{aligned}
& \text{perplexity}(\mathbf{w}_{\tilde{m}}, \mathbf{t}_{\tilde{m}} | \mathbf{a}_{\tilde{m}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi) \\
& = \exp \left[ - \frac{\ln P(\mathbf{w}_{\tilde{m}}, \mathbf{t}_{\tilde{m}} | \mathbf{a}_{\tilde{m}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi)}{N_{\tilde{m}}} \right]
\end{aligned} \tag{9}$$

TABLE 3: An illustration of 8 topics from a 100-topic solution for the NIPS collection. The titles are our own interpretation of the topics. Each topic is shown with the 10 words and authors that have the highest probability conditioned on that topic. Histograms show how the topics are distributed over time; the fitted beta PDFs is shown also.

Topic 87		Topic 37		Topic 11		Topic 88	
SVM and Kernel methods		Neural networks		Reinforcement learning		EM and mixture models	
Word	Prop.	Word	Prop.	Word	Prop.	Word	Prop.
set	0.0188195	learning	0.01106740	state	0.0468466	density	0.0279477
support	0.0187117	network	0.00948016	learning	0.0252876	log	0.0217790
vector	0.0186039	neural	0.00780503	belief	0.0213999	distribution	0.0186946
kernel	0.0160163	input	0.00682192	policy	0.0182191	mixture	0.0178379
function	0.0146146	model	0.00681643	function	0.0175122	method	0.0144108
svm	0.0138060	training	0.00604202	action	0.0150383	gaussian	0.0142394
training	0.0129974	data	0.00597611	states	0.0148615	likelihood	0.0140681
problem	0.0124583	figure	0.00594316	reinforcement	0.0118574	entropy	0.0132113
space	0.0119731	networks	0.00560813	actions	0.0118574	gaussians	0.0123546
solution	0.0115957	function	0.00554222	mdp	0.0102670	form	0.0113264
Author	Prop.	Author	Prop.	Author	Prop.	Author	Prop.
Scholkopf_B	0.949692	Reggia_J	0.979832	Zhang_N	0.629412	Barron_A	0.608507
Crisp_D	0.888975	Todorov_E	0.976750	Rodriguez_A	0.578235	Wainwright_M	0.372871
Laskov_P	0.706170	Horne_B	0.974146	Dietterich_T	0.342954	Mukherjee_S	0.340927
Steinhage_V	0.634973	Thmn_S	0.973083	Sallans_B	0.228042	Li_J	0.337108
Chapelle_O	0.610385	Weigend_A	0.972806	Walker_M	0.189143	Jebara_T	0.253203
Li_Y	0.513418	McCallum_R	0.969777	Koller_D	0.1885150	Millman_K	0.171569
Herbrich_R	0.454384	Camana_R	0.969388	Yeung_D	0.1213730	Fisher_J	0.148230
Gordon_M	0.425090	Slaney_M	0.969382	Thrun_S	0.0842081	Ihler_A	0.128369
Vapnik_V	0.330421	Miikkulainen_R	0.968541	Konda_V	0.0680365	Beal_M	0.126578
Dom_B	0.286036	Bergen_J	0.968358	Parr_R	0.0468006	Hansen_L	0.0849109

