A Probabilistic Framework for Temporal User Modeling on Microblogs

Jitao Sang[1,3], Dongyuan Lu[2], Changsheng Xu[1,3]
[1] Institute of Automation, Chinese Academy of Sciences
[2] School of Information Technology & Management, University of International Business and Economics
[3] China-Singapore Institute of Digital Media
jtsang@nlpr.ia.ac.cn, ludy@uibe.edu.cn, csxu@nlpr.ia.ac.cn

ABSTRACT

In social media, users have contributed enormous behavior data online which can be leveraged for user modeling and conduct personalized services. Temporal user modeling, which incorporates the timestamp of these behavior data and understands users' interest evolution, have attracted attention recently. With the recognition that user interests are vulnerable to transient events, many current temporal user modeling solutions propose to first identify the transient events and then consider the identified events into user behavior modeling. In this work, in the context of microblogs, we propose a unified probabilistic framework to simultaneously model the process of transient event detection and temporal user tweeting. The outputs of the framework include: (1) one long-term topic space spanning over general categories, (2) one short-term topic space for each time interval corresponding to the transient events, and (3) users' interest distributions over the long- and short-term topic spaces. Qualitative and quantitative experimental evaluation are conducted on a large-scale Twitter dataset, with more than 2 million users and 0.3 billion tweets. The promising results demonstrate the advantage of the proposed topic models.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

temporal user modeling, topic model, event detection, microblogs

1. INTRODUCTION

The huge amount of User Generated Content (UGC) online has made the exploration and discovery of interesting resources extremely difficult. Traditional "one-to-all" strategy is no more adequate towards users' customized demands. Understanding the customized interests by building user profiles has stood out for solutions, and enabled "one-to-one" personalized information services. One of the most critical issues in building user profiles is the dynamicity problem, i.e., users' interests vary over time and should be reflected in their profiles [1]. Modeling users' temporal profiles can help providing more qualified services, e.g., recommending timely news to users by capturing their temporal interest shifting [2, 3], providing the right ad at the right time by analyzing users' recent shopping-related activities [4, 5].

One fundamental solution for temporal user modeling is based on the assumption that users' dynamic preferences are affected by both the long-term and the short-term interests [6, 2, 7, 8, 9, 10]. Long-term interests indicate users' stable preferences and distribute over general topic, e.g., the intrinsic interests in politics and sports. While short-term interests are generally consistent with long-term interests, they usually distribute over more specific topics and are changeable over time, e.g., the focuses on "Crimean crisis" at March, 2014, and "FIFA World Cup" around July, 2014. Successfully capturing the short-term interest evolvement will facilitate user preference understanding both in a timely fashion and at a fine-grained level, which is critical for personalized information services.

On microblogging websites, such as Twitter and Weibo, users' temporal behaviors are recognized to be affected by transient events [11], such as new product release and social breaking news. The dominant solutions conduct short-term interest modeling by investigating the interplay between users' temporal behaviors with the transient events detected in advance. For example, [2] first recognizes trending entities at specific periods and then represent users' short-term interests over these trending entities; [12] models user posting behavior as a generative process influenced by the breaking news, which are identified in advance by examining the bursty keywords. The current solutions separating the transient event detection and temporal user modeling suffer from two problems: (1) As illustrated from the example in Figure 1, the transient events and users' temporal behaviors are mutually influenced [13]: on one hand, transient events are identified from aggregated user behaviors; on the other hand, users' temporal behaviors are largely affected by the transient events, which results in the short-term interest evolvement. It is difficult to say influence in which direction happens first and is more significant. (2) The identified transient events beforehand are not well compatible with the task of temporal user modeling. For example, in [2, 12], transient events are represented by a set of bursty entities/keywords at each time interval. Different events mix with each other if they happen within the same time interval. However, for short-term interest modeling, users' interplay with the respective transient events is desired. Moreover, independently performing event detection will lead to loss of transient events which are important to understand

CIKM'15, October 19–23, 2015, Melbourne, Australia.

DOI: http://dx.doi.org/10.1145/2806416.2806470.

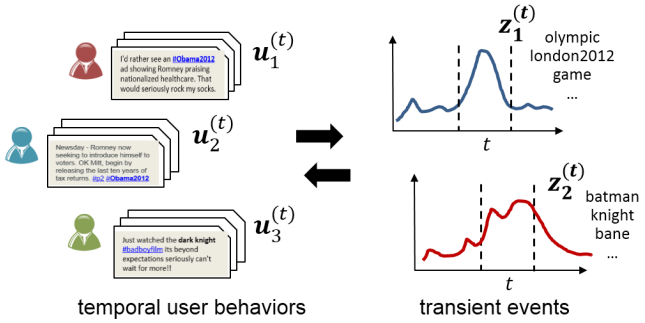

Figure 1: The interplay between users' temporal behaviors and the transient events.

short-term interest evolvement. For example, a local event, though significantly affecting the online behaviors of local users, will be drowned in a global fashion and thus cannot be identified to facilitate the task of temporal user modeling.

To address the above problems, in this paper, we propose a probabilistic framework for temporal user modeling on microblogs, where transient event detection and user tweeting modeling are conducted simultaneously. The user tweets are modeled in a fully generative way: one tweet by a specific user at a specific timestamp is generated either from his/her long-term interest or from his/her short-term interest at this time interval. The long-term interest is expected to distribute over the general topics in a global timeline. While, the short-term interest distributes over the temporal topics discovered at each corresponding time interval, which basically indicates the transient events. The model operates in a unified way, by inputting the user tweets attached with the posted timestamps, and outputting one long-term topic space, N short-term topic spaces [1], and users' long-term and short-term interest distributions. The advantages of this framework include: (1) Event detection and user modeling are conducted simultaneously in a unified topic-based framework. The simultaneity is consistent with the event-user behavior interplay that happened in real world. (2) The unified framework makes it possible to describe events and model users at the topic level. From the perspective of event detection, multiple transient events within one time interval are allowed and the compactness for each event representation is guaranteed. From the perspective of user modeling, both user interests over the general topics and user responses to the transient events are obtained.

The remainder of this paper is organized as follows: section 2 provides a brief review of related work on temporal user modeling, section 3 formally presents the proposed probabilistic topic models and elaborates the model learning and update solutions, followed by the experimental results and analysis in section 4. Finally in section 5, we conclude this work with future directions.

2. RELATED WORK

The goal of temporal user modeling is to capture the dynamic characteristics of users' interests over time. Researchers have proposed many solutions towards this goal. One straightforward idea is to record users' behaviors in time order and build user profiles at each time interval for temporal information services. For example, Zimdars [14] conducted an early work by extending the traditional collaborative filtering (CF) with time order information. With the aim to predict future user behaviors based on the history temporal data, this research line is much promoted by the famous Netflix Prize competition. The Netflix award winning algorithm *timeSVD++* [15] records the history user ratings in a factorization model and conducts prediction by setting bias at each specific time interval. Another popular temporal solution on the Netflix dataset is a Bayesian Probabilistic Tensor Factorization (BPTF) model [16], where users, items and time are represented in three-order tensors, and prediction is conducted in the shared low-dimensional factor space. In the context of microblogs, Abel et al. [17] proposed to represent temporal Twitter user profiles as a set of weighted concepts at each time interval, and conducted personalized website recommendation directly based on the cosine similarity. A recent work on Weibo incorporates the time factor into the matrix factorization model, SocialMF [18], and expresses user dynamic interests as a series of temporal matrices [11].

Another line for temporal user modeling is to analyze the user interest evolvement. The basic idea in most of the work is to emphasize on the new data and reduce the data's importance by time. In [19], a modified collaborative filtering algorithm weighted by time is proposed for temporal recommendation. An exponential time decay function is designed to calculate the time weights for different history user behaviors. A similar temporal user modeling solution is introduced in [20], where the interests' weights are reduced by time if they are not involved by the user until they disappear. Michlmayr and Cayzer [21] modeled user interest evolvement from two perspectives. They added evaporation and reinforcement operations to reduce the weight of old tagging data and increase the weight of repeated tagging data, respectively. An online evolutionary collaborative filtering model is proposed in [22], where temporal information is incorporated into an incremental updating algorithm to track the user interest evolvement over time. Recently, Ceren et al. [23] introduced their user interest inference work from microblogs. For modeling the interest evolution over time, a Markov like assumption is made: the current user interest is a function of the interests in previous time intervals and the estimate of interest at the current time interval.

Our work belongs to the third research line, which explicitly separates the dynamic component of user interests and build temporal user profiles from both long- and short-term interests. Xiang et al. [6] proposed a session-based temporal graph model to capture users' dynamic preferences on the social bookmarking websites, where all items viewed by a user construct his/her long-term interests and the items viewed at a given time interval construct the short-term interests. In [7], regarding history queries and clicked documents, the interaction between short- and long-term behaviors is investigated. An effective hybrid model is proposed for search personalization. Yang et al. [8] proposed a local implicit feedback model for temporal music recommendation, where local and global information are represented by implicit feedback and combined to capture users' stable and local changeable preferences.

As mentioned in *Introduction*, users' temporal behaviors are largely affected by the transient events, especially in the context of microblogs. In [2], the interaction between user dynamic interests and public trends on Twitter are investigated. The public trend at a given time interval is identified as a set of weighted entities in advance. Similarly in [12], the authors first detected the breaking news on Twitter from emerging bursty keywords, and then model the user tweeting behavior as influenced by the detected breaking news. Recently, Deng et al. [10] presented a cross-network solution to model the short-term interests, by discovering user-specific transient events on Twitter and conducting personalized

[1] N is the number of time intervals. Each short-term topic space corresponds to the identified temporal topics (transient events) within this time interval. In the rest of this paper, we will mix using "temporal topics" and "transient events" when no ambiguity is caused.

video recommendation on YouTube. These work largely ignore the mutual influence between transient events and users' temporal behaviors. This leads to the mixed and biased events towards the task of temporal user modeling. The most related work to this paper is [9], which models the fully generative process of user ratings as influenced both by the user-oriented topics and the time-oriented topics. The user- and time-oriented topics generally correspond to the long- and short-term interests. However, in their proposed topic model: (1) The proposed topic model in [9] is designed and evaluated on the bookmarking websites, e.g., MovieLens, Digg, Douban Movie, etc. The rating behaviors from all time shared the same time-oriented topic space, making it difficult to identify the real transient events at each time interval, especially in the context of noisy microblogs. For example, some long-standing words will be mixed into time-oriented topics and disturb the short-time interest modeling; (2) Within certain time interval, all users share the same temporal context, i.e., time-oriented topic distribution, and no user-specific temporal topic distribution is obtained. This actually assumes that all users take unique responses to the discovered transient events. In this paper, we obtain topic spaces for different time intervals and short-term topic distribution is assumed for each user: the derived short-term topics indicates the transient events at each time interval, and the user-specific topic distribution over the discovered short-term topics reflects user's response/interest to the transient events.

3. THE APPROACH

3.1 Preliminaries

To model the generation of user' tweets, the proposed generative model is based on the standard topic model, or Latent Dirichlet Allocation (LDA [24]). Fig. 2(a) shows the graphical model of LDA. The generative process is assumed in a corpus-document-word structure, where the corpus consists of D documents and document d has N_d words. α and β are fixed parameters of symmetric Dirichlet priors for the D document-topic multinomials θ and the K topic-word multinomials ϕ. For each document d, the N_d words are generated by drawing a topic k from the document-topic distribution $p(z|\theta_d)$ and then drawing a word w from the topic-word distribution $p(w|z=k,\phi_k)$.

Since we intend to model the generation of user' tweets as influenced by alternative sources, an important extension to the standard topic model, LDA with switch variable is also introduced (as shown in Fig. 2(b)). The latent variable x acts as a switch: if $x=1$, the previously described standard topic mechanism is used to generate the word, whereas if $x=0$, words are sampled from a background distribution specific for the corpus [25]. The corpus-specific background distribution can be viewed as a general topic consisting of words that are commonly used across a broad range of documents in the corpus. The full generation process is described in Table 1. The conditional probability of generating a word w given a document d can be written as:

$$p(w|d) = \lambda_d \cdot \psi_w + (1-\lambda_d) \cdot \sum_{k=1}^{K} \theta_{d,z_k} \phi_{z_k,w} \qquad (1)$$

where $\lambda_d = p(x=0|d)$, $1-\lambda_d = p(x=1|d)$, ψ_w is the probability of generating w from the background distribution, $\theta_{d,z_k} = p(z_k|\theta_d)$ is the document-topic distribution that document d selects the k^{th} topic, and $\phi_{z_k,w} = p(w|z_k)$ is the topic-word distribution that generates w from the k^{th} topic. By sampling the switch variable x for each word token, the common words can be identified (with sampled switch variable $x=0$) and separated

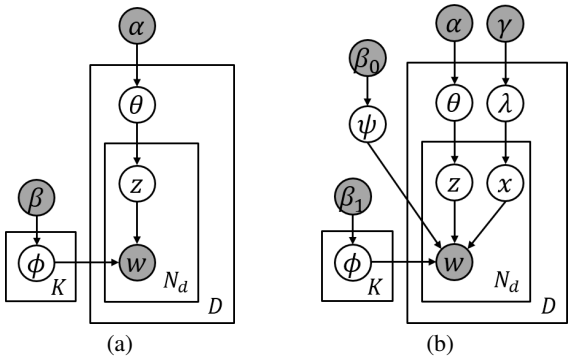

Figure 2: Graphical models for (a) the standard LDA topic model; and (b) topic model with switch variable enabling alternative generation sources.

Table 1: The generation process of LDA with switch variable.

1. Draw background multinomial distribution $\psi \sim \text{Dir}(\beta_0)$.
2. For each topic $k = 1, \cdots, K$:
   (a) draw topic-word multinomial distribution $\phi_k \sim \text{Dir}(\beta_1)$.
3. For each document d:
   (a) draw document-topic multinomial distribution $\theta_d \sim \text{Dir}(\alpha)$;
   (b) draw bernoulli distribution $\lambda_d \sim \text{Dir}(\gamma)$;
   (c) for each word $w_{d,i}$ in document d:
      i. draw a topic $z_{d,i} \sim \text{Multi}(\theta_d)$;
      ii. draw a switch variable $x \sim \text{Bernoulli}(\lambda_d)$;
      iii. draw $w_{d,i} \sim \text{Multi}(\psi)$ if $x=0$, and $w_{d,i} \sim \text{Multi}(\phi_{z_{d,i}})$ if $x=1$.

to construct the background topic, making the K other topics focused on specific aspects of the corpus. We can see that, by introducing the switch variable, topic model is able to explain the observed words in alternative ways. Extensive work have exploited this potential of switch variable with topic models to discover the broader structure of data [26, 27, 25].

3.2 The Proposed Topic Model

This subsection introduces the proposed probabilistic topic model for temporal user modeling on microblogs. Specifically, we take tweets from Twitter as running example for elaboration. We follow the notations used in the standard LDA when possible.

3.2.1 Temporal User Modeling (TUM) model

As mentioned in *Introduction*, we explain users' temporal profiles as decomposed into long-term interests and short-term interests. From a generative perspective, user tweets can be assumed as an aggregated result from users' long-term and short-term interests. Inspired by the above introduced topic model, we realize this assumption by setting a similar binary switch variable x to control the generation source of the observed words w, i.e., either from a long-term topic z^L or from a short-term topic z^S. Since short-term interests should distribute over different time intervals, an observed variable t is further introduced to record the posted timestamp of the tweet. That is, once the switch variable of generation source is sampled as short-term interest, the short-term topic spaces ψ_t of

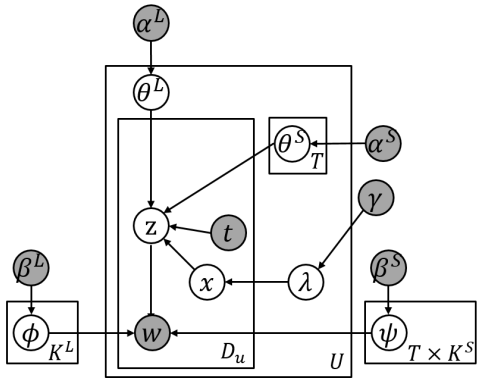

Figure 3: Probabilistic generative model of TUM.

the specific time interval t that the tweet was posted is selected to generate the word. In this case, since the topic spaces obtained for different time intervals are expected to indicate the transient events, users' short-term interests actually distribute over the transient events at the corresponding time intervals, i.e., the responses to the transient events reflect users' short-term preferences and affect their temporal tweeting behaviors.

Fig. 3 illustrates the graphical structure of the proposed Temporal User Modeling (TUM) model. The full generation process is described in Table 2. This model includes three sets of variables, i.e., the parameters $\{\Phi, \Psi, \Theta^L, \Theta^S, \boldsymbol{\lambda}\}$, the latent variables $\{X, Z\}$, and the observation W. We use Gibbs Sampling to generate samples for the latent variables and then calculate the desired parameters. Given the graphical model and generative process, it is straightforward to derive the full conditional probability of the latent variables for each word token $w_{u,i}$:

$$p(x_{u,i}=0, z_{u,i}=k|X_{u,\neg i}, Z_{u,\neg i}, W; \cdot) \propto \frac{C_{U,X}(u,0)+\gamma}{C_U(u)+2\gamma} \cdot \frac{C_{U,X,Z}(u,0,k)+\alpha^L}{C_{U,X}(u,0)+K^L\alpha^L} \cdot \frac{C_{X,Z,W}(0,k,w_{u,i})+\beta^L}{C_{X,Z}(0,k)+|\mathcal{V}|\beta^L} \quad (2)$$

$$p(x_{u,i}=1, z_{u,i}=k|X_{u,\neg i}, Z_{u,\neg i}, W; \cdot) \propto \frac{C_{U,X}(u,1)+\gamma}{C_U(u)+2\gamma} \cdot \frac{C_{U,X,T,K^S}(u,1,t_{u,i},k)+\alpha^S}{C_{U,X,T}(u,1,t_{u,i})+K^S\alpha^S} \cdot \frac{C_{X,T,K^S,W}(1,t_{u,i},k,w_{u,i})+\beta^S}{C_{X,T,K^S}(1,t_{u,i},k)+|\mathcal{V}|\beta^S} \quad (3)$$

where $|\mathcal{V}|$ denotes the size of the word vocabulary, $t_{u,i}$ denotes the time interval when the word $w_{u,i}$ was posted, and $C(\cdot)$ stores the number of samples satisfying certain requirements during the iterative sampling process. For example, $C_{U,X,T,K^S}(u,1,t_{u,i},k)$ indicates the number of words for user u that are supposed to be generated from the short-term topic z^k at time interval $t_{u,i}$. Note that for model derivation simplification, we assume all parameters follow symmetric Dirichlet priors [2].

After a sufficient number of Gibbs sampling iterations, the approximate posterior can be used to obtain estimates of the desired parameters of topic spaces and user-topic distributions, by examining the counts of sampled latent variables of Z, X.

[2] The assumption for symmetric prior is easy to relax [28].

Table 2: The full generation process of TUM.

1. For each long-term topic $k = 1, \cdots, K^L$:
   (1) draw topic-word multinomial distribution $\boldsymbol{\phi}_k \sim \text{Dir}(\boldsymbol{\beta}^L)$.
2. For each time interval $t = 1, \cdots, T$:
   (1) for each short-time topic $k = 1, \cdots, K^S$ at the t^{th} time interval:
      a. draw topic-word multinomial distribution $\boldsymbol{\psi}_{t,k} \sim \text{Dir}(\boldsymbol{\beta}_S)$.
3. For each user $u = 1, \cdots, U$:
   (1) draw long-term user-topic multinomial distribution $\boldsymbol{\theta}_u^L \sim \text{Dir}(\boldsymbol{\alpha}^L)$;
   (2) for each time interval $t = 1, \cdots, T$:
      a. draw short-term user-topic multinomial distribution $\boldsymbol{\theta}_{u,t}^S \sim \text{Dir}(\boldsymbol{\alpha}^S)$.
   (3) draw bernoulli distribution $\lambda_u \sim \text{Dir}(\gamma)$;
   (4) for each word $i = 1, \cdots, D_u$:
      a. draw a switch variable $x_{u,i} \sim \text{Bernoulli}(\lambda_u)$;
      b. if $x_{u,i} = 0$, first draw a topic $z_{u,i} \sim \text{Multi}(\boldsymbol{\theta}_u^L)$, then draw $w_{u,i} \sim \text{Multi}(\boldsymbol{\phi}_{z_{u,i}^L})$;
      c. if $x_{u,i} = 1$, first draw a topic $z_{u,i} \sim \text{Multi}(\boldsymbol{\theta}_{u,t}^S)$, then draw $w_{u,i} \sim \text{Multi}(\boldsymbol{\psi}_{t,z_{u,i}^S})$. t is the time interval that the word $w_{u,i}$ was posted.

Specifically, the MAP estimates are as follows:

$$\begin{aligned} \phi_{k,w} &= \frac{C_{X,Z,W}(0,k,w)+\beta^L}{C_{X,Z}(0,k)+|\mathcal{V}|\beta^L}, \\ \psi_{t,k,w} &= \frac{C_{X,T,K^S,W}(1,t,k,w)+\beta^S}{C_{X,T,K^S}(1,t,k)+|\mathcal{V}|\beta^s}, \\ \theta_{u,k}^L &= \frac{C_{U,X,Z}(u,0,k)+\alpha^L}{C_{U,X}(u,0)+K^L\alpha^L}, \\ \theta_{u,t,k}^S &= \frac{C_{U,X,T,K^S}(u,1,t,k)+\alpha^S}{C_{U,X,T}(u,1,t)+K^S\alpha^S} \end{aligned} \quad (4)$$

3.2.2 TwitterTUM model

We also pursued a variant of TUM. In the context of microblogs, e.g., Twitter, single tweet is recognized to usually involve with one single topic [29]. In the above TUM model, the words in the same tweet are sampled separately from different topics. In practical implementation, it is reasonable to assume that users express unique interest in the same tweet. Therefore, we modify the TUM model by introducing an additive tweet plate and assuming that one tweet can be generated from only one long-term and one short-term topic. For example, user posted a tweet about the semifinal match between Germany and Brazil expressing his/her long-term interest in sports and short-term interest in 2014 FIFA World Cup.

Specifically, in the generative process, one long-term topic $z_{u,e}^L$ and one short-term topic $z_{u,e}^S$ are firstly sampled for the tweet e_u. The words in e_u are then generated from either $z_{u,e}^L$ or $z_{u,e}^S$ according to the sampling of switch variable x. This constraint is consistent with user tweeting behavior and improves the compactness of the derived topics, which helps discover more meaningful transient events. The modified model is referred to as TwitterTUM, whose graphical structure is shown in Fig. 4.

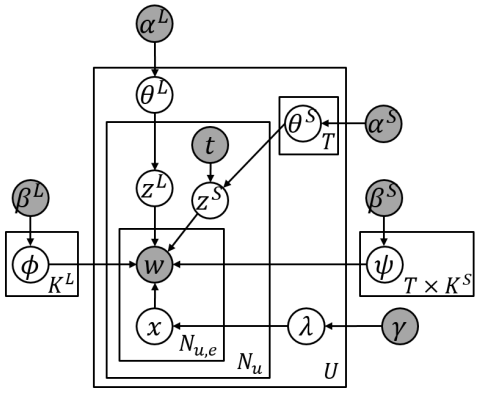

Figure 4: Probabilistic generative model of TwitterTUM.

Table 3: The generation process of one user in TwitterTUM.

For each user $u = 1, \cdots, U$:

(1) (2) (3): same with *TUM* in Table 2;

(4) for each tweet $e = 1, \cdots, N_u$:

a. draw long-term topic $z^L_{u,e} \sim \text{Multi}(\boldsymbol{\theta}^L_u)$;

b. draw short-term topic $z^S_{u,e} \sim \text{Multi}(\boldsymbol{\theta}^S_{u,t})$. t is the time interval that the tweet e was posted.

c. for each word $i = 1, \cdots, N_{u,e}$:

i. draw a switch variable $x_{u,e,i} \sim \text{Bernoulli}(\lambda_u)$;

ii. draw $w_{u,e,i} \sim \text{Multi}(\boldsymbol{\phi}_{z^L_{u,e}})$ if $x_{u,e,i} = 0$, and $w_{u,e,i} \sim \text{Multi}(\boldsymbol{\psi}_{t,z^S_{u,e}})$ if $x_{u,e,i} = 1$.

Since the generation of the long-term and short-term topics remain unchanged, we only summarize the generation process within the user and tweet plates in Table 3. The following equations list the main update rules in the Gibbs Sampling for TwitterTUM:

$$p(x_{u,e,i} = 0 | X_{u,e,\neg i}, Z^L, W; \cdot) \propto \frac{C_{U,X}(u,0) + \gamma}{C_U(u) + 2\gamma} \cdot \frac{C_{Z^L,W}(z^L_{u,e}, w_{u,e,i}) + \beta^L}{C_{Z^L}(z^L_{u,e}) + |\mathcal{V}|\beta^L};$$

$$p(x_{u,e,i} = 1 | X_{u,e,\neg i}, Z^S, W; \cdot) \propto \frac{C_{U,X}(u,1) + \gamma}{C_U(u) + 2\gamma} \cdot \frac{C_{T,Z^S,W}(t_{u,i}, z^S_{u,e}, w_{u,e,i}) + \beta^S}{C_{T,Z^S}(z^S_{u,e}) + |\mathcal{V}|\beta^S};$$

$$p(z^L_{u,e} = k | X, Z^L_{u,\neg e}, W; \cdot) \propto \frac{C_{U,Z^L}(u,k) + \alpha^L}{C_U(u) + K^L\alpha^L} \cdot \prod_{i=1}^{N_{u,e}} \left(\frac{C_{Z^L,W}(k, w_{u,e,i}) + \beta^L}{C_{Z^L}(k) + |\mathcal{V}|\beta^L} \right)^{1 - x_{u,e,i}};$$

$$p(z^S_{u,e} = k | X, Z^S_{u,\neg e}, W; \cdot) \propto \frac{C_{U,Z^S}(u,k) + \alpha^S}{C_U(u) + K^S\alpha^S} \cdot \prod_{i=1}^{N_{u,e}} \left(\frac{C_{T,Z^S,W}(t_{u,i}, k, w_{u,e,i}) + \beta^S}{C_{T,Z^S}(k) + |\mathcal{V}|\beta^S} \right)^{x_{u,e,i}}.$$

3.3 Batch Training v.s. Incremental Update

In the above introduced model learning, we assume the user tweets at all time intervals are observed simultaneously, and the long- and short-term parameters are estimated from all the observed data. We refer to this learning strategy as *batch training*. While, in practical implementation, data arrive in a stream. Moreover, the addressed temporal user modeling task needs the capability to update model continuously. These all necessitate the introduction of an incremental learning strategy to the proposed topic models. To this end, in this subsection, using TUM as example, we present an incremental learning strategy to scan new data on the basis of time interval, which is referred to as *incremental update*.

In the incremental update learning strategy, batch training is initially run on data of the first T_0 time intervals, and the model parameters are initialized as $\{\Phi^{(T_0)}, \Psi_{1:T_0}, \Theta^{L(T_0)}, \Theta^S_{1:T_0}\}$ according to Eqn. (4). Here the superscript $\cdot^{(T_0)}$ indicates the parameter value after observing T_0 time intervals, and the subscript $\cdot_{1:T_0}$ denotes the parameters from the first to the $T_0{}^{th}$ time intervals. Henceforth, with the arrival of data $W^{(t+1)}$ in every new time interval, fixing the value of old switch and topic latent variables $\{X_{1:t}, Z_{1:t}\}$, we sample $\{X_{(t+1)}, Z_{(t+1)}\}$ for each new word token by applying Eqn. (2) and Eqn. (3) conditioning on the words observed so far. We can see that for long-term topics, all the three components in Eqn. (2) are vulnerable to the samples in the previous time intervals, while for short-term topics the second and third components in Eqn. (3) are only decided by the analyzed samples at the current time intervals.

Subsequently, the MAP estimates of the parameters are updated using the sampled assignments of topic and switch variables to the data at the new time interval. According to the generative process described in Table 2, the short-term topics $\Psi_{1:t}$ and users' short-term interest distribution $\Theta^S_{1:t}$ at the previous t time intervals are not dependent on the incoming data, and thus should be fixed during update. This is easy to understand as $\Psi_{1:t}$ corresponds the occurred transient events and $\Theta^S_{1:t}$ indicates users' responses to these events. Therefore, after Gibbs sampling on the data at $t+1$ time interval, the model updates $\Phi^{(t)} \rightarrow \Phi^{(t+1)}$, $\Theta^{L(t)} \rightarrow \Theta^{L(t+1)}$ and estimates $\Psi_{t+1}, \Theta^S_{t+1}$.

Note that the sampled latent variables at the previous time intervals actually add up to the counts on the number of times words are sampled under certain conditions [3]. This plays the same role as hyper-parameters. Therefore, the generated model at the previous time intervals can be interpreted as a prior for the model at the successive time intervals. Specifically, users' long-term topic distribution is modified as follows:

$$\begin{aligned} \phi^{(t+1)}_{k,w} &\sim Dir(\beta^L_w + \beta^{L(t)}_w), \\ \theta^{L(t+1)}_{u,k} &\sim Dir(\alpha^L_k + \alpha^{L(t)}_k) \end{aligned} \tag{5}$$

where $\beta^{L(t)}_w \triangleq C^{1:t}_{X,K^L,W}(0,k,w), \alpha^{L(t)}_k \triangleq C^{1:t}_{U,X,K^L}(u,0,k)$ denote at the previous t time intervals the number of times that word w are generated from the k^{th} long-term topic, and the number of words for user u that are generated from the k^{th} long-term topic, respectively. To modify the contribution of history samples in computing the priors for the incoming data, a time decay term can be further added to define $\beta^{L(t)}_w$ and $\alpha^{L(t)}_k$ as follows:

$$\begin{aligned} \beta^{L(t)}_w &\triangleq \sum_{i=1}^{t} (1 - e^{-\frac{t-i}{\delta}}) \cdot C^{(i)}_{X,K^L,W}(0,k,w), \\ \alpha^{L(t)}_k &\triangleq \sum_{i=1}^{t} (1 - e^{-\frac{t-i}{\delta}}) \cdot C^{(i)}_{U,X,K^L}(u,0,k) \end{aligned} \tag{6}$$

[3] Since the vocabulary enlarges over time due to the new words from the incoming time intervals, we assume these words have zero count for all topics at the previous time intervals.

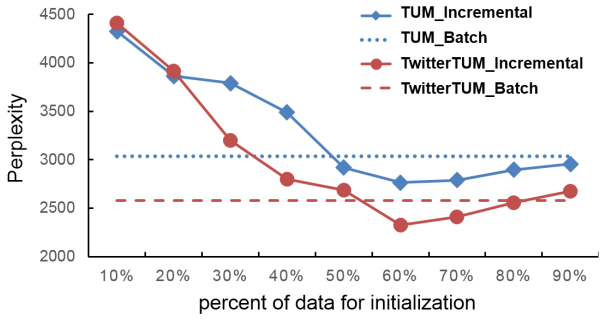

Figure 5: Perplexities for different learning strategies.

Table 4: Runtimes of different learning strategies (in seconds).

	Batch Training	Incremental Update		
		20%	50%	80%
TUM	9,325	4,658	6,525	8,621
TwitterTUM	14,772	6,639	9,383	12,199

where $C^{(i)}_{X,K^L,W}(0,k,w)$ denotes the number of times that word w are generated from the k^{th} long-term topic at the i^{th} time interval, δ is a tuning parameter to control the time decay speed [4].

In batch training, to simplify the learning process, we assume the short-term topic spaces of different time intervals share the same topic number K^S. In incremental update, respective short-term topic numbers are selected for each time interval by Bayesian model selection [30]. Specifically, for each time interval t, the likelihood $p(W^{(t)}|K^{S(t)})$ is calculated with different short-term topic number $K^{S(t)}$. The $K^{S,(t)}$ that obtains the largest likelihood is considered to best account for the structure of the data and thus set as the short-term topic number for this time interval [**?**].

Different from the short-term topics, the long-term topic structure is fixed in the initialization stage, i.e., the number of long-term topics has been decided after the first T_0 time intervals, and only the topic-word distributions are modified during incremental update. For this reason, the model performance depends critically on the accuracy of the topics inferred during the initialization stage. To compare the performance with different time intervals for initialization, we evaluate the held-out perplexity on a separate validation set from the collected 10-month Twitter dataset. By splitting the data from one month as one time interval, different number of initialization time intervals $T_0 = 1, \cdots, 9$ is examined, using the best short- and long-term topic number settings. Figure 5 shows the results for both TUM and TwitterTUM. The perplexity of the batch training strategy is also examined.

It is shown that the perplexity of incremental update learning strategy initially decreases as a function of the data utilized for initialization. This is due to the fact that more initialization time intervals will lead to more accurate long-term topic structure. The perplexity reaches the lowest point around 60% initialization data and then increases thereafter. With the increase of initialization time intervals, the flexibility in optimizing short-term topic number reduces. Therefore, a balance between the long- and short-term topic structures is critical to the final model performance. The results suggest that, by setting proper percent of data for initialization, the incremental update strategy can find a solution as good as the batch training strategy. Moreover, we can see that, by introducing an extra tweet plate, TwitterTUM generally obtains lower perplexity than TUM, showing its ability to find better topic structure. Table 4 summarizes the total runtimes before converge for different learning strategies. We can see that the batch training strategy generally converges slower than the incremental learning strategy, by costing more computation time on resampling all the latent variables in each iteration. For similar reason, TwitterTUM is slower than TUM in sampling additive latent variables.

[4] In our experiments on a 10-month Twitter dataset, we set $\delta \to \infty$ to remove the temporal difference.

4. EXPERIMENTS

4.1 Data Set

Twitter API is used to collect the dataset for the experiments. We started from a random Twitter user and crawled his followees using Breadth First Search. All the examined users' public tweets from Feb.1, 2012 to Nov.30, 2012 are collected. After removing non-English tweets, this results in 852,800 Twitter users with 599,818,231 tweets. To focus on the active Twitter users for temporal modeling, we further removed Twitter users with less than 1,000 tweets within the examined 10 month period. The final dataset contains 228,921 Twitter users and 362,217,995 tweets, with an average of 40 tweets for each user per week.

4.2 Perplexity Results

We first examine the performance of the proposed topic modesl in terms of perplexity. The perplexity in the context of this study measures the accuracy in predicting the coming of new tweets, which can be calculated over all test tweets:

$$\text{Perplexity}(\mathcal{D}_{test}) = \exp\left(-\frac{\sum_{d\in\mathcal{D}_{test}}\sum_{i=1}^{N_d}\log p_\Theta(w_{d,i})}{\sum_{d\in\mathcal{D}_{test}} N_d}\right) \quad (7)$$

where $\mathcal{D}_{test}$ is the test tweet set, $p_\Theta(w_{d,i})$ is the predictive probability of a word according to the derived model parameters. We randomly split the tweets of each user per month into 90% training tweets and 10% test tweets.

We compare the perplexities of the proposed TUM and TwitterTUM with two non-temporal and two temporal user modeling methods:

- **Author-Topic model (AT)** [31]: assuming the tweets are generated considering the authorship, i.e., the user static interests.
- **TwitterLDA** [29]: adding an additional tweet plate to the *AT* model, and constraining that only one topic is allowed within the same tweet.
- **Mixture Latent Topic model (Mixture)** [12]: a temporal user modeling solution assuming user posting is influenced by breaking news, social friends, as well as users' intrinsic interests. Since social influence is not the focus of this paper, we implemented a modified version of *Mixture* for comparison that removes the social friend influence.
- **TTCAM** [9]: a temporal user modeling solution that integrates the discovery of long-term and short-term topics in a unified model.

To simplify the performance comparison, we learned the proposed TUM and TwitterTUM using the batch training strategy. The examined topic models are run ten times and the obtained perplexity is averaged over the ten times and shown in Fig. 6. Time intervals for the temporal user modeling methods are changed from $T = 1$ day to $T = 2$ month. It is shown that: (1) The perplexities remain unchanged for the *AT* and *TwitterLDA* models. Basically, considering the temporal characteristics, the four temporal user modeling methods obtain better predictive performances than the

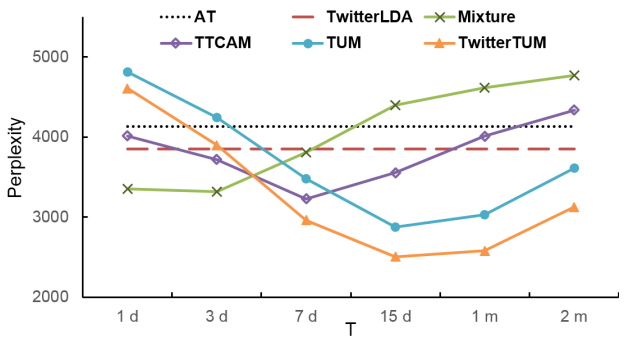

Figure 6: Perplexity with different time interval settings.

Table 5: Examples of discovered long-term topics by *TwitterTUM*.

Topic No.	Top words
Long 3	app apple google android kinect
Long 7	film movie festival films trailer
Long 14	game team live cup sports
Long 19	social twitter marketing search brand

non-temporal user modeling methods in terms of perplexity. (2) *Mixture* obtains the best predictive performance when the time interval is relative short ($T = 1 - 3$ day). This is due to that when T becomes longer, the breaking news identified by *Mixture* are more likely to be the mixtures of several transient events. (3) When T increases, the perplexity of *TTCAM* first decreases, obtains the lowest value when $T = 7$ day, and then increases. The reason for the early decreasing perplexity is that longer time intervals contribute to adequate user data and improved topic discovery. As the time interval gets longer, the reduced temporal influence leads to the increasing perplexity. (4) The proposed *TUM* and *TwitterTUM* follow the similar trend with *TTCAM*, and obtain the lowest perplexities when $T = 15$ day. However, by setting respective short-term topics at each time interval and allowing users to have different responses to the short-term topics, *TUM* and *TwitterTUM* achieve even lower perplexities than *TTCAM*. This demonstrates that the proposed topic models are more consistent with the users' temporal tweeting behaviors on microblogs.

4.3 Discovered Topic Results

4.3.1 Long-term Topics

We first examine the discovered long-term topics using the proposed topic models. In the following, unless specified, the results of *TUM* and *TwitterTUM* are obtained with time interval $T = 15$ day and using the batch training strategy. Therefore, the number of time interval is 20. The number of long-term and short-term topics is selected by Bayesian model selection, that $K^L = 25$, $K^S = 10$.

Table 5 shows four of the discovered 25 long-term topics by *TwitterTUM*, with each topic represented by the five most probable words. We can see that within each topic, the probable words are semantic-consistent with each other. The discovered long-term topics basically describe some general topics like *digital goods*, *movie*, *sports*, *social marketing*, etc, which is consistent with our understanding and expectations.

To analyze the steadiness of long-term topics, we examine the popularity of the discovered long-term topics at different time intervals. The popularity of a long-term topic k at time interval t is measured by the proportion of words that sampled from this topic, i.e., $C_{T,Z^L,X}(t,k,0)/C_{T,X}(t,0)$. Fig. 7 visualizes the popularity evolvement of the long-term topics. The colorbar on the right shows the corresponding popularity values. We can see that users' macro interests over the long-term topics are generally steady, with slight fluctuation at some time intervals. For example, the interest over the *topic* 14 increases at time interval 12 and 13, which is possibly due to the 2012 London Olympics between July 27 to August 12. This inter-relationship between the long-term and short-term topics will also be reflected and discussed in the later experiments.

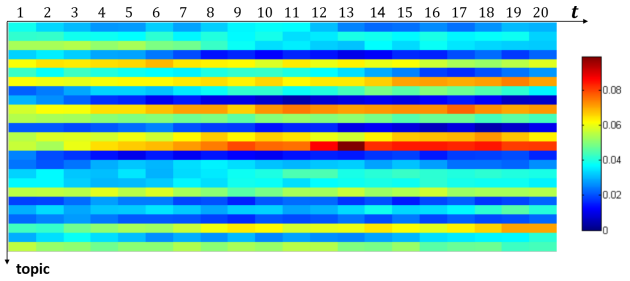

Figure 7: The popularity of long-term topics over time.

For individual user's long-term topical interest evolvement, we examine into the results from *incremental update* learning with three month (30%) for initialization and 15 day as one time interval for update. Therefore, for each user, we obtained 15 long-term topic distributions, one derived from initialization and the other 14 derived at each update time interval. 3,000 users who have posted more than 100 tweets at each of the examined time intervals, are randomly selected to construct an active test user set $\mathcal{U}_{active}$. For each of the active users, we fitted his/her 15 topical distribution vectors with a multivariate normal distribution. The long-term topic distribution variance is defined as follows:

$$\text{Long-term interest variance}(u) = \frac{\sum_{i=1}^{15} d(\mathbf{u}_i^L, \boldsymbol{\mu}_u)}{15} \quad (8)$$

where $d(\cdot,\cdot)$ is the Euclidean distance, $\mathbf{u}_i^L$ is user u's i^{th} topical distribution vector over the long-term topics, $\boldsymbol{\mu}_u$ is the mean vector of u's fitted multivariate normal distribution. The results find that over 85% users have a long-term interest variance less than 0.002. This demonstrates that most users hold stable long-term interests and validates the steadiness of the long-term topic distribution at micro user level.

4.3.2 Short-term Topics

We then investigate into the discovered short-term topics. At each time interval t, a short-term topic space with $K^S = 10$ topics are obtained. These topics are ranked by their popularity, which is defined as the proportion of words that sampled from the topic, i.e., $C_{T,Z^S,X}(t,k,1)/C_{T,X}(t,1)$. The most popular short-term topics are expected to represent the transient events. Table 6 presents the three most popular short-term topics at time interval 12, with each topic represented by the five most probable words. It is easy to see that three transient events are identified and well described, i.e., "2012 London Olympics", "The Dark Knight Rises", "Syrian civil war". The identified short-term topics derive from the co-burst usage of tweet words and reflect users' increasing interests during this period.

For comparison, Table 7 presents the bursty words detected by *Mixture* at the same time interval, which is derived by examining the word frequency. The first row shows the top bursty words

Table 6: Examples of discovered short-term topics by *TwitterTUM* at time interval 12: from July 16 to July 31.

Topic No.	Top words
Short 3	olympic london2012 game ceremony beijing
Short 4	batman knight bane dark nolan
Short 7	syrian war bomb arab uprising

Table 7: Busty words by *Mixture* at the time interval 12.

Time interval	Top words
July 16-31	olympic london2012 vincente sports batman
July 27	london2012 game olympic heywood sports

between July 16-31, the same time span with the results from Table 6. It is shown that the words representative of different events are mixed, due to the fact that multiple transient events happen during the 15-day time interval. The second row shows the detected bursty words on July 27, when the opening ceremony of 2012 London Olympics took place. We can see, even reduce the time interval to one day, the top words detected by *Mixture* still have a very loose structure that different events mix with each other. The reason for this result is that, the word frequency is considered independent and the word-word co-occurrence in tweet usage is ignored. Moreover, without the separation from long-term topics, some general words also emerges as the bursty words, e.g., "game", "sports", which limits the capability to represent the transient events.

We further quantitatively evaluate the effectiveness of the discovered short-term topics in identifying transient events. Specifically, from the trending searches revealed by Google Zeitgeist 2012 [5], we selected 20 transient events with different categories happened from Feb.1, 2012 to Nov.30, 2012. By examining the top probable words, we found that 19 out of the 20 transient events were successfully identified from the top three ranked short-term topics at different time intervals. Table 8 shows four of the transient events and the corresponding short-term topics. We can see that the top probable words are consistent with each other and together serve as good indication for the transient events. In addition to the global events, we also found some local transient events. For example, at time interval 2 a short-term topic is discovered with the top words of "NUS, dog, singaporean, china, scholarship", which indicates the event of "NUS student insulting Singaporean" happened at Singapore on Feb.18, 2012. We examined the users who have high topical probability on this topic and found that many of them come from Singapore. The configuration of short-term topic spaces at each time intervals make it possible to identify the transient events popular in local users. This type of local events are neglected in a global scale by the methods of *Mixture* and *TTCAM*.

4.4 User Modeling Results

4.4.1 Personalized Information Recommendation

One significant application of user modeling on microblogs is to recommend users with interested information. On Twitter, retweet is the behavior to broadcast users' interested tweets to their followers and therefore can serve as the ground-truth for users' interests. In this subsection, we evaluate the different user modeling methods in the context of predicting whether a tweet will be retweeted by a test user.

[5] https://www.google.com/zeitgeist/2012.

Table 8: Transient events and the corresponding short-term topics.

Transient events	Time interval /Topic No.	Top words
Whitney Houston's death	1/3	whitneyhouston, whitney, drug, bobby, tragic
Gangnam Style upsurge	15/3	gangnamstyle, gangnam, spy, korean, youtube
Hurricane Sandy	18/1	sandy, hurricane, cyclone, pacific, victim
US presidential election 2012	19/5	obama, election, romney, democratic, immigration

Specifically, two experimental settings are conducted. (1) In the first setting, for each user in every time interval t, five random tweets that were retweeted by this user are assumed as relevant documents and construct the positive testing set. Since we focus on modeling the tweet content and ignore the social structure, the selected tweets are required to contain at least 50 characters. In this way, we identified 19,271 test users. For each test user, the selected five retweets at each time interval are removed from the topic model learning. 20 other tweets that were not retweeted at the same time interval are added as the negative test samples. (2) In the second setting, the goal is to simulate the process of personalized news recommendation. Within the collected tweet dataset, we identified 90,883 retweets to 75 tweets that (a) were originally posted by the official accounts of BBC or CNN, (b) were retweeted totally more than 10,000 times, and (c) contain more than 50 characters. These identified tweets are more likely in reporting some transient events, and the retweets from the examined users indicate the interest/response to the events. We selected 500 users who have retweeted at least five out of the 75 tweets to construct the test user set for the second setting. The average number of relevant tweets for each test user is 6.6.

Therefore, the tasks for the first (denoted as *retweet prediction*) and the second (denoted as *news recommendation*) settings, are to retrieve the five and on average 6.6 relevant retweets from the tweet collection consisting of 25 and 75 tweets, respectively. The predictive probability of user u retweeting d is estimated as:

$$p(d) = \frac{1}{N_d} \sum_{i=1}^{N_d} p_{\Theta_u}(w_{d,i})$$

where $p_{\Theta_u}(w_{d,i})$ is the probability of seeing word $w_{d,i}$ according to the learned model parameters for user u. The tweets are ranked according to the above predictive probability. We use Average Precision (AP) for evaluation, which is calculated as:

$$AP = \frac{\sum_{i=1}^{N^+ + N^-} (\sum_{j=1}^{i} \frac{rel(j)}{i} \times rel(i))}{N^+} \quad (9)$$

where N^+, N^- are the number of positive and negative tweets, i, j are the rank of the tweet, $rel(i)$ equalizes 1 if the tweet at rank i is relevant and 0 otherwise. The final result is shown in mean Average Precision (mAP), which is averaged over 500 users (for *news recommendation*), or 19,271 users and 20 time intervals (for *retweet prediction*).

The comparison results with the four baseline methods of the two experimental settings are shown in Fig. 8. It is easy to have the following observations: (1) The experimental results in terms of mAP are consistent with that of perplexity. The four temporal user modeling methods generally outperforms the two non-temporal

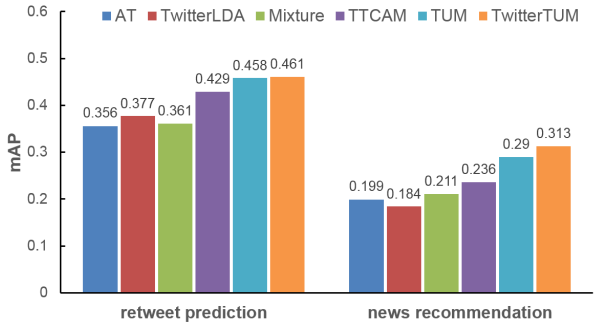

Figure 8: mAP of personalized information recommendation.

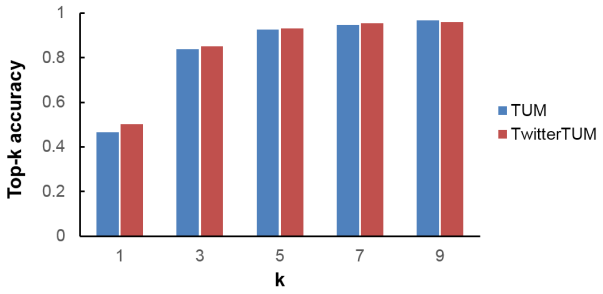

Figure 9: Consistency between long- and short-term interests.

user modeling methods. The proposed two topic models obtain the best performance in both settings with a fixed time interval of 15 day. (2) The improvement over the baseline methods is more significant in *news recommendation* than *retweet prediction*. In the task of *news recommendation*, users' responses to the transient events are expected to contribute more to the final recommendation accuracy. However, *Mixture* mixes different transient events, and *TTCAM* assumes all users share the same interest distribution over time-oriented topics. This significant improvement shows the advantage of the proposed topic models in capturing users' responses to respective transient events.

4.4.2 *Long-term and Short-term Interests*

In this subsection, to understand the advantage in personalized information recommendation, we further investigated into the results of users' long-term and short-term interests. As shown in the previous experimental results, the discovered long-term and short-term topics are in different granularities and thus may semantically mix with each other. We invited three graduate students who are very familiar with Twitter to match each of the discovered 20*10=200 short-term topics to one of the 25 long-term topics, according to the most probable words of the topic. The final short-term and long-term topic matching pairs are obtained by aggregating the three labelers' votes, and recorded as $\text{Long}(z_i^{S(t)})$. For example, the short-term topic 12/3 and 12/4 (shown in Table 5) are labeled as matching the long-term topic 14 and 7 (shown in Table 4), respectively. These matching pairs are recorded as $\text{Long}(z_3^{S(12)}) = 14$, $\text{Long}(z_3^{S(12)}) = 7$.

In the proposed *TwitterTUM*, for each tweet, only one long-term topic and one short-term topic are allowed. We first examined the consistency of the sampled long- and short-term topics within the same tweet. We randomly selected 1,000,000 tweets and found that 74.4% were sampled with matched long- and short-term topics according to the labeled matching pairs. We then examined this consistency at user level. For each test user $u \in \mathcal{U}_{active}$, we investigated how his/her most significant short-term interest at each time interval matches the significant long-term interests. *Top-k accuracy* is utilized as the evaluation metric, which is calculated as follows:

$$\text{Top-k accuracy} = \frac{\sum_{t=1}^{20} \mathbf{I}(\tau_u(Long(\mathbf{u}^{S(t)}.max)) \leq k)}{20} \quad (10)$$

where $\mathbf{u}^{S(t)}.max$ denotes user u's maximum short-term topic at time interval t, $\tau_u(z^L)$ denotes the ranking position of z^L in user u's long-term topic distribution, $\mathbf{I}(\cdot)$ is indicator function returning 1 if it is true and 0 otherwise.

The final results are averaged over all the test users and shown in Fig. 9. We can see that for both proposed topic models, about half of the users/time intervals have consistent short-term and long-term interests, i.e., the most significant short-term topic exactly matches the most significant long-term topic. Over 80% of users/time intervals have the matched long-term interest within the top-3 ranks. This indicates that users do not equally follow all the transient events as assumed by *TTCAM*, but choose to concern more about those consistent with their long-term interests. This result also interprets the phenomenon that long-term topic popularity fluctuates with the short-term topics as shown in Fig. 7. Moreover, one limitation for the proposed topic models is that, the accurate short-term topic distribution is conditioned on the adequate tweeting behavior at each time interval. This observation can be leveraged to make up for this limitation. For example, if users' tweeting behavior at one time interval is sparse, the discovered short-term topic distribution will be modified by its matching relation with the discovered long-term topic distribution.

5. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a probabilistic framework to model the mutual interaction between users' temporal behavior and the transient events. At each time interval, a short-term topic space is discovered from users' co-burst tweeting patterns, which corresponds to the transient events popular during this period. We evaluated the performance of the proposed framework from three perspectives: perplexity, the discovered topic investigation, and the user modeling performance in personalized information recommendation. In the future, we will be working on: (1) addressing the limitation of requiring adequate behavior data within each time interval by leveraging the observed long- and short-term interest consistency, and (2) implementing the incremental learning strategy in real applications and exploring the potentials in transient event tracking.

6. ACKNOWLEDGEMENT

This work is supported in part by National Basic Research Program of China (No. 2012CB316304), National Natural Science Foundation of China (No. 61225009, 61332016, 61303176, 61432019, 61272256), and Beijing Natural Science Foundation (No. 4131004).

7. REFERENCES

[1] Ahmad Abdel-Hafez and Yue Xu. A survey of user modelling in social media websites. *Computer and Information Science*, 6(4):59–71, 2013.

[2] Qi Gao, Fabian Abel, Geert-Jan Houben, and Ke Tao. Interweaving trend and user modeling for personalized news recommendation. In *Web Intelligence*, pages 100–103, 2011.

[3] Ming Yan, Jitao Sang, and Changsheng Xu. Mining cross-network association for youtube video promotion. In *ACM Multimedia*, pages 557–566, 2014.

[4] Amr Ahmed, Yucheng Low, Mohamed Aly, Vanja Josifovski, and Alexander J. Smola. Scalable distributed inference of dynamic user interests for behavioral targeting. In *KDD*, pages 114–122, 2011.

[5] Jitao Sang, Tao Mei, and Changsheng Xu. Activity sensor: Check-in usage mining for local recommendation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):41, 2015.

[6] Liang Xiang, Quan Yuan, Shiwan Zhao, Li Chen, Xiatian Zhang, Qing Yang, and Jimeng Sun. Temporal recommendation on graphs via long- and short-term preference fusion. In *KDD*, pages 723–732, 2010.

[7] Paul N. Bennett, Ryen W. White, Wei Chu, Susan T. Dumais, Peter Bailey, Fedor Borisyuk, and Xiaoyuan Cui. Modeling the impact of short- and long-term behavior on search personalization. In *The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, Portland, OR, USA, August 12-16, 2012*, pages 185–194, 2012.

[8] Diyi Yang, Tianqi Chen, Weinan Zhang, Qiuxia Lu, and Yong Yu. Local implicit feedback mining for music recommendation. In *RecSys*, pages 91–98, 2012.

[9] Hongzhi Yin, Bin Cui, Ling Chen, Zhiting Hu, and Zi Huang. A temporal context-aware model for user behavior modeling in social media systems. In *SIGMOD Conference*, pages 1543–1554, 2014.

[10] Zhengyu Deng, Ming Yan, Jitao Sang, and Changsheng Xu. Twitter is faster: Personalized time-aware video recommendation from twitter to youtube. *ACM Trans. Multimedia Comput. Commun. Appl.*, 2014.

[11] Hongyun Bao, Qiudan Li, Stephen Shaoyi Liao, Shuangyong Song, and Heng Gao. A new temporal and social pmf-based method to predict users' interests in micro-blogging. *Decision Support Systems*, 55(3):698–709, 2013.

[12] Zhiheng Xu, Yang Zhang, Yao Wu, and Qing Yang. Modeling user posting behavior on social media. In *SIGIR*, pages 545–554, 2012.

[13] Jitao Sang, Changsheng Xu, and Jing Liu. User-aware image tag refinement via ternary semantic analysis. *TMM*, 14(3):883–895, 2012.

[14] Andrew Zimdars, David Maxwell Chickering, and Christopher Meek. Using temporal data for making recommendations. In *UAI*, pages 580–588, 2001.

[15] Yehuda Koren. Collaborative filtering with temporal dynamics. In *KDD*, pages 447–456, 2009.

[16] Liang Xiong, Xi Chen, Tzu-Kuo Huang, Jeff G. Schneider, and Jaime G. Carbonell. Temporal collaborative filtering with bayesian probabilistic tensor factorization. In *SDM*, pages 211–222, 2010.

[17] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. In *WebSci*, page 2, 2011.

[18] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *RecSys*, pages 135–142, 2010.

[19] Yi Ding and Xue Li. Time weight collaborative filtering. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 485–492, 2005.

[20] Eugene Santos Jr and Hien Nguyen. Modeling users for adaptive information retrieval by capturing user intent. *Collaborative and Social Information Retrieval and Access: Techniques for Improved User Modeling. IGI Global*, pages 88–118, 2009.

[21] Elke Michlmayr and Steve Cayzer. Learning user profiles from tagging data and leveraging them for personal (ized) information access. In *Proceedings of the Workshop on Tagging and Metadata for Social Information Organization, 16th International World Wide Web Conference (WWW2007)*, pages 1–7, 2007.

[22] Nathan N Liu, Min Zhao, Evan Xiang, and Qiang Yang. Online evolutionary collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 95–102, 2010.

[23] Ceren Budak, Anitha Kannan, and Rakesh Agrawal Jan Pedersen. Inferring user interests from microblogs. 2014.

[24] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[25] Jitao Sang and Changsheng Xu. Right buddy makes the difference: An early exploration of social relation analysis in multimedia applications. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 19–28, 2012.

[26] Michael Paul and Roxana Girju. Cross-cultural analysis of blogs and forums with mixed-collection topic models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, pages 1408–1417, 2009.

[27] Liangjie Hong, Byron Dom, Siva Gurumurthy, and Kostas Tsioutsiouliklis. A time-dependent topic model for multiple text streams. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 832–840, 2011.

[28] Hanna M. Wallach, David M. Mimno, and Andrew McCallum. Rethinking LDA: why priors matter. In *Advances in Neural Information Processing Systems*, pages 1973–1981, 2009.

[29] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. Comparing twitter and traditional media using topic models. In *Advances in Information Retrieval*, pages 338–349. 2011.

[30] Thomas L Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National academy of Sciences of the United States of America*, 101(Suppl 1):5228–5235, 2004.

[31] Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. The author-topic model for authors and documents. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 487–494, 2004.