A Dynamic Users’ Interest Discovery Model with Distributed Inference Algorithm

Shuo Xu, Qingwei Shi, Xiaodong Qiao, Lijun Zhu, Han Zhang, Hanmin Jung, Seungwoo Lee, and Sung-Pil Choi

1 Information Technology Support Center, Institute of Scientific and Technical Information of China, No. 15 Fuxing Road, Haidian District, Beijing 100038, China
2 School of Software, Liaoning Technical University, No. 188 Longwan Street South, Huludao, Liaoning 125105, China
3 College of Software, Northeast Normal University, 5268 Renmin Street, Changchun, Jilin 130024, China
4 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: Goreti 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...
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 multi-author 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 $\varphi_k$ from Dirichlet($\beta$);
   - (2) for each author $a \in [1, A],$
     - (i) draw a multinomial $\varphi_{a,k}$ from Dirichlet($\alpha$);

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.

(i) draw a multinomial $\varphi_n$ from 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 $a_n$;
   - (ii) draw a topic assignment $z_{m,n}$ from Multinomial($\theta_{x_{m,n}}$);
   - (iii) draw a word $w_{m,n}$ from Multinomial($\varphi_{z_{m,n}}$).
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.

(1) For each topic \( k \in [1, K] \),
   (i) draw a multinomial \( \varphi_k \) from Dirichlet(\( \beta \));

(2) for each document \( m \in [1, M] \),
   (i) draw a multinomial \( \vartheta_m \) from Dirichlet(\( \alpha \));
   (ii) for each word \( n \in [1, N_m] \) in document \( m \),
      (a) draw a topic assignment \( z_{m,n} \) from Multinomial(\( \vartheta_{x_{m,n}} \));
      (b) draw a word \( w_{m,n} \) from Multinomial(\( \varphi_{z_{m,n}} \));
      (c) draw a timestamp \( t_{m,n} \) from Beta(\( \psi_{z_{m,n}} \)).

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

\[
\begin{align*}
\vartheta_a | \alpha & \sim \text{Dirichlet}(\alpha) \\
\varphi_k | \beta & \sim \text{Dirichlet}(\beta) \\
z_{m,n} | \vartheta_{x_{m,n}} & \sim \text{Multinomial}(\vartheta_{x_{m,n}}) \\
w_{m,n} | \varphi_{z_{m,n}} & \sim \text{Multinomial}(\varphi_{z_{m,n}}) \\
x_{m,n} | A_m & \sim \text{Multinomial}(1/A_m) \\
t_{m,n} | \psi_{z_{m,n}} & \sim \text{Beta}(\psi_{z_{m,n}}).
\end{align*}
\]

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 AToT model: (1) \( \Phi = \{\varphi_{k,v}\}_{k=1}^{K} \), \( \Theta = \{\theta_{k}\}_{k=1}^{K} \), and \( \Psi = \{\psi_{a,k}\}_{a=1}^{A} \) and (2) the corresponding topic and author assignments \( z_{mn} \), \( x_{mn} \) for each word token \( w_{mn} \). 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 of estimates from several local maxima of the posterior distribution.

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

\[
\psi_{k,1} = \frac{t_{k}}{s_{k}} \left( \frac{1 - t_{k}}{s_{k}} - 1 \right),
\]

\[
\psi_{k,2} = (1 - \bar{t}_{k}) \left( \frac{t_{k} (1 - \bar{t}_{k})}{s_{k}^{2}} - 1 \right),
\]

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 AToT model is summarized in Algorithm 1. The procedure itself uses only seven larger data structures, the count variables \( n_{a}^{(k)} \) and \( n_{v}^{(k)} \), which have dimension \( A \times K \) and \( K \times V \), respectively, their row sums \( n_{a} \) and \( n_{v} \) with dimensions \( A \) and \( K \), Beta parameters \( \Psi \) with dimension \( K \times 2 \), and the state variable \( z_{mn}, x_{mn} \) with dimension \( W = \sum_{m=1}^{M} N_{m} \).

