

Received October 7, 2020, accepted November 16, 2020, date of publication November 24, 2020, date of current version December 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039963

Exploiting Long-Term Dependency for Topic Sentiment Analysis

FALIANG HUANG¹, CHANGAN YUAN², YINGZHOU BI¹, AND JIANBO LU¹

¹School of Computer and Information Engineering, Nanning Normal University, Nanning 530001, China

²Guangxi Academy of Sciences, Nanning 530022, China

Corresponding author: Faliang Huang (huangfliang@163.com)

This work was supported in part by the Natural Science Foundation of China under Grant 61962038 and Grant 61962006, and in part by the Guangxi Bagui Teams for Innovation and Research under Grant 201979.

ABSTRACT Most existing unsupervised approaches to detect topic sentiment in social texts consider only the text sequences in corpus and put aside social dynamics, as leads to algorithm's disability to discover true sentiment of social users. To address the issue, a probabilistic graphical model LDTSM (Long-term Dependence Topic-Sentiment Mixture) is proposed, which introduces dependency distance and uses the dynamics of social media to achieve the perfect combination of inheriting historical topic sentiment and fitting topic sentiment distribution underlying in current social texts. Extensive experiments on real-world SinaWeibo datasets show that LDTSM significantly outperforms JST, TUS-LDA and dNJST in terms of sentiment classification accuracy, with better inference convergence, and topic and sentiment evolution analysis results demonstrate that our approach is promising.

INDEX TERMS Sentiment analysis, topic detection, probabilistic graphical model, long-term dependency.

I. INTRODUCTION

With the rapid development of social media and intelligent mobile devices, the public can conveniently express their opinions and share their feelings regarding social, economic or political issues through online social platforms like Twitter and Facebook. The various types of user generated contents (UGCs), such as microblogs and online product reviews, are usually opinionated and topic-oriented, and present a valuable source of information for human intelligent decisions. For instance, government departments would like to effectively manage public opinions of important political events via mining text content in social media systems. Manufacturing enterprises may improve their production strategy through analyzing users' valuable feedback from online product reviews. In this context, sentiment analysis in social media has received significant attentions from governments, enterprises and academic institutions.

Great efforts in techniques to automate sentiment identification in diversified text data have flourished in the recent years [1]–[3]. Most of the existing works treat text sentiment identification as a text classification problem and exploit supervised machine learning techniques to estimate

the sentiment distribution of texts. Those supervised methods usually do not take text topic into consideration while analyzing text sentiment, as ignores the fact that sentiment polarities of texts are closely related to their topics in social media. For instance, the adjective “complicated” may be negative in sentiment space when it occurs in a review about a movie character, whereas it conveys positive sentiment in a message commenting on movie plot. The ignorance tends to degrade the performance of text sentiment analysis. To alleviate the drawback, generative models are proposed to model the joint distribution of sentiment and topic with parameters that can be interpreted as reflecting latent associations between topics and sentiments in the data. These generative models are robust and require no or little human effort in topic sentiment distribution estimation, thus are widely used in sentiment analysis tasks.

In online social networks, public sentiments towards entities such as events, services and organizations are usually time-varying. The social media data shares some characteristics of time series, as makes it appealing for researchers to leverage time series analysis tools for sentiment tracking. Various sliding time window strategies are proposed, which can obtain sentiment distribution of different time periods by applying some off-the-shelf sentiment analysis methods to UGCs in different windows and further acquire sentiment

The associate editor coordinating the review of this manuscript and approving it for publication was Feng Xia¹.

WinID	TwID	Tweets	Data
W1	M1	I just planted some small trees. Hope tomorrow will be a rainy day! 🌧️	2018/4/2
W2	M2	Please, please, rain, rain, help my pitiful little trees! 🌧️	2018/4/5
W3	M3	Oh my goodness, another sunny day, what should my lovely trees do! 🌞	2018/4/8

FIGURE 1. Topic sentiment dependencies in different time windows.

evolution patterns via time series analysis techniques. Most of the sliding-window-based sentiment analysis approaches take an assumption that sentiment patterns in different time windows are independent and identically distributed. However, the assumption is often not true. For example, sentiments of Twitter posts in adjacent sliding time windows are to some extent dependent. When a typical generative model, such as JST [4] or TSM [5], is chosen to estimate sentiment distribution of Twitter posts in a given time window, it may not achieve satisfactory results. Specially, true topic sentiment patterns of social media content in current time window may not be discovered, because how sentiments of UGCs in current time window impact those in current time window is not taken into account in most of the existing generative models. And on the other hand, prior Dirichlet distributions initialized with random hyperparameters may be not good start points to search for the optimal parameters of the generative models, and thus slow down convergence of model inference.

Considering an example as shown in Fig 1, there are three tweets posted by the same twitter user in different days. According to a given sliding window size, tweets M1, M2 and M3 belong respectively to windows W1, W2 and W3. If utilize a traditional generative models to determine sentiment polarities of the three tweets, we may have that M2 and M3 are negative and positive respectively, since sentiment polarity of word “*pitiful*” in M2 is negative, and sentiment polarity of words “*lovely*” and “*sunny*” in M3 is positive, and there is no explicit sentiment word in M1. However, if taking into consideration context of tweet M3, i.e., sentiment polarity of its previous tweet M2 is inherited while inferring sentiment polarity of tweet M3, it is not difficult to determine that sentiment polarity of M3 is negative.

From the example we can see that the sentiment topic independence assumption about UGCs in different time windows may lead to unsatisfactory sentiment analysis results.

To address the issue, in this paper we propose a weakly-supervised generative approach based on sentiment consistency in sociology [6]. In particular, motivated by natural immediacy of social media and subtle dependency between text topics and text sentiments, and based on assumptions that 1) topics of texts are generated according to sentiment

distributions of texts; 2) roughly similar is sentiment and topic conveyed in social media content posted by the same user over adjacent periods of time. We devise a novel topic sentiment model dubbed LDTSM (Long-term time-Dependent Topic Sentiment Model), which embeds an additional sentiment layer into LDA (Latent Dirichlet Allocation) [7] and takes topic and sentiment dependency existing in adjacent periods of time into consideration while inferring text sentiment. We estimate the model parameters using the Gibbs sampling approach. For performance evaluation, we apply the LDTSM model to four real-life datasets from SinaWeibo (<http://weibo.com/>) and compare its performance with state-of-the-art sentiment analysis models JST, TUS-LDA and dNJST [38]. The experimental results indicate that the proposed LDTSM outperforms the other models in terms of sentiment classification accuracy and can efficiently discover topics with good interpretability and efficiency. Moreover, LDTSM can effectively achieve evolution analysis of topics and sentiment.

Our contributions in the paper mainly include

- Based on sentiment consistence theory and natural immediacy of social text stream, a unified weakly-supervised generative learning framework is proposed for sentiment analysis in social media.
- Based on the rationale of social text sentiment long-distance dependency, we invent a novel generative model dubbed LDTSM under the proposed framework and make inferences about model LDTSM with Gibbs sampling techniques.
- Through an extensive set of experiments on real-world data sets, we verify that our method can give very promising results in text sentiment analysis in social media.

The rest of the paper is organized as follows. We review related work in Section 2. Related definitions are simply introduced in Section 3. Section 4 and Section 5 describe the proposed framework and LDTSM model respectively. Experimental results are discussed in Section 6. Finally, we conclude the paper in Section 7.

II. RELATED WORK

Our approach is closely related to sentiment evolution analysis and generative approaches for text sentiment analysis, and we review some of the most relevant work here.

A. SENTIMENT EVOLUTION ANALYSIS

Sentiment evolution analysis has received significant attention due to its increasing importance, and numerous methods for sentiment evolution analysis have emerged recently. Existing methods for sentiment evolution analysis are roughly classified as online sentiment analysis and offline sentiment analysis.