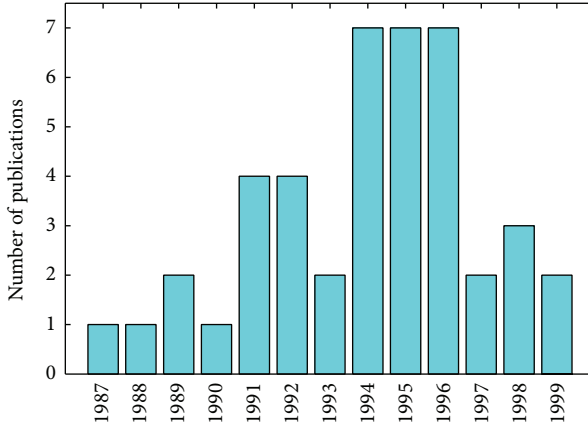





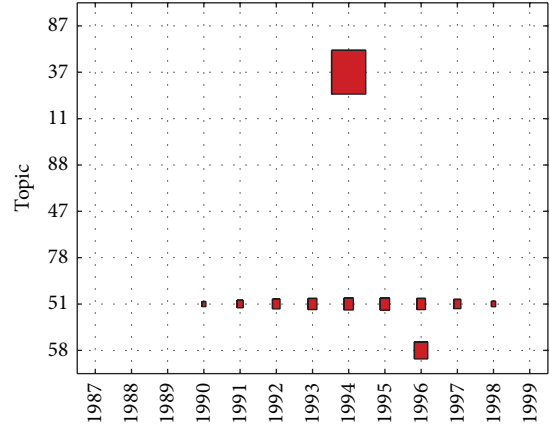
Topic 47		Topic 78		Topic 51		Topic 58	
Speech recognition		Bayesian learning		Eye recognition and factor analysis		Data model and learning algorithm	
Word	Prop.	Word	Prop.	Word	Prop.	Word	Prop.
hmm	0.0415364	bayesian	0.0243032	sejnowski	0.0265409	learning	0.00904655
speech	0.0392921	sampling	0.0184560	eye	0.0265409	model	0.00752741
hmms	0.0216579	prior	0.0178563	ica	0.0183324	neural	0.00705102
mixture	0.0179708	distribution	0.0148578	vor	0.0159531	data	0.00700339
suffix	0.0104362	monte	0.0127588	disparity	0.0153583	function	0.00683930
probabilistic	0.00995527	carlo	0.0118592	head	0.0135738	network	0.00624646
probabilities	0.00947434	model	0.0109597	position	0.0125031	input	0.00593946
singer	0.00883310	posterior	0.0105099	eeg	0.0119083	set	0.00561128
acoustic	0.00883310	priors	0.00946041	parietal	0.0109566	networks	0.00556365
saul	0.00867279	sample	0.00901063	salk	0.0105997	figure	0.00545249

TABLE 3: Continued.

Author	Prop.	Author	Prop.	Author	Prop.	Author	Prop.
Rigoll_G	0.460882	Schuermans_D	0.651505	Sejnowski_T	0.410459	Gray_M	0.974482
Singer_Y	0.437547	Sykacek_P	0.495506	Pouget_A	0.269781	Dimitrov_A	0.973538
Nix_D	0.192342	Andrieu_C	0.413324	Anastasio_T	0.112957	Galperin_G	0.97094
Saul_L	0.170699	Rasmussen_C	0.344185	Horiuchi_T	0.0328485	Malik_J	0.968536
Hermansky_H	0.0795602	Zlochin_M	0.244745	Albright_T	0.0099278	Davies_S	0.966534
Roweis_S	0.0391364	Beal_M	0.157807	Jousmaki_V	0.00791139	Cook_G	0.96519
Attias_H	0.0357538	Hansen_L	0.122773	Fredholm_H	0.00681818	Ghosn_J	0.964184
Movellan_J	0.033414	Herbrich_R	0.0882701	Bohr_J	0.00643777	Orponen_P	0.964184
Schuster_M	0.0293324	Downs_O	0.0694726	Ramanujam_N	0.00621891	Yen_S	0.963001
Muller_K	0.028258	Williams_C	0.0652069	Dixon_L	0.00585938	Chatterjee_C	0.962627



(a) Distribution of number of publications over time



(b) Research interest evolution

FIGURE 6: The distribution of number of publications and research interest evolution for Sejnowski\_T.

