ECAI 2016
G.A. Kaminka et al. (Eds.)

doi:10.3233/978-1-61499-672-9-338

A Joint Model for Sentiment-Aware Topic Detection on Social Media

Kang Xu and Guilin Qi and Junheng Huang and Tianxing Wu [1]

Abstract. Joint sentiment/topic models are widely applied in detecting sentiment-aware topics on the lengthy review data and they are achieved with Latent Dirichlet Allocation (LDA) based model. Nowadays plenty of user-generated posts, e.g., tweets and E-commerce short reviews, are published on the social media and the posts imply the public's sentiments (i.e., positive and negative) towards various topics. However, the existing sentiment/topic models are not applicable to detect sentiment-aware topics on the posts, i.e., short texts, because applying the models to the short texts directly will suffer from the context sparsity problem. In this paper, we propose a Time-User Sentiment/Topic Latent Dirichlet Allocation (TUS-LDA) which aggregates posts in the same timeslice or user as a pseudo-document to alleviate the context sparsity problem. Moreover, we design approaches for parameter inference and incorporating prior knowledge into TUS-LDA. Experiments on the Sentiment140 and tweets of electronic products from Twitter7 show that TUS-LDA outperforms previous models in the tasks of sentiment classification and sentiment-aware topic extraction. Finally, we visualize the sentiment-aware topics discovered by TUS-LDA.

1 Introduction

With the rapid growth of Web 2.0, a mass of user-generated posts, e.g., tweets and E-commerce short reviews, which capture people's interests, thoughts, sentiments and actions. The posts have been accumulating on the social media with each passing day. Sentiment analysis attempts to find user preference, likes and dislikes from the posts on social media, such as reviews, blogs and microblogs [21] and topic modeling attempts to discover the topics or aspects from from reviews, blogs and microblogs etc [3]. Topic modeling and sentiment analysis on the posts are two significant tasks which can benefit many people. For example, we can discover a topic about "Apple Inc." and the overall sentiment of the topic. The sentiment of the topic about "Apple Inc." is implicitly associated with the stock trading of "Apple Inc.", because negative sentiments towards the company on social media can fall sales and financial gains but positive sentiments can improve sales [2]. Topic modeling [1] focuses on extracting word-level or document-level topics, while sentiment analysis [23] is to analyze the sentiments of words or documents.

Topic modeling and sentiment analysis on the social media are complementary where sentiments on the social media often change over different topics and topics on the social media are always related to public sentiments. So jointly modeling topics and sentiments on the social media is a feasible and significative task and it can reflect people's sentiment on different topics. However, unlike the normal documents (e.g., news and long reviews), the short and informal characteristic of the posts, e.g., tweets and short reviews, on the social media makes the tasks of topic modeling and sentiment analysis more challenging.

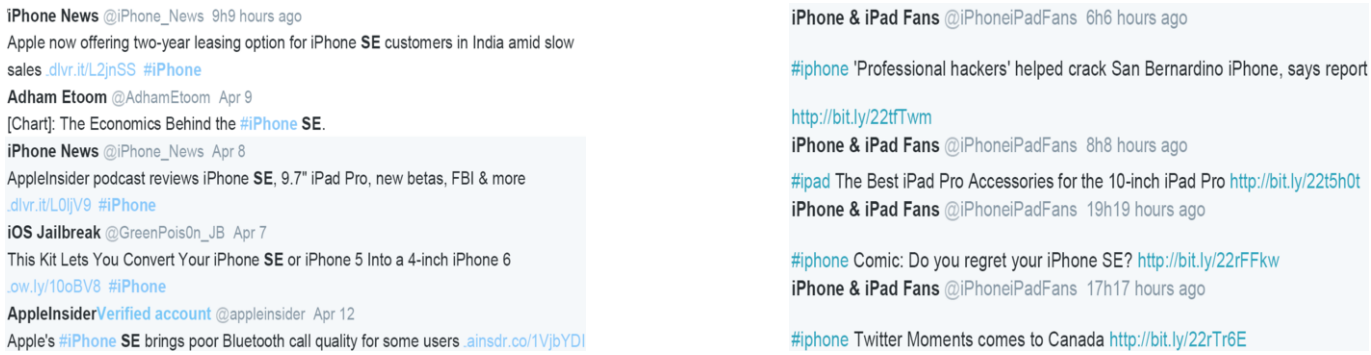
By jointly modeling topics and sentiments on social media, we want to obtain sentiment-aware topics from the posts, e.g., a topic about "Apple Inc." ('ipad', 'iphone', 'itouch', 'imac', 'beautiful' and 'popular') with the overall sentiment polarity "positive". Topic models, e.g., LDA [1] and pLSA [10], originally focus on mining topics from texts, but the models can also be extended to extract an extra aspect of texts, i.e., sentiment. Conventional sentiment-aware topic models, like Joint Sentiment/Topic Model (JST) [15] and Aspect/Sentiment Unification Model (ASUM) [11], are utilized for uncovering the hidden topics and sentiments from text corpus where each document is a mixture of sentiment/topics and each sentiment/topic is a mixture of words. Thereinto, each sentiment label in the models is viewed as a special kind of topic where topics are unknown and data-driven but sentiments are known and specified. However, for the short and informal characteristic of the posts, applying the models to the short posts on the social media directly always suffers from the context sparsity problem. So the models fail to recognize the accurate sentiments and senses of words in the posts.

One simple and effective way to alleviate the sparsity problem is to aggregate short posts into lengthy pseudo-documents [5, 31]. Here we assume that the posts on the social media are a mixture of two kinds of topics: temporal topics which are related to current events (e.g., tweets about a topic "Announcement of iphone SE" in Fig 1(a) which are produced in a timeslice) and stable topics which are related to personal interests (e.g., tweets about a topic "Apple products" in Fig 1(b) which are produced by a user). Thereinto, temporal topics are sensitive to time. If posts belong to temporal topics, we aggregate the posts in the same timeslice as a single document. We assume each timeslice is a mixture of sentiment-aware topics, i.e., each sentiment in the timeslice corresponds to several topics. Similar to temporal topics, stable topics are related to specific users and each user is a mixture of sentiment-aware topics. If a post belongs to a temporal topic, the post is assigned to a sentiment-aware topic in its publishing timeslice; otherwise, it is assigned to a sentiment-aware topic in its publishing user.

Moreover, based on the analysis of the characteristics of topics and sentiments, we exploit the important observation of topics: A single post always talks about a single topic [31]. Although a post usually talks about a single topic, a post may talk about multiple aspects of the topic with different sentiment polarities [12, 18].

For example, while the following short review of cannon camera from Amazon.com expresses the overall sentiment polarity of *Camera*, which corresponds to the part in italics, as positive, it addi-

[1] Southeast University, Nanjing, China
Email: {kxu,gqi,jhhuang,wutianxing}@seu.edu.cn



(a) Announcement of iPhone SE

(b) Apple products

Figure 1. (a) A temporal topic (b) A stable topic

tionally expresses a negative opinions towards the camera's lenses which corresponds to the part in bold.