Due to the demands of real time analysis, majority of online sentiment analysis methods is lexicon-based. Core idea of lexicon-based sentiment analysis is to tag discrete labels such as positive, negative and neutral or assign a continuous

sentiment intensity value for each word in documents to be handled via retrieving a certain handy sentiment lexicon, and compute a final sentiment label or sentiment intensity for the documents with some fusion strategies such as high-voting or linear average. It is a formidable challenge to construct a high-quality sentiment lexicon. Taboada *et al.* [8] proposed a lexicon-based approach called SO-CAL, which takes into account valence shifters (intensifiers, downtoners, negation, and irrealis markers) when calculating semantic orientation of the extracted sentiment-bearing words. Ortega *et al.* [9] utilized WordNet and SentiWordNet for sentiment polarity detection and managed to achieve good results on the SemEval-2013 dataset. Al-Ayyoub *et al.* [10] construct a sentiment lexicon of about 120,000 Arabic terms and built a lexicon-based sentiment analysis tool. Trinh *et al.* [11] proposed a lexicon-based method for sentiment analysis in Facebook comments in Vietnamese language. Hazimeh *et al.* [12] applied a lexicon-based approach to track sentiment of large-scale social events with essential features and auxiliary features. When determining sentiment scores of social text sequences with lexicon-based approaches, each word is assigned with a fixed sentiment. However, the actual sentiment behind a word may be depended on its context of the complete sentence and may change across time.

Considering that social media data is usually time-ordered, researchers attempted to make the utmost out of timestamp information from social media for offline sentiment analysis. Specially, it mainly consists of steps as follows. First, the original social corpus is partitioned into several text subsets based on a particular time granular; secondly, a suitable machine-learning method is chosen to compute sentiment score of each text subset; finally, sentiment time series is constructed with the computed sentiment scores and used for subsequent sentiment analysis tasks such as prediction, clustering and so. To explore quantifiable relationship between overall public mood and macroscopic social and economic indicators, Bollen *et al.* [13] constructed two sentiment time series to analyze long-term sentiment drifts and short-term sentiment fluctuations. Based on the different propensity users have on disclosing positive and extreme feelings. Guerra *et al.* [14] proposed a feature representation strategy that focuses on terms which appear at spikes in the social stream for sentiment evolution analysis. Bifet and Frank [15] developed a sliding window based on Kappa statistic for quantifying the predictive performance of twitter text stream sentiment classifiers. Si *et al.* [16] applied a continuous Dirichlet Process Mixture model and a moving prediction process under sliding windows to regress the stock index and the twitter sentiment time series. Sentiment time series is constructed using sentiment scores of sentiment words over a sliding time window to achieve multi-document sentiment prediction [17]. Ranco *et al.* [18] explored dependency between the sentiment polarities of twitter event windows and stock price returns. Giachanou *et al.* [19] attempted to utilize conventional time series analysis techniques, such as frequency analysis, outlier

detection and time series decomposition, to track sentiment in social media. Yang *et al.* [20] designed a location-based dynamic sentiment-topic model which can jointly model topic, sentiment, time and geolocation information, with an attempt to track sentiment shifts in different geographical regions. Giachanou *et al.* [21] proposed to leverage time series outlier detection technique for sentiment spike identification, and combine LDA and relative entropy to extract the topics and compute their contribution to the sentiment spikes. Wang *et al.* [22] proposed an iterative algorithm SentiDiff to improve Twitter sentiment analysis using sentiment diffusion patterns. All the above-mentioned offline sentiment analysis approaches are efficient and intuitive, but are built on an unconvincing assumption that social data in different time windows are independent and identically distributed.

B. GENERATIVE MODELS FOR SENTIMENT ANALYSIS

As a statistical learning paradigm, generative model excels at expressing directly observable and even indirect relationships between the observed and target variables. This property is of particular importance in sentiment analysis, since there exists subtle dependency between sentiment and text sequence content.

Various generative models have been proposed for sentiment analysis of different granularity levels, including document-level, topic-level, aspect-level and term-level, based on characteristics of social media such as small world effect of social network, preference of social content users, multi-modality of social media, to name a few. Zhao *et al.* [23] proposed a MaxEnt-LDA hybrid model to jointly discover both aspects and aspect-specific opinion words. Lin *et al.* [4] designed a joint sentiment-topic model JST by adding sentiment layer in LDA to reflect dependency between word topic and word sentiment. Aiming to discover users' interests about different sentiment topics in texts, Almars *et al.* [24] presented a hierarchical user sentiment topic model HUSTM, where each word in a document is associated with three latent variables: a user, a topic, and a sentiment. He *et al.* [25] proposed a weakly-supervised latent sentiment model, which replaces topic layer in LDA with sentiment layer and acquires sentiment prior information from English sentiment lexicons via machine translation. Hai *et al.* [26] presented a supervised joint topic model SJASM, which leverages the inter-dependency between the aspect-based sentiment and overall sentiment, and estimate the overall sentiments of reviews via a normal linear model. Liang *et al.* [27] designed a universal affective model, consisting of topic-level and term-level sub-models, to identify readers' emotions hidden in short texts. Motivated by the item response theory, Lin *et al.* [28] proposed a topical user item response model, which assumes that user's review towards a specific item is jointly determined by the individuality of the user and attribute of the item. By taking into account user preferences and thought patterns, Poddar *et al.* [29] built an author-aware aspect topic sentiment model Author-ATS, which treats a word as a combination of a hierarchy

of aspect, topic and sentiment. Rahman *et al.* [30] came up with a hidden topic sentiment model HTSM to extract latent aspects and corresponding sentiment polarities by imposing constraints topic coherence and sentiment consistency. Tang *et al.* [31] proposed a joint aspect-based sentiment topic model JABST to simultaneously extract multi-grained aspects and corresponding sentiment polarities. Borrowing the idea of lifelong machine learning, Wang *et al.* [32] developed jointly aspect sentiment model LAST (Lifelong Aspect-based Sentiment Topic), which can simultaneously identify aspects, opinions, opinion polarity and opinion generality, and achieve knowledge transfer in different domains. With an attempt to address data sparse problem in social short texts, Xiong *et al.* [33] constructed a word-pair sentiment-topic model WSTM, where generation elements are word-pairs from the corpus, rather than words in traditional models. Inspired by the observations that users in the same community usually travel on nearby regions and share common activities and topics, Zhang and Chow [34] proposed a LDA-based model called CRATS to jointly mine the latent communities, regions, activities, topics, and sentiments based on the important dependencies among these latent variables. Wang *et al.* [35] provided a probabilistic model PSEM (Public Sentiment Evolution Model) to simultaneously model the evolution of opinion for social incidents. Huang *et al.* [36] devised a multimodal joint sentiment topic model, which utilizes emoticons to perform unsupervised sentiment analysis. Gui *et al.* [37] proposed a multitask learning framework which jointly learns a sentiment classifier and a topic model by making the word level latent topic distributions in the topic model to be similar to the word level attention vectors in sentiment classifiers through mutual learning.

The aforesaid generative approaches do not take natural dynamic of social media into consideration, thus cannot succeed in revealing true sentiment patterns hidden in the social media data to be analyzed. Very few efforts have been paid for utilizing time information related with social media content to detect text sentiment polarities. Fu *et al.* [38] proposed a dynamic non-parametric joint sentiment topic mixture model dNJST, which adds a sentiment layer and time decay dependencies of historical epochs to the current epochs in hierarchical Dirichlet process topic model. Different from dNJST, which only takes into account historical topic distribution and stick-breaking construction sentiment parameters while inferring topic sentiment distribution of social media data in current time window, our model LDTSM consider both sentiment distributions, topic-sentiment distributions, and topic-word distributions from historical time windows.

III. DEFINITION OF TERMINOLOGIES

To facilitate presentation, we define the basic terminologies we will use in this paper.

Definition (Timeslice): Each user-generated post has a timeslice that specifies when the user publishes the post, e.g., an hour or a day. Size of timeslice is determined by knowledge engineers according to the handy task.