with

$$\begin{aligned}
 & P(\mathbf{w}_{\bar{m}}, \mathbf{t}_{\bar{m}} \mid \mathbf{a}_{\bar{m}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi) \\
 &= \frac{1}{[A_{\bar{m}}]^{N_m}} \times \sum_{\mathbf{z}_{\bar{m}}} \text{Beta}(\psi_{z_{\bar{m},1}}, \psi_{z_{\bar{m},2}} \mid \mathcal{D}^{\text{train}}) \\
 &\quad \times \int p(\Phi \mid \boldsymbol{\beta}, \mathcal{D}^{\text{train}}) \sum_{\mathbf{z}_{\bar{m}}} \varphi_{z_{\bar{m}}, w_{\bar{m},n}} d\Phi \\
 &\quad \times \int p(\Theta \mid \boldsymbol{\alpha}, \mathcal{D}^{\text{train}}) \sum_{\mathbf{x}_{\bar{m}}} \vartheta_{x_{\bar{m}}, z_{\bar{m},n}} d\Theta.
 \end{aligned} \tag{10}$$

We approximate the integrals over  $\Phi$  and  $\Theta$  using the point estimates obtained in (4)-(5) for each sample  $s \in \{1, 2, \dots, 10\}$  of assignments  $\mathbf{x}$ ,  $\mathbf{z}$  and then average over samples. Figure 7 shows the results for the ATot model and

AT model in a post hoc fashion on 102 single-authored papers. It is not difficult to see that the perplexity of ATot model is smaller than that of AT model when the number of topics  $> 10$ , which indicates that ATot model outperforms AT model.

## 6. Conclusions

With a dynamic users' interest discovery model, one can answer many important questions about the content of information uploaded or shared to SNS. Based on our previous work, Author-Topic over Time (ATot) model [19], for documents using authors and topics with timestamps, this paper proposes a dynamic users' interest discovery model with distributed inference algorithm following the main idea of MapReduce, named as Distributed ATot (D-ATot) model.



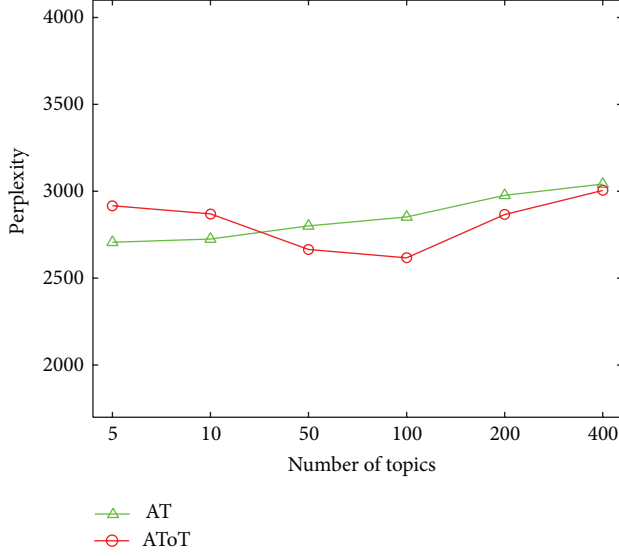


FIGURE 7: Perplexity of the 102 single-authored test documents.

The D-AToT model combines the merits of AT and ToT models. Specifically, it can automatically detect latent topics, users' interests, and their changing patterns from large-scale social network information. The results on NIPS dataset show the increase of salient topics and more reasonable users' interest changing patterns.

One can generalize the approach in the work to construct alternative dynamic models from other static users' interest discovery models and ToT model with distributed inference algorithm. As a matter of fact, our work currently is limited to deal with the users and latent topics with timestamps in SNS. Though NIPS proceeding dataset is a benchmark data for academic social network, the D-AToT model ignores the links in SNS. In ongoing work, novel topic model, considering the links in SNS, will be constructed to identify the users with similar interests from social networks.