Camera is great, but lenses are crap and cheap and don't work on auto focus. Buy body and lenses separately.

For a tweet, it can express a positive, a negative or neutral sentiment, and it can also express both positive and negative sentiments[24].

So, for sentiment polarities, we exploit the observation that words in a single post may correspond to multiple sentiment polarities [12, 18]. A post can talk about the same topic with different sentiments. For better modeling topics and sentiments respectively, we follow the assumption that words in the same post shouldbelong to the same topic, but they can have different sentiments.

Moreover, we add a sentiment label for each post. The sentiment label represents the overall sentiment polarities of the post and is determined by the sentiment polarities of words in the post. If words of a post express both positive and negative sentiments, the overall sentiment polarities of the post should be judged as the stronger one [24]. The sentiment label is utilized to model the association between sentiments and topics.

In this paper, we propose a novel Time-User Sentiment/Topic Latent Dirichlet Allocation (TUS-LDA) to mine sentiment-aware topics from the user-generated posts on social media.

There exist four main contributions of TUS-LDA:

1) TUS-LDA aggregates posts in the same timeslice or user as a single document to alleviate the context sparsity problem.

2) We design different ways to model topics and sentiments based on the characteristics of topics and sentiments. Thereinto, the sentiments of a post and the words in the post are all drawn from document-level sentiment distribution. Within the chosen sentiment of the post, the topic of the post is drawn from a user-level or timeslice-level sentiment/topic distribution.

3) We design approaches of parameter inference and incorporating prior sentiment knowledge for TUS-LDA.

4) We implement experiments on two datasets to evaluate the effectiveness of sentiment classification and topic extraction in TUS-LDA and visualize sentiment-aware topics discovered by TUS-LDA.

The rest of the paper is organized as follows: in Section 2, we introduce the related work about topic models on short texts and joint sentiment/topic models; in Section 3, we give the definitions of the basic terminologies we will use in our paper; in Section 4 we present our proposed model Time-User Sentiment/Topic Latent Dirichlet Allocation (TUS-LDA); Experimental settings and results are shown in Section 5. Finally, in Section 6, we conclude this paper and lists the future work.

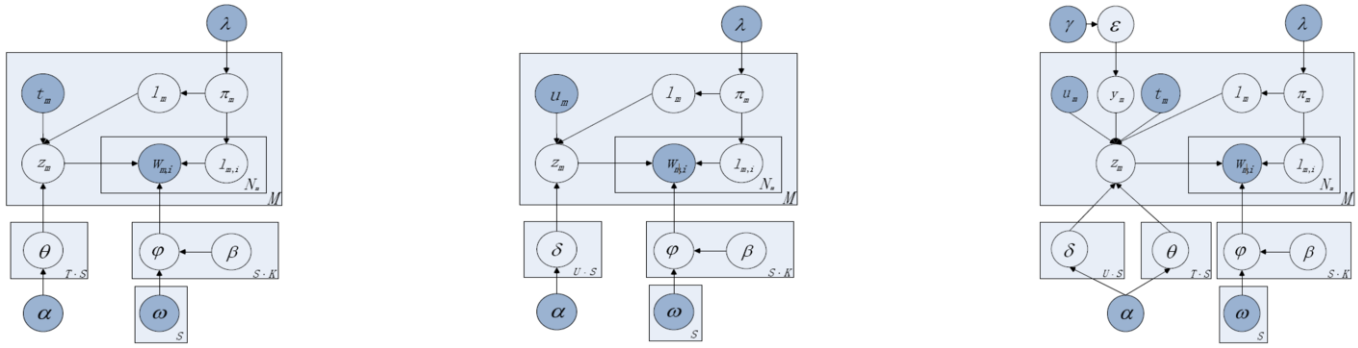
2 Related Work

2.1 Topic Models on Short Texts

LDA [1] and PLSA [10] originally focus on mining topics from lengthy documents. Recently topic modeling in the posts on social media is popular, however, it also suffers from the context sparsity problem of the posts. To overcome the sparsity problem of posts on the social media, there exist some work of aggregating posts into pseudo-documents. In [31], Twitter-LDA aggregated posts published by a user into one lengthy pseudo-document and made words in the same post belong to the same topic. In [5], posts in TimeUserLDA were aggregated by timeslices or users for finding bursty topics where posts belong to two kinds of topics: personal topics and temporal topics. Similar to TimeUserLDA, posts in TUK-TTM [29] were also aggregated by timeslices or users and TUK-TTM was utilized for time-aware personalized hashtag recommendation. Although these models can alleviate the problem of the context sparsity of posts on social media, they did not model an extra aspect of posts, i.e., sentiment.

2.2 Joint Sentiment/Topic Models

Recently, some topic models have been extended to model topics and sentiments jointly. The first work of topic and sentiment modeling is Topic-Sentiment Mixture model TSM [19]. In TSM, a sentiment is a special kind of topic and each word is generated from either a sentiment or a topic. The relation between sentiments and topics cannot be mined by TSM. At the same time, TSM is based on PLSA and suffers from the problems of inferencing on new documents and overfitting the data. To overcome these shortcomings, **Joint Sentiment-Topic model (JST)** [15] which is a two-level sentiment-topic model based on Latent Dirichlet Allocation (LDA) was proposed. In JST, sentiment labels are associated with documents, under which topics are associated with sentiment labels and words are associated with both sentiment labels and topics. Reverse-JST (RJST) [16] is a variant of JST where the position of sentiment and topic layer is swapped. In JST, topics were generated conditioned on a sentiment polarity, while in RJST sentiments were generated conditioned on a topic. **Aspect/Sentiment Unification Model (ASUM)** [11] is similar to JST. In ASUM, words in the same sentence belong to the same sentiment and topic. Sentiment Topic Model with Decomposed Prior (STDP) [32] is anthor variant of JST. STDP first determined whether the word is used as a sentiment word or ordinary topic words and then chose the accurate sentiments for sentiment words. Time-aware



(a) TS-LDA (b) US-LDA (c) TUS-LDA

Figure 2. The graphical representation of the proposed model (TS-LDA (a), US-LDA (b), TUS-LDA (c)). Shaded circles are observations or constants. Unshaded ones are hidden variables.

Topic-Sentiment Model (TTS) [4] extracted the hidden topics from texts and modeled the association between topics and sentiments and tracked the strength of topic-sentiment association over time. In TTS, time is viewed as a special word to bias the topic-sentiment distributions. But in our model, we use time to aggregate short texts and generate pseudo documents for modeling topics and sentiments. JST, RJST, ASUM, STDP and TTS are designed for normal texts where each piece of text has rich context to infer topics and sentiments, but our work models posts (i.e., short and informal texts) on social media and all of these models lose efficacy in the short and informal texts. MaxEnt-LDA [30] jointly discovers both aspects and aspect-specific opinion words by integrating a supervised maximum entropy algorithm to separate opinion words from objective ones. However, it does not further discover aspect-aware sentiment polarities of opinion words, which are very useful for sentiment analysis.