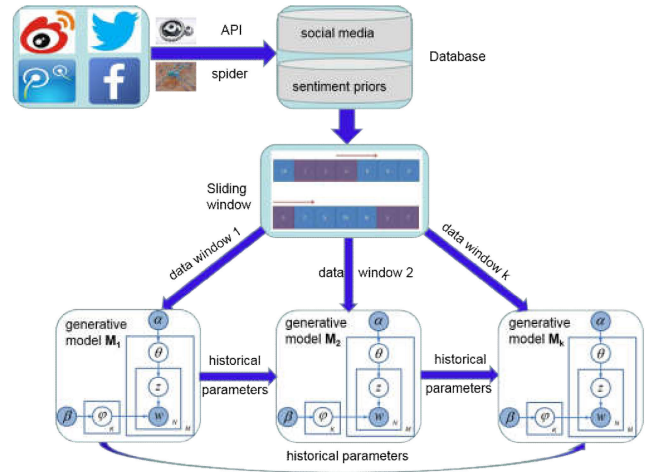


FIGURE 2. The proposed framework.

Definition (Topic): A topic is a subject of conversation or discussion, which is mathematically formalized as a set of words. It is also represented as a discrete distribution over words in a fixed vocabulary.

Definition (Sentiment): Sentiment represents a user’s emotional feelings about a particular entity. It can be denoted by a discrete random variable taking value from $\mathbb{S} = \{l_1, l_2, \dots, l_{|\mathbb{S}|}\}$, e.g., positive or negative, or a random continuous variable in some value range, e.g., $[0, 1]$ or $[-1, 1]$.

Definition (Sentiment-aware topic): A sentiment-aware topic is a topic labeled with a sentiment polarity. For example, the overall sentiment of the topic “riots in Hong Kong” is negative, hence the topic “riots in Hong Kong” is a negative topic.

Based on the above definitions, we strive to develop a probabilistic generative framework (Section IV) and design a long-term time-dependent topic sentiment model (Section V) based on weakly-supervised learning for automatic detection of sentiment-aware topics in social media.

IV. THE PROPOSED FRAMEWORK

In this section, we illustrate our sentiment analysis framework in Figure 2.

For a sentiment analysis task, for instance, to investigate into public opinions of political event “Ethiopian Airlines Boeing 737 MAX crash”, we first collect the relevant posts and their timestamps from various sources such as Tweeter and Facebook with the help of API (Application Programming Interface) provided by the social media service or the web crawler from some third party. Following that, some noisy posts are removed from the collected posts with preprocessing techniques like keyword-based matching. The pre-processed posts and sentiment priors (sentiment dictionaries) are then saved in database systems. Afterwards, according to post’s timeslice, sliding window technique is applied to divide the collected data in the database into several small datasets. Finally, some probabilistic generative approach is chosen to learn the topic sentiment hidden in each small dataset.

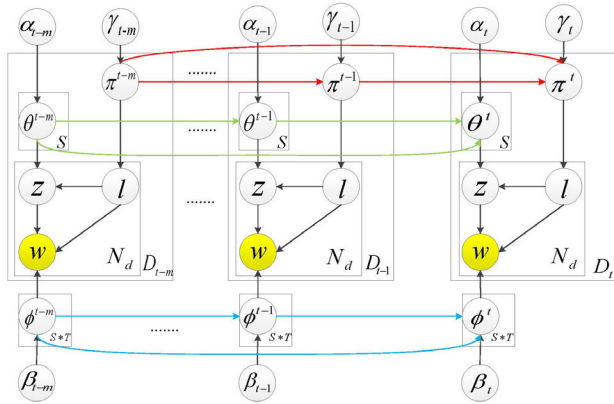


FIGURE 3. Graphical representation of LDTSM.

It is well worthy of highlighting that parameters of historical generative models exert some influence on corresponding ones of the current model. Specially, parameters of model M_t is initialized with parameters of previous L models $\{M_{t-1}, M_{t-2}, \dots, M_{t-L}\}$.

V. THE PROPOSED MODEL

In this section, we firstly introduce the notation and formally formulate our model. Then, we give the method for learning parameters. Finally, we present the method to conduct topic sentiment analysis with the learned distributions. To allow for detailed illustration, we list the relevant symbols and notations in Table 1.

A. MODEL DESCRIPTION

Key assumption of JST model is that a document is a compound distribution generated by different member probability distributions over words, and each member probability distribution corresponds to a sentiment. The existing framework of JST consists of hierarchical layers as follows: document layer, sentiment layer, topic layer and word layer. Through the four layers, documents are assigned with sentiment labels, under which topics are associated with sentiment labels and words are associated with both sentiment labels and topics. Although JST has received a great success in text sentiment analysis, estimating JST model parameters is very time-consuming. Moreover, time information is not taken into consideration when JST attempts to learn topic sentiment patterns in texts.

With the aim of better modeling document sentiments, a long-term time-dependent topic sentiment model (LDTSM) model (Figure 3) is proposed and integrated into the above framework. From Figure 3, we can see that LDTSM uses JST as member component by concatenating the individual JST model. Specially, when LDTSM performs sentiment analysis of posts in sliding window t , sentiment-topic-word distributions $\{\phi_{t-m}\}_{m=1}^L$ of previous continuous L well-trained JST are used to generate the distribution sentiment-topic-word ϕ_t . Similarly, sentiment distributions $\{\pi_{t-m}\}_{m=1}^L$ and sentiment-topic distributions $\{\theta_{t-m}\}_{m=1}^L$ are applied to generate the

TABLE 1. Symbols and notations.

Type	Symbol	Description
random variables	l	sentiment label index
	z	topic index
	w	word index
	t	timeslice index
	m	historical window index
	d_t	index of document in window t
distributions & hyperparameters	θ_t	distribution sentiment-topic in timeslice t
	ϕ_t	distribution sentiment-topic-word in timeslice t
	π_t	distribution sentiment in timeslice t
	$\alpha_t, \beta_t, \gamma_t$	priors of Dirichlet distributions α_t, β_t and γ_t
predetermined parameters	S, T	predetermined number of distinct sentiment polarities, topics hidden in corpus
	L	predetermined number of historical windows exerting effect on current window
	V	size of vocabulary
abbreviations	$\overline{\pi_{t-m}}$	$\pi_{t-m,l}$
	$\overline{\gamma_{t-m+1}}$	$\gamma_{t-m+1,l}$
	$\overline{\alpha_{t-m+1}}$	$\alpha_{t-m+1,l,z}$
	$\overline{\theta_{t-m}}$	$\theta_{t-m,l,z}$
	$\overline{\beta_{t-m+1}}$	$\beta_{t-m,l,z}$
	$\overline{\phi_{t-m}}$	$\phi_{t-m,l,z,v}$

sentiment distribution π_t and the sentiment-topic distribution θ_t respectively.

It is assumed that there exists a social corpus D consisting of timestamped posts, denoted as $D = \{d_1, d_2, \dots, d_{|D|}\}$. Vocabulary of D is denoted as V . Each post d_i in the corpus is generated by a user within a timeslice and the post d_i contains a bag of words, denoted as $\{w_i^1, w_i^2, \dots, w_i^{|d_i|}\}$. With sliding window preprocessing operation, D is represented as $D = \{D_1, \dots, D_t, \dots, D_W\}$, where D_t denotes the posts posted in timeslice t , W denotes index of the last time window. Formally, the generative process for each post in time window t is formalized as a procedure named Post_Generator.

From the graphical representation and generative process of LDTSM, we can see that LDTSM has the following characteristics: (1) In terms of topic detection and sentiment identification, LDTSM takes into social inheritance of sentiment and topic when learning sentiment topic patterns of posts in current time window, which is different from traditional probabilistic generative models; (2) As far as efficiency of model parameter estimation, LDTSM may be more efficient to find optimal parameter configuration, since parameter initialization based on historical information is better than completely random guess.