4. Distributed Inference Algorithm

Our distributed inference algorithm, named as D-AToT, 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 AToT 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-AToT partitions document \( [w], [a], \) and \( [t] \) into \( \{|[w]_{p}\}_{p=1}^{P} \), \( \{|[a]_{p}\}_{p=1}^{P} \), and \( \{|[t]_{p}\}_{p=1}^{P} \) and corresponding topic and author assignments \( [z] \) and \( [x] \) into \( \{|[z]_{p}\}_{p=1}^{P} \) and \( \{|[x]_{p}\}_{p=1}^{P} \) where \( [w]_{p}, [a]_{p}, [t]_{p}, [z]_{p}, \) and \( [x]_{p} \) exist only on mapper \( p \). The Author-Topic count \( n_{a}^{(v)} \) and topic-word count \( n_{v}^{(a)} \) are likewise distributed, denoted as \( n_{a}^{(v)} \) and \( n_{v}^{(a)} \) on mapper \( p \), which are used to temporarily store local Author-Topic and topic-word counts.
International Journal of Distributed Sensor Networks

Algorithm AToT Gibbs((w), (a), (t), α, β, ψ, K)
Input: word vectors (w), author vector (a), time vector (t), hyperparameters α, β. Beta parameters ψ, topic number K
Global data: count statistics \( n^{(k)}_a, n^{(v)}_a \) and their sums \( n_a, n_v \)
Output: topic associations \( z \), author associations \( x \), multinomial parameters \( \Phi \) and \( \Theta \), Beta parameter estimates \( \psi \), hyperparameter estimates \( \alpha, \beta \)

// initialization
zero all count variables, \( n^{(k)}_a, n^{(v)}_a, n_a, n_v \)
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 \( x_{m,n} \sim \text{Multinomial}(p) \) with \( p_a \)
    // increment counts and sums
    \( n^{(z)}_{m,n} + 1; n^{(x)}_{m,n} + 1; n^{(w)}_{m,n} + 1; n^{(k)}_{m,n} + 1 \)
  // 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
        \( n^{(z)}_{m,n} - 1; n^{(x)}_{m,n} - 1; n^{(w)}_{m,n} - 1 \)
        sample author index \( \hat{a} \) according to (2)
        sample topic index \( \hat{z} \) according to (3)
        // increment counts and sums
        \( n^{(a)}_{\hat{a}} + 1; n^{(v)}_{\hat{a}} + 1; n^{(w)}_{\hat{z}} + 1; n^{(k)}_{\hat{z}} + 1 \)
        update \( \psi \) according to (6)
    if converged and \( L \) sampling iterations since last readout 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.
In each Gibbs sampling iteration, each mapper \( p \) updates \( \{z_{m,n} \} \) and \( \{x_{m,n} \} \) by sampling \( z_{m,n|p} \) and \( x_{m,n|p} \) from the following posterior distributions:

\[
P \left( x_{m,n|p} | x_{-m,n|p}, z_{m,n|p}, a_{m,n}, \alpha \right) \propto \frac{n(z_{m,n}) + \alpha_{z_{m,n}} - 1}{\sum_{k=1}^{K} \left( n(k)_{z_{m,n}} + \alpha_{k} \right) - 1},
\]

\[
P \left( z_{m,n|p} | w_{m,i}, z_{-m,n|p}, x_{m,n|p}, t_{m,n}, \alpha, \beta, \Psi \right) \propto \frac{n(w_{m,n}) + \beta_{w_{m,n}} - 1}{\sum_{i=1}^{V} \left( n(v)_{w_{m,n}} + \beta_{v} \right) - 1} \times \frac{n(a)_{z_{m,n}} + \alpha_{z_{m,n}} - 1}{\sum_{k=1}^{K} \left( n(k)_{z_{m,n}} + \alpha_{k} \right) - 1} \times \text{Beta} \left( \Psi_{z_{m,n}} \right)
\]

and updates local \( n_{w_{m,n}}^{(k)} \) and \( n_{a_{m,n}}^{(v)} \) 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 \( n_{w_{m,n}}^{(k)} \), \( n_{a_{m,n}}^{(v)} \), 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

\[
P(z,t | a) = P(z | a) p(z | t) = \Theta_{a,z} \times \text{Beta} \left( \Psi_{z} \right).
\]

