A Sparse Topic Model for Bursty Topic Discovery in Social Networks

Lei Shi, Junping Du, and Feifei Kou
Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing

Abstract: *Bursty topic discovery aims to automatically identify bursty events and continuously keep track of known events. The existing methods focus on the topic model. However, the sparsity of short text brings the challenge to the traditional topic models because the words are too few to learn from the original corpus. To tackle this problem, we propose a Sparse Topic Model (STM) for bursty topic discovery. First, we distinguish the modeling between the bursty topic and the common topic to detect the change of the words in time and discover the bursty words. Second, we introduce "Spike and Slab" prior to decouple the sparsity and smoothness of a distribution. The bursty words are leveraged to achieve automatic discovery of the bursty topics. Finally, to evaluate the effectiveness of our proposed algorithm, we collect Sina weibo dataset to conduct various experiments. Both qualitative and quantitative evaluations demonstrate that the proposed STM algorithm outperforms favorably against several state-of-the-art methods.*

Keywords: *Bursty topic discovery, topic model, "Spike and Slab" prior.*

Received August 15, 2017; accepted January 28, 2019
https://doi.org/10.34028/iajit/17/5/15

1. Introduction

Online social networks have become the most popular platform for people to establish online social relationship, share information, ranging from politics, economics, and entertainments. In China, the most popular social networks platform is Sina weibo. Recent statistics show Sina weibo (China's largest microblog platform) has more than 500 million registered users and maintains the more growth rate. These platforms have many times been the first publisher of significant bursty topics, such as natural disasters and violent terrorist incidents. If the bursty topics can be discovered from the social networks, which is conducive to guide public opinions and control network rumors. Therefore, this study has not only theoretical significance but also has abundant social, practical value [27].

However, the bursty topic discovery in social networks has the following challenges:

1. The contents are particularly short in social networks. How to extract high-quality topics from the short text is a much-watched challenge.
2. Social network topics are noisy and diverse, with many misleading information and meaningless topics.

Thus, it is necessary and challenging to distinguish the bursty topics from common contents.

In the previous study, the bursty topic discovery focused on leveraging Latent Dirichlet Allocation (LDA) to model text. However, the conventional topic models are designed to modellong text [2, 10], which are not directly applied for the bursty topic discovery in social networks. Although some extensive topic model [3, 16, 18, 20, 26, 31, 34, 39] can be used to alleviate topic sparsity in short texts, it is not effective for the bursty topic discovery. To overcome this problem, many researchers have proposed online topic model [9] and temporal topic model [4, 7, 23, 28]. Unfortunately, they rely on the post-processing steps for bursty topic discovery.

Another idea is to detect the bursty topic by clustering [13, 17, 24, 25, 29, 30]. These methods focus on detecting the bursty topics via the bursty words clustering. However, these methods still have sparsity and post-processing, since the bursty features are noisy and fragmentary [36]. Therefore, it is difficult to distinguish between the two similar topics take place at the same time.

In this paper, we propose a novel sparse topic model for bursty topic discovery, named Sparse Topic Model (STM), which apply the burstiness of word pair to discover the bursty topics, and then introduce "Spike and Slab" prior to decouple the smoothness and sparsity of a bursty topic distribution. According to the actual situation of the social networks, a topic is considered to be bursty in a time slice if it is widely shared and discussed in a time slice. But it has a little discussion at other times. The basic idea of STM is to exploit the burstiness of word pair as the prior knowledge incorporate into the topic model for the bursty topic modeling. Meanwhile, the "Spike and Slab" prior are leveraged to decouple the smoothness

and sparsity of a bursty topic distribution. It can not only implement bursty topic discovery without any post-processing but also overcome data sparsity of the short texts and topic scatter problem.

We have conducted extensive experiments over a Sina weibo dataset. The experimental results suggest that our propose STM obtains better results the state-of-the-art methods.

2. Related Work

At present, the typical way of topic detection focuses on the topic model and the incremental clustering [38].

In the topic-model-based, the traditional topic model is designed to detect the topic of news events and does not consider the short text. To overcome this problem, many researchers have improved the traditional topic model [9, 15, 16, 20, 32, 33]. Cheng *et al.* [3] proposed a word pair topic model, named BTM based on the mixture of unigrams, which effectively solves the sparseness problem of the short text topic in social networks. Zuo *et al.* [39] proposed a pseudo-document topic model, named PTM for the short text topic modeling. Yang *et al.* [34] leveraged the conceptual level of the N-level concept to capture the dependency of the words to detect the multi-document topic. In the topic-model-based, the bursty topic can also be detected by tackling a global optimization problem. Huang *et al.* [11] applied the local weighted linear regression to estimate the word novelty, which can highlight the word novelty of expressing a bursty topic.