B. PARAMETERS ESTIMATION

Exact inference is intractable in probabilistic graph models and approximate inference is often necessary. Popular

Procedure 1: Post_Generator

// Step 1. generate distribution sentiment-topic-word

1. for each sentiment $l \in \{l_1, l_2, \dots, l_S\}$
2. for each topic $z \in \{z_1, z_2, \dots, z_T\}$
3. for each word $w \in \{w_1, w_2, \dots, w_V\}$
4. Draw $\phi_{t,l,z,w} \sim Dir\left(\sum_{m=1}^M \frac{\beta_{t-m+1}}{\gamma_{t-m+1}} \frac{\phi_{t-m}}{\pi_{t-m}}\right)$

//Step 2. generate distributions sentiment and sentiment-topic

5. for each document $d \in D_t$
6. for each sentiment $l \in \{l_1, l_2, \dots, l_S\}$
7. Draw $\pi_{t,d,l} \sim Dir\left(\sum_{m=1}^M \frac{\gamma_{t-m+1}}{\alpha_{t-m+1}} \frac{\pi_{t-m}}{\theta_{t-m}}\right)$
8. for each topic $z \in \{z_1, z_2, \dots, z_T\}$
9. Draw $\theta_{t,d,l,z} \sim Dir\left(\sum_{m=1}^M \frac{\alpha_{t-m+1}}{\beta_{t-m+1}} \frac{\theta_{t-m}}{\phi_{t-m}}\right)$

// Step 3. generate words using distributions ϕ, θ and π

10. for each word w in document d
11. Draw $l \sim Mul(\pi_{d,t})$
12. Draw $z \sim Mul(\theta_{d,t,l})$
13. Draw $w \sim Mul(\phi_{t,l,z})$

approximate inference methods mainly include sampling (Monte Carlo) methods, variational methods and loopy belief propagation. Here Gibbs sampling technique is adopted to estimate parameters of LDTSM due to its intuitive simplicity and high efficiency.

According to the probability graphic model (depicted in Figure 3) and Bayes chain rule, the joint probability of words, topics and sentiments can be factored in Eq. (1). For the three terms in right hand of Eq.(1), we can obtain the three probabilities computation by integrating out π_t, ϕ_t and θ_t , as can be formalized as equations (2), (3) and (4).

$$P(w_t, l_t, z_t | \phi'_{t-m}, \theta'_{t-m}, \pi'_{t-m}, \alpha'_{t-m+1}, \beta'_{t-m+1}, \gamma'_{t-m+1}) = P(w_t | l_t, z_t, \phi'_{t-m}, \beta'_{t-m+1}) \times P(z_t | l_t, \theta'_{t-m}, \alpha'_{t-m+1}) \times P(l_t | \pi'_{t-m}, \gamma'_{t-m+1}) \quad (1)$$

where $\theta'_{t-m} = \{\theta_{t-m}\}_{m=1}^L, \pi'_{t-m} = \{\pi_{t-m}\}_{m=1}^L, \alpha'_{t-m+1} = \{\alpha_{t-m+1}\}_{m=1}^L, \beta'_{t-m+1} = \{\beta_{t-m+1}\}_{m=1}^L$ and $\gamma'_{t-m+1} = \{\gamma_{t-m+1}\}_{m=1}^L$.

$$P(w_t | l_t, z_t, \phi'_{t-m}, \beta'_{t-m+1}) = \prod_{l=1}^S \prod_{z=1}^T \frac{\Gamma\left(\sum_{v=1}^V \sum_{m=1}^M \frac{\beta_{t-m+1}}{\gamma_{t-m+1}} \frac{\phi_{t-m}}{\pi_{t-m}}\right)}{\prod_{v=1}^V \Gamma\left(\sum_{m=1}^M \frac{\beta_{t-m+1}}{\gamma_{t-m+1}} \frac{\phi_{t-m}}{\pi_{t-m}}\right)} \times \prod_{l=1}^S \prod_{z=1}^T \frac{\prod_{v=1}^V \Gamma\left(n_{t,l,z,v} + \sum_{m=1}^M \frac{\beta_{t-m+1}}{\gamma_{t-m+1}} \frac{\phi_{t-m}}{\pi_{t-m}}\right)}{\Gamma\left(\sum_{v=1}^V \left(n_{t,l,z,v} + \sum_{m=1}^M \frac{\beta_{t-m+1}}{\gamma_{t-m+1}} \frac{\phi_{t-m}}{\pi_{t-m}}\right)\right)} \quad (2)$$

where $n_{t,l,z,v}$ denotes the frequency of word v assigned to both sentiment l and topic z in timeslice $t, n_{t,l,z}$ denotes the total

frequency of all words assigned to both sentiment l and topic z in timeslice t .

$$P(l_t | \pi'_{t-m}, \gamma'_{t-m+1}) = \frac{\Gamma\left(\sum_{l=1}^S \sum_{m=1}^M \frac{\gamma_{t-m+1}}{\alpha_{t-m+1}} \frac{\pi_{t-m}}{\theta_{t-m}}\right)}{\prod_{l=1}^S \Gamma\left(\sum_{m=1}^M \frac{\gamma_{t-m+1}}{\alpha_{t-m+1}} \frac{\pi_{t-m}}{\theta_{t-m}}\right)} \frac{\prod_{l=1}^S \Gamma\left(n_{t,l} + \sum_{m=1}^M \frac{\gamma_{t-m+1}}{\alpha_{t-m+1}} \frac{\pi_{t-m}}{\theta_{t-m}}\right)}{\Gamma\left(\sum_{l=1}^S \left(n_{t,l} + \sum_{m=1}^M \frac{\gamma_{t-m+1}}{\alpha_{t-m+1}} \frac{\pi_{t-m}}{\theta_{t-m}}\right)\right)} \quad (3)$$

where $n_{t,l}$ denotes total frequency of all words with sentiment label being l in timeslice t .

$$P(z_t | l_t, \theta'_{t-m}, \alpha'_{t-m+1}) = \prod_{l=1}^S \frac{\Gamma\left(\sum_{z=1}^T \sum_{m=1}^M \frac{\alpha_{t-m+1}}{\beta_{t-m+1}} \frac{\theta_{t-m}}{\phi_{t-m}}\right)}{\prod_{z=1}^T \Gamma\left(\sum_{m=1}^M \frac{\alpha_{t-m+1}}{\beta_{t-m+1}} \frac{\theta_{t-m}}{\phi_{t-m}}\right)} \times \prod_{l=1}^S \frac{\prod_{z=1}^T \Gamma\left(n_{t,l,z} + \sum_{m=1}^M \frac{\alpha_{t-m+1}}{\beta_{t-m+1}} \frac{\theta_{t-m}}{\phi_{t-m}}\right)}{\Gamma\left(\sum_{z=1}^T \left(n_{t,l,z} + \sum_{m=1}^M \frac{\alpha_{t-m+1}}{\beta_{t-m+1}} \frac{\theta_{t-m}}{\phi_{t-m}}\right)\right)} \quad (4)$$

where n_t denotes the word-granular size of document d .

With the joint distribution, posterior distribution $P(l_{t,i} = l, z_{t,i} = z | l_{t,-i}, z_{t,-i}, w_t, \phi'_{t-m}, \theta'_{t-m}, \pi'_{t-m}, \alpha'_{t-m+1}, \beta'_{t-m+1}, \gamma'_{t-m+1})$ is estimated via formula (5), where the subscript $-i$ denotes a count excluding the current assignment.

$$P(l_{t,i}, z_{t,i} | l_{t,-i}, z_{t,-i}, w_t, \phi'_{t-m}, \theta'_{t-m}, \pi'_{t-m}, \alpha'_{t-m+1}, \beta'_{t-m+1}, \gamma'_{t-m+1}) \propto \frac{n_{t,l} + \sum_{m=1}^M (\gamma_{t-m+1} \frac{\pi_{t-m}}{\theta_{t-m}}) - 1}{n_t + \sum_{l=1}^S \sum_{m=1}^M (\gamma_{t-m+1} \frac{\pi_{t-m}}{\theta_{t-m}}) - 1} \times \frac{n_{t,l,z} + \sum_{m=1}^M (\alpha_{t-m+1} \frac{\theta_{t-m}}{\phi_{t-m}}) - 1}{n_{t,l} + \sum_{l=1}^S \sum_{m=1}^M (\alpha_{t-m+1} \frac{\theta_{t-m}}{\phi_{t-m}}) - 1} \times \frac{n_{t,l,z,v} + \sum_{m=1}^M (\beta_{t-m+1} \frac{\phi_{t-m}}{\pi_{t-m}}) - 1}{n_{t,l,z} + \sum_{l=1}^S \sum_{m=1}^M (\beta_{t-m+1} \frac{\phi_{t-m}}{\pi_{t-m}}) - 1} \quad (5)$$