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:

\[
\text{perplexity} \left( w_{m,i}, t_{m,i} | a_{m,i}, \alpha, \beta, \Psi \right) = \exp \left[ -\frac{\ln P \left( w_{m,i}, t_{m,i} | a_{m,i}, \alpha, \beta, \Psi \right)}{N_{m}} \right]
\]

\[
\sum_{w_{m,i}, t_{m,i}} \frac{N_{m}}{P \left( w_{m,i}, t_{m,i} | a_{m,i}, \alpha, \beta, \Psi \right)}
\]

\[
\sum_{w_{m,i}, t_{m,i}} \frac{N_{m}}{P \left( w_{m,i}, t_{m,i} | a_{m,i}, \alpha, \beta, \Psi \right)}
\]
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 | 0.01016746 | state | 0.0468466 | density | 0.0279477 |
| support | 0.0187117 | network | 0.00948016 | learning | 0.0025826 | 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 | 0.0146416 | 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 | 0.0115957 | 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 | 0.974164 | Dietterich_T | 0.342954 | Mukherjee_S | 0.349072 |
| 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 | Jebraa_T | 0.253203 |
| Li_Y | 0.531418 | McCallum_R | 0.969777 | Koller_D | 0.188510 | Millman_K | 0.171569 |
| Herbrich_R | 0.454384 | Camana_R | 0.969388 | Yeung_D | 0.121373 | 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 |
| hmm_m | 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 | 0.0062464 |
| probabilities | 0.00974734 | model | 0.0109597 | position | 0.0125031 | input | 0.00593946 |
| singer | 0.00883310 | posterior | 0.0105099 | eeg | 0.019083 | set | 0.00561128 |
| acoustic | 0.00883310 | priors | 0.00946041 | parietal | 0.0109566 | networks | 0.00565365 |
| saul | 0.00867279 | sample | 0.00901063 | salk | 0.0105997 | figure | 0.00545249 |

Histograms show how the topics are distributed over time; the fitted beta PDFs is shown also.
Table 3: Continued.

| Author  | Prop.  | Author  | Prop.  | Author  | Prop.  | Author  | Prop.  |
|---------|--------|---------|--------|---------|--------|---------|--------|
| Rigoll  | 0.460882 | Schuurmans  | 0.651505 | Sejnowski  | 0.410459 | Gray  | 0.974482 |
| Singer  | 0.437547 | Sykacek  | 0.495506 | Pouget  | 0.269781 | Dimitrov  | 0.973538 |
| Nix  | 0.192342 | Andrieu  | 0.413324 | Anastasio  | 0.112957 | Davies  | 0.966534 |
| Saul  | 0.0795602 | Rasmussen  | 0.344185 | Horiiuchi  | 0.0328485 | Malik  | 0.968536 |
| Hermansky  | 0.0795602 | Zlochin  | 0.244745 | Galperin  | 0.97094 | |
| Roweis  | 0.0391364 | Beal  | 0.157807 | Jousmaki  | 0.00791139 | Cook  | 0.96519 |
| Attias  | 0.0357538 | Hansen  | 0.122773 | Fredholm  | 0.0068185 | Ghosn  | 0.964184 |
| Movellan  | 0.033414 | Herbrich  | 0.0882701 | Bohr  | 0.00643777 | Orponen  | 0.964184 |
| Schuster  | 0.0293324 | Downs  | 0.0694726 | Ramanujam  | 0.00621891 | Yen  | 0.963001 |
| Muller  | 0.028258 | Williams  | 0.0652069 | Dixon  | 0.00585938 | Chatterjee  | 0.962627 |

Figure 6: The distribution of number of publications and research interest evolution for Sejnowski T.