In our model, we focus on short and informal texts on social media. There exists some work about LDA-based sentiment analysis on social media. Twitter Opinion Topic Model (TOTM) [14] aggregated or summarized opinions of a product from tweets, which can discover target specific opinion words and improve opinion prediction. Topic Sentiment Latent Dirichlet Allocation (TSLDA) [22] utilized sentiments on social media for predicting stock price movement. TSLDA distinguished topic words and opinion words where topic words were drawn from the topic-word distribution and opinion words were drawn from the sentiment-topic-word distribution. Although these two work focuses on posts on social media, they do not consider and solve the context sparsity problem of posts.

3 Problem Definition

In this section, we define the basic terminologies we will use in this paper.

- **Post**: A post contains a sequence of words which express the opinions and thoughts of people towards different things (e.g., a tweet or a review).
- **User**: Each user-generated post has a user identification that specifies who publishes the post.
- **Timeslice**: Each user-generated post has a timeslice that specifies when the user publishes the post, in this paper, the length of timeslice is a day.
- **Topic**: A topic is a discrete piece of content that is about a specific subject, has an identifiable purpose (e.g., an event, a current hot problem and a product). Here, a topic is represented as a list of words.
- **Aspect**: An aspect refers to a distinct ratable facet of an entity. For a product, an aspect is an attribute or a component of the product that has been commented on in a review, e.g., "screen" for a digital camera. For an event or other kinds of topics, an aspect can be participants of the topic [25], e.g., "Obama" in the event of "Obama's visit to cuba".
- **Sentiment**: Sentiment is a label which refers to the polarity in which a concept or opinion is interpreted [17], i.e., "positive" and "negative". For example, "positive" is a sentiment for the post "Tom was glad to visit his friends.".
- **Sentiment-aware topic**: A sentiment-aware topic is a topic labeled with a sentiment polarity. For example, the overall sentiment of the topic "Obama's visit to cuba" is positive, so the topic "Obama's visit to cuba" is a positive topic.

4 The Proposed Models

In this section, we firstly introduce the notation and formally formulate our problem. Then, we describe the method utilized for learning parameters. Finally, we present the method of incorporating prior knowledge into our model.

4.1 The Generation Process

It is assumed that there exists a stream of M posts, denoted as $d_1, d_2, ..., d_M$. Each post d_m is generated by a user u_m within a timeslice t_m and the post d_m contains a bag of words, denoted as $\{w_{m,1}, w_{m,2}, ..., w_{m,N_m}\}$.

In LDA, a document is viewed as a multinomial distribution over topics and a topic is a multinomial distribution over words. In JST, each document is associated with the sentiment/topic distribution, i.e., each sentiment in the document has a topic distribution; the document also has a sentiment distribution for document-level sentiment-classification and a sentiment/topic is a multinomial distribution over words. LDA and JST only work well for lengthy documents, because the lengthy document have rich contexts. Based on the analysis of posts on the social media, words in the same post tend to be about a single topic [31]. However, the sentiment polarities of words in the same post can be different [12]. At the same time, to model the association between sentiments and topics, we also add a sentiment label for each post which is determined by the overall sentiment of all the words in the post.

On social media, a part of posts talks about stable topics which are related to users' personal interests with certain sentiments, so we

introduce a global sentiment/topic distribution δ for each user to capture personal long-term topical interests and sentiment preferences. Another part of posts is about temporal topics which are related to current events with the corresponding sentiments, so we add a time-dependent sentiment/topic distribution θ for each timeslice to capture temporal topics and the sentiments towards the topics.

Here, we construct the generative process of all the posts in the stream. When a user u_m publishes a post d_m within a timeslice t_m, the user first utilizes the variable y_m , which is drawn from the global user-timeslice switch distribution ε, to decide whether the post talks about a stable topic or a temporal topic. Then the user chooses a sentiment label l_m for the post from the document-sentiment π_m. If the user chooses a stable topic u_m and a sentiment label l_m, the user then selects a topic z_m from δ_{u_m,l_m}; otherwise, the user selects a topic z_m from θ_{t_m,l_m}. For each word $w_{m,i}$ in the post d_m, the user first chooses a sentiment label $l_{m,i}$; with the chosen topic z_m and sentiment label $l_{m,i}$, the word is drawn from the sentiment-topic word distribution $\varphi_{l_{m,i},z_m}$.

The notations in this paper are summarized in Table 1. Fig 2(c) shows the graphical representation of the generation process. Formally, the generative story for each post is as follows:

1. Draw $\varepsilon \sim Beta(\gamma)$
2. For each timeslice $t = 1, ..., T$
   i. For each sentiment label $s = 0, 1, 2$
      a. Draw $\theta_{t,s} \sim \text{Dir}(\alpha)$
3. For each user $u = 1, ..., U$
   i. For each sentiment label $s = 0, 1, 2$
      a. Draw $\delta_{u,s} \sim \text{Dir}(\alpha)$
4. For each sentiment label $s = 0, 1, 2$
   i. For each topic $k = 1, ..., K$
      a. Draw $\varphi_{s,k} \sim \text{Dir}(\beta)$
5. For each post d_m, $m = 1, ..., M$
   i. Draw $\pi_m \sim \text{Dir}(\lambda)$
   ii. Draw $l_m \sim \text{Multi}(\pi_m)$
   iii. Draw $y_m \sim Bernoulli(\varepsilon)$
   iv. if y_m=0, Draw $z_m \sim \text{Multi}(\theta_{u_m,l_m})$ or if y_m=1, Draw $z_m \sim \text{Multi}(\delta_{t_m,l_m})$
   v. For each word w $i = 1, ...N_m$
      a. Draw $l_{m,i} \sim \text{Multi}(\pi_m)$
      b. Draw $w_{m,i} \sim \text{Multi}(\varphi_{z_m,l_{m,i}})$

There are two degenerate variations of our model which are shown in the experiments. The first one is depicted in Fig 2(a), which considers the temporal topic-sentiment distribution. The second one is depicted in Fig 2(b), which only considers the stable topic-sentiment distribution. We refer to our complete model as TUS-LDA, the model in Fig 2(a) as TS-LDA and the model in Fig 2(b) as US-LDA.

4.2 Parameters Inference

Like LDA, exact inference is intractable in our models. Hence approximate estimation approaches, such as Gibbs Sampling [9], are utilized to solve the problem. Gibbs Sampling, a special case of Markov Chain Monte Carlo (MCMC) [6], is a relatively simple algorithm of approximate inference for our models. Due to space limitation, only the final formulas are given here.