With formula (5), document-sentiment distribution $\pi_{t,d,l}$ in timeslice t , document sentiment-topic distribution $\theta_{t,d,l,z}$ in timeslice t , and sentiment-topic word distribution $\phi_{t,l,z,v}$ in timeslice t (formulae (6), (7) and (8)) can be approximated respectively using samples obtained from the Markov chain.

$$\pi_{t,d,l} = \frac{n_{t,l} + \sum_{m=1}^M \frac{\gamma_{t-m+1}}{\alpha_{t-m+1}} \frac{\pi_{t-m}}{\theta_{t-m}}}{n_t + \sum_{l=1}^S \sum_{m=1}^M \frac{\gamma_{t-m+1}}{\alpha_{t-m+1}} \frac{\pi_{t-m}}{\theta_{t-m}}} \quad (6)$$

$$\theta_{t,l,z} = \frac{n_{t,l,z} + \sum_{m=1}^M \overline{\alpha_{t-m+1}} \overline{\theta_{t-m}}}{n_{t,l} + \sum_{z=1}^T \sum_{m=1}^M \overline{\alpha_{t-m+1}} \overline{\theta_{t-m}}} \quad (7)$$

$$\phi_{t,l,z,w} = \frac{n_{t,l,z,w} + \sum_{m=1}^M \overline{\beta_{t-m+1}} \overline{\phi_{t-m}}}{n_{t,l,z} + \sum_{w=1}^V \sum_{m=1}^M \overline{\beta_{t-m+1}} \overline{\phi_{t-m}}} \quad (8)$$

Gibbs sampling is one of the simplest Monte Carlo sampling procedures. It starts with a random setting of hidden states and then updates each hidden state, according to the probability distribution conditioned on all the other states and the fixed parameters. Similar to JST_Gibbs [4], a complete overview of Gibbs sampling procedure corresponding to LDTSM is given in Algorithm 1.

Algorithm 1: LDTSM_Gibbs

Input: Corpus D_t , historical parameters $(\pi'_{t-m}, \theta'_{t-m}, \phi'_{t-m})$

Output: Current model parameters $\pi_t, \theta_{t,l}$ and $\phi_{t,l,z}$

```

1 Iter = 1;
2 while Iter < MaxIterations do
3   for each document  $d \in D_t$  do
4     for each word  $w \in [1, V]$  do
5       Exclude word  $w$  with being given sentiment
        polarity and topic, and update count
        variables  $n_{t,l-} = 1, n_{t,l,z-} = 1$  and
         $n_{t,l,z,w-} = 1$ ;
6       Randomly draw a new sentiment topic pair
         $(l^*, z^*)$  for  $w$  using Eq. (5);
7       Update the count variables related with word
         $w$   $n_{t,l^*} + = 1, n_{t,l^*,z^*} + = 1$  and
         $n_{t,l^*,z^*,w} + = 1$ ;
8     end
9   end
10  if Iter > 800 then
11    Update matrices  $\pi, \theta$  and  $\phi$  using Eq 6, 7,
        8 every 10 iterations;
12  end
13 end

```

C. SENTIMENT TOPIC ANALYSIS

Based on the proposed framework and model, we formalize the procedure of sentiment topic analysis as an algorithm, named LDTSM_Analyzer (Algorithm 2). LDTSM_Analyzer mainly consists of three subprocedures: 1) text preprocessing and corpus partition (step1-step2), 2) estimating LDTSM parameters via Gibbs sampling (step4-step9), and 3) topic detection, sentiment classification and evolution analysis of topic-sentiment with the estimated sentiment distribution and sentiment-topic distribution (step10-step11).

Algorithm 2: LDTSM_Analyzer

Input: Corpus D , timeslice granularity g , dependency distance M

Output: topic label and sentiment polarity of each document in D

```

1 Construct vocabulary for  $D$  via NLP preprocessing
  techniques such as word segmentation and stemming ;
2 Partition  $D$  into  $\mathbb{D} = \{D_1, D_2, \dots, D_p\}$  with parameter  $g$ 
  ;
3 for  $i = 1$  to  $|\mathbb{D}|$  do
4   Utilize sentiment dictionary to initialize sentiment
    distribution for words in  $D_i$ ;
5   if  $t < M$  then
6     Apply JST_Gibbs to estimate parameters  $\pi_t, \theta_{t,l}$ 
    and  $\phi_{t,l,z}$ ;
7   else
8     Apply LDTSM_Gibbs to estimate parameters
     $\pi_t, \theta_{t,l}$  and  $\phi_{t,l,z}$ ;
9   end
10  Use the estimated sentiment distribution  $\pi$  to
    determine the sentiment polarity of each document
     $d$  in  $D_i$ : if  $\pi_{d1} > \pi_{d2}$ , then  $d$  is positive, otherwise
     $d$  is negative ;
11  Use the estimated  $\phi$  and  $\theta$  to extract topics and
    conduct evolution analysis topic and sentiment ;
12 end

```

Time complexity of LDTSM_Analyzer can be computed as follows. Vocabulary construction (step 1) and corpus partition (step 2) are one-shot deals for LDTSM_Analyzer, for simplicity, and can be excluded as data preprocessing during time complexity analysis of LDTSM_Analyzer. Time complexity of document sentiment polarity determination (step 10) and topic extraction (step 11) is $O(|D|)$. Parameter estimation(step5-step9) is the most computation-intensive. By comparing the computation process of JST_Gibbs and LDTSM_Gibbs, it is not difficult to find that computation of LDTSM_Gibbs is more expensive. With the formula (6), (7) and (8), size of tensors π, θ and ϕ is $L^*|D|^*S, L^*S^*T$ and $L^*S^*T^*V$, respectively. Apparently, the inequality $L^*S^*T \ll L^*|D|^*S \ll L^*S^*T^*V$ is true. Therefore, time complexity of parameter estimation (step5-step9) is $O(L^*S^*T^*V)$. From the above analysis, we conclude that time complexity of LDTSM_Analyzer is $O(L^*S^*T^*V)$.

VI. EXPERIMENTAL STUDY

A. DATASETS

Considering few manually-labeled corpuses with temporal information are available, we constructed four real-world datasets to evaluate the proposed model. Specially, message content, update time and message author are automatically collected through using the search API of Sinaweibo.

For the collected microblogs, we perform some preprocessing steps such as removing duplicate microblogs and

TABLE 2. Description of the four datasets used.

Datasets	#Instance	PNRatio	TimeRange	Focused
lindan	12485	0.31	2016/11/17-2016/11/23	yes
thaad	25108	0.72	2017/03/06-2017/03/25	yes
data1	25954	0.57	2012/1/1-2012/1/15	no
data2	12719	0.54	2014/11/6-2014/1/15	no

TABLE 3. Consistency of sentiment annotation.

Datasets	EmployeePair (1-2)	EmployeePair (1-3)	EmployeePair (2-3)
lindan	0.857	0.898	0.873
thaad	0.846	0.835	0.858
data1	0.743	0.772	0.776
data2	0.802	0.815	0.821

zombie users, filtering out the too-short microblogs. To obtain convincingly sentiment label of the collected microblogs, we employ three volunteers to independently assign a sentiment polarity tag for each message, and conduct check consistency for the labelled result with Kappa test (Table 3). From Table 3, we can see that, sentiment distribution of data1 is fuzzier than the others, and compared to the other two datasets, consistency of sentiment annotation is higher in the hot-topic-focused datasets linda and thaad.

For the inconsistent annotations from the volunteers, we determine the final sentiment polarity labels, according to high-voting principle. The four datasets are detailed in Table 2, where PNRatio column denotes the ratio of the number of positive instances to the number of negative instances, and column ‘‘Focused’’ denotes the fetching method in constructing the four datasets. In particular, dataset lindan and thaad is topic-driven and the other two are not topic-focused. Dataset lindan is focused on the event ‘‘Extramarital affair of badminton great Lin Dan’’, and dataset thaad is focused on the event ‘‘THAAD Deployment in South Korea’’.