with

\[
P(w_m, t_m | a_m, \alpha, \beta, \Psi) = \frac{1}{A_m^{N_m}} \sum_{z_{m, \cdot}} \text{Beta}(\psi_{z_{m,1}}, \psi_{z_{m,2}} | \mathcal{D}_{\text{train}}) \times \int p(\phi | \beta, \mathcal{D}_{\text{train}}) \sum_{x_{m, \cdot}} \phi_{x_{m, \cdot}, w_{m, \cdot}} d\phi \times \int p(\Theta | \alpha, \mathcal{D}_{\text{train}}) \sum_{z_{m, \cdot}} \Theta_{z_{m, \cdot}, \cdot} z_{m, \cdot} d\Theta.
\]

(10)

We approximate the integrals over \( \phi \) and \( \Theta \) using the point estimates obtained in (4)-(5) for each sample \( s \in \{1, 2, \ldots, 10\} \) of assignments \( x, z \) 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.
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 \( P(w, z, x, t \mid a, \alpha, \beta, \Psi) \). We can take advantage of conjugate priors to simplify the integrals. Consider

\[
P(w, z, x, t \mid a, \alpha, \beta, \Psi) = P(w \mid z, \beta) P(t \mid \Psi, z) P(z \mid x, \alpha) P(x \mid a)
\]

\[
= \int P(w \mid \Phi, z) P(\Phi \mid \beta) d\Phi \times P(t \mid \Psi, z)
\]

\[
\times \int P(z \mid x, \Theta) P(\Theta \mid \alpha) d\Theta \times P(x \mid a)
\]

\[
= \frac{1}{M_n} \sum_{m=1}^{M_n} \prod_{m=1}^{M_n} \prod_{n=1}^{N_m} \prod_{k=1}^{K} \int P(w_{m,n} \mid \varphi_{z_{m,n}}) P(\varphi_{z_{m,n}} \mid \beta) d\Phi
\]

\[
\times \prod_{m=1}^{M_n} \prod_{n=1}^{N_m} \prod_{a=1}^{A} P(t_{m,n} \mid \psi_{z_{m,n}}) P(\psi_{z_{m,n}} \mid a_m)
\]

\[
\times P(x_{m,n} \mid a_m) \prod_{m=1}^{M_n} \prod_{n=1}^{N_m} \prod_{k=1}^{K} \Gamma(z_{m,n})^{-1}
\]

\[
\times \prod_{m=1}^{M_n} \prod_{a=1}^{A} \prod_{k=1}^{K} \Gamma(z_{m,n})^{-1}
\]

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

\[
P(z_{m,n}, x_{m,n} \mid w, z-(m,n), x-(m,n), t, a, \alpha, \beta, \Psi)
\]

\[
= (P(z_{m,n}, x_{m,n} \mid w_{m,n}, t_{m,n} \mid w-(m,n)),
\]

\[
\times P(t_{-(m,n)}, z_{-(m,n)}, x_{-(m,n)} \mid a, \alpha, \beta, \Psi)
\]

\[
\times P(w_{m,n}, t_{m,n} \mid w-(m,n), t-(m,n), x_{-(m,n)}, z_{-(m,n)} \mid a, \alpha, \beta, \Psi)^{-1}
\]

\[
\times \int P(w \mid z, \beta) P(t \mid \Psi, z) P(z \mid x, \alpha) P(x \mid a)
\]

\[
\times \int P(w \mid \Phi, z) P(\Phi \mid \beta) d\Phi \times P(t \mid \Psi, z)
\]

\[
\times \int P(z \mid x, \Theta) P(\Theta \mid \alpha) d\Theta \times P(x \mid a)
\]
\[
\alpha \sum_{v=1}^{V} \left( n_{x,mn}^{(v)} + \beta_v \right) - 1 \times \beta \sum_{k=1}^{K} \left( n_{x,mn}^{(k)} + \alpha_k \right) - 1 \times \text{Beta} \left( \psi_{m,n} \right).
\]  

(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. Chemudugunta, 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-ccs-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 Rijsbergen, “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.