Table 1. Notation used in the TUS-LDA model

Symbol	Description
M,K	number of documents,topics
V,U,T	number of vocabulary,users,timeslices
Z,**W**,**Y**	all the topics, words, user-timeslice switches
T,**U**	all the timeslices and users
L,$\bar{\mathbf{L}}$	all the sentiments of posts and words
N_m	number of word tokens in post d_m
u_m, t_m y_m, l_m	user,timeslice,user-timeslice switch and sentiment of post d_m
$l_{m,i}$	sentiment of word $w_{m,i}$
ε	beta distribution of stable topics and temporal topics
π_m	document-sentiment distribution, $\Omega = \{\pi_m\}_{m=1}^{M}$
$\theta_{t,s}$	timeslice-sentiment topic distribution, $\Theta = \{\theta_{t,s}\}_{t=1,s=1}^{T*S}$
$\delta_{u,s}$	user-sentiment topic distribution, $\Phi = \{\delta_{u,s}\}_{u=1,s=1}^{U*S}$
$\varphi_{s,k}$	sentiment-topic word distribution, $\Psi = \{\varphi_{s,k}\}_{s=1,k=1}^{S*K}$
α	hyperparameters of $\theta_{t,s}$ and $\delta_{u,s}$
β, λ	hyperparameters of $\varphi_{s,k}, \pi_m$
γ	hyperparameters of ε
ω_s	prior knowledge of $\varphi_{s,k}$

4.2.1 Joint Distribution

The joint probability of words, users, timeslices, timeslices-user switches, topics and sentiments can be factored in Eq 1, where ε, π, φ, δ and θ are integrated and $\overrightarrow{n}_m$ counts the number of three sentiment labels of a post and the words in the post (All the notations are illustrated in Table 1.).

$$
\begin{aligned}
&P_{TUS-LDA}(\mathbf{Z},\mathbf{W},\mathbf{T},\mathbf{U},\mathbf{Y},\mathbf{L},\bar{\mathbf{L}}|\alpha,\gamma,\lambda,\beta,\omega) = \\
&P(\mathbf{Y}|\gamma)P(\mathbf{L}|\lambda)P(\mathbf{Z}|\mathbf{Y},\mathbf{L},\alpha)P(\bar{\mathbf{L}}|\lambda)P(\mathbf{W}|\mathbf{Z},\bar{\mathbf{L}},\beta,\omega) = \\
&\frac{\Delta(\overrightarrow{n}_y+\overrightarrow{\gamma})}{\Delta(\overrightarrow{\gamma})} \times \prod_{m=1}^{M} \frac{\Delta(\overrightarrow{n}_m+\overrightarrow{\lambda})}{\Delta(\overrightarrow{\lambda})} \times \prod_{u=1}^{U}\prod_{s=1}^{S} \frac{\Delta(\overrightarrow{n}_{u,s}+\overrightarrow{\alpha})}{\Delta(\overrightarrow{\alpha})} \\
&\times \prod_{t=1}^{T}\prod_{s=1}^{S} \frac{\Delta(\overrightarrow{n}_{t,s}+\overrightarrow{\alpha})}{\Delta(\overrightarrow{\alpha})} \times \prod_{s=1}^{S}\prod_{k=1}^{K} \frac{\Delta(\overrightarrow{n}_{s,k}+\overrightarrow{\beta})}{\Delta(\overrightarrow{\beta})}; \\
&\Delta = \frac{\prod_{k=1}^{dim\,\overrightarrow{x}} \Gamma(x_k)}{\Gamma(\prod_{k=1}^{dim\,\overrightarrow{x}} x_k)}, \overrightarrow{n}_y = \{n_y^0, n_y^1\}, \overrightarrow{n}_m = \{n_m^{pos}, n_m^{neg}\} \\
&\overrightarrow{n}_{u,s} = \{n_{u,s}^k\}_{k=1}^K, \overrightarrow{n}_{t,s} = \{n_{t,s}^k\}_{k=1}^K, \overrightarrow{n}_{s,k} = \{n_{s,k}^v\}_{v=1}^V
\end{aligned}
\tag{1}
$$

4.2.2 Posterior Distribution

Posterior distribution is estimated as follows: for the i-th post, the user u_i and timeslice t_i are known. y_i, z_i and l_i can be jointly sampled given all other variables. Here, we use $\mathbf{y}$ to denote all the hidden variables y and $\mathbf{y}_{-i}$ to denote all the other y except y_i. All the hyperparameters are omitted.

$$
\begin{aligned}
&P(y_i=0, z_i=k, l_i=s|\mathbf{y}_{-i},\mathbf{z}_{-i},\mathbf{l}_{-i},\bar{\mathbf{l}},\mathbf{w}) \propto \frac{\gamma_0+n_{y,-i}^0}{\sum_{p=1}^{2}\gamma_p+n_{y,-i}^p} \\
&\times \frac{\lambda_s+n_{m,-i}^s}{\sum_{s'=1}^{S}\lambda_{s'}+n_{m,-i}^s} \times \frac{\alpha_k+n_{u,s,-i}^k}{\sum_{k=1}^{K}\alpha_{k'}+n_{u,s,-i}^{k'}} \\
&\times \frac{\prod_{v=1}^{V}\prod_{n_v=0}^{N_{(v)}-1}(\beta_{s,k}+n_{s,k,-i}^v+n_v)}{\prod_{n=0}^{N-1}(\sum_{v'=1}^{V}(\beta_{s,k}+n_{s,k,-i}^{v'})+n)}
\end{aligned}
\tag{2}
$$

If y_i=0, the i-th post talks about a stable topic, the sampling formula is shown in Eq 2; otherwise, the i-th post talks about a temporal topic, the sampling formula is shown in Eq 3.

$$P(y_i=1,z_i=k,l_i=s|\mathbf{y}_{-i},\mathbf{z}_{-i},\mathbf{l}_{-i},\bar{\mathbf{l}},\mathbf{w}) \propto \frac{\gamma_1+n^1_{y,-i}}{\sum_{p=1}^{2}\gamma_p+n^p_{y,-i}} \times \frac{\lambda_s+n^s_{m,-i}}{\sum_{s'=1}^{S}\lambda_{s'}+n^s_{m,-i}} \times \frac{\alpha_k+n^k_{t,s,-i}}{\sum_{k'=1}^{K}\alpha_{k'}+n^{k'}_{t,s,-i}} \times \frac{\prod_{v=1}^{V}\prod_{n_v=0}^{N_{(v)}-1}(\beta_{s,k}+n^v_{s,k,-i}+n_v)}{\prod_{n=0}^{N-1}(\sum_{v'=1}^{V}(\beta_{s,k}+n^{v'}_{s,k,-i})+n)} \tag{3}$$

For the j-th word in the i-th post, the sample formula of is shown in Eq 4.