B. SENTIMENT CLASSIFICATION

In this section, we evaluate sentiment classification performance of our approach: 1) comparative analysis between LDTSM and other state-of-the-art methods in terms of sentiment classification accuracy; 2) LDTSM classification performance’s sensitiveness to the number of topics and dependency distance.

1) METRIC

To evaluate performance of methods for text sentiment classification, we adopt the widely-used metric accuracy.

$$Accuracy = \frac{M}{N} \tag{9}$$

where M is the number of correctly classified samples, N is the size of corpus to be analyzed.

TABLE 4. Accuracy comparison on four datasets.

Datasets	SVM	JST	dNJST	TUS-LDA	LDTSM
lindan	0.803	0.744	0.753	0.761	0.786
thaad	0.782	0.741	0.742	0.753	0.771
data1	0.695	0.626	0.632	0.664	0.687
data2	0.736	0.646	0.657	0.683	0.715

2) SENTIMENT CLASSIFICATION ACCURACY ANALYSIS

In view of non-supervision, nature of LDTSM, we select three state-of-the-art probabilistic graphical models ([4], dNJST [38], TUS-LDA [39]) and a typical libSVM-based (features: 1-grams and 2-grams) supervised learning model as competitors. Experimental results on four datasets are listed in Table 4. According to Table 4, we have the following findings: 1) In the four probabilistic graphical models, sentiment accuracy of JST is lower than those of the other three, namely, dNJST, TUS-LDA and LDTSM. This demonstrates that suitable utilization of time information is beneficial to improve sentiment detection performance. 2) LDTSM outperforms all the other unsupervised learning competitors in terms of sentiment classification accuracy on the four datasets, especially dNJST and TUS-LDA, as indicates that, compared to dNJST and TUS-LDA, sentiment inheritance is proposed in LDTSM to make time information get more reasonable utilization. 3) There is a small gap in classification accuracy between LDTSM and SVM. If take into consideration high cost of acquiring sentiment polarity labeled training corpuses, the gap is acceptable in most cases. It is worthy of noting that this observation agrees with Pang’s conclusion [40], i.e., SVMs based on bag-of-unigram features can make excellent result in sentiment classification in short text corpus.

3) SENTIMENT CLASSIFICATION ACCURACY VS. NUMBER OF TOPICS

Considering that text sentiment polarities are intertwined with its topics and the number of topics is predetermined in LDTSM, we attempt to explore how the predetermined parameter exerts influence on sentiment classification of LDTSM. A group of experiments on LDTSM with topic number $T \in \{1, 5, 10, 20, 30, 40\}$ is conducted and the performances of sentiment classification are shown in Figure 4. In Figure 4, topic number has diverse effects on sentiment classification performance of LDTSM on different datasets. Specially, classification accuracy of LDTSM reaches its maximum when topic number is set as 5 in datasets lindan and thaad, 30 and 40 in datasets data1 and data2. The observations may be explained as follows: too small T pays less attention to the correlation between topic and sentiment and may degrade LDTSM into LDA focused on sentiment detection, and leads to low sentiment classification accuracy. And too great T may arbitrarily break some intact topics into some fragments of noisy topics, and make LDTSM ineffectively recognize

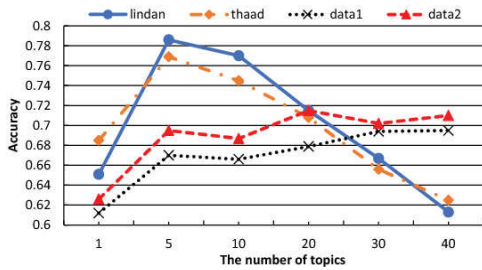


FIGURE 4. Sentiment classification accuracy with different #topics.

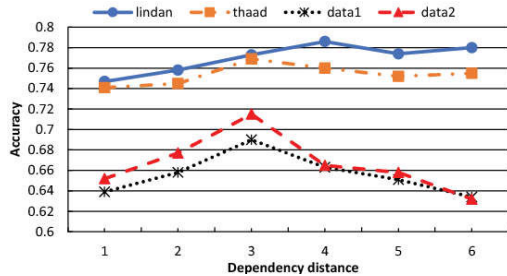


FIGURE 5. Sentiment classification accuracy with dependency distance (timeslice unit is day).

underlying sentiment patterns in text sequences. Examination of Figure 4 indicates that, accuracy of LDTSM on datasets lindan and thaad gradually ascends and descends with the increase of the number of topics, and in datasets data1 and data2, classification accuracy slowly grows with the greater T , except for a few cases: from $T = 20$ to $T = 30$ in data2, and a sharp increase from $T = 1$ to $T = 5$ in data1 and data2. This is also explained like this: datasets lindan and thaad are topic-focused and may contain less latent subtopics than datasets data1 and data2.

It is difficult to accurately set T for a social text sentiment analysis task, but the above experimental results may hint that T should be set as a less value in a topic-focused corpus and vice versa.

4) SENTIMENT CLASSIFICATION ACCURACY VS. DEPENDENCY DISTANCE

Long-term dependency mechanism is one of LDTSM key characteristics, this section aims to explore the impact of dependency distance on sentiment classification accuracy. The experimental results in the first 6 timeslices of four datasets are shown in Figure 5.

As shown in the Figure 5, we can observe that, 1) classification accuracy is highest when dependency distance is set as 3 or 4. Too large dependency distance may prevent LDTSM from effectively inheriting historical sentiment knowledge, and too small one may introduce noisy sentiment, both cases will evidently incur loss of sentiment classification performance. 2) in datasets lindan and thaad, classification performance is basically stable after LDTSM reaches its maximal classification accuracy, and yet decreases significantly in datasets data1 and data2. That topic-focused datasets lindan and thaad are often highly emotional contagious may make sense this observation.

C. TOPIC DETECTION

In this section, we concern about the quality of topics with different sentiment polarities detected by LDTSM. We run LDTSM on the topic-focused dataset lindan and topic-diversified dataset data2, and select top10 topic words from each sentiment polarity category. The experimental results are listed in Table 5. From Table 5, it is not difficult to conclude that the extracted topics seem to be fairly informative and coherent, and can reflect the underlying concerns of microbloggers.

For example, positive topic words such as “羽毛球 (badminton), 运动员 (player), 伟大 (great), 比赛(game)” imply that the topic discussed by microbloggers is about “LinDan is a great badminton player”, and “天猫 (Tmall), 十一 (Singles Day), 优惠(discount), 商家(merchant)” may be about “Big promotion from Tmall in Singles Day”, and the microbloggers hold positive attitudes, such as support, respect, joy and happiness, towards the above two topics.

Similarly, through the negative topic words “垃圾 (rubbish), 恶心(nausea), 家暴(domestic violence), 出轨 (having an affair)” and “假(fake), 慢(slow), 差(bad), 骗子(cheater)”, we can see that microbloggers show negative emotions to LinDan in the gossip “extramarital affairs” and merchants with low-quality services and commodities.

D. EVOLUTION ANALYSIS OF TOPIC AND SENTIMENT

1) TOPIC EVOLUTION

Tracking topic evolution helps decision makers to understand how hot topics produce and develop, and make reasonable predictions. Taking dataset lindan as a sample, we analyze how the extracted topics evolve through Jaccard similarity (formula (10)). Similarity evolution and content evolution of an example topic (topic 4, which is randomly chosen from all the topics extracted from dataset lindan) are shown in Figure 6 and Table 6. From Figure 6, we can see that topic similarity from day 2 to day 5 is high and stable, and there is a sharp fluctuation in time buckets(1-2 and 5-6). Shift of topic words in Table 6 may give a rational explanation for this: topic words “出轨(having an affair), 男人(man), 老婆(wife), 道歉(apology), 林丹(Lin Dan)” in day 1 indicates that microbloggers firstly discuss the event “LinDan apologized for betraying his wife”, “Apology from Lin Dan” becomes a continuous focus in the next 4 days, and yet the concerned topic changes from “LinDan’s betrayal” to “hating the mistress Zhao Yaqi(赵雅淇)” in day 6.

$$J(A_{t-1}, A_t) = \frac{|A_{t-1} \cap A_t|}{|A_{t-1} \cup A_t|} \quad (10)$$

where A_t denotes the set consisting of probabilistically top 100 key words in topics extracted from timeslice t .