## Appendix

### Gibbs Sampling Derivation for AToT

We begin with the joint distribution  $P(\mathbf{w}, \mathbf{z}, \mathbf{x}, \mathbf{t} \mid \mathbf{a}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi)$ . We can take advantage of conjugate priors to simplify the integrals. Consider

$$\begin{aligned}
 P(\mathbf{w}, \mathbf{z}, \mathbf{x}, \mathbf{t} \mid \mathbf{a}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi) &= P(\mathbf{w} \mid \mathbf{z}, \boldsymbol{\beta}) p(\mathbf{t} \mid \Psi, \mathbf{z}) P(\mathbf{z} \mid \mathbf{x}, \boldsymbol{\alpha}) P(\mathbf{x} \mid \mathbf{a}) \\
 &= \int P(\mathbf{w} \mid \Phi, \mathbf{z}) p(\Phi \mid \boldsymbol{\beta}) d\Phi \times p(\mathbf{t} \mid \Psi, \mathbf{z}) \\
 &\quad \times \int P(\mathbf{z} \mid \mathbf{x}, \Theta) p(\Theta \mid \boldsymbol{\alpha}) d\Theta \times P(\mathbf{x} \mid \mathbf{a})
 \end{aligned}$$

$$\begin{aligned}
 &= \int \prod_{m=1}^M \prod_{n=1}^{N_m} P(w_{m,n} \mid \varphi_{z_{m,n}}) \prod_{k=1}^K P(\varphi_k \mid \boldsymbol{\beta}) d\Phi \\
 &\quad \times \int \prod_{m=1}^M \prod_{n=1}^{N_m} P(z_{m,n} \mid \boldsymbol{\vartheta}_{x_{m,n}}) \prod_{a=1}^A P(\boldsymbol{\vartheta}_a \mid \boldsymbol{\alpha}) d\Theta \\
 &\quad \times \prod_{m=1}^M \prod_{n=1}^{N_m} P(t_{m,n} \mid \boldsymbol{\psi}_{z_{m,n}}) \times \prod_{m=1}^M \prod_{n=1}^{N_m} P(x_{m,n} \mid \mathbf{a}_m) \\
 &= \frac{1}{\prod_{m=1}^M A_m^{N_m}} \\
 &\quad \times \int \prod_{k=1}^K \prod_{v=1}^V \varphi_{k,v}^{n_k^{(v)}} \prod_{k=1}^K \left( \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \prod_{v=1}^V \varphi_{k,v}^{\beta_v-1} \right) d\Phi \\
 &\quad \times \int \prod_{a=1}^A \prod_{k=1}^K \boldsymbol{\vartheta}_{a,k}^{n_a^{(k)}} \prod_{a=1}^A \left( \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \boldsymbol{\vartheta}_{a,k}^{\alpha_k-1} \right) d\Theta \\
 &\quad \times \prod_{m=1}^M \prod_{n=1}^{N_m} P(t_{m,n} \mid \boldsymbol{\psi}_{z_{m,n}}) \\
 &= \frac{1}{\prod_{m=1}^M A_m^{N_m}} \left( \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \right)^K \left( \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \right)^A \\
 &\quad \times \prod_{m=1}^M \prod_{n=1}^{N_m} P(t_{m,n} \mid \boldsymbol{\psi}_{z_{m,n}}) \times \prod_{k=1}^K \frac{\prod_{v=1}^V \Gamma(n_k^{(v)} + \beta_v)}{\Gamma(\sum_{v=1}^V (n_k^{(v)} + \beta_v))} \\
 &\quad \times \prod_{a=1}^A \frac{\prod_{k=1}^K \Gamma(n_a^{(k)} + \alpha_k)}{\Gamma(\sum_{k=1}^K (n_a^{(k)} + \alpha_k))}. \tag{A.1}
 \end{aligned}$$

Using the chain rule, we can obtain the conditional probability conveniently as follows:

$$\begin{aligned}
 &P(z_{m,n}, x_{m,n} \mid \mathbf{w}, \mathbf{z}_{-(m,n)}, \mathbf{x}_{-(m,n)}, \mathbf{t}, \mathbf{a}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi) \\
 &= (P(z_{m,n}, x_{m,n}, w_{m,n}, t_{m,n} \mid \mathbf{w}_{-(m,n)}, \\
 &\quad \mathbf{t}_{-(m,n)}, \mathbf{z}_{-(m,n)}, \mathbf{x}_{-(m,n)}, \mathbf{a}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi)) \\
 &\quad \times (P(w_{m,n}, t_{m,n} \mid \mathbf{w}_{-(m,n)}, \mathbf{t}_{-(m,n)}, \\
 &\quad \mathbf{x}_{-(m,n)}, \mathbf{z}_{-(m,n)}, \mathbf{a}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi))^{-1} \\
 &\propto \frac{P(\mathbf{w}, \mathbf{t}, \mathbf{z}, \mathbf{x} \mid \mathbf{a}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi)}{P(\mathbf{w}_{-(m,n)}, \mathbf{t}_{-(m,n)}, \mathbf{z}_{-(m,n)}, \mathbf{x}_{-(m,n)} \mid \mathbf{a}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \Psi)} \\
 &\propto \frac{n_{z_{m,n}}^{(w_{m,n})} + \beta_{w_{m,n}} - 1}{\sum_{v=1}^V (n_{z_{m,n}}^{(v)} + \beta_v) - 1} \times \frac{n_{x_{m,n}}^{(z_{m,n})} + \alpha_{z_{m,n}} - 1}{\sum_{k=1}^K (n_{x_{m,n}}^{(k)} + \alpha_k) - 1} \\
 &\quad \times \frac{1}{A_m} \times P(t_{m,n} \mid \boldsymbol{\psi}_{z_{m,n}})
 \end{aligned}$$

$$\propto \frac{n_{z_{m,n}}^{(w_{m,n})} + \beta_{w_{m,n}} - 1}{\sum_{v=1}^V (n_{z_{m,n}}^{(v)} + \beta_v) - 1} \times \frac{n_{x_{m,n}}^{(z_{m,n})} + \alpha_{z_{m,n}} - 1}{\sum_{k=1}^K (n_{x_{m,n}}^{(k)} + \alpha_k) - 1} \times \text{Beta}(\psi_{z_{m,n}}).$$

(A.2)

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This work was funded partially by the Key Technologies R&D Program of Chinese 12th Five-Year Plan (2011–2015), Key Technologies Research on Large-Scale Semantic Calculation for Foreign STKOS, and Key Technologies Research on Data Mining from the Multiple Electric Vehicle Information Sources under Grant nos. 2011BAH10B04 and 2013BAG06B01, respectively.

## References

- [1] F. Qiu and J. Cho, "Automatic identification of user interest for personalized search," in *Proceedings of the 15th International Conference on World Wide Web (WWW '06)*, pp. 727–736, ACM, Edinburgh, UK, May 2006.
- [2] J. Kim, D.-H. Jeong, D. Lee, and H. Jung, "User-centered innovative technology analysis and prediction application in mobile environment," *Multimedia Tools and Applications*, 2013.
- [3] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," in *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI '04)*, pp. 487–494, AUAI Press, Arlington, Va, USA, 2004.
- [4] M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths, "Probabilistic author-topic models for information discovery," in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04)*, pp. 306–315, ACM, Seattle, Wash, USA, August 2004.
- [5] M. Rosen-Zvi, C. Chedudugunta, T. Griffiths, P. Smyth, and M. Steyvers, "Learning author-topic models from text corpora," *ACM Transactions on Information Systems*, vol. 28, no. 1, article 4, pp. 1–38, 2010.
- [6] A. McCallum, A. Corrada-Emmanuel, and X. Wang, "The author-recipient-topic model for topic and role discovery in social networks: experiments with enron and academic email," Tech. Rep. um-cs-2004-096, Department of Computer Science, University of Massachusetts Amherst, 2004.
- [7] A. McCallum, A. Corrada-Emmanuel, and X. Wang, "Topic and role discovery in social networks," in *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pp. 786–791, Morgan Kaufmann, San Francisco, Calif, USA, 2005.
- [8] A. McCallum, X. Wang, and A. Corrada-Emmanuel, "Topic and role discovery in social networks with experiments on enron and academic email," *Journal of Artificial Intelligence Research*, vol. 30, no. 1, pp. 249–272, 2007.
- [9] D. Mimno and A. McCallum, "Expertise modeling for matching papers with reviewers," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '07)*, pp. 500–509, San Jose, Calif, USA, August 2007.
- [10] N. Kawamae, "Author interest topic model," in *Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '10)*, pp. 887–888, ACM, Geneva, Switzerland, July 2010.
- [11] N. Kawamae, "Latent interest-topic model: finding the causal relationships behind dyadic data," in *Proceedings of the 19th International Conference on Information and Knowledge Management and Co-located Workshops (CIKM '10)*, pp. 649–658, ACM, Toronto, Canada, October 2010.
- [12] J. Tang, J. Zhang, R. Jin et al., "Topic level expertise search over heterogeneous networks," *Machine Learning*, vol. 82, no. 2, pp. 211–237, 2011.
- [13] X. Wang, N. Mohanty, and A. McCallum, "Group and topic discovery from relations and their attributes," in *Advances in Neural Information Processing Systems 18*, Y. Weiss, B. Schölkopf, and J. Platt, Eds., pp. 1449–1456, MIT Press, Cambridge, Mass, USA, 2006.
- [14] X. Wang and A. McCallum, "Topics over time: a non-markov continuous-time model of topical trends," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*, pp. 424–433, August 2006.
- [15] S. Xu, L. Zhu, Q. Xiaodong, S. Qingwei, and G. Jie, "Topic linkages between papers and patents," in *Proceedings of the 4th International Conference on Advanced Science and Technology*, pp. 176–183, Science & Engineering Research Support soCiety, Daejeon, Republic of Korea, 2012.
- [16] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 113–120, ACM, June 2006.
- [17] C. Wang, D. Blei, and D. Heckerman, "Continuous time dynamic topic models," in *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI '08)*, pp. 579–586, July 2008.
- [18] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008.
- [19] S. Xu, Q. Shi, X. Qiao et al., "Author-topic over time (AToT): a dynamic users' interest model," in *Mobile, Ubiquitous, and Intelligent Computing: The 2nd International Conference on Ubiquitous Context-Awareness and Wireless Sensor Network*, vol. 274, pp. 227–233, Springer, Berlin, Germany, 2014.
- [20] Q. Shi, X. Qiao, S. Xu, and G. Nong, "Author-topic evolution model and its application in analysis of research interests evolution," *Journal of the China Society for Scientific and Technical Information*, vol. 32, no. 9, pp. 912–919, 2013.
- [21] J. M. Winn, *Variational message passing and its applications [Ph.D. thesis]*, University of Cambridge, 2004.
- [22] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, no. 4–5, pp. 993–1022, 2003.
- [23] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pp. 362–369, Morgan Kaufmann, San Francisco, Calif, USA, 2001.
- [24] T. Minka and J. Lafferty, "Expectation-propagation for the generative aspect model," in *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pp. 352–359, 2002.

- [25] J. Zeng, "A topic modeling toolbox using belief propagation," *Journal of Machine Learning Research*, vol. 13, pp. 2233–2236, 2012.
- [26] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, supplement 1, pp. 5228–5235, 2004.
- [27] G. Heinrich, "Parameter estimation for text analysis," Tech. Rep. version 2.9, vsonix GmbH and University of Leipzig, 2009.
- [28] C. B. Owen, *Parameter estimation for the Beta distribution [M.S. thesis]*, Brigham Young University, 2008.
- [29] D. Newman, A. Asuncion, P. Smyth, and M. Welling, "Distributed inference for latent Dirichlet allocation," in *Advances in Neural Information Processing Systems 20*, J. C. Platt, D. Koller, Y. Singer, and S. Roweis, Eds., pp. 1081–1088, MIT Press, Cambridge, Mass, USA, 2008.
- [30] D. Newman, A. Asuncion, P. Smyth, and M. Welling, "Distributed algorithms for topic models," *Journal of Machine Learning Research*, vol. 10, pp. 1801–1828, 2009.
- [31] L. Azzonpardi, M. Girolami, and K. van Risjbergen, "Investigating the relationship between language model perplexity and IR precision-recall measures," in *Proceedings of the 26th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '03)*, pp. 369–370, ACM, Toronto, Canada, August 2003.