$$P(\bar{l}_{ij}=s|\mathbf{z},\bar{\mathbf{l}}_{-ij},\mathbf{w},\mathbf{y},\mathbf{l}) \propto \frac{\lambda_s+n^s_{m,-ij}}{\sum_{s'=1}^{S}(\lambda_{s'}+n^{s'}_{m,-ij})} \times \frac{\beta^v_{s,k}+n^v_{s,k.-ij}}{\sum_{v'=1}^{V}(\beta^{v'}_{s,k}+n^{v'}_{s,k.-ij})} \tag{4}$$

Samples obtained from MCMC are then utilized for estimating the distributions π (Eq 5), δ (Eq 6) and θ (Eq 7), ϕ (Eq 8).

$$\pi^s_m = \frac{\lambda_s+n^s_m}{\sum_{s'=1}^{S}(\lambda'_s+n^{s'}_m)} \tag{5}$$

$$\delta^k_{u,s} = \frac{\alpha_k+n^k_{u,s}}{\sum_{k'=1}^{K}(\alpha_{k'}+n^{k'}_{u,s})} \tag{6}$$

$$\theta^k_{t,s} = \frac{\alpha_k+n^k_{t,s}}{\sum_{k'=1}^{K}(\alpha_{k'}+n^{k'}_{t,s})} \tag{7}$$

$$\varphi^v_{s,k} = \frac{\beta^v_{s,k}+n^v_{s,k}}{\sum_{v'=1}^{V}(\beta^{v'}_{s,k}+n^{v'}_{s,k})} \tag{8}$$

4.2.3 Gibbs Sampling Algorithm

A complete overview of Gibbs sampling procedure is given in Algorithm 1 (All the notations are listed in Table 1).

4.3 Incorporating Prior Knowledge

Drawing on the experience of JST and RJST [16], we also add an additional dependency link of φ on the matrix ω of size $S*V$, which is utilized for encoding word prior sentiment information into the TUS-LDA and its variants. To incorporate prior knowledge into TUS-LDA and its variants, we first set all the values of ω as 1. Then the matrix ω is updated with a sentiment lexicon which contains words with the corresponding sentiment labels, i.e., positive and negative. For each term $w \in \{1,...,V\}$ in the corpus, if w is found in the sentiment lexicon with the sentiment label $l \in \{1,...,S\}$, the element ω_{lw} is set as 1 and other elements of the word w are set as 0. The element lw is updated as follows:

$$\omega_{lw} = \begin{cases} 1 & \text{if S(w)=l} \\ 0 & \text{otherwise} \end{cases}$$

The Dirichlet prior β of the size $S*K*V$ are multiplied by the matrix ω (a transformation matrix) to capture the word prior sentiment polarity.

Algorithm 1: Inference on TUS-LDA

Input: α,γ,λ,β,ω

Initialize matrices Ω, Θ, Φ, Ψ and ε.
for *iteration c=1 to numIterations* **do**
  for *post m=1 to M* **do**
    Exclude post m and update count variables.
    Sample a timeslice-user switch, topic and sentiment label for post m.
    if *y=0* **then**
      Use Eq 2
    if *y=1* **then**
      Use Eq 3
    Update count variables with new timeslice-user switch, topic and sentiment label.
    for *n=1 to n_m* **do**
      Exclude word w_n and update count variables.
      Sample the sentiment label for word w_n using Eq 4.
      Update count variables with new sentiment label.
Update matrices Ω, Φ, Θ, Ψ using Eq 5, 6, 7, 8

5 Experiment Analysis

5.1 Dataset Description and Preprocessing

For experiments, we performed sentiment-aware topic discovery and sentiment classification on tweets, which are characterized by their limited 140 characters text. We selected tweets, which are related to electronic products such as camera and mobile phones, from Tweet7[2] and all the queried words are listed in Table 2). These tweets contain the description and reviews of various electronic products and correspond to multiple sentiment-aware topics. Besides, each tweet contains the content, the release timeslice, the user information.

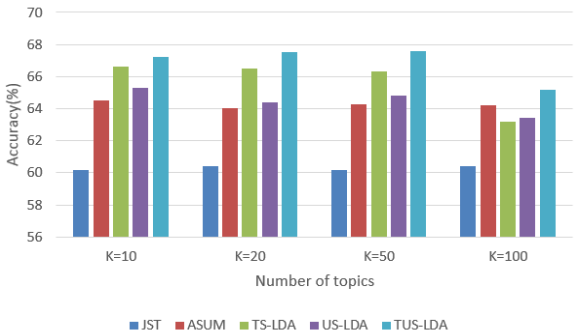
Table 2. Selected Words for Extracting Tweets Related to Electronics Products

iphone, blackberry, nokia, palmpre, sony, motorola, canon, nikon, dell, lenovo, toshiba, acer, asus, macbook, hp,alienware, camera, laptop, tablet, netbook, ipad, ipod, xbox,playstation, wii, phone, nintendo, printer, panasonic, epson,samsung, kyocera, ibm, sony, microsoft, lg, hitachi, scanner,computer, fujitsu, kodak, gameboy, sega, squareenix, android,ios, windows, operatingsystem, apple

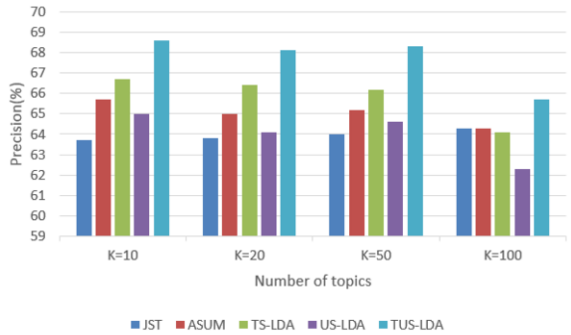
Due to the lack of sentiment labels on the Tweet7, we utilized the Sentiment140[3] [8], which contains 1.6 million tweets, for sentiment classification evaluation. Each tweet in Sentiment140 has the content, a release timeslice, a user and the overall polarity label (positive or negative). The number of positive and negative tweets are nearly identical.

We followed the preprocessing steps in BTM [28]. To improve the quality of our model, we added two extra steps: (1) Part-of-speech tagging of tweet contents using the Part-of-speech tagger[4] specially trained on tweets [7], retaining the words tagged as nouns, verbs or adjectives; (2) Lemmatizing words tagged as noun, verb, which was used to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. After preprocessing,

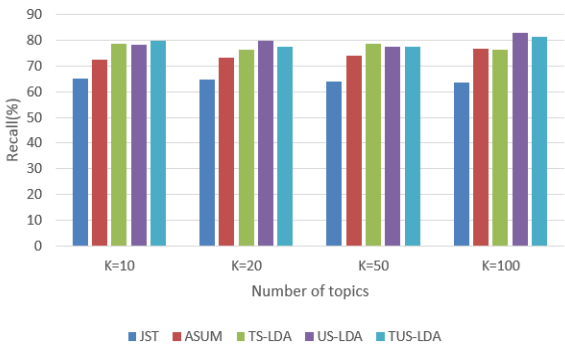
[2] https://snap.stanford.edu/data/twitter7.html
[3] http://help.sentiment140.com/for-students/
[4] http://www.ark.cs.cmu.edu/TweetNLP/



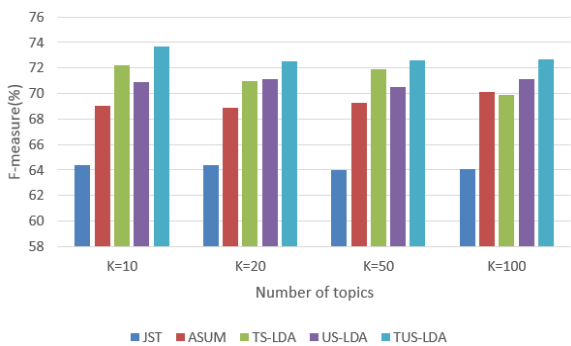
(a) Accuracy



(b) Precision



(c) Recall



(d) F1

Figure 3. (a) Accuracy (b) Precision (c)Recall (d) F1 of sentiment classification

as is shown in Table 3, we left 2,766,325 valid tweets, 80,083 distinct words, 174 timeslices (days) and 572,238 users in Tweet7, and we left 258,268 valid tweets, 29,486 distinct words, 48 timeslices (days) and 21,815 users in Sentiment140.

Table 3. Corpus Statistics

	Electronic	Senti140
Number of tweets	2,766,325	258,268
Users	572,238	21,815
Timeslices	174	48

5.2 Sentiment Lexicon

In JST [15] and our models, each sentiment label is viewed as a special kind of topic that we have known in advance. To improve the accuracy of sentiment detection, we need to incorporate prior knowledge or subjectivity lexicon (i.e., words with positive or negative polarity). Here, we chose PARADIGM [26], which consists of a set of positive and negative words, e.g., happy and sad. It defines the positive and negative semantic orientation of words. Moreover, emoticons are also strong emotion indicators on social media. The entire list of emoticons is taken from Wikipedia[5]. To adjust to our scenario on social media, we just chose a subset of the emoticons in Table 4.

5.3 Parameter Settings

To optimize the number of topics K, we empircally ran the models with four values of K: 10, 20, 50 and 100 in Sentiment140 and ran the models with three values of K: 10, 20, 50 in Twitter7 (In Twitter7, these tweets only contain a small number of electronic product-related topics). In our model, we simply selected symmetric Dirichlet prior vectors as is empircally done in JST and ASUM. For JST and ASUM, $\alpha = \frac{50}{K}, \beta = 0.01$ and $\gamma = 0.01$. For TUS-LDA, we set $\alpha = 0.5, \gamma = 0.01, \lambda = 0.01$ and $\beta = 0.01$. These LDA-based models are not sensitive to the hyperparameters [27]. In all the methods, Gibbs sampling was run for 1,000 iterations with 200 burn-in periods.

[5] https://en.wikipedia.org/wiki/List_of_emoticons

Table 4. Emoticons

Positive	Negative
:-) :o) :] :3 :c)	>:-(>:[:-(:c
:> =] 8) : } :-D	:@ >:(;(;-(:'-(
;-D :D 8-D \o/^^	:'(D; (T_T) (;_;)
:} (ò)/ ()/	(;_;) T.T !_!

5.4 Quantitative Evaluations

5.4.1 *Sentiment Classification*

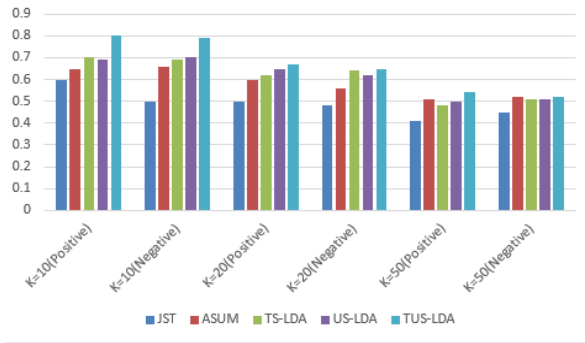
In this section, we performed a sentiment classification task to predict the sentiment labels of the test data in Sentiment140. Note that the Sentiment140 tweets do not contain neutral tweets. We determined the polarity of a tweet m by selecting the polarity s that has a higher probability in π_m^s (π_m is the sentiment distribution of the m-th post), the function is shown in Eq 9.

$$polarity(m) = \underset{s=\{neg,pos\}}{\text{argmax}} \ \pi_m^s \tag{9}$$

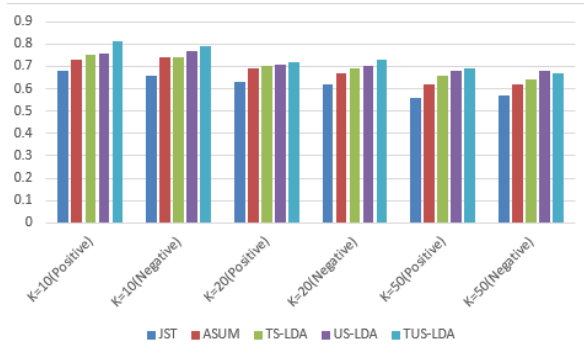
We present the results of sentiment classification with $Accuracy$, $Precision$, $Recall$ and $F1$ which are defined in the following.

Table 5. Average coherence score on the top T words in the K topics discovered on tweets of electronic products

T	Top 5			Top 10			Top 20		
K	10	20	50	10	20	50	10	20	50
JST	-39.88	-42.08	-41.68	-242.74	-246.79	-251.97	-1139.43	-1145.36	-1142.01
ASUM	-38.02	-39.86	-39.58	-240.47	-243.97	-246.47	-1135.44	-1131.96	-1135.27
TS-LDA	-35.53	-37.66	-38.91	-238.29	-240.87	-244.71	-1136.04	-1133.02	-1130.81
US-LDA	-36.42	-36.84	-36.59	-237.01	-238.67	-245.29	-1134.07	-1130.45	-1132.23
TUS-LDA	-33.91	-35.7	-35.61	-233.08	-234.72	-241.78	-1030.83	-1127.64	-1130.02



(a) Proportion of *coherent* topics generated by each model in $K = 10, 20, 50$



(b) Average Precision @20 (p @20) of words in *coherent* topics generated by each model in $K = 10, 20, 50$

Figure 4. (a) Proportion of *coherent* topics (b) Average Precision @20 (p @20) of words in *coherent* topics

Accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined in the binary classification.
Precision is the proportion of the true positives against all the predicated positive results (both true positives and false positives) in the binary classification.
Recall is the proportion of the true positives against all the actual positive results (both true positives and false negatives) in the binary classification.
F1 is the harmonic mean of **Precision** and **Recall**.

Based on the results of sentiment classification, we can see that TUS-LDA outperformed JST, ASUM, TS-LDA and US-LDA in $F1$ (Fig 3(d)). For $Recall$ (3(c)), ASUM, TS-LDA, US-LDA and TUS-LDA performed equally well, JST performed worst. For $Accuracy$ (Fig 3(a)) and $Precision$ (Fig 3(b)), TUS-LDA performed best and TS-LDA performed better than US-LDA. There exist 48 timeslices and 21,815 users, the number of users is far more than that of timeslices which causes that modeling tweets aggregated in timeslices performed better than tweets aggregated in users. Aggregating tweets in timeslices or users (i.e., TUS-LDA) with $K = 10$ performed best in Sentiment140.

5.4.2 Topic Coherence

Another goal of TUS-LDA is to extract coherent sentiment-aware topics from user-generated post collection and evaluate the effectiveness of topic and sentiment captured by our models. In order to conduct quantitative evaluation of topic coherence, we used an automated metric proposed in [20], which is shown in Eq 10, where topic coherence, denoted as $D(v)$, is the document frequency of word v, $D(v, v^{'})$ is the co-document frequency of word v and $v^{'}$ and $V^{(k)} = (v_1^{(k)}, ..., v_T^{(k)})$ is a list of the T most probable words in topic k. The key idea of the coherence score is that if a word pair is related to the same topic, they will co-occur frequently in the corpus. In order to quantify the overall coherence of the discovered topics, the average coherence score, $\frac{1}{K}\sum^{k} C(z_k; V^{(z_k)})$, was utilized. We conducted and evaluated the topic extraction experiments on the tweets of electronic products. Here we also compared TUS-LDA with four sentiment-topic models: JST, ASUM, TS-LDA and US-LDA. In this collection, we set the number of topics $K = 10, 20, 50$ for all the methods. The result is listed in Table 5. From the topic coherent results, it is clear that aggregating tweets in timeslices or users (TUS-LDA) directly leads to significant improvement of topic coherent. Note that TUS-LDA also performed best in the topic coherent and the performance of TS-LDA (aggregating tweets in timeslices) was similar to US-LDA (aggregating tweets in users).

$$C(t; V^{(t)}) = \sum_{m=2}^{M}\sum_{l=1}^{m-1} log\frac{D(v_m^{(t)}, v_l^{(t)}) + 1}{D(v_l^{(t)})} \tag{10}$$

5.4.3 Human Evaluation

As our objective is to discover more coherent sentiment-aware topics, so we chose to evaluate the topics manually which is based on human judgement. Without enough knowledge, the annotation will not be credible. Following [20], we asked two human judges, who are familiar with common knowledge and skilled in looking up the test tweet dataset, to annotate the discovered sentiment-aware topics manually. To ensure the annotation reliable, we labeled the generated topics by all the baseline models and our proposed model at learning iteration 10.

Topic Labeling: Following [20], we asked the judges to label each sentiment-aware topic as *coherent* or *incoherent*. Each sentiment-aware topic is represented as a list of 20 most probable words in word distribution φ of the topic. Here they annotated a sentiment-aware topic as *coherent* when at least half of top 20 words were

Table 6. Example of topics extracted by TUS-LDA

Positive sentiment label			Negative sentiment label		
Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3
camera	ipod	xbox	printer	window	phone
digit	song	game	ink	vista	problem
canon	phone	live	print	us	information
nikon	listen	sale	cartridge	microsoft	security
new	music	console	toner	install	strange
len	love	plai	laser	download	risk
photograph	new	playstat	color	software	finiance
review	play	ps3	laserjet	file	mobile
panason	shuffle	microsoft	paper	free	digit
slr	good	new	scanner	server	on-line

related to the same semantic-coherent concept (e.g., an event, a hot topic) and the sentiment polarities of the words are accurate, others were $incoherent$..

Word Labeling: Then we chose $coherent$ sentiment-aware topics which were judged before and asked judges to label each word of the top 20 words among these $coherent$ sentiment-aware topics. When a word was in accordance with the main semantic-coherent concept that represents the topic, the word was annotated as $correct$ and others were $incorrect$. After topic labeling, the judges had known the concept of each sentiment-aware topic and the overall sentiment of the topic, it is easy to label words of each sentiment-aware topic. As is shown in Table 7, the annotation of both judges in $Precision@20$ (or $p@20$) also have good agreements (Cohen's Kappa score is greater than 0.8 [13]).

Table 7. Cohen's Kappa for pairwise inter-rater agreements

	Topic Labeling	Word Labeling		
		p@5	p@10	p@20
Kappa	0.820	0.911	0.821	0.816

Figure 4(a) shows that TUS-LDA can discover more $coherent$ topics than JST, ASUM, TS-LDA and US-LDA. Thereinto, TUS-LDA can discover the nearly equal number of positive and negative topics. Figure 4(b) gives the average $Precision@20$ of all coherent topics. TUS-LDA performed better than other four models and performed best in $K = 10$.

From the above, we can observe that aggregating posts in the same timeslice or user as a single document can indeed improve the performance in sentiment classification and sentiment-aware topic extraction in user-generated posts as TUS-LDA consistently outperformed the baseline models except in $K = 50(Negative)$. Also the empirical results reveal that the most likely number of topic for tweets of electronic products in Twitter7 is 10.

5.5 Qualitative Analysis

To investigate the quality of topics discovered by TUS-LDA, we randomly choose some topics for visualization. We randomly selected six topics, i.e., three positive topics and three negative topics. For each topic, we choose the top 10 words which can most represent the topic.

Table 6 presents the top words of the selected topics. The three topics with a positive sentiment label respectively talk about "Camera", "apple music product" and "game" and these topics are listed in the left columns of Table 6; the three negative topics are related to "printer", "window product" and "phone" are listed in right columns of Table 6. As we can see clearly from Table 6, the six topics are quite explicit and coherent, where each of them tried to capture the topic of a kind of electronic product. In terms of topic sentiment, by checking each of the topics in Topic 6, it is clear that Topic 2 under the positive sentiment label and Topic 3 under the negative sentiment label indeed bear positive and negative sentiment labels respectively. However, other topics under positive and negative sentiment label carry fewer sentiment words than the above two topics. By manually examining the tweet data, we observe that the sentiment labels of these topics are accurate. The analysis of these topics shows that TUS-LDA can indeed discover coherent sentiment-aware topics.

6 Conclusion and Future Work

In this paper, we studied the problem of sentiment-aware topic detection from the user-generated posts on the social media. The existing work is not suitable for the short and informal posts, we proposed a new sentiment/topic model that considers the time, user information of posts to jointly model topics and sentiments. Based on the different characteristics of sentiments and topics, we limited that words in the same post belong to the same topic, but they can belong to different sentiments. We compared our model with JST, ASUM as well as two degenerate variations of our model on two Twitter datasets. Our quantitative evaluation showed that our model outperformed other models both in sentiment classification and topic coherence. At the same time, we asked two judges to evaluate our models and baseline methods and the result also showed that our model TUS-LDA performed best in sentiment-aware topic extraction. Moreover, we used six examples to visualize some sentiment-aware topics. In the future work, we want to further mine sentiment-aware events in the posts which can monitor the sentiment variation over time of each event. Moreover, we can also utilize the user's topic and sentiment information to cluster similar users. We will also consider to expand our model for aspect-based opinion mining.

ACKNOWLEDGEMENTS

We would like to thank the reviewers for their comments, which helped improve this paper considerably. This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant No.61272378 and the 863 Program under Grant No.2015AA015406.

REFERENCES

[1] David M Blei, Andrew Y Ng, and Michael I Jordan, 'Latent dirichlet allocation', *Journal of Machine Learning Research*, **3**, 993–1022, (2003).

[2] Johan Bollen, Huina Mao, and Xiaojun Zeng, 'Twitter mood predicts the stock market', *Journal of Computational Science*, **2**(1), 1–8, (2011).
[3] Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh, 'Leveraging multi-domain prior knowledge in topic models', in *Proc. of IJCAI*, pp. 2071–2077. AAAI, (2013).
[4] Mohamed Dermouche, Julien Velcin, Leila Khouas, and Sabine Loudcher, 'A joint model for topic-sentiment evolution over time', in *Proc. of ICDM*, pp. 773–778. IEEE, (2014).
[5] Qiming Diao, Jing Jiang, Feida Zhu, and Ee-Peng Lim, 'Finding bursty topics from microblogs', in *Proc. of ACL*, pp. 536–544. ACL, (2012).
[6] Charles J Geyer, 'Practical markov chain monte carlo', *Statistical Science*, 473–483, (1992).
[7] Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A Smith, 'Part-of-speech tagging for twitter: Annotation, features, and experiments', in *Proc. of ACL*, pp. 42–47. ACL, (2011).
[8] Alec Go, Richa Bhayani, and Lei Huang, 'Twitter sentiment classification using distant supervision', *CS224N Project Report, Stanford*, **1**, 12, (2009).
[9] Gregor Heinrich, 'Parameter estimation for text analysis', Technical report, Technical report, (2005).
[10] Thomas Hofmann, 'Probabilistic latent semantic indexing', in *Proc. of SIGIR*, pp. 50–57. ACM, (1999).
[11] Yohan Jo and Alice H Oh, 'Aspect and sentiment unification model for online review analysis', in *Proc. of WSDM*, pp. 815–824. ACM, (2011).
[12] Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad, 'Nrc-canada-2014: Detecting aspects and sentiment in customer reviews', in *SemEval*, pp. 437–442. ACL, (2014).
[13] JR Landis and GG Koch, 'The measurement of observer agreement for categorical data.', *Biometrics*, **33**, 159–174, (1977).
[14] Kar Wai Lim and Wray Buntine, 'Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon', in *Proc. of CIKM*, pp. 1319–1328. ACM, (2014).
[15] Chenghua Lin and Yulan He, 'Joint sentiment/topic model for sentiment analysis', in *Proc. of CIKM*, pp. 375–384. ACM, (2009).
[16] Chenghua Lin, Yulan He, Richard Everson, and Stefan Rüger, 'Weakly supervised joint sentiment-topic detection from text', *IEEE Transactions on Knowledge and Data Engineering*, **24**(6), 1134–1145, (2012).
[17] Bing Liu, *Web data mining: exploring hyperlinks, contents, and usage data*, Springer Science & Business Media, 2007.
[18] Bin Lu, Myle Ott, Claire Cardie, and Benjamin K Tsou, 'Multi-aspect sentiment analysis with topic models', in *Proc. of ICDMW*, pp. 81–88. IEEE, (2011).
[19] Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai, 'Topic sentiment mixture: modeling facets and opinions in weblogs', in *Proc. of WWW*, pp. 171–180. ACM, (2007).
[20] David Mimno, Hanna M Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum, 'Optimizing semantic coherence in topic models', in *Proc. of EMNLP*, pp. 262–272. ACL, (2011).
[21] Subhabrata Mukherjee, Gaurab Basu, and Sachindra Joshi, 'Joint author sentiment topic model', in *SDM*, pp. 370–378. SIAM, (2014).
[22] Thien Hai Nguyen and Kiyoaki Shirai, 'Topic modeling based sentiment analysis on social media for stock market prediction', in *Proc. of ACL*, pp. 1354–1364. ACL, (2015).
[23] Alexander Pak and Patrick Paroubek, 'Twitter as a corpus for sentiment analysis and opinion mining.', in *Proc. of LREC*, volume 10, pp. 1320–1326, (2010).
[24] Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif M Mohammad, Alan Ritter, and Veselin Stoyanov, 'Semeval-2015 task 10: Sentiment analysis in twitter', *SemEval*, (2015).
[25] Kim Schouten and Flavius Frasincar, 'Survey on aspect-level sentiment analysis', *IEEE Transactions on Knowledge & Data Engineering*, **28**(3), 813–830, (2016).
[26] Peter D Turney and Michael L Littman, 'Measuring praise and criticism: Inference of semantic orientation from association', *ACM Transactions on Information Systems*, **21**(4), 315–346, (2003).
[27] Hanna M Wallach, David M Mimno, and Andrew McCallum, 'Rethinking lda: Why priors matter', in *NIPS*, pp. 1973–1981, (2009).
[28] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng, 'A biterm topic model for short texts', in *Proc. of WWW*, pp. 1445–1456. Springer, (2013).
[29] Qi Zhang, Yeyun Gong, Xuyang Sun, and Xuanjing Huang, 'Time-aware personalized hashtag recommendation on social media', in *Proc. of COLING*, pp. 203–212. ACL, (2014).
[30] Wayne Xin Zhao, Jing Jiang, Hongfei Yan, and Xiaoming Li, 'Jointly modeling aspects and opinions with a maxent-lda hybrid', in *Proc. of EMNLP*, pp. 56–65. ACL, (2010).
[31] Wayne Xin Zhao, Jiang Jing, Weng Jianshu, He Jing, Lim Ee-Peng, Yan Hongfei, and Li Xiaoming, 'Comparing twitter and traditional media using topic models', in *Proc. of ECIR*, pp. 338–349. Springer, (2011).
[32] Chen Zheng, Li Chengtao, Sun Jian-Tao, and Jianwen Zhang, 'Sentiment topic model with decomposed prior', in *Proc. of SDM*, pp. 767–775. SIAM, (2013).