2) SENTIMENT EVOLUTION

Generally, microbloggers’ attitudes to different topics develop along different traces. Understanding topic content evolution alone is insufficient for a decision maker, grasping

TABLE 5. Topic examples detected by LDTSM.

Datasets	Positive	Negative
lindan	羽毛球(badminton), 羽超(badminton super league), 运动员(player), 鼓励(encourage), 好壮(buff), 舞台(arena), 加油(come on), 俱乐部(club), 伟大(great), 比赛(game)	垃圾(rubbish), 恶心(nausea), 家暴(domestic violence), 离婚(divorce), 帽子(cuckold), 出轨(having an affair), 小三(mistress), 三观(three outlooks), 娱乐圈(entertainment world), 退出 (retire)
data2	红包(red envelope), 降价(cut prices), 天猫(Tmall), 十一(Singles Day), 优惠(discount), 商家(merchant), 漂亮(beautiful), 礼物(gift), 好评(positive review), 送(give away), 价格(price)	客服(customer service), 发货(delivery), 物流(logistics), 假(fake), 慢(slow), 差(bad), 骗子(cheater), 订单(order), 态度(attitude), 包裹(parcel)

TABLE 6. Content evolution of topic 4.

Timeslice	Key words
Day 1	出轨(having an affair), 男人(man), 老婆(wife), 道歉(apology), 林丹(Lin Dan), 怀孕(pregnant), 孩子(kid), 垃圾(rubbish), 恶心(nausea), 评论(review)
Day 2	谢杏芳(Xie Xingfang), 家人(family), 出轨(having an affair), 原谅(forgive), 林丹(Lin Dan), 女人(woman), 男人(man), 离婚(divorce), 担当(responsibility), 孩子(kid)
Day 3	出轨(having an affair), 孩子(kid), 女人(woman), 男人(man), 原谅(forgive), 林丹(Lin Dan), 谢杏芳(Xie Xingfang), 担当(responsibility), 天理不容(abhorrent), 拒绝(refuse), 参加(take part in)
Day 4	出轨(having an affair), 林丹(Lin Dan), 原谅(forgive), 男人(man), 女人(woman), 孩子(kid), 老婆(wife), 马蓉(Ma Rong), 恶心(nausea), 三观(three outlooks)
Day 5	加油(come on), 出轨(having an affair), 委屈(wronged), 马蓉(Ma Rong), 折磨(torment), 痛恨(hate), 原谅(forgive), 林丹(Lin Dan), 老婆(wife), 恶心(nausea)
Day 6	道歉(apology), 鼓励(encourage), 林丹(Lin Dan), 恶心, 炒作(hype), 原谅(forgive), 伤害(hurt), 矫情(preentious), 婊子(bitch), 戏子(actress), 赵雅淇(Zhao Yaqi)
Day 7	新赛季(new season), 林丹(Lin Dan), 朋友(friends), 问候(abuse), 羽毛球 (badminton), 再接再厉(make persistent efforts), 俱乐部(club), 道歉(apology), 原谅(forgive), 运动员(player)

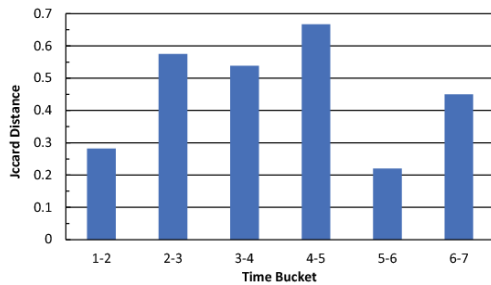


FIGURE 6. Topic distance in dataset lindan.

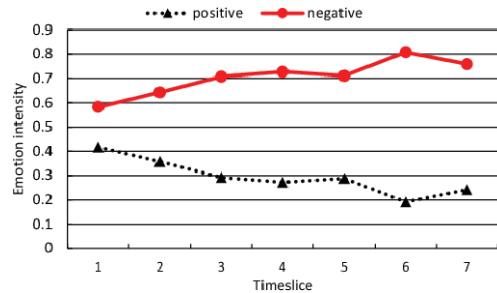


FIGURE 7. Sentiment evolution in dataset lindan.

situations of public sentiment hidden in the hot topics is also indispensable. Similar to the topic evolution analysis, here we analyze how microbloggers’ sentiment evolves in the hot event “Extramarital affair of badminton great Lin Dan” via sentiment knowledge discovery in dataset lindan. Experimental results are depicted in Figure 7. From Figure 7, we can see that, on the whole, sentiment polarity of microbloggers is negative, and negative intensity gradually increases in the event lifecycle, and yet positive intensity gradually decreases, but there is an exception for timeslice 6, where the intensity of sentiment is reduced for negative polarity and enlarged for positive polarity. The fluctuations of sentiment intensity are attributed to the discussed event: it is widely believed that having an extramarital affair is immoral, so most microbloggers hold negative attitudes to Lin Dan throughout

the episode. From timeslice 1 to timeslice 3, intensity of negative sentiment climbs slowly, as may be that the anger sentiment of netizens was gradually fermenting, and from timeslice 3 to timeslice 5, microbloggers gradually reconcile to the fact “Lin Dan’s having an extramarital affair”, and key words in Row 6 of Table 6 indicates that apology from the mistress Zhao Yaqi(赵雅淇) stung the injured heart of netizens again, and so negative sentiment intensity sharply increases at day 6.

E. EFFICIENCY ANALYSIS OF LDTSM

In this section, we investigate efficiency of LDTSM, by comparing running time of LDTSM and the other three probabilistic graphical models, i.e., JST, dNJST and TUS-LDA.

TABLE 7. Time efficiency comparison on four datasets.

Datasets	JST	dNJST	TUS-LDA	LDTSM
lindan	7335.41	7223.37	7461.37	7212.75
thaad	11876.22	111256.74	12128.53	11025.38
data1	12268.49	11987.88	12205.15	11886.93
data2	7678.36	7391.38	7693.45	7376.66

Obviously, the number of iterations cannot be used as a time measure, since parameter estimations of different models demand different amount of work in their inner loops. For the sake of fairness, we choose the elapsed CPU time as a measure instead of number of iterations, and record the running time of the four algorithms according to the following rule: if the classification accuracy of an algorithm has achieved the corresponding value in Table 4 before the user-defined maximum number of iterations is met, then we break the loop and record the time used in the loop.

The experiment results are listed in Table 7. From the table, we can see that, 1) compared to JST and TUS-LDA, LDTSM performs better on the four datasets in terms of inference efficiency, 2) CPU time of LDTSM is as good as dNJST on datasets with easily discernable sentiment polarity patterns, such as lindan and thaad, and yet LDTSM exhibits remarkable advantages on data1, which is with indistinguishable boundary of sentiment polarities, 3) LDTSM and dNJST surpass JST and TUS-LDA in time efficiency of estimating model parameters. Explanation to this observation is obvious. Inheritance of sentiment is taken into account while dNJST and LDTSM learn sentiment topic distribution of social texts posted in current timeslice, as gives a better initialization of model parameters and further produces higher-quality sampling convergence, and datasets with complex and ambiguous sentiment structure will benefit the optimization ability of LDTSM and dNJST. Especially for LDTSM, the advantage is more significant, since LDTSM takes into consideration both sentiment inheritance and topic inheritance, and only sentiment inheritance is considered in dNJST.

VII. CONCLUSION

In this paper, we have presented a new topic-sentiment modelling framework with long-term dependency for both sentiment classification and topic extraction. We evaluate our model on four real-world microblog datasets and show that our model achieves superior performance compared with several strong baselines on sentiment classification. Evolution analysis results on experimental datasets demonstrate that our approach is effective and promising. In future work, we will explore extending this model for sentiment community dynamics analysis, and will apply heterogeneous scheduling algorithms to speed up text sentiment analysis in distributed computing environment.

REFERENCES

[1] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.

[2] R. Wang, D. Zhou, M. Jiang, J. Si, and Y. Yang, "A survey on opinion mining: From stance to product aspect," *IEEE Access*, vol. 7, pp. 41101–41124, 2019.

[3] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowl. Inf. Syst.*, vol. 60, pp. 617–663, Jul. 2018.

[4] C. Lin, Y. He, R. Everson, and S. Ruder, "Weakly supervised joint sentiment-topic detection from text," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 1134–1145, Jun. 2012.

[5] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, "Topic sentiment mixture: Modeling facets and opinions in weblogs," in *Proc. 16th Int. Conf. World Wide Web WWW*, 2007, pp. 171–180.

[6] R. P. Abelson, "Whatever became of consistency theory?" *Personality Social Psychol. Bull.*, vol. 9, no. 1, pp. 37–54, Mar. 1983.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[8] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Comput. Linguistics*, vol. 37, no. 2, pp. 267–307, Jun. 2011.

[9] R. Ortega, A. Fonseca, and A. Montoyo, "SSA-UO: Unsupervised twitter sentiment analysis," in *Proc. 2nd Joint Conf. Lexical Comput. Semantics*, vol. 2, 2013, pp. 501–507.

[10] M. Al-Ayyoub, S. B. Essa, and I. Alsmadi, "Lexicon-based sentiment analysis of Arabic tweets," *Int. J. Social Netw. Mining*, vol. 2, no. 2, pp. 101–114, 2015.

[11] S. Trinh, L. Nguyen, M. Vo, and P. Do, "Lexicon-based sentiment analysis of Facebook comments in Vietnamese language," in *Recent Developments in Intelligent Information and Database Systems*. Berlin, Germany: Springer, 2016, pp. 263–276.

[12] H. Hazimeh, M. Harissa, E. Mugellini, and O. A. Khaled, "Temporal sentiment tracking and analysis on large-scale social events," in *Proc. 8th Int. Conf. Softw. Comput. Appl. - ICSCA*, 2019, pp. 17–21.

[13] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 1–10.

[14] P. C. Guerra, W. Meira, and C. Cardie, "Sentiment analysis on evolving social streams: How self-report imbalances can help," in *Proc. 7th ACM Int. Conf. Web Search Data Mining - WSDM*, 2014, pp. 443–452.

[15] A. Bifet and E. Frank, "Sentiment knowledge discovery in twitter streaming data," in *Proc. Int. Conf. Discovery Sci.* Canberra, ACT, Australia: Springer, 2010, pp. 1–15.

[16] J. Si, A. Mukherjee, B. Liu, Q. Li, H. Li, and X. Deng, "Exploiting topic based twitter sentiment for stock prediction," in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, 2013, pp. 24–29.

[17] L. Li, Y. Wu, Y. Zhang, and T. Zhao, "Time+ user dual attention based sentiment prediction for multiple social network texts with time series," *IEEE Access*, vol. 7, pp. 17644–17653, 2019.

[18] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič, "The effects of Twitter sentiment on stock price returns," *PLoS ONE*, vol. 10, no. 9, Sep. 2015, Art. no. e0138441.

[19] A. Giachanou and F. Crestani, "Tracking sentiment by time series analysis," in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. - SIGIR*, 2016, pp. 1037–1040.

[20] M. Yang, J. Mei, H. Ji, Z. Wei, Z. Zhao, and X. Chen, "Identifying and tracking sentiments and topics from social media texts during natural disasters," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 527–533.

[21] A. Giachanou, I. Mele, and F. Crestani, "Explaining sentiment spikes in Twitter," in *Proc. 25th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2016, pp. 2263–2268.

[22] L. Wang, J. Niu, and S. Yu, "SentiDiff: Combining textual information and sentiment diffusion patterns for Twitter sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 10, pp. 2026–2039, Oct. 2020.

[23] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid," in *Proc. Conf. Empirical Methods Natural Lang. Process., Assoc. Comput. Linguistics*, 2010, pp. 56–65.

[24] A. Almars, X. Li, and X. Zhao, "Modelling user attitudes using hierarchical sentiment-topic model," *Data Knowl. Eng.*, vol. 119, pp. 139–149, Jan. 2019.

[25] Y. He, "Latent sentiment model for weakly-supervised crosslingual sentiment classification," in *Proc. Eur. Conf. Inf. Retr.* Dublin, Ireland: Springer, 2011, pp. 214–225.

[26] Z. Hai, G. Cong, K. Chang, P. Cheng, and C. Miao, "Analyzing sentiments in one go: A supervised joint topic modeling approach," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 6, pp. 1172–1185, Jun. 2017.

- [27] W. Liang, H. Xie, Y. Rao, R. Y. K. Lau, and F. L. Wang, "Universal affective model for Readers' emotion classification over short texts," *Expert Syst. Appl.*, vol. 114, pp. 322–333, Dec. 2018.
- [28] L. Lin, L. Gong, and H. Wang, "Learning personalized topical compositions with item response theory," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Jan. 2019, pp. 609–617.
- [29] L. Poddar, W. Hsu, and M. L. Lee, "Author-aware aspect topic sentiment model to retrieve supporting opinions from reviews," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 472–481.
- [30] M. M. Rahman and H. Wang, "Hidden topic sentiment model," in *Proc. 25th Int. Conf. World Wide Web - WWW*, 2016, pp. 155–165.
- [31] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Inf. Sci.*, vol. 488, pp. 190–204, Jul. 2019.
- [32] S. Wang, Z. Chen, and B. Liu, "Mining aspect-specific opinion using a holistic lifelong topic model," in *Proc. 25th Int. Conf. World Wide Web - WWW*, 2016, pp. 167–176.
- [33] S. Xiong, K. Wang, D. Ji, and B. Wang, "A short text sentiment-topic model for product reviews," *Neurocomputing*, vol. 297, pp. 94–102, Jul. 2018.
- [34] J.-D. Zhang and C.-Y. Chow, "CRATS: An LDA-based model for jointly mining latent communities, regions, activities, topics, and sentiments from geosocial network data," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 11, pp. 2895–2909, Nov. 2016.
- [35] Y. Wang, H. Li, and C. Lin, "Modeling sentiment evolution for social incidents," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 2413–2416.
- [36] F. Huang, S. Zhang, J. Zhang, and G. Yu, "Multimodal learning for topic sentiment analysis in microblogging," *Neurocomputing*, vol. 253, pp. 144–153, Aug. 2017.
- [37] L. Gui, L. Jia, J. Zhou, R. Xu, and Y. He, "Multi-task learning with mutual learning for joint sentiment classification and topic detection," *IEEE Trans. Knowl. Data Eng.*, early access, Jun. 9, 2020, doi: [10.1109/TKDE.2020.2999489](https://doi.org/10.1109/TKDE.2020.2999489).
- [38] X. Fu, K. Yang, J. Z. Huang, and L. Cui, "Dynamic non-parametric joint sentiment topic mixture model," *Knowl.-Based Syst.*, vol. 82, pp. 102–114, Jul. 2015.
- [39] K. Xu, G. Qi, J. Huang, T. Wu, and X. Fu, "Detecting bursts in sentiment-aware topics from social media," *Knowl.-Based Syst.*, vol. 141, pp. 44–54, Feb. 2018.
- [40] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. EMNLP*, 2002, pp. 79–86.



FALIANG HUANG received the Ph.D. degree in computer science from the South China University of Technology, Guangzhou, China, in 2011. He is currently an Associate Professor with Nanning Normal University, China. His current research interests include data mining and machine learning.



CHANGAN YUAN received the Ph.D. degree in computer application technology from Sichuan University, China, in 2006. He is currently a Full-Time Professor with the Guangxi Academy of Sciences, Nanning, China. His research interests include computational intelligence and data mining.



YINGZHOU BI received the Ph.D. degree in computer science from Wuhan University, Wuhan, China, in 2008. He is currently a Professor with Nanning Normal University, Nanning, China. His research interests include intelligent computing, intelligent information processing, and social computing.



JIANBO LU received the M.S. degree in computer science from Guangxi University, Nanning, China, in 2003. He is currently an Associate Professor with Nanning Normal University, Nanning. His research interests include intelligent computing and software engineering.

...