In the clustering method, the documents are usually clustered according to the topic similarity of the corpus. The typical method is incremental clustering [1, 6, 24, 35, 37] and dictionary learning [5, 13, 17, 19]. Zhang *et al.* [37] utilized the term frequency and user's social relation to discover the bursty events from social networks and predict the popular events. Fang *et al.* [6] used multiview with the semantic relations, social tag relations, and temporal relations clustering to detect the topic. Becker *et al.* [1] proposed an incremental clustering method to detect emergency events in social networks. Other similar research applied the dictionary learning method [14] to discover the new topics. However, this method has a great dependence on the knowledge base and may omit some topics or events. Huang *et al.* [12] proposed a novel approach to detect and track the bursty topic. However, the above methods require complex heuristic adjustments and processing. Since the detected bursty characteristics are noisy and ambiguous, so it is not easy to cluster.

3. Model Introduction and Inference

We assume a bursty strong word pair, is more likely to be produced by a bursty topic; on the contrary, a burstiness weak word pair is more likely to be generated by a common topic. When the bursty event breaks out, the word pair may be observed more frequently than usual. For instance, the word pair such as "Kunming violent" and "Wenchuan earthquake" became much more frequent than usual in Sina weibo when these events took place. Such high-frequency word pair provides us crucial clues for bursty topic discovery.

Based on the above assumptions, we propose a sparse bursty topic model, which introduces the bursty term to guide bursty topic discovery, and adopts the weak smoothing prior based on "Spike and Slab" prior to decouple the sparsity and smoothness of a bursty topic. The bursty topic discovery is quantified by incorporating into the topic model. It is important to note that, our model models the generation of each word pair in a document set to learn the topic, unlike traditional topic model by modelling document generation. So, we assume that the two words in each word pair are generated independently from the same topic, and the topic is generated from a global topic distribution.

3.1. The "Spike and Slab" Prior

The "Spike and Slab" prior is a very effective established approach in Statistics and Mathematics, which is originally introduced by Wang and Blei [31] into the topic model to implement sparse topic-word distribution. It can decouple the distribution of sparse and smooth. Especially Bernoulli variables are introduced into the prior, which determine "on" or "off" status of switch variables. Therefore, the model can judge whether a corresponding variable appears or not. In our approach, the switch variable indicates whether a topic is focused in the dataset. Since the "Spike and Slab" prior can produce null selection, which will lead to the probability distribution to be ill-defined. To tackle this problem, Lin *et al.* [18] proposed a weak smoothing prior to avoid the ill-defined distribution by the direct application of the "Spike and Slab" prior. Therefore, we also apply the weak smoothing prior to avoid an ill-defined and simpler reasoning process, which can ensure the scalability of our model.

3.2. Model Formulation

Assume the word pair P occurs n_w^t times in a time slice T. Since a word pair may be identified either using normally or in some bursty topics, so we decompose a word pair n_w^t into two parts: $n_{w,0}^t$ is the number of the word pair P occurs in normal usage, while $n_{w,1}^t$ is the number of the word pair P occurs in bursty topic. Where $n_{w,0}^t + n_{w,1}^t = n_w^t$, Such $n_{w,0}^t$ almost is constant over time, while $n_{w,1}^t$ may continuously change at different time slices. When some bursty

topics related to the word pair break out, $n_{w,1}^t$ might sharply increase, while there is no bursty topic to generate in other time slices, and $n_{w,1}^t$ will be nearly 0. Therefore, we estimate $n_{w,0}^t$ by the mean value of n_w^t in the last *M* time slices $\bar{n}_w^t = \frac{1}{M}\sum_{M=1}^{M} n_w^{t-m}$. Then we can obtain $\hat{n}_{w,1}^t = \max\left[\left(n_w^t - \bar{n}_w^t\right), \tau\right]$ at the same time, where $n_{w,0}^t$ and $n_{w,1}^t$ cannot be observed, τ is a relatively small positive number to avoid the 0 value. We can apply the time and frequency to approximate the probabilistic of the word pair generated from a bursty topic in time slices t. The process is as Equation (1):

$$\mu_w^t = \frac{\max\left[\left(n_w^t - \bar{n}_w^t\right), \tau\right]}{n_w^t} \quad (1)$$

Where μ_w^t is the bursty probability of the word pair P in the time slice *T*. It suggests that the word pair P appears more frequently than in a time slice than other times, and it will be more likely to be generated from the bursty topics. Table 1 lists the key notations of our proposed STM model.

Table 1. Variables and notations.

Notation	Meaning
D	collection of short documents
N_P	number of word pair
K	number of topics
P	set of word pair
ϕ_0	background word distribution
θ	bursty topic distribution
b_z	topic selector
μ_w^t	bursty probability of word pair
z	topic assignment
α	bursty Topic smoothing prior
$\bar{\alpha}$	Weak topic smoothing prior
γ_0, γ_1	hyperparameter
π	binary variable
A_Z	set of its focused topics
$I[\cdot]$	Indicator function

- *Definition* 1: Corpus contains two types of topics: the bursty topic and the common topic, the content of a bursty topic increases rapidly in the current time slice, while the common topics are almost constant over time.
- *Definition* 2: Given the short text corpus $D = \{d_1, d_2, ..., d_{N_d}\}$, a topic selector b_z is a binary switch variable that indicates whether the topic is a focused topic. b_z is sampled from the Bernoulli distribution.
- *Definition* 3: The Smoothing Prior α is Dirichlet hyperparameter to smooth the topic which is selected by the topic selector, while the weak Smoothing Prior $\bar{\alpha}$ is another Dirichlet hyperparameter to smooth the topic which does not appear in the topic. Since $\bar{\alpha} << \alpha$, the hyperparameter $\bar{\alpha}$ is called weak smoothing prior.
- *Definition* 4: If the topic selector b_z =1, the topic is a focused topic. For the dataset $A_z = \{z : b_z = 1, z \in \{1, \cdots, K\}\}$, it is defined as the focus topic.

3.3. A Sparse Topic Model

Based on the above analysis, the word pair is generated by the topic. Therefore, the burstiness of the word pair relates directly with the burstiness of the topic, and we assume that a word pair is identified either normal usage or in some bursty topic. Our STM model learning the burstiness of the word pair to discover the bursty topics based on the above assumption. We define a binary switch variable π to represent the source of occurring a word pair. Where π=0 indicates the word pair is generated from the normal topic, while π=1 indicates the word pair is generated from the bursty topic. So, we apply the bursty probability of a word pair to encode the prior knowledge from the bursty topic and define a Bernoulli distribution with parameter μ_w^t as the prior distribution of π. Moreover, we introduce θ to denote the bursty topics distribution in the collection, and ϕ_k to denote the word distribution for the bursty topics in the collection. A normal word distribution ϕ_c denotes the normal usage. Then we apply Smoothing Prior and Weak Smoothing Prior to decouple the topic distribution of sparse and smooth. Given a short text data $D = \{d_1, d_2, ..., d_{N_d}\}$, the corresponding set of the word pair is $P = \{p_1, p_2, ..., p_{Np}\}$, where p_i=($w_{i,1}$,$w_{i,2}$). Figure 1 is the graphical representation of our STM.

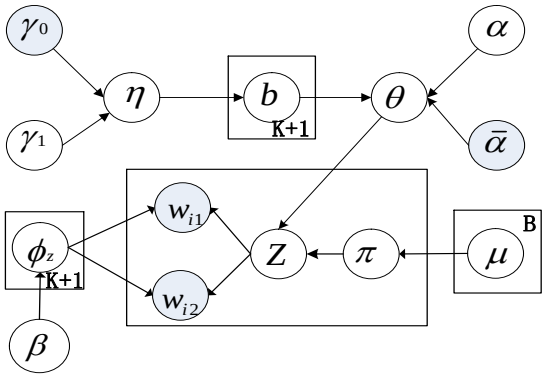

Figure 1. The graphical model of STM.

The generative process in the time slice, which is then defined as follows:

1. For the collectionsample $\pi \sim Beta(\gamma_0, \gamma_1)$ sample the topic selector $b_z \sim Bernoulli(\eta), \vec{b} = \{b_z\}_{k=0}^{K}$
   sample a bursty topic distribution $\theta \sim Dir(\alpha\vec{b} + \bar{\alpha}\vec{1})$

2. For each bursty topicsample a word distribution: $\phi_k \sim \mathrm{Dir}(\beta)$ sample a normal word distribution $\phi_c \sim \mathrm{Dir}(\beta)$
3. For each word pair $p_i \in P$ sample a binary switch $\pi \sim Bernoulli(\mu_w)$
   If π=0
   sample two words $w_{i,1}, w_{i,2} \sim Multi(\phi_c)$
   If π=1
   sample a bursty topic $z \sim Multi(\theta)$
   sample two words $w_{i,1}, w_{i,2} \sim Multi(\phi_z)$

3.4. Parameter Estimation

We employ the collapsed Gibbs sampling algorithm [8] to approximate to obtain samples of the latent variables and estimate the unknown parameters in STM, which is simple to derive, comparable in speed to other estimators and can approximate a global maximum. The key idea is to alternately estimate the random variables for posterior sampling, where each random variable is sampled based on the assignment of other random variables.

In STM, we sample a topic for each word pair. Integrating out θ, ϕ, and η analytically, the latent variables needed by the Gibbs sampling algorithm are switching variables π and the topic selector b_z. We also sample Dirichlet hyper-parameter α, β and hyper-parameter γ_1, and fix $\bar{\alpha}$ equal to 10^{-8} and γ_0 equal to 1. According to sampling the Equations (2) and (3):

$$P(\pi=0 \mid rest) \propto (1-\mu_i)\frac{\left(n_{0,w_{i,1}}^{-i}+\beta\right)\left(n_{0,w_{i,2}}^{-i}+\beta\right)}{\left(n_{0,\cdot}^{-i}+W\beta\right)\left(n_{0,\cdot}^{-i}+1+W\beta\right)} \quad (2)$$

$$P(\pi=1, z_i=k \mid rest) \propto \mu_i \frac{\left(n_k^{-i}+b_z\alpha+\bar{\alpha}\right)}{\left(n_{\cdot}^{-i}+|A_z|\alpha+K\bar{\alpha}\right)}\frac{\left(n_{k,w_{i,1}}^{-i}+\beta\right)\left(n_{k,w_{i,2}}^{-i}+\beta\right)}{\left(n_{k,\cdot}^{-i}+W\beta\right)\left(n_{k,\cdot}^{-i}+1+W\beta\right)} \quad (3)$$

Where $\pi=\{\pi_i\}_{i=0}^{N_P}$, $Z=\{z_i\}_{i=0}^{N_P}$, $\mu=\{\mu_i\}_{i=0}^{N_P}$, $n_{0,w}$ is the number of times that the word pair is assigned to the normal word distribution, $n_{0,\cdot}=\sum_{w=1}^{W} n_{0,w}$ is the total number of the words assigned to the normal word distribution, n_k is the number of the word pair assigned to the bursty topics, $A_z=\{z: b_z=1, z\in\{1,\cdots,K\}\}$ is the set of indices of $\vec{b}$ that is "on", $|A_z|$ is the size of A_z, $n_{\cdot}=\sum_{k=1}^{K} n_k$ is the total number of the word pair assigned to the bursty topics, α is topic smoothing prior, $\bar{\alpha}$ is weak topic smoothing prior, $n_{k,w}$ is the number of times that the word w is assigned to the bursty topic K, $n_{k,\cdot}=\sum_{w=1}^{W} n_{k,w}$ is the total number of the words assigned to the bursty topic k, and $\neg i$ means ignoring the word pair .

Sampling the topic selector b_z: For sampling, we leverage π as an auxiliary variable. Give the joint conditional distribution as the Equations (4):

$$P(\eta, \vec{b}_z \mid rest) \propto \prod_z P(b_z \mid \eta) P(\eta \mid \gamma_0, \gamma_1) \frac{I[B_l]\Gamma(|A_z|\alpha+K\bar{\alpha})}{\Gamma(n_{\cdot}+|A_z|\alpha+K\bar{\alpha})} \quad (4)$$

With the joint conditional distribution, we iteratively sample b_z condition on η and eventually obtain a sample for b_z. Then we integrate out π and sample b_z using the reverse method [18]. For hyper-parameter α, we apply Metropolis-Hastings with a symmetric Gaussian as the proposal distribution. For the concentration parameter γ_1, we apply previously developed approaches for Gamma priors [29], $I[\cdot]$ is an indicator function. $B_l=\{z: n_k>0, z\in\{1,\cdots,K\}\}$.

The Gibbs sampling procedure is shown in Algorithm 1. We randomly assign a topic to each word as the initial state. Then, we sample the latent variables according to Equations (2) and (3) in each iteration process. After enough iterations, we can estimate the parameters by the learned parameter mean. The distributions are obtained by the Equations (5) and (6):

$$\theta_k = \frac{n_k^{-i}+b_z\alpha+\bar{\alpha}}{n_{\cdot}^{-i}+|A_z|\alpha+K\bar{\alpha}} \quad (5)$$

$$\phi_{k,w} = \frac{n_{k,w}+\beta}{\left(n_{k,\cdot}+W\beta\right)} \quad (6)$$

Algorithm 1 Gibbs sampling algorithm for STM

Input: topic number K, α, $\bar{\alpha}$, β, γ_0, *word pair set P*
Output: ϕ_K *and* θ
Initialize topic assignments for each word pair randomly
for $iter=1$ *to* N_{iter} *do*
for each word pair do if π=0 *then*
sample π, b_Z *from Eqs.(2-4)*
if then
Update $n_{0,w_{i,1}}$, $n_{0,w_{i,2}}$
else
Update n_k, $n_{0,w_{i,1}}$, $n_{0,w_{i,1}}$, $|A_Z|$
end for
end for
Compute the ϕ_K *and* θ *by Eqs. (5-6)*

3.5. Extension and Discussion

In this paper, we emphasize "likely" because some burstiness strong words pair may still be generated by the common topics. Meanwhile, the burstiness weak word pair may still be generated by the bursty topic, but because it does not appear many times in the bursty

topic. In the above analysis, we ignore the randomness of the occurrences of words pair. Moreover, even in the common topic, the number of occurrences of a word pair will fluctuate in each time slices. This fluctuation has little effect on high-frequency words pair, but it will lead to the high μ_w^t value of low-frequency word pair. Because these low-frequency word pairs have low frequency and weak burstiness, the probability of generating bursty topic should be smaller. Therefore, we take the largest between $n_w^t - \bar{n}_w^t$ and τ.

4. Experiment

4.1. Dataset

We collect data from Sina weibo, which is the largest microblog platform in China. A total of about 2 million microblog data were collected from February 26, 2014 to March 15, 2014. Then

1. Removing the duplicate documents.
2. Chinese sentences are processed by Chinese word segmentation tools based on deep learning [14].
3. Removing the stop words.
4. Removing the number of occurrences less than 8.
5. Removing the documents with less than 3 words.

4.2. Basicline Method

- *OnlineLDA*: Online Latent Dirichlet Allocation (Online LDA) is a typical bursty topic discovery method based on the topic learning [15], which model the text by dividing the text stream into a set of textbooks with sequential relationships in successive time slices.[1]
- *Twevent*: Twevent [17] is the latest method of emergency detection based on feature clustering.[2]
- SATM: Self-Aggregation Topic Model (SATM) [26] aggregates the short texts into pseudo documents without the auxiliary information.[3]
- *BBTM*: A bursty topic discovery model [33] based on the Biterm Topic Model (BTM) model, it introduces binary switching variables to determine whether the topic is a bursty topic based on the burstiness of the word.[4]

4.3. Parameter Setting

In our experiments, the length of the time slice is set to 1 day, $\alpha = 0.1$, $\bar{\alpha} = 10^{-12}$, $\beta = 0.01$, $\gamma_0 = 0.1$ and the number of the bursty topic K varies from 10 to 50. The parameter settings for the other algorithms are based on the default parameters described in their paper.

4.4. The Accuracy of Bursty Topic Discovery

First, we evaluate the accuracy of the bursty topic discovery for each approach. Five volunteers are invited to manually label the discovered bursty topics as true or false by all these methods. Criteria for identifying the bursty topics: a topic is labeled true if the bursty topic presented is both meaningful and the bursty appears in the current slice but does not appear in the previous slice. Besides, if a topic contains the words that come from different topics or daily communication, it will be judged "false". A bursty topic is correctly detected if more than half of the volunteers label it "true". Finally, we evaluate the accuracy of the bursty topic based on the average precision at *K* (P@K) for different methods. Table 2 lists the accuracy of all the methods with different settings of the bursty topic number *K*.

Table 2. The accuracy of the bursty topics discovery.

	P@10	P@30	P@50
STM	0.751	0.812	0.783
BBTM	0.720	0.732	0.724
Twevent	0.711	0.725	0.689
OnlineLDA	0.228	0.213	0.186
SATM	0.563	0.478	0.449

From Table 2, we can see:

1. The accuracy of the proposed STM model is always greater than 0.75, which is significantly better than the baseline methods. It indicates that our STM can more accurately discover the bursty topic. Compared with the accuracy of all topic-model-based methods with different settings of the different bursty topic K, we also find that the proposed STM method is slightly less effective at $K = 10$, which is mainly because the number of the topics is too small, and it leads the topic to be more dispersed.
2. Bursty Biterm Topic Model (BBTM) achieves higher accuracy than the other three baseline methods, but compared to the proposed STM method, BBTM is relatively poor. It shows that our proposed model is helpful for the discovery of the bursty topic by leveraging "Spike and Slab" prior to decouple the sparsity and smoothness of a distribution.
3. OnlineLDA and SATM that based on the common topic model always perform the worst. This is because the common topic model fails to model the burstiness of the topic, and cannot effectively distinguish between the common topics and bursty topics.

[1] https://github.com/jhlau/online_twitter_lda
[2] https://github.com/KeithYue/Twevent
[3] https://github.com/WHUIR/SATM
[4] https://github.com/xiaohuiyan/BurstyBTM

4.5. Novelty of Bursty Topics Discovery

In social networks, the bursty topic is constantly changing. We introduce Novelty [33] to evaluate the sensitivity and novelty of different algorithms for discovering the bursty topics. We collect a more likely word from the topic Z to construct a set of keywords in each time slice, $W^{(t)}$ and $W^{(t-1)}$ is the word pair set of two adjacent time slices, the Novelty of the bursty topics is calculated by using Equation (7):

$$Novelty\left(Z^{(t)}\right)=\frac{\left|W^{(t)}\right|-\left|W^{(t)}\cap W^{(t-1)}\right|}{T*K} \tag{7}$$

Where $|\cdot|$ is the number of elements in the sets, and T is the number of words contained in each topic. In the experiment, only top10 (i.e., T = 10) terms of each topic are used for calculating Novelty.

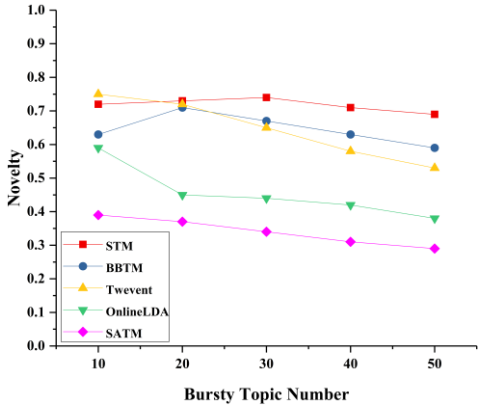

Figure 2. Novelty of the bursty topic discovery.

Figure 2 is comparison of the novelty with different settings of the bursty topic number K. From the results, we can observe that:

1. Our proposed STM always outperforms other baseline methods on Novelty, especially when the K is large. This is because the proposed STM model is more sensitive to the bursty topics by incorporating the burstiness of the word pair as prior and introducing "Spike and Slab" prior than the baseline methods.
2. Twevent obtains better performance when the K is small, since it detects the bursty topic only by the bursty word. However, the performance of Twevent decreases fasts with the increasing bursty topic number K. This is because more noisy topics are generated with increasing in the number of the bursty topics.
3. BBTM significantly outperforms Twevent. This is because the BBTM employs the word pair to model the bursty topic, and effectively improves the handling ability on the short texts and the discovered topics.

4.6. Topic coherence

We apply Pointwise Mutual Information Score (PMI-Score) topic coherence to evaluate the topic model [22]. The PMI-Score uses the point mutual information to evaluate the topic coherence [21]. Given the topic z, we choose the top-N possible words, w_1, w_2,…, and calculate the PMI scores for each word pair [18]. The higher the PMI, the more relevant the words. So, if the higher the PMI-score of a topic, the better the expandability of the topic, the *PMI* is calculated by using Equation (8):

$$PMI\left(z\right)=\frac{2}{N\left(N-1\right)}\sum_{1\le i\le j\le N}\log\frac{p\left(w_i,w_j\right)}{p\left(w_i\right)p\left(w_j\right)} \tag{8}$$

Where $p(w_i, w_j)$ is the joint probability distribution of the word pair w_i and w_j co-occurring in the same sliding window, $p(w_i)$ is the marginal probability of the word w_i appears in the sliding window within the edge probability distribution. We estimate the value of the relevant probability by Wikipedia. In our experiment, the value of N is set to10.

We calculate the average PMI of the top 10 words by using Chinese Wikipedia articles as an auxiliary corpus. Figure 3 is the results of the coherence with K varying from 10 to 50.

We can make the following conclusions:

1. Our proposed STM consistently outperforms other existing state-of-the-art topic models and indicates good coherence of the learned bursty topics from social networks.
2. BBTM also works better, but it performs poorer than ours STM. The major reason is that more focus bursty topics are generated by STM.
3. Twevent is always the lowest. This is due to the fact that Twevent simply clustering the bursty words might be the noisy topic and less topic coherent.

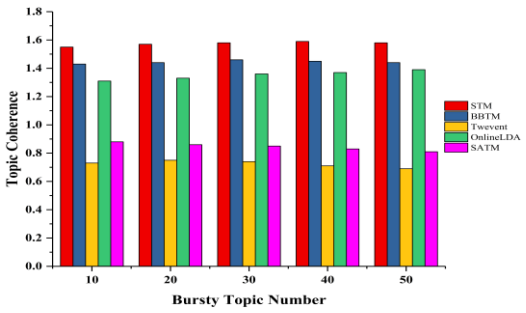

Figure 3. Coherence of the bursty topics discovered.

To further analyze the effectiveness of our proposed model, we will qualitatively analyze the bursty topic discovery. We first randomly select the hot bursty topics and high-frequency hashtags. The hashtags are "#KunMing Railway Station violent terrorist event #", which occurred in on March. 1, 2014. For each hashtag, extracting the microblogs that contain these hashtags, statistical word frequency, and

normalization. Then, for each method, we select the bursty topic closest to the empirical word distribution of the hashtag. Table 3 lists the top-10 words of the most similar topics with the hashtag, where the second line represents the hashtag corresponding to the topic content.

We can see that:

1. The word in STM is the most similar to the word distribution corresponding to the hashtag.
2. BBTM is also closer to the topic hashtag word distribution.
3. Twevent contains some irrelevant words. It indicates that the bursty word clustering is more sensitive to the noise.
4. The topics discovered by Online LDA contain many common words, and only part of the words is related to "# Kunming Railway Station violent terrorist event #", it shows that the similarity is the lowest.

This shows that the basic topic model cannot distinguish well between the bursty topic and common topic.

Table 3. The bursty topic discovered by each method mostly relates to "#昆明火车站暴恐案（KunMing Railway Station violent terrorist event）#" on March .1, 2014. The first column list the most frequent words in the Sina weibo with the hashtag "#昆明火车站暴恐案（KunMing Railway Station violent terrorist case）#".

Empirical	STM	BBTM	TWevent	OnlineLDA	SATM
昆明(kunminng)	火车站(railway station)	嫌疑人(suspect)	暴力 (violence)	暴力(violence)	火车站(railway station)
火车站(railway station)	遇难(victims)	火车站(railway station)	昆明(kunming)	危险(danger)	袭击(attack)
暴力(violence)	昆明(kunminng)	救治(treatment)	砍人(killing)	昆明(kunming)	新疆(xinjiang)
恐怖(terror)	暴力(violence)	警察(police)	袭击(attack)	情况(situation)	手机(mobile phone)
袭击(attack)	嫌疑人(suspect)	嫌疑犯(suspect)	进站口(Entrance)	救护车（ambulance）	乘客(passenger)
遇难(victims)	救护车(ambulance)	新疆(xinjiang)	恐怖(terror)	乘务员(attendant)	旅游(tourism)
现场(scene)	死亡(death)	遇难(victims)	购物(shopping)	警察(police)	现场(scene)
嫌疑人(suspect)	救治(treatment)	祈祷(pray)	美食(delicious food)	百货大楼(department store)	景点(tourist attractions)
打击(combat)	紧急(emergency)	亲人(relatives)	祈祷(pray)	新疆(xinjiang)	遇难(victims)
救治(treatment)	砍人(killing)	进站口(Entrance)	云南(yunnan)	晚点(late)	事件(event)

5. Conclusions

In this paper, we propose a sparse topic model to discover the bursty topic in social networks.

The key idea is to exploit the burstiness of the word pair as the prior knowledge incorporate into the topic model for the bursty topic model, then the "Spike and Slab" prior is used to decouple the smoothness and sparsity of a bursty topic distribution, which aims to model the bursty topics, the common topics and eliminate the irrelevant topics. STM uses the frequency of the words as a prior to guide the discovery of the bursty topic. Our approach can not only overcome the data sparsity of the short texts in social networks but also can effectively discover the bursty topic.

Extensive experiments on the real-world data sets demonstrate that the significant superiority of STM to some state-of-the-art methods. However, social networks including unstructured metadata in multiple modalities and our STM cannot model the multi-modal property of the social topic. In our future work, we will focus on introducing the social visual modality to achieve the discovery of the bursty topic based on the cross-media topic model.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (No. 61902037, No.61532006, No. No.61772083) and Science and Technology Major Project of Guangxi (GuikeAA18118054).

References

[1] Becker H., Naaman M., and Gravano L., "Beyond Trending Topics: Real-World Event Identification on Twitter," *in Proceedings of the 5th International Conference on Weblogs and Social Media*, Barcelona, pp. 438-441, 2011.

[2] Blei D., Ng A., and Jordan M I., "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.

[3] Cheng X., Yan X., Guo J, and Lan Y., "BTM: Topic Modeling over Short Texts," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 2928-2941, 2014.

[4] Diao Q., Jiang J., LIM E, and Zhu F., "Finding Bursty Topics from Microblogs," *in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Association for Computational Linguistics*, Jeju Island, pp. 536-544, 2012.

[5] Dong X., Mavroeidis D., Calabrese F, and Frossard P., "Multiscale Event Detection in Social Media," *Data Mining and Knowledge Discovery*, vol. 29, no. 5, pp. 1374-1405, 2015.

[6] Fang Y., Zhang H., Ye Y., and Li X., "Detecting Hot Topics from Twitter: A Multiview Approach," *Journal of Information Science*, vol. 40, no. 5, pp. 578-593, 2014.

[7] Guille A. and Favre C., "Event Detection, Tracking, and Visualization in Twitter: A Mention-Anomaly-Based Approach," *Social*

Network Analysis and Mining, vol. 5, no. 1, pp. 1-18, 2015.

[8] Griffiths T. and Steyvers M., "Finding Scientific Topics," *in Proceedings of the National academy of Sciences*, pp. 5228-5235, 2004.

[9] Hoffman M., Bach F., and Blei D., "Online Learning for Latent Dirichlet Allocation," *in Processing of Advances in Neural Information Processing Systems*, Vancouver, pp. 856-864, 2010.

[10] Hofmann T., "Probabilistic Latent Semantic Indexing," *in Proceedings of the 22nd annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, pp. 50-57, 1999.

[11] Huang J., Peng M., Wang H., Cao J., and Gao W., Zhang X., "A Probabilistic Method for Emerging Topic Tracking in Microblog Stream," *World Wide Web-internet and Web Information Systems*, vol. 20, no. 2, pp. 325-350, 2017.

[12] Huang W., Wang T., Chen W., and Wang Z., "Category-Level Transfer Learning From Knowledge Base to Microblog Stream for Accurate Event Detection," *in Proceedings of International Conference on Database Systems for Advanced Applications*, Suzhou, pp. 50-67, 2017.

[13] Kasiviswanathan S., Melville P., Banerjee A., and Sindhwani V., "Emerging Topic Detection Using Dictionary Learning," *in Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, Glasgow, pp. 745-754, 2011.

[14] Lample G., Ballesteros M., Subramanian S., Kawakami., and Dyer C., "Neural Architectures for Named Entity Recognition," *in Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, pp. 260-270, 2016.

[15] Lau J., Collier N., and Baldwin T., "On-line Trend Analysis with Topic Models: #twitter Trends Detection Topic Model Online," *in Proceedings of COLING*, Mumbai, pp. 1519-1534, 2012.

[16] Lin T., Zhang S., and Cheng H., "Understanding Sparse Topical Structure of Short Text via Stochastic Variational-Gibbs Inference," *in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, Indianapolis, pp. 407-416, 2016.

[17] Li C., Sun A., and Datta A., "Twevent: Segment-Based Event Detection from Tweets," *in Proceedings of ACM International Conference on Information and Knowledge Management*, Maui, pp. 155-164, 2012.

[18] Lin T., Tian W., Mei Q., and Cheng H., "The Dual-Sparse Topic Model: Mining Focused Topics and Focused Terms in Short Text," *in Proceedings of International Conference on World Wide Web*, Seoul, pp. 539-550, 2014.

[19] Mcminn A. and Jose J., "Real-Time Entity-Based Event Detection for Twitter," *in Proceedings of International Conference of the Cross-Language Evaluation Forum for European Languages*, Toulouse, pp. 65-77, 2015.

[20] Mehrotra R., Sanner S., Buntine W., and Xie L., "Improving Lda Topic Models for Microblogs Via Tweet Pooling and Automatic Labeling," *in Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Ireland, pp. 889-892, 2013.

[21] Mimno D., Wallach H., Talley E., Leenders M., and McCallum A., "Optimizing Semantic Coherence in Topic Models," *in Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, pp. 262-272, 2011.

[22] Newman D., Lau J., Grieser K., and Baldwin T., "Automatic Evaluation of Topic Coherence," *in Proceedings of Human Language Technologies: Annual Conference of the North American Chapter of the Association for Computational Linguistics. Association for Computational Linguistics*, Boulder, pp. 100-108, 2010.

[23] Parikh R. and Karlapalem K., "Et: Events From Tweets," *in Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, pp. 613-620, 2013.

[24] Petrovic S., Osborne M., and Lavrenko V., "Streaming First Story Detection with Application to Twitter," *in Proceedings of Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings*, Los Angeles, pp. 181-189, 2010.

[25] Petrovi., Osborne M., and Lavrenko V., "Using Paraphrases for Improving First Story Detection in News and Twitter," *in Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics*, Montreal, pp. 338-346, 2012.

[26] Quan X., Kit C., Ge Y., and Pan S., "Short and Sparse Text Topic Modeling via Self-Aggregation," *in Proceedings of the 24th International Joint Conference on Artificial Intelligence*, Buenos Aires, pp. 2270-2276, 2015.

[27] Selvam S., Balakrishnan R., and Ramakrishnan B., "Social Event Detection-A Systematic Approach Using Ontology and Linked Open Data with Significance to Semantic Links," *The International Arab Journal of Information Technology*, vol. 15, no. 4, pp. 729-738, 2018.

[28] Stilo G. and Velardi P., "Efficient Temporal Mining of Micro-Blog Texts and Its Application to Event Discovery," *Data Mining and*

Knowledge Discovery, vol. 30, no. 2, pp. 372-402, 2016.

[29] Teh Y., Jordan M., Beal M., and Blei D., "Sharing Clusters Among Related Groups: Hierarchical Dirichlet Processes," *in Proceedings of Advances in Neural Information Processing Systems 17 Neural Information Processing Systems*, Vancouver, pp. 1385-1392, 2004.

[30] Wang Y., Liu J., Huang Y., and Feng V., "Using Hashtag Graph-Based Topic Model To Connect Semantically-Related Words Without Co-Occurrence In Microblogs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 7, pp. 1919-1933, 2016.

[31] Wang C. and Blei D., "Decoupling Sparsity and Smoothness in the Discrete Hierarchical Dirichlet Process," *in Processing of Advances in Neural Information Processing Systems*, Vancouver, pp. 1982-1989, 2009.

[32] Xie W., Zhu F., Jiang J., Lim E., and Wang K., "Topicsketch: Real-Time Bursty Topic Detection From Twitter," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 8, pp. 2216-2229, 2016.

[33] Yan X., Guo J., Lan Y., Xu J., and Cheng X., "A Probabilistic Model for Bursty Topic Discovery in Microblogs," *in Proceedings of the 29th AAAI Conference on Artificial Intelligence*, Austin, pp. 353-359, 2015.

[34] Yang G., Wen D., Chen N., Sutinen N., and Kinshuk., "A Novel Contextual Topic Model for Multi-Document Summarization," *Expert Systems with Applications*, vol. 42, no.3, pp. 1340-1352, 2015.

[35] Yin J. and Wang J., "A Dirichlet Multinomial Mixture Model-Based Approach for Short Text Clustering," *in Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, pp. 233-242, 2014.

[36] Zarrinkalam F. and Bagheri E., "Event Identification in Social Networks," *Encyclopedia with Semantic Computing*, vol. 1, no.1, pp. 1-8, 2016.

[37] Zhang, X., and Chen X., Chen Y., Wang S., and Xia J., "Event Detection and Popularity Prediction in Microblogging," *Neurocomputing*, vol. 149, no. 2, pp. 1469-1480, 2015.

[38] Zhou X. and Chen L., "Event Detection over Twitter Social Media Streams," *The Very Large Data Bases journal*, vol. 23, no.3, pp. 381-400, 2014.

[39] Zuo Y., Wu J., Zhang H., Lin H., Wang F., Xu K., and Xiong H., "Topic Modeling of Short Texts: A Pseudo-Document View," *in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, pp. 2105-2114, 2016.

Lei Shi was born in 1986. He received the M.S. degree in Control Engineering from Inner Mongolia University of Science and Technology. He is now a Ph.D. candidate in Computer Science and Technology of Beijing University of Posts and Telecommunications. His research interests include social network search, data mining and cross-media search

Junping Du was born in 1963. She is now a professor and Ph.D. tutor at the School of Computer Science and Technology, Beijing University of Posts and Telecommunications. Her research interests include artificial intelligence, image processing and pattern recognition.

Feifei Kou was born in 1989. She received her M.S. degree in Computer technology from Beijing Technology and Business University. She is now a Ph.D. candidate in Computer Science and Technology of Beijing University of Posts and Telecommunications. Her research interests include social network search, semantic analysis and semantic